

Assessment of Genetic Diversity in *Pisum sativum* L. Using Nucleotide Sequences of Meristematic Genes

Yasna Abbasi

11th-Grade Student, Allameh Iqbal High School, District 4, Shiraz, Iran (fraei181158@gmail.com)*

Sasan Mohsenzadeh

Professor, Department of Biology, Shiraz University of Medical Sciences, Shiraz, Iran

Sara Bostanian

M.Sc. in Developmental–Molecular Biology, Dr. Hesabi Research Center, District 4, Shiraz, Iran

Abstract

Pea (*Pisum sativum* L.), as one of the most important legumes, plays a pivotal role in ensuring food and nutritional security in many countries. The genetic diversity of this species serves as a fundamental basis for breeding programs, conservation of genetic resources, and the development of cultivars resistant to biotic and abiotic stresses. This study investigates the genetic diversity of pea using nucleotide sequences of meristematic genes extracted from individuals of this species available in the NCBI¹ database. Precise characterization of this genetic diversity—particularly within meristematic genes that play a key role in plant growth and development—is of particular significance. The primary objective of the study is to determine the extent and pattern of genetic diversity within selected meristematic

Introduction

Legumes play a significant role in meeting human nutritional needs and represent a primary source of protein in developing countries. Therefore, they can make a substantial contribution to food production in these regions[1]. Evaluating the status of legume production is essential due to their importance in global nutrition and their role in designing cropping patterns. Pea (*P. sativum* L.) is an annual, bush-type plant belonging to the Fabaceae family, well adapted to cold climates, and cultivated during the

genes in pea genotypes using sequencing data and bioinformatic analyses to support breeding improvement programs. In this study, meristematic gene sequences from multiple genotypes were collected, aligned using bioinformatic tools, and analyzed through phylogenetic tree construction to determine evolutionary relationships and genetic proximity among samples, thereby clarifying genetic structure and relatedness. Results showed that out of a total of 4,144 nucleotides, 52.31% similarity and 47.69% genetic variation were observed among sequences, indicating a high level of genetic diversity in meristematic genes. The phylogenetic structure revealed that some genotypes exhibited greater genetic proximity while others displayed wider evolutionary divergence, indicating the presence of distinct genetic groups. The high degree of genetic diversity observed can provide a suitable basis for identifying superior genotypes in terms of meristematic growth, stress tolerance, and agricultural performance. These findings underscore the importance of sequencing data and bioinformatic analyses for improving our understanding of genetic structure, enhancing pea breeding programs, and developing strategies for the conservation and utilization of genetic resources in similar crop species.

Keywords: *Pisum sativum*, Genetic Variation, Nucleotides, Meristem Tissue, Bioinformatics cold seasons in many countries from winter to early summer depending on the cultivation region. This plant contains high levels of crude protein and starch and is also rich in energy[2, 3]. As one of the most important legumes contributing to global food security, peas play a vital role in both nutrition and agriculture. Assessment of genetic diversity in this species forms the foundation for breeding programs, conservation of genetic resources, and the development of cultivars resistant to biotic and abiotic stresses[4]. Previous studies using molecular markers such as ISSR² and SSR³ have

¹ National Center for Biotechnology Information

² Inter Simple Sequence Repeat

³ Simple Sequence Repeat

identified considerable diversity within pea germplasm, highlighting the need for modern bioinformatic approaches to achieve deeper analysis of this variation[5]. Although these methods have provided valuable information, they have been insufficient for precisely identifying genetic differences at the level of key genes.

Meristematic genes, due to their essential roles in plant growth and development and in the regulation of cell division processes, represent a valuable target for identifying genetic diversity and selecting superior genotypes. Genetic diversity, as the foundation of breeding programs and plant adaptation to environmental conditions, is crucial for maintaining production stability and developing stress-resistant cultivars. Advances in sequencing technologies and bioinformatic tools have made it possible to investigate genetic structure and diversity at the level of specific genes with greater accuracy and detail. Nevertheless, insufficient information is currently available regarding the extent and pattern of genetic diversity in meristematic genes across pea genotypes, representing a significant gap in the genetic study of this species.

Therefore, it is essential to precisely examine the extent and pattern of genetic diversity in pea using sequencing data from meristematic genes and bioinformatic analyses, so that these data may be applied in breeding programs, identification of resistant genotypes, and conservation of genetic resources. This study addresses this need by analyzing the genetic diversity of meristematic genes in pea using nucleotide data and advanced bioinformatic tools[6].

Genetic diversity is a level of biodiversity that encompasses all genetic characteristics within the genetic structure of a species and refers to the heterogeneous combination of genes that elicit

different responses to environmental conditions. Genetic diversity is considered the most critical factor for the survival of organisms, including plants, under changing environmental conditions and biotic stresses. Awareness of the extent of hereditary variation and the genetic relationships among germplasm constitutes a primary requirement for the improvement of plant species. Indeed, genetic diversity forms the foundation of breeding studies in plant species. Given the role of genetic diversity in advancing the objectives of plant breeding programs, its identification through molecular and non-molecular methods is of significant importance[7].

Genetic diversity also pertains to intraspecific variation, including differences in DNA, which is essential for the adaptability and resilience of species and ecosystems. It plays a crucial role in enabling species to adjust to changing environments and resist environmental disturbances and climate change. Genetic diversity is also essential for successful ecological restoration and food security, as it supports the ability of species to survive under diverse conditions[8]. Maintaining a balance between conserving genetic diversity and achieving specific breeding objectives remains a key focus in conservation and agricultural practices. Genetic diversity results from the accumulation and combination of genetic mutations, which generate changes in genes and ultimately increase biological variation within populations[9].

A genetic mutation is a permanent alteration in one or more bases in the DNA sequence, which can affect the structure and function of the resulting protein. Such changes may occur as deletions, insertions, or substitutions of nucleotides within a gene, contributing to genetic diversity in living organisms. This phenomenon plays a significant role in evolution and the manifestation of diseases

and is therefore of considerable importance in genetic studies [10].

Nucleotides are organic molecules composed of a nucleoside and a phosphate group. As monomeric units, they form nucleic acid polymers, including deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), both of which are essential biomolecules in all forms of life on Earth[11]. At the cellular level,

nucleotides play a central role in metabolism, providing energy for many cellular functions through the chemical energy stored within these molecules[12]. Nucleotides are linked via sugar-phosphate bonds to form the DNA and RNA chains (Figure 1). They constitute the foundation of cell division and growth processes in meristems, and without them, the vital activities of these actively dividing regions would not be possible [13].

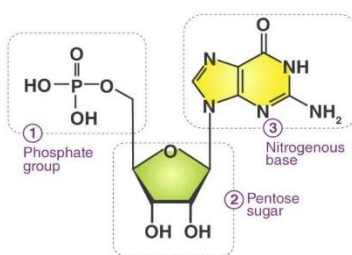


Figure 1 – Nucleotide Structure

The meristem is a plant tissue composed of embryonic cells with rapid division, responsible for plant growth and development. Meristematic cells, often referred to as plant stem cells, are undifferentiated cells located in specific regions of plants known as meristems. These cells remain in a meristematic state and enable the indeterminate growth of vascular plants. Meristematic cells are small, polygonal, and possess a large central nucleus. They typically lack vacuoles or contain small vacuoles and do not exhibit tissue specialization. These cells have thin primary cell walls without secondary walls and contain protoplasts. Additionally, they possess numerous mitochondria and undifferentiated plastids. In plants, cell division occurs exclusively within meristems[14, 15].

The function of meristems is tightly regulated by various genes. These genes, which play a key role

in maintaining and regulating meristem activity, are known as meristematic genes. They are involved in multiple processes, including maintaining meristem cell identity, regulating cell division, and controlling cell differentiation[13].

Sequencing data of meristematic genes available in NCBI provide structural and evolutionary information and allow the identification of mutation patterns and similarities among genotypes through bioinformatic and phylogenetic tools. These approaches not only reveal diversity within conserved meristematic sequences but also determine genetic relationships among different ecotypes with greater accuracy than conventional molecular markers[4].

The National Center for Biotechnology Information (NCBI) is a branch of the United States National Library of Medicine. It manages tools and databases such as Entrez, BLAST, PubMed, and

GenBank. NCBI is among the world's most important biological data repositories, operating under the National Institutes of Health (NIH)⁴, with the mission to advance scientific research, education, and data sharing in molecular biology, genetics, bioinformatics, and medicine. Integrating NCBI data with bioinformatic analyses enables the identification of superior genotypes carrying beneficial alleles associated with meristem growth, stress tolerance, and agricultural performance. Such analyses complement SSR- and ISSR-based studies and provide a robust genetic foundation for breeding programs[16].

Using NCBI data and performing sequence alignment allows researchers to more precisely identify alleles related to meristem growth and stress tolerance, optimize the selection of superior genotypes based on genetic similarities and differences, and integrate sequencing data with SSR and ISSR results for more comprehensive analyses[17].

In bioinformatics, alignment refers to the process of comparing and matching biological sequences such as DNA, RNA, or proteins to identify conserved regions or genetic variations. This process aids in detecting beneficial alleles, mutations, and genetic differences, which is critical for genotype analysis and identifying genes associated with important agronomic traits. Consequently, sequence alignment is a key bioinformatic tool that facilitates the integration of NCBI data, strengthens the genetic basis for crop improvement, and enables the selection of superior genotypes [18].

For accurate and efficient multiple sequence alignment, the online tool Multialin uses advanced algorithms to compare multiple DNA, RNA, or protein sequences simultaneously and presents results graphically. This greatly assists researchers

in identifying conserved regions, functional motifs, and evolutionary relationships[19].

Literature Review

Next-generation sequencing (NGS)⁵ based studies have enabled faster and more precise identification of genetic diversity in meristematic genes. For instance, Kreplak et al. (2019), using whole-genome sequencing of pea, identified genetic regions associated with meristem growth traits and disease resistance, emphasizing the importance of utilizing genomic data for the genetic improvement of this species. Additionally, studies on meristem gene expression under varying environmental conditions have demonstrated that the regulation of these genes plays a critical role in the adaptation of pea to biotic and abiotic stresses[20]. These findings underscore the importance of detailed analysis of meristem gene sequences and expression in breeding programs aimed at developing more resilient cultivars. Furthermore, comparative analyses of meristem gene sequences in pea and other legumes have shown that certain conserved motifs and regions can serve as molecular markers for selecting superior genotypes. Overall, recent studies highlight the value of integrating genomic and bioinformatic data to identify and exploit genetic diversity in meristem genes of pea, positioning these approaches as powerful complements to traditional plant genetics methods[21].

Materials and Methods

Nucleotide sequence data of meristematic genes present in linear mRNA of *Pisum sativum* L. were collected from the NCBI plant databases. The degree of similarity and variation among these sequences was analyzed using the Multialin

⁴ National Institutes of Health

⁵Next Generation Sequencing

website. Percent similarity and divergence were calculated (Figure 2). The reference identifiers for the collected nucleotide sequences of meristematic genes are as follows:

- 1(FG534345.1): Pisum sativum Shoot apical meristem ESTs
- 2(FG534344.1): Pisum sativum Shoot apical meristem ESTs
- 3(FG534316.1): Pisum sativum Shoot apical meristem ESTs

- 4(FG534298.1): Pisum sativum Shoot apical meristem ESTs
- 5(FG534287.1): Pisum sativum Shoot apical meristem ESTs
- 6(FG534215.1): Pisum sativum Shoot apical meristem ESTs
- 7(FG534210.1): Pisum sativum Shoot apical meristem ESTs
- 8(FG534199.1): Pisum sativum Shoot apical meristem ESTs

Results

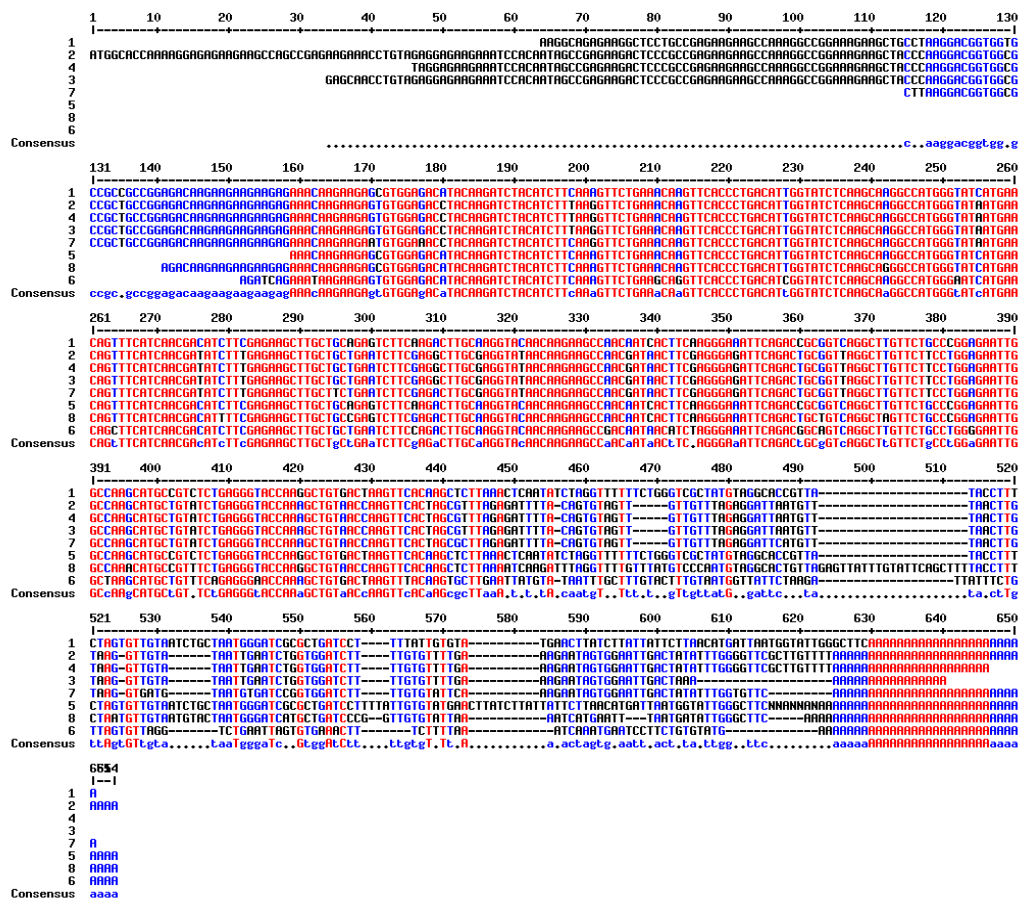


Figure 2 – Multiple Sequence Alignment of Pea Meristematic Gene Nucleotide Sequences

The genetic diversity of Pisum sativum L. was analyzed through a detailed multiple sequence alignment (MSA) of meristematic gene nucleotide sequences across eight selected genotypes (Figure 2). This alignment, performed using precise

bioinformatic algorithms, provides an essential foundation for both quantitative and qualitative characterization of conservation patterns and variability in these critical plant sequences.

Statistical analysis and position-by-position examination of the aligned sequences in Figure 2 allowed for the extraction of precise genetic diversity indices. The calculations were based on a total of 4,144 valid nucleotides aligned across all genotypes.

The visual representation of the alignment in Figure 2 reveals spatial patterns of sequence conservation and various nucleotide mutations according to bioinformatic color-coding conventions indicating

degrees of conservation: red represents complete similarity, black indicates complete divergence, and blue denotes partial or relative similarity.

Nucleotides colored red at a specific position across all eight genotypes were counted as conserved nucleotides, resulting in the identification of 2,168 conserved nucleotides. The genetic conservation rate (C) was calculated using Formula 1:

$$C(\text{Percentage of similarity}) = \frac{(2168) \text{ Number of identical}}{(4144) \text{ Total number of aligned nucleotides}} \times 100 = 52.31\%$$

Formula 1 – Calculation of Genetic Conservation Rate

This rate indicates the presence of strongly conserved regions within the structure of meristematic genes, whose integrity must be maintained under natural selection pressure for proper biological function.

Variable or differing nucleotides (black indicating complete difference and blue indicating partial

similarity) refer to positions where, in at least one genotype, the nucleotide differs from the others (arising from substitution or insertion/deletion mutations). These variations encompassed a total of 1,976 nucleotides. The genetic diversity rate (D) was calculated using Formula 2:

$$D(\text{Percentage of variation}) = \frac{(1976) \text{ Number of variable}}{(4144) \text{ Total number of aligned nucleotides}} \times 100 = 47.69\%$$

Formula 2 – Calculation of Genetic Diversity Rate

The observed genetic diversity rate of 47.69% confirms that meristematic genes possess a broad and significant genetic reservoir at the nucleotide level. The consensus sequence, displayed at the end of the alignment, represents the nucleotide at each position with the highest frequency across the eight genotypes. This sequence directly confirms

level among the examined genotypes, with high potential for novel allelic variants.

conserved regions (52.31%) and provides a hypothetical sequence with maximal evolutionary stability for the meristematic gene. It serves as a

primary reference for identifying nucleotide variants in other genotypes.

The color patterns in Figure 2 directly illustrate mutation patterns. The predominance of red in many sequence regions clearly indicates complete nucleotide conservation, where nucleotides fully match the consensus sequence. Sudden shifts from red to blue or black along the sequence represent variable regions; in this color scheme, blue and black do not necessarily indicate specific nucleotides but denote positions where the nucleotide deviates from the consensus (most frequent) nucleotide. These color changes provide evidence of single-nucleotide substitutions, which are the principal contributors to the genetic diversity observed in meristematic genes.

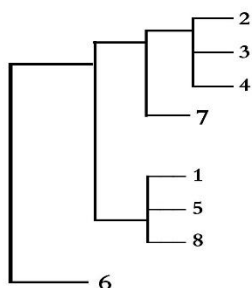
Notable evidence of insertion/deletion mutations is visible as dashes (-) in the sequences of some genotypes. These gaps indicate historical events of nucleotide insertion or deletion during the evolution of those genotypes.

Comprehensive analysis of Figure 2, considering the high genetic diversity rate of 47.69% and the observed color patterns, demonstrates that the meristematic genes of this species harbor a broad and dynamic genetic reservoir. Genotypes such as Sample 6, which exhibit the greatest deviation from the consensus sequence, represent valuable sources of genetic variation for selection in breeding programs. These findings provide a solid scientific basis for identifying gene alleles associated with improved agronomic traits and enhanced tolerance to environmental stresses.

Table 1 – Summary of Nucleotide Sequencing Results for Pea Meristematic Genes

Total Nucleotides	Conserved Nucleotides	Variable Nucleotides	Percent Similarity(%)	Percent Variation(%)	Database	Analysis Tool
4144	2168	1976	52.31%	47.69	NCBI	multialin

Phylogenetic Tree



Interpretation of the Phylogenetic Tree

The phylogenetic tree illustrates the evolutionary relationships among the eight *Pisum sativum* L. genotypes, labeled 1 through 8. The branching structure (clades) indicates which genotypes are more closely related and which are more evolutionarily distant.

Three genotypes 1, 5, and 8 form a closely related clade, demonstrating the highest genetic similarity, with their divergence originating from a relatively recent common ancestor. Another clade is formed by genotypes 2, 3, and 4. Within this group, genotypes 2 and 3 are the most closely related, and together they cluster with genotype 4, which is slightly more distant but still genetically close to 2 and 3.

Genotype 7 occupies a separate branch near the 2, 3 and 4 clade, indicating that it shares some genetic similarity with this group but exhibits a greater

genetic distance relative to the internal members of the clade.

Genotype 6 is positioned on the most basal branch, showing the greatest genetic distance from all other genotypes. This placement indicates that it diverged earliest from the common ancestor of the group and is the most evolutionarily distinct among the eight genotypes.

Overall, the phylogenetic tree, based on nucleotide sequence alignment data from Figure 2, effectively represents evolutionary relationships. Proximity of branches reflects genetic similarity, while branch separation indicates greater genetic divergence and earlier evolutionary separation. This structural insight is particularly valuable for plant breeding programs, as selecting genetically distant genotypes, such as genotype 6, can maximize genetic variation and enhance the potential for developing improved cultivars.

Table 2 – Summary of Evolutionary Relationships

Group / Genotypes	Description of Genetic Relationship
2, 3, 4	Closest relatives: 2 and 3 are most closely related, then cluster with 4
7	Close to the 2, 3, 4 group but with greater genetic distance
1, 5, 8	Closely clustered clade, relatively independent from other groups
6	Most distant genotype relative to all others, occupying the earliest-diverging branch

Discussion and Conclusion

The findings of this study, based on meristematic gene sequencing data and advanced bioinformatic

analyses, provide a precise depiction of the structure and extent of genetic diversity among

Pisum sativum L. genotypes. The results indicated that out of 4,144 nucleotides analyzed, 52.31% were conserved and 47.69% exhibited genetic variation among the genotypes, reflecting a high level of diversity in key genes associated with plant growth and development. This level of genetic variation, encompassing nearly half of the total sequence, has significant implications for breeding programs, identification of superior genotypes, and the preservation of genetic resources.

The inclusion of a phylogenetic tree in the analysis allowed for a deeper examination of evolutionary relationships and the genetic structure of the pea population. The tree's branching pattern revealed that some genotypes (e.g., 1, 5, and 8) clustered closely, demonstrating the highest genetic similarity, whereas others (e.g., genotype 6) formed a distinct branch with the greatest genetic distance from the rest. These patterns indicate the existence of distinct genetic subgroups within the pea population, which can be strategically used for parent selection in breeding programs to enhance genetic variation in future generations.

Previous studies primarily using molecular markers such as SSR and ISSR have reported substantial diversity in pea germplasm[5, 6]. However, the present study, focusing on meristematic genes and nucleotide sequence analysis, revealed genetic differences at a more precise and targeted level. Furthermore, constructing a phylogenetic tree based on sequencing data provides higher resolution in identifying evolutionary relationships and genetic structure compared to traditional methods, as also highlighted in recent studies[21].

From a practical perspective, the observed genetic diversity and the identification of distinct genetic clusters offer valuable opportunities for plant breeders to select suitable parents for developing new cultivars with desirable traits, including

enhanced meristematic growth, increased resistance to biotic and abiotic stresses, and improved agronomic performance. Additionally, the data generated in this study can serve as a foundation for the targeted conservation and management of pea germplasm.

Overall, this research demonstrates that integrating genomic sequencing data with bioinformatic and phylogenetic analyses constitutes a powerful approach for identifying mutation patterns and exploiting genetic diversity in crop species. The preservation and strategic utilization of this genetic variation are crucial not only for improving the yield and stability of pea production but also for ensuring food security and adaptation to climate change. Sequencing and bioinformatic analyses provide a precise tool for detecting mutation patterns and genetic differences, offering greater accuracy and depth than traditional methods.

A comprehensive understanding of meristematic gene function requires not only sequence analysis but also knowledge of the conditions under which these genes are active or inactive. RNA sequencing (RNA-seq) enables the analysis of gene expression patterns in meristematic tissues under various environmental conditions such as drought, salinity, heat, or cold. This approach allows the identification of genes that play key roles in environmental adaptation and can serve as targets for future genetic improvement. In recent years, integrating biological data with machine learning algorithms has emerged as a novel approach in molecular biology. These algorithms can uncover hidden patterns among genetic information, gene expression, and phenotypic traits, thereby accelerating and refining the selection of resilient and high-performing genotypes. Native genetic resources often harbor hidden genetic diversity and natural adaptations to local climates that are less common in domesticated varieties. Comparing

meristematic genes in wild and native Iranian peas with cultivated genotypes can reveal valuable alleles that have not yet been utilized and hold substantial potential for plant breeding. To increase the precision of early-stage genotype selection in breeding programs, gene-specific molecular markers based on meristematic genes can be

designed for Marker-Assisted Selection (MAS). These markers, directly derived from meristematic gene sequences, enable targeted and scientific selection, allowing breeders to identify and choose superior genotypes without the need for prolonged phenotypic evaluation.

References

- 1-Bagheri E, Nezami A, Porsa H. An analysis of pulse research strategies in Iran based on the first national pulse symposium approaches. *Journal of Field Crop Research*. 2006;4(1):1-13.
- 2-Santalla M, Amurrio JM, Deron AM. Symbiotic interaction between *Rhizobium leguminosarum* strains and elite cultivars of *Pisum sativum* L. *Journal of Agricultural Science*. 2001;187(1):59068.
- 3-Abbasi A, Nakhzari Moghadam A, Pirdashti H, Gholamalipour Alamdari E. Effect of different garlic (*Allium sativum*) and pea (*Pisum sativum*) intercropping patterns on weed population indices. *Journal of Weed Science*. 2017;12(2):222-235.
- 4-Samimi Fard R, Hagh Nazari A, Najafi J, Mardi M, Shahnejat Bushehri AA. Evaluation of genetic diversity in pea (*Pisum sativum* L.) germplasm using SSR markers. 4th National Biotechnology Conference. 1384.
- 5-Naderi H, Shakarpour M, Asghari A, Kanouni H, Esfandiari AA. Assessment of genetic diversity in pea lines using ISSR molecular markers. *Iranian Journal of Crop Sciences*. 1393;45(4):509-519.
- 6-Kreplak J, Sempéré G, Wu J, Zhang C, Bendahmane A, Šimoník J, et al. A reference genome for pea provides insight into legume genome evolution. *Nature Genetics*. 2019;51(9):1411-1422.
- 7-Sadeghi R. Genetic diversity. 1399. Available from: <http://agrinet.ir>
- 8-Kershaw F. Protecting genetic diversity to benefit nature and society. 2024.
- 9-Rauf S, Khan A, Zafar MM, Imtiaz B, Manzoor H, Khan Z. The molecular basis of genetic diversity and its role in crop improvement. *Front Plant Sci*. 2019;10:14.
- 10-Lodish H, Berk A, Kaiser CA, Krieger M, Bretscher A, Ploegh H, et al. *Molecular cell biology*. 9th ed. New York: W.H. Freeman; 2021.
- 11-Zaharevitz DW, Anderson LW, Malinowski NM, Hyman K, Strong JM, Cysyk RL. Contribution of de-novo and salvage synthesis to the uracil nucleotide pool in mouse tissues and tumors in vivo. *Eur J Biochem*. 1992;210(1):293-296.
- 12-Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular biology of the cell*. 4th ed. New York: Garland Science; 2002. p. 120-121.
- 13-Taiz L, Zeiger E, Møller IM, Murphy A. *Plant physiology and development*. 7th ed. Sunderland: Sinauer Associates; 2018.
- 14-Fletcher JC. Shoot and floral meristem maintenance in *Arabidopsis*. *Annual Review of Plant Biology*. 2002;53:45-66.
- 15-Galun E. *Plant patterning: Structural and molecular genetic aspects*. Singapore: World Scientific Publishing; 2007.
- 16-Bagheri M, Tahmasebi Z, Geravandi M. Genetic diversity and relationships between some agronomic traits of Desi chickpea genotypes (*Cicer arietinum* L.) under autumn sowing conditions. *Journal of Crop Breeding*. 2024;16(3):114-124.
- 17-Kamali S, Singh A. Genomic and transcriptomic approaches to developing abiotic stress-resilient crops. *Agronomy*. 2023;13(12):2903.
- 18-Ahmad MF, Khaliq N, Khan M, Qadri T. Study on advancement in bioinformatics for crop improvement: integrating genomics, phenomics, and machine learning.

BOHR Journal of Biocomputing and Nano Technology.
2023;2(1):19-30.

19-Nayak D, Behera BK, Mishra P, Nayak RK. An enhanced web-based tools for multiple sequence alignment: A comparative approach. Journal of Bioprocessing and Biotechnology. 2023;13(2):1-8.

20-Chen Y, Wang Y, Ding Y, Wang J, Yu Y. Transcriptome analysis of shoot apical meristem in response to drought stress in pea (*Pisum sativum* L.). Plant Physiology and Biochemistry. 2021;162:1-10.

21-Yildiz M, Yildirim B, Aydin S. Comparative analysis of meristem gene sequences in legumes: Implications for marker development and breeding. Plant Mol Biol Rep. 2022;40(3):512-524.