



Recent Research in Genetics and Genomics 2022; 4(2): 1-16

Original article:

Auto Calculation of Genetic Diversity Measures Based on Molecular Markers Data

"Auto Calculation Measures Genetic Diversity"

Alaa Ahmed Bader ELDeen^{1*}, Ibrahim ElAwady Abd ElHameed ELSemman², Ameer Effat Mohamed Mahmoud ELFarash³, Sara Salah Mohamed⁴

¹: Department of Biostatistics of Molecular Biology Research & Studies Institute, Assuit University, Egypt; ²: Department of information systems of computer and information Faculty, Assuit university, Egypt; ³: Department of Genetics of Agriculture Faculty, Assuit University, Egypt; ⁴: Department of mathematics of science Faculty, new valley university,

Egypt. *: Correspondence to: aalaaahmed379@gmail.com

DOI: 10.21608/rrgg.2022.156544.1010. Received: 15 August 2022; accepted: 26 August 2022; published 30 August 2022

Abstract

Ahmed Bader EL-Deen calc is a free computer software program that offers a simple interface for measuring genetic diversity to help researchers achieve accurate genetic diversity estimates to satisfy the demands of geneticists and breeders by using data from molecular markers. to easily calculate gene diversity (Gd) indices (such as polymorphism information content (PIC), resolving power (Rb), polymorphism percent (P%), the diversity index (Di), the molecular index (MI), observed heterozygosity (Ho), expected heterozygosity (He), and nucleotide diversity of the dominant and co-dominant markers' Gd) by using an appropriate programming language (C-sharp) for the design. To evaluate our program, we collected data on different shapes of dominant and co-dominant molecular markers from other researchers and used them to compare the efficacy and, most commonly, for genetic diversity measures. For ease of interpretation, we presented the results in a tabular format. In conclusion, the computer-based solution can potentially reduce the time necessary for manual computations while simultaneously making data analysis simple.

Keywords: Genetics diversity measures; software; nucleotide diversity.

1. Introduction

One of the most critical concerns for geneticists and breeders is assessing genetic diversity, which allows for the quick selection of proper genotypes for improvement efforts (Etminan *et al.*, 2016). We can accomplish the assessment of genetic diversity in a variety of ways, the most efficient of which is molecular markers (Schulman, 2007). There are various types of molecular markers, but the majority of them use polymerase chain reaction (PCR) and DNA amplification (Tautz, 1989; Williams *et al.*, 1990; Vos *et al.*, 1995; Li and Quiros 2001; Collard and Mackill 2009). Based on their ability to distinguish between homozygous and heterozygous individuals, DNA-based molecular markers are divided into two categories: those that cannot distinguish between shapes of genotypes (AA and Aa) as a dominant marker, and those that do so as a co-dominant marker. Any molecular marker's final product is a binary matrix that includes "1" or "0" for the presence or absence of a band, respectively. This matrix is used to determine numerous polymorphisms and genetic diversity measures among the individuals being studied. The main problem that faced previous research is a scarcity of intelligent software that can aid in the automatic computation of genetic diversity indices. For example, there is no free online tool for calculating genetic diversity (GD) using a formula for dominant and co-dominant

markers. Such tools are important since calculating GD manually might be difficult depending on the number of alleles and markers examined (Abuzayed *et al.*, 2017).

There are a variety of software options to measure genetic diversity, such as the PolyPicker software (Botstein *et al.*, 1980); GenAlEx software (Peakall and Smouse 2006); the GDA software (Lewis, 2001); PopGene software (Yeh *et al.*, 1999); Power Marker software (Liu and Muse 2005); Cervus software (Kalinowski *et al.*, 2007); Arlequin software (Excoffier *et al.*, 2005); GDdom software (Abuzayed *et al.*, 2017); and the software Gene calc (Miks *et al.*, 1955); and although the previous programs were concerned with measuring Polymorphism information content (PIC), polymorphism percent (P), and expected heterozygosity (He), Ahmed Bader's EL-Deen calc program sheds light on PIC, P, and He in addition to resolving power (RP), the diversity index (DI), molecular index (MI), observed heterozygosity (Ho), and nucleotide diversity (ND).

As a result, we created an auto-computation tool for distinct genetic diversity measures using molecular marker data. Our program is free computer software that offers easy, rapid, and precise computation for dominant and co-dominant markers using a frequently used formula. For ease of use, the results are given in tabular format. It can calculate PIC, RP, P, DI, and MI for the dominant

marker and PIC, RP, Ho, He, P, DI, and MI for the co-dominant marker (bi-allelic, diploid, and tetraploid for only 3 alleles), and it can also calculate nucleotide diversity for only homologous sequence.

2. Material and methods

We used Microsoft Visual Studio Page for creating our program and indicating the results as tables; Visual Studio is a server-side scripting language for web development, created by Microsoft and widely used, free source. We created an algorithm that takes an Excel file as input and produces the tables. To treat the Excel file as a matrix, our method first parses the contents of the file. The values of the columns and rows are then used to perform calculations on this matrix, we can obtain data to use in our program by; **Firstly**, we obtain the image of the gel (either a dominant or co-dominant marker) from research. **Secondly**, despite whether the marker is dominant or co-dominant, we will code the information as a binary matrix; by using gene profiler software, 1D scan software, or a manual method, we can convert the image of the gel into a binary matrix. **Thirdly**, we enter the binary number (matrix) 0 or 1 into our program to measure genetic diversity based on molecular marker data, and it can also calculate nucleotide diversity.

We scored alleles for the presence or absence of each dominant and co-dominant marker, and we computed GD using the following formulas:

According to (**Chesnokov and Artemyeva 2015**), **He** can be defined as the likelihood that an individual in the population is heterozygous for the locus. It can be computed using the following formula (**Liu, 1998**):

$$H = 1 - \sum p_i^2 \quad (1)$$

Where the summing is over all possible alleles and p_i is the allele frequency for the (I) allele (**Amiryousefi *et al.*, 2018**).

Ho is the fraction of heterozygous genes in a population. It is calculated by dividing the total number of heterozygotes by the sample size for each locus (**Chesnokov and Artemyeva 2015**). It can be computed using the following formula (**Liu, 1998**):

$H_o = \text{Number of heterozygote individuals} / \text{size of the sample} \quad (2)$

PIC is the ability of a marker to establish polymorphism in the population based on the number of alleles discovered and on their distribution frequency (**Chesnokov and Artemyeva 2015**). PIC can be computed for co-dominant markers using the following formula (**Guo and Elston 1999**):

$$PIC = 1 - \sum p_i^2 - \sum \sum p_i^2 p_j^2 \quad (3)$$

Where, the first summation reflects the entire number of alleles (i-th and j-th alleles have

population frequencies of p_i and P_j , respectively), whereas the two subsequent summations represent all i and j , where $i \neq j$ (Amiryousefi *et al.*, 2018).

PIC for dominant markers can be calculated using the following formula (Roldan Ruiz *et al.*, 2000):

$$PIC = 2f_i(1-f_i) \quad (4)$$

Where, the frequency of the amplified allele (band presence) is f_i , while the frequency of the null allele is $1-f_i$ (Abuzayed *et al.*, 2017).

R_p is the ability to identify changes between a large number of genotypes (Chesnokov and Artemyeva, 2015) and can be computed using the following formula (Prevost and Wilkinson 1999):

$$R_p = \sum I_b \quad (5)$$

Where I_b (band informativeness) is calculated as $I_b = 1 - (2|0.5-p|)$, p being the percentage of samples containing the observed band (Chesnokov and Artemyeva 2015).

AP% locus in a population is a measure of genetic variation. can be calculated using the formula (Ford, 1940):

$$\text{Number of polymorphic loci} / \text{total number of loci} \quad (6)$$

Where the ratio is derived by dividing the number of polymorphic loci by the total number of loci

The **MI** is calculated by multiplying the polymorphism information content value (or predicted heterozygosity) by the effective multiplex ratio (Nagaraju *et al.*, 2001). The formula can be computed using the following formula (Powellet *al.*, 1996):

$$MI = PIC \times EMR \quad (7)$$

EMR (Amiryousefi *et al.*, 2018) is defined as $n_p / (n_p + n_{np})$, where p and np represent the polymorphic and non-polymorphic percentages of the markers, respectively, and PIC is polymorphism information content.

DI has been employed frequently in research on genetic diversity as a measure of genetic variation. It can be computed using the following formula (Shannon and Weaver 1949):

$$DI = - \sum P_i \ln P_i \quad (8)$$

Where p_i is the frequency of a band's presence or absence in a locus for all (Laurentin,2009).

ND is the average number of nucleotide changes per site between two DNA sequences in all possible pairings in the sample population (Nei and Tajima 1981). can be calculated using the formula (Nei and Li 1979):

$$\pi X = n / (n - 1) \sum X_i X_j \pi_{ij} \quad (9)$$

X_i and X_j are the i and j sequences' respective frequencies, π_{ij} is the number of nucleotide changes per nucleotide site between the i and j sequences, and n is the total number of sequences in the sample (Nei and Tajima 1981).

According to those formulas, the PIC value for dominant markers ranges from 0 to 0.5, the PIC value for co-dominant markers ranges from 0 to 1, and I_b ranges from zero to one (Chesnokov and Artemyeva 2015). A locus is called P% when the band is present at a frequency ranging from 5% to 95% (Mukhopadhyay and Bhattacharjee 2016)

or from 95% to higher (Pitoyo and Mulyani 2015); the procedure is more successful when the MI is higher (Chesnokov and Artemyeva 2015); the DI obtained from the data is normally between 1.5 and 3.5; however, it can sometimes exceed 4 (Margalef, 1972) or range from 0 to 5 (Launchbaugh, 2009); He and Ho values range from 0 (no heterozygosity) to nearly 1 (high heterozygosity; a large number of alleles with equal frequency) (Chesnokov and Artemyeva 2015), and ND values range from 0 to 1 (De Vicente *et al.*, 2004).

3. Result

To help researchers analyze data for dominant and co-dominant markers, we designed the Ahmed Bader EL-Deen calc program to be simple to use, save time, and prevent errors. For ease of comprehension, the results are provided in

3.1 Interface of our program

a simple table. Our program offers a simple interface for getting genetic diversity measures as well as supporting researchers in acquiring accurate genetic diversity estimates. Our program is a free and straightforward program for calculating genetic diversity measures that are either dominant or co-dominant (bi-allelic, diploid, and tetraploid). In the case of bi-allelic and diploid species, we can enter the number of variations between 1 to infinity and the number of alleles for each variant between 2 to infinity. But in the case of tetraploids, we can enter the number of variations between 1 to infinity and the number of alleles for only three alleles, in the general, the number of variations refers to the columns in the excel sheet (number of individual or genotypes) and number of alleles refer to rows in excel sheet and also calculate nucleotide diversity.

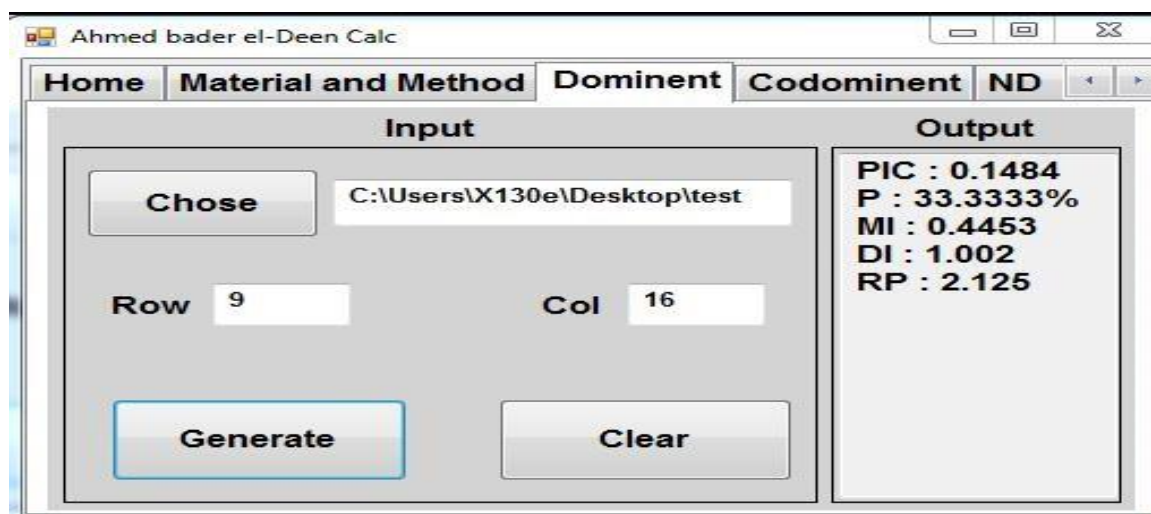


Figure 1: depicts our program's graphical user interface (GUI). In the case of a dominant marker, the user determines the measure of genetic diversity by entering a binary number (matrix) of 0 or 1 into an excel sheet and then determining the number of rows (refer to the number of bands or alleles) and columns (refer to the number of genotypes).

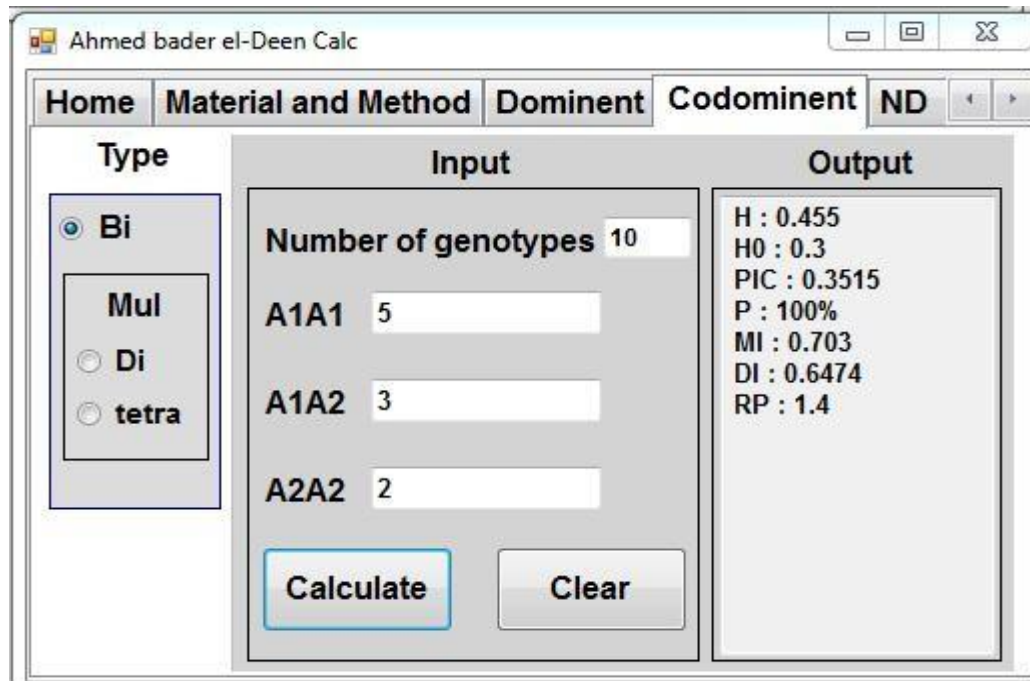


Figure 2: shows the graphical user interface (GUI) of our program; in the case of a co-dominant marker (bi-allelic), the user determines the number of total genotypes and then the number of individuals (the sum of the number of individuals equals the total genotypes of an individual).

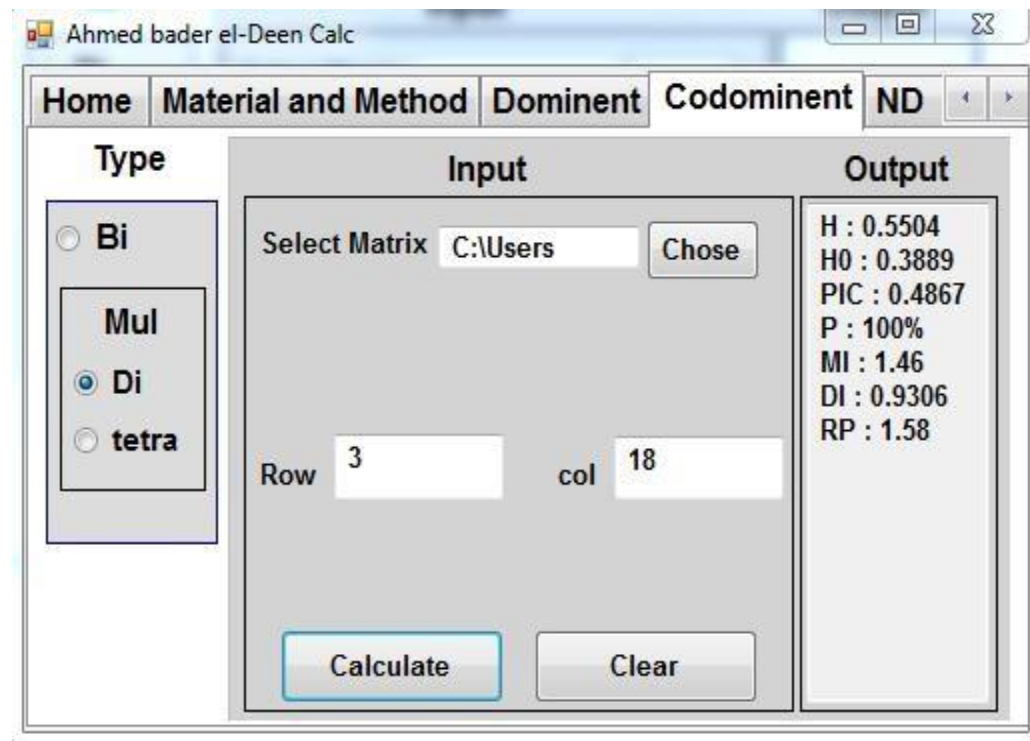


Figure 3: depicts our program's graphical user interface (GUI). In the case of a co-dominant marker (Diploid), the steps are the same as in **Figure 1**.

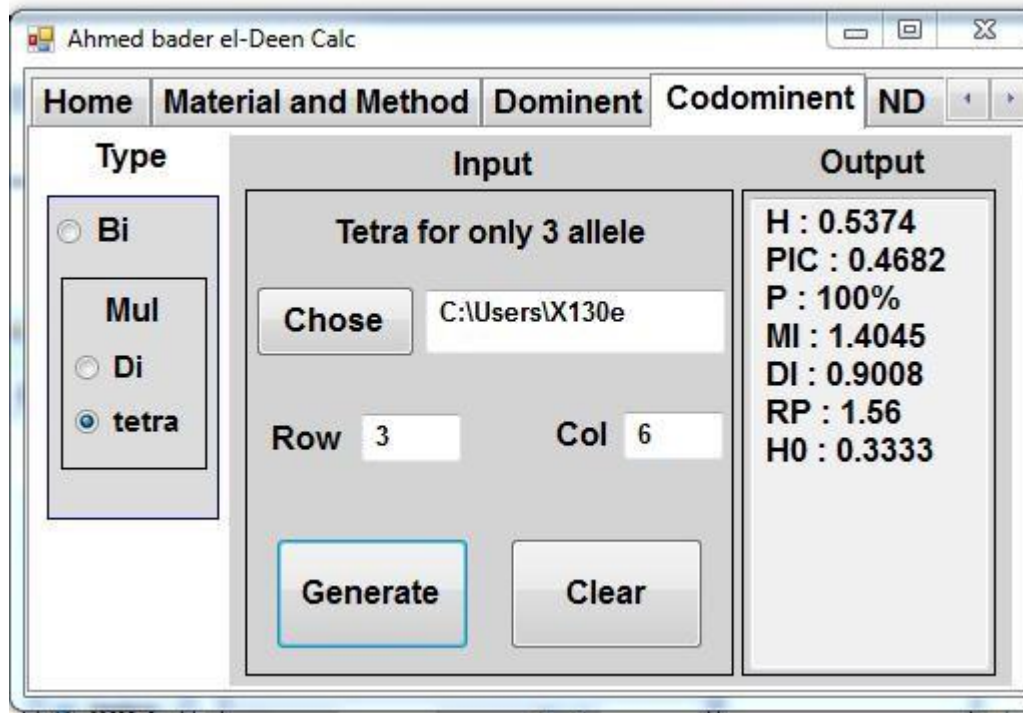


Figure 4: demonstrates the graphical user interface (GUI) of our program; In the case of co-dominant marker (Tetraploid), the steps are the same as in **Figure.1**.

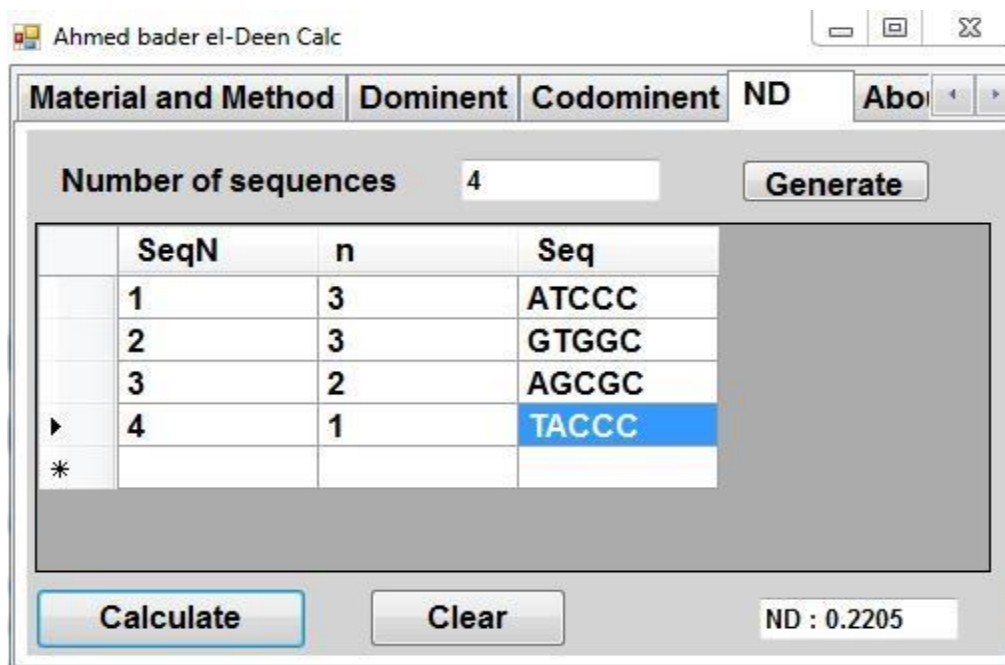


Figure 5: depicts our program's graphical user interface (GUI), which allows the user to enter the number of sequences, the number of polymorphic sites (n), and the sequence by only homologous sequence to calculate nucleotide diversity.

4. Discussion

In this work, we created a program for calculating GD. The application generates a tabular summary of all the data. Despite the lack of software-calculated data for co-dominant markers

(tetraploid for only three alleles), our program can easily and quickly compute data for a co-dominant marker (tetraploid).

Table 1: Comparison of different characteristics of the most frequently used software

Software	Type (Dominant/Co-dominant)	Description
Poly picker	co-dominant markers	This is an online tool for quickly calculating H and PIC to select the number of variations (1–5) only and the number of alleles for each variant (2–20) only.
GenALEX	co-dominant markers	This is calculation software for He and P, but it also has other applications for genetic diversity.
GDA	co-dominant markers	This is He's calculation software, but it also has genetic diversity applications.
Popgene	co-dominant markers	This is calculation software of He and P, but it also has other applications for genetic diversity.
Power Marker	co-dominant markers	This is calculation software of He, PIC, and P, but it also has other applications for genetic diversity.
Cervus	co-dominant markers	This is a calculation program of He and PIC, but it also has other applications for genetic diversity.
Arlequin	co-dominant markers	This is He's calculation software but also has applications for genetic diversity.
GDdom	dominant markers	This is software for calculating PIC, and created an algorithm that accepts a CSV file (Comma delimited)" or "CSV (MS-DOS)" as inputs and produces the tables that appear on the web page as outputs.
Gene calc	dominant markers & co-dominant markers	That is a He and PIC calculation software, but it also has other applications for genetic diversity.

4.1 Validation of data set of Ahmed calc program

Regardless of their intensity, the bands were graded as present (1) or missing (0). The resulting binary data was utilized to calculate PIC, P%, Rp, MI, DI, Ho, and He by using a data set of dominant markers and co-dominant markers. The final result of any molecular marker is a binary

matrix that comprises "1" and "0" for whether or not a band is present; and in the case of a co-dominant marker, a binary matrix should be "translated" into a matrix of genotypes (homozygous or heterozygous) for each individual at each locus.

4.1.1 First Dominant Marker

Table.2: DNA polymorphism data matrix based on SRAP-PCR, where the sequence-related amplified polymorphism (SRAP-3) primer combinations in 16 genotypes and 14 bands (Mahmoud and Abd El-Fatah, 2020).

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0
2	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0
5	0	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0
6	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0
7	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1
9	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1
10	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
11	0	1	0	0	1	0	1	1	1	0	1	0	0	0	0	1
12	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
14	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1

Table 3: Comparison of research (Mahmoud and Abd El-Fatah 2020); Gene Calc and our program to determine genetic diversity parameters by using the matrix in Table.2

Types	Software	PIC	Rp	P%	MI	DI
Dominant	research	0.28	5.25	92.86%	3.66	N/A
	Gen calc	0.2812	N/A	N/A	N/A	N/A
	Ahmed calc	0.2812	5.25	92.8571%	3.6562	3.3799

4.1.2 Second Co-dominant Marker:

4.1.2.1 Bi-allelic

Table 4: shows a data matrix of DNA polymorphism based on the INDEL marker, with the INDEL primer (chr09.0755) combinations in 17 genotypes and 2 alleles (Niihama et al., 2015).

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table.5: Comparison between Gene Calc and our program to determine genetic diversity parameters by using the matrix in Table.4

Types	Software	PIC	Rp	P%	MI	DI	Ho	He
Co-dominant (Bi-allelic)	Gen calc	0.1889	N/A	N/A	N/A	N/A	N/A	0.2112
	Ahmed calc	0.1889	0.48	100%	0.3778	0.3669	Zero	0.2112

4.1.2.2 Diploid:

Table 6: shows a data matrix of DNA polymorphisms based on the STS marker, with sequence tagged sites primer (Sb01) combinations in 5 genotypes and 5 alleles (Perry and Bousquet 1998).

N	1	2	3	4	5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	1	1	1	1	1
5	0	0	0	0	1

Table.7: Comparison of research (Perry and Bousquet 1998), Gene Calc, and our program to determine genetic diversity parameters by using the matrix in Table.6

Types	Software	PIC	Rp	P%	MI	DI	Ho	He
Co-dominant (diploid)	research	N/A	N/A	N/A	N/A	N/A	0.77	N/A
	Gen calc	0.57	N/A	N/A	N/A	N/A	N/A	0.6
	Ahmed calc	0.57	1.6	80%	2.28	1.2275	0.8	0.6

4.1.2.3 Tetraploid:

Table.8: is a data matrix of DNA polymorphism based on the INDEL marker (Settles et al. 2014) used the INDEL primer combinations in 6 genotypes and 3 alleles.

N	1	2	3	4	5	6
1	0	0	0	1	0	1
2	1	0	1	1	1	1
3	0	1	0	1	0	1

Table.9: Our program was used to determine genetic diversity parameters by using the matrix in Table 8

Types	Software	PIC	Rp	P%	MI	DI	Ho	He
Co-dominant (tetraploid)	Ahmed calc	0.4682	1.56	100%	1.4045	0.9008	0.3333	0.5374

4.1.3 Third Data set of Nucleotide diversity:

Sequence	N	Freq. xi
Seq1 : TCC T CGAT T ATT C CAGGGTGCC GATG A AT	5	5/10=0.5
Seq2 : TCC A CGAT T ATT G CAGGGTGCC GATG A AT	2	2/10=0.2
Seq3 : TCC A CGAT C ATT C CAGGGTGCA GATG G AT	1	1/10=0.1
Seq4 : TCC G CGAT T ATTCT CAGGGTGCC GATG A AT	2	2/10=0.2

Figure 6: shows a nucleotide diversity data set with ten people in population X. We looked at one 30-nucleotide sequence for each person and discovered that the sequences differed at five nucleotide places. There are four different sequences for those 30 nucleotides in the population. The number of people (n) who have the specific sequence alternative is listed in the first column. The number of nucleotide differences in each sequence pair within the population is then calculated. For example, $1,2 = 2/30$ denotes that there are two nucleotide differences between sequences 1 and 2. Then, for the total population, we compute X. The number obtained from the sequence examined in a sample of ten individuals is 0.037, or 3.7 percent nucleotide diversity (De Vicente *et al.*, 2004).

Table 10: Comparison for calculating nucleotide diversity between the manual method and our program by using Figure.6

	Manual Method	Ahmed Calc
Nucleotide diversity	0.037	0.037

5. Conclusion

Our software was created depending on how well it handled the challenges the researcher faced. The results must be interpreted in light of the morphological traits. Biological traits of the species under study must obviously be taken into consideration when evaluating genetic determinants because, in different species, the same parameter may vary not only in phylogeny but also in ontogeny (Chesnokov and Artemyeva 2015).

6. Acknowledgements

We are extremely grateful to the Institute of Molecular Biology Research & Studies for their assistance.

7. Conflict of interest

The author disclosed no possible conflicts of interest (s).

8. Authors contribution

I ELA, A.E.M and S.S.M: reviewing and editing the manuscripts; A ELD: designing and conducting the program and writing manuscripts.

9. Funding

This research received no specific funding from government, commercial, or non-profit funding entities.

10. References

- Abuzayed, M., El-Dabba, N., Frary, A., & Doganlar, S. (2017). GDdom: an online tool for calculation of dominant marker gene diversity. *Biochemical genetics*, 55(2), 155-157.
- Amiryousefi, A., Hyvönen, J., & Poczai, P. (2018). iMEC: Online marker efficiency calculator. *Applications in plant sciences*, 6(6), e01159.
- Botstein, D. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*, 32, 314-331.
- Chesnokov, Y. V., & Artemyeva, A. M. (2015). Evaluation of the measure of polymorphism information of genetic diversity. , (5 (eng)), 571-578.
- Collard, B. C., & Mackill, D. J. (2009). Start codon targeted (SCoT) polymorphism:

a simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant molecular biology reporter*, 27(1), 86.

De Vicente, M. C., Lopez, C., & Fulton, T.

(2004). Genetic diversity analysis with molecular marker data: learning module. *International Plant Genetic Resources Institute (IPGRI)*.

Etminan, A., Pour-Aboughadareh, A.,

Mohammadi, R., Ahmadi-Rad, A., Noori, A., Mahdavian, Z., & Moradi, Z. (2016). Applicability of start codon targeted (SCoT) and inter-simple sequence repeat (ISSR) markers for genetic diversity analysis in durum wheat genotypes. *Biotechnology & Biotechnological Equipment*, 30(6), 1075-1081.

Excoffier, L., Laval, G., & Schneider, S. (2005).

Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, 1, 117693430500100003.

Ford, E. B. (1940). Polymorphism and taxonomy.

The new systematics. *Ed. J. Huxley*, 493-513.

Guo X, Elston RC (1999) Linkage informative

content of polymorphic genetic markers. *Hum Hered*49:112–118.

Kalinowski, S. T., Taper, M. L., & Marshall, T.

C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16(5), 1099-1106.

Launchbaugh, K. (2009). Principles of vegetation

measurement and assessment and ecological monitoring and analysis. *University of Idaho Website* https://www.webpages.idaho.edu/veg_measure.

Lewis, P. O. (2001). Genetic Data Analysis: A

computer program for the analysis of allelic data. Version 1.0 (d16c). http://lewis.web.UConn.edu/lewis_home/software.HTML.

Li, G., & Quiros, C. F. (2001). Sequence-related

amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. *Theoretical and applied genetics*, 103(2-3), 455-461.

Liu, B. H. (1998). Statistical genomics: Linkage,

mapping, and QTL analysis. CRC Press, Boca Raton, Florida, USA.

Liu, K., & Muse, S. V. (2005). Power Marker: an

integrated analysis environment for genetic marker

analysis. *Bioinformatics*, 21(9), 2128-2129.

Mahmoud, A. F., & Abd El-Fatah, B. E. (2020).

Genetic diversity studies and identification of molecular and biochemical markers associated with fusarium wilt resistance in cultivated faba bean (*Vicia faba*). *The plant pathology journal*, 36(1), 11.

Margalef, R. (1972). Homage to Evelyn Hutchinson or why is there an upper limit to diversity? *Transitions in Connection Academic Arts and Science* 44: 211-235.

Miks, G. C. S., Bińkowski, J., & Calc, G. (1955). Materials & Methods.

Mukhopadhyay, T., & Bhattacharjee, S. (2016). Genetic Diversity: Importance and Measurements. *Conserving Biological Diversity: A Multiscaled Approach. Research India Publications, New Delhi*, 251-295.

Nagaraju, J., Reddy, K. D., Nagaraja, G. M., & Sethuraman, B. N. (2001). Comparison of multilocus RFLPs and PCR-based marker systems for genetic analysis of the silkworm, *Bombyx mori*. *Heredity*, 86(5), 588-597.

Nei, M., & Li, W. H. (1979). A mathematical model for studying genetic variation in

terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10), 5269-5273.

Nei, M., & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics*, 97(1), 145-163.

Niihama, M., Mochizuki, M., Kurata, N., & Nonomura, K. I. (2015). PCR-based INDEL markers co-dominant between *Oryza sativa*, japonica cultivars, and closely-related wild *Oryza* species. *Breeding Science*, 65(4), 357-361.

Peakall, R. O. D., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular ecology notes*, 6(1), 288-295.

Perry, D. J., & Bousquet, J. (1998). Sequence-tagged-site (STS) markers of arbitrary genes: development, characterization, and analysis of linkage in black spruce. *Genetics*, 149(2), 1089-1098.

Pitoyo, A., & Mulyani, S. (2015). Assessment of genetic diversity among soursop (*Annona muricata*) populations from Java, Indonesia using RAPD markers. *Biodiversitas: Journal of Biological Diversity*, 16(2), 247-253.

- Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. (1996). The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2: 225–238.
- Prevost, A., and M. J. Wilkinson. (1999). A new system of comparing PCR primers.
- 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.
- Schulman, A. H. (2007). Molecular markers to assess genetic diversity. *Euphytica*158(3),313-321.
- Settles, A. M., Bagadion, A. M., Bai, F., Zhang, J., Barron, B., Leach, K., ... & Braun, D. M. (2014). Efficient molecular marker design using the MaizeGDB Mo17 SNPs and Indels track. *G3: Genes, Genomes, Genetics*, 4(6), 1143-1145.
- Shannon, C. E., & Weaver, W. (1949). The mathematical theory of information. *Urbana: University of Illinois Press*, 97(6), 128-164.
- Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids research*, 17(16), 6463-6471.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. V. D., Hornes, M., & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic acids research*, 23(21), 4407-4414.
- Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research*, 18(22), 6531-6535.
- Yeh, F. C., Yang, R. C., Boyle, T., Ye, Z. H., & Mao, J. X. (1999). POPGENE, version 1.32: the user-friendly software for population genetic analysis. *Molecular biology and biotechnology center, University of Alberta, Edmonton, AB, Canada*.