

Evaluation of Population Structure in Some Bread and Durum Wheat Genotypes Using SNP Markers and PCA and DAPC Methods

Hossein Abdi¹, Hadi Alipour^{2,*}, Iraj Bernousi² and Jafar Jafarzadeh³

1- Ph.D. Student, Department of Plant Production and Genetics, Faculty of Agriculture, Urmia University, Urmia, Iran

2- Associate Professor, Department of Plant Production and Genetics, Faculty of Agriculture, Urmia University, Urmia, Iran

3- Assistant Professor, Dryland Agricultural Research Institute (DARI), Agriculture Research, Education and Extension Organization (AREEO), Maragheh, Iran

*Corresponding author ✉: ha.alipour@urmia.ac.ir

Citation: Abdi, H., Alipour, H., Bernousi, I. and Jafarzadeh, J. (2023). Evaluation of population structure in some bread and durum wheat genotypes using SNP markers and PCA and DAPC methods. *Plant Genetic Researches*, **10(1)**: 95-110. <http://dx.doi.org/10.22034/pgr.10.1.6>

(Received: April 2, 2023; Final Revised: July 29, 2023; Accepted: August 8, 2023; Published online: September 21, 2023)

Extended abstract

Introduction

Wheat is the most important crop in the world, and the evaluation of its population structure is essential for understanding diversity patterns, choosing suitable parents for mating, accurate identification of genomic regions controlling traits, and evolutionary and kinship relationship studies. Until now, various statistical methods have been introduced to evaluate the genetic structure of the population. Principal component analysis (PCA) and discriminant analysis of principal component (DAPC) are important techniques for evaluating the population structure using molecular markers. PCA is a tool whose purpose is to reduce the complexity of the data and summarizes the diversity in the population, including the diversity between groups and within groups, in a number of components. While, DAPC is a method that maximizes the diversity between groups, the aim of this research was to compare the PCA and DAPC approaches in determining the population structure of tetraploid and hexaploid Iranian wheat genotypes.

Materials and methods

In total, 383 Iranian wheat genotypes, including 370 hexaploid accessions (100 cultivars and 270 landraces), and 13 tetraploid genotypes were studied. These genotypes were previously genotyped using the genotyping by sequencing (GBS) technique, where 16,270 SNP markers were used for population structure analysis. Analysis of molecular variance (AMOVA), genetic differentiation index (PhiPT), and fixation index (Fst) were performed between the groups using GenAlex software. Also, PCA and DAPC analysis were performed in adegenet R-package. Finally, various statistics such as observed heterozygosity (Ho), expected heterozygosity (Hs), expected heterozygosity in random mating (Ht), corrected Ht (Htp), gene diversity among samples (Dst), corrected Dst (Dstp), fixation index (Fst), corrected Fst (Fstp), and inbreeding coefficient (Fis) were estimated in adegenet R-package.

Results and discussion

AMOVA results showed that 24.12% and 75.88% of the calculated variation was found between and within groups, respectively. Low diversity between populations can be explained by high gene flow among genotypes. Between the two hexaploid subpopulations, the Fst statistic was low (0.1544), while this index was significantly



©2023 The author(s). This is an open access article distributed under [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source.

higher between the hexaploid and tetraploid genotypes. This can be explained by the fact that the tetraploid wheat genotypes lack the D genome. The estimation of different statistics for three wheat genomes showed that compared to A and B genomes, genome D had the lowest number of SNP markers and the lowest values of H_o , H_s , D_{st} and F_{st} , but D genome had a high F_{is} value. Chromosomes 5B, 7B and 6A had the highest expected heterozygosity with values of 0.2217, 0.2216 and 0.2210, respectively. In terms of D_{st} index, chromosomes 4B (0.1026), 7B (0.0883) and 3A (0.0868) had the highest values, and chromosome 6D (0.0238) had the lowest values. Chromosomes 4B (0.2514), 4A (0.2200), 3A (0.2192) and 7B (0.2165) had the highest F_{st} index, respectively. Finally, chromosomes 4D (0.8730) and 1D (0.8723) had the highest and chromosomes 4B (0.7586) and 7B (0.7632) had the lowest inbreeding coefficient. High genetic diversity and low inbreeding coefficient indicate extensive hybridization and germplasm exchange. Therefore, based on result of this study, chromosome 4B comprises the highest genetic diversity and genetic differentiation. The low diversity coefficient and the number of SNP markers as well as the high inbreeding coefficient in the D genome are consistent with the evolutionary history of wheat. The first two components of PCA explained 22.88% of the variation in the genotyping data. Based on the biplot analysis of the first two components, the hexaploid landraces clearly had a different distribution from the hexaploid cultivars. However, such a distinction was not observed among the tetraploid genotypes, illustrating that they had a similar distribution to the hexaploid cultivars. According to the results of the DAPC method, the first function differentiated between tetraploid and hexaploid genotypes, whereas the second function led to the differentiation of cultivars from landraces. The PCA method with the first two components could not make the distinction made by the DAPC method. PCA searches for the direction that shows the largest total variance, whereas DAPC maximizes the distinction between groups and minimizes the within-group variations. The distinction between cultivars and landraces can be related to the change of allelic compositions as a result of different selection processes.

Conclusion

The genetic characteristics of 383 wheat genotypes allowed us to evaluate their structure. The DAPC method was able to distinguish the predetermined groups, i.e. tetraploid genotypes, cultivars, landraces and hexaploid genotypes. Considering the distinction between landraces and cultivars, it is expected that landraces will be used more in the future breeding programs to increase the genetic diversity in the gene pool of wheat cultivars. The information presented in this paper can be used in combination with phenotypic data to identify gene loci through GWAS. Finally, the result of present study suggest that different statistical methods should be used side by side for the detailed analysis of the genetic structure of the wheat populations in the future genetics studies.

Keywords: Cultivars, Landraces, Fixation index, Tetraploid and hexaploid wheat, SNP markers



تجزیه ساختار جمعیت در برخی از ژنوتیپ‌های گندم نان و دوروم با استفاده از نشانگرهای SNP و روش‌های PCA و DAPC

حسین عبدی^۱، هادی علی‌پور^{۲*}، ایرج برنوسی^۲ و جعفر جعفرزاده^۳

۱- دانشجوی دکتری، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه

۲- دانشیار، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه

۳- استادیار، مؤسسه تحقیقات کشاورزی دیم کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، مراغه

(تاریخ دریافت: ۱۴۰۲/۰۱/۱۳؛ تاریخ آخرین ویرایش: ۱۴۰۲/۰۵/۰۷؛ تاریخ پذیرش: ۱۴۰۲/۰۵/۱۷؛ تاریخ انتشار برخط: ۱۴۰۲/۰۶/۳۰)

چکیده

ارزیابی ساختار جمعیت، جهت درک الگوهای تنوع، انتخاب والدین مناسب برای تلاقی، شناسایی دقیق مکان‌های ژنومی کنترل‌کننده صفات، مطالعات تکاملی و روابط خویشاوندی ضروری است. در جریان تحقیق حاضر، ساختار ژنتیکی در جمعیتی متشکل از ۳۸۳ ژنوتیپ گندم ایرانی از گونه‌های هگزاپلوئید (ارقام زراعی و توده‌های بومی) و تتراپلوئید بر اساس روش‌های مبتنی بر فاصله (تجزیه به مؤلفه‌های اصلی و تجزیه تابع تشخیص مؤلفه‌های اصلی) مورد مطالعه قرار گرفت. بدین منظور از ۱۶۲۷۰ نشانگر چندشکلی تک نوکلئوتیدی (SNP) به‌دست آمده توسط روش GBS، استفاده شد. بر اساس نتایج، تقریباً یک‌چهارم از واریانس کل به تنوع بین جمعیت‌ها تعلق داشت و ضریب F_{st} بین ارقام زراعی و توده‌های بومی برابر با ۰/۱۵ به‌دست آمد. در حالی که ضریب فوق بین نمونه‌های تتراپلوئید و توده‌های بومی هگزاپلوئید بالا و برابر با ۰/۴۴ بود. ژنوم D کمترین مقدار شاخص F_{st} را به خود اختصاص داد و کروموزوم 4B بیشترین میزان ضریب F_{st} و سایر شاخص تنوع ژنی را نشان داد. گرچه بای‌پلات PCA، ارقام زراعی و توده‌های بومی گندم‌های هگزاپلوئید را به‌خوبی از هم تفکیک نمود، اما نتوانست تمایز روشنی بین نمونه‌های تتراپلوئید با سایر ژنوتیپ‌ها ایجاد نماید. ارزیابی دقیق ساختار جمعیت با روش DAPC نتوانست به‌طور موفقیت‌آمیزی گروه‌های از پیش تعیین شده را شناسایی و تفکیک کند. این نتیجه بدان دلیل است که رویکرد DAPC تمایز بین گروه‌ها را به حداکثر و تغییرات درون گروه را به حداقل می‌رساند. اختلاط جزئی بین ارقام زراعی و توده‌های بومی هگزاپلوئید را می‌توان به تبادل ژنی بین این دو گروه و یا شاید برچسب‌گذاری اشتباه آنان در زمان جمع‌آوری ارتباط داد. به‌طور کلی، نتایج حاصل از این مطالعه اطلاعات مفیدی درباره تمایز ژنتیکی نمونه‌های گندم تتراپلوئید و هگزاپلوئید ایرانی ارائه داد که می‌تواند در برنامه‌ریزی‌های آتی به‌نژادی گندم مورد استفاده قرار گیرد. همچنین حفاظت از این نمونه‌ها در بانک‌های ژن برای استراتژی‌های مختلف ضروری می‌باشد.

واژگان کلیدی: ارقام زراعی، توده‌های بومی، شاخص تثبیت، گندم تتراپلوئید و هگزاپلوئید، نشانگرهای SNP

گندم با تولید سالانه بیش از ۶۰۰ میلیون تن بیشترین سطح زیر کشت را در جهان دارد و از ارزش تغذیه‌ای و اقتصادی بالایی برای انسان برخوردار است (Sansaloni et al., 2020). حدود ۹۵ درصد از تولید جهانی گندم مربوط به گندم نان (*Triticum aestivum* L. با ساختار ژنوم هگزاپلوئید ($2n=6x=42$), (AABBDD) می‌باشد و تقریباً پنج درصد از تولید آن به گندم دوروم (*Triticum turgidum* L. *durum*) با ساختار ژنوم تتراپلوئید ($2n=4x=28$, AABB) و سایر گونه‌ها اختصاص دارد (Peng et al., 2011).

آگاهی از میزان تنوع ژنتیکی گام مهم در راستای حفظ ژرم-پلاسم و طراحی برنامه‌های به‌نژادی در گیاهان است. مکانیسم‌های تشخیص و تجزیه و تحلیل تنوع ژنتیکی به تدریج از بررسی مندلی صفات مورفولوژیکی به بررسی‌های مولکولی تنوع در سطح DNA پیشرفت کرده‌اند. در بخشی از مطالعات تنوع ژنتیکی، به‌نژادگران در پی ساختار جمعیت برای یافتن پاسخ به طیفی از پرسش‌های مهم نظیر روابط خویشاوندی، بررسی‌های تکاملی، میزان تشابه ژنوتیپ‌ها و تفسیر الگوهای تنوع هستند (MirMohammadi Maibody and Golkar, 2019). ساختار ژنتیکی یک جمعیت را می‌توان توسط چندین روش و با استفاده از نرم‌افزارهای مختلفی ارزیابی کرد (Sharifi-Tehrani, 2018). یکی از محبوب‌ترین روش‌ها، رویکرد STRUCTURE با استفاده از خوشه‌بندی Bayesian می‌باشد (Pritchard et al., 2000; Pritchard et al., 2010). این حال، روش مذکور مبتنی بر مفروضاتی است که تأیید آن‌ها دشوار می‌باشد و تخمین پارامترهای مربوط به آن نیاز به زمان محاسباتی قابل‌توجهی در مجموع داده‌های پیچیده دارد (Jombart et al., 2010)؛ بنابراین به روش‌هایی با قابلیت عملیاتی بالا و انعطاف‌پذیر نیازمند هست. در همین راستا، تجزیه به مؤلفه‌های اصلی (PCA: Principal component analysis) و تجزیه تابع تشخیص مؤلفه‌های اصلی (DAPC: Discriminant analysis of principal component) از فنون جایگزین جهت ارزیابی ساختار ژنتیکی جمعیت با استفاده از نشانگرهای مولکولی می‌باشند (Price et al., 2006; Patterson

et al., 2010; Jombart et al., 2006). برخلاف روش STRUCTURE که مبتنی بر مدل می‌باشد، روش‌های PCA و DAPC بر مبنای فاصله هستند (Tehseen et al., 2022).

تجزیه به مؤلفه‌های اصلی یکی از مهم‌ترین، ساده‌ترین و پرکاربردترین ابزارهای چندمتغیره می‌باشد که هدف آن کاهش پیچیدگی داده‌ها در تعدادی مؤلفه اصلی است. معمولاً از چند مؤلفه اول این روش به‌عنوان متغیرهای کمکی ساختار جمعیت در مدل تحلیل ارتباط در گستره ژنوم (GWAS: Genome-wide association study) استفاده می‌کنند (Ghorbani et al., 2023; Abdi et al., 2023). علی‌رغم این‌که روش PCA ساختار ژنتیکی موجود در داده‌های بسیار بزرگ را طی زمان محاسباتی ناچیز و بدون هیچ مفروضاتی درباره مدل ژنتیکی جمعیت شناسایی می‌کند، اما نتایج این روش می‌تواند در تحقیقات ژنتیکی جمعیت همراه با اریب باشد (Elhaik, 2022). این مورد بدان دلیل است که گاهی PCA به اشتباه ساختار عدم تعادل پیوستگی (LD: Linkage disequilibrium) را به‌جای ساختار جمعیت در نظر می‌گیرد (Privé et al., 2020) و به شدت به مشاهدات پرت حساس می‌باشد (Liu et al., 2013). همچنین این روش، تمام تنوع موجود در افراد جمعیت از جمله تنوع بین گروه‌ها (تنوع ژنتیکی ساختاریافته) و تنوع درون گروه‌ها (تنوع ژنتیکی تصادفی) را در تعدادی مؤلفه خلاصه می‌کند. در نتیجه به روشی نیاز هست که تنوع بین گروه‌ها را به حداکثر برساند. تجزیه DAPC یک روش چندمتغیره است که امکان استخراج اطلاعات از داده‌های ژنتیکی، تخصیص افراد به گروه‌ها، ارزیابی بصری تمایز بین جمعیت‌ها و مشارکت آلل‌های فردی در ساختار جمعیت را فراهم می‌کند (Jombart et al., 2010). توابع تشخیصی در این روش، ترکیب‌های خطی از آلل‌ها هستند که به بهترین وجه گروه‌ها را تفکیک می‌کنند؛ بنابراین، آلل‌هایی که بیشتر به این تشخیص کمک می‌کنند، آن‌هایی هستند که به‌طور قابل‌توجهی در بین گروه‌ها متفاوت هستند (Deperi et al., 2018). گزارش شده است که رویکرد DAPC همانند روش‌های خوشه‌بندی بیزی امکان تخصیص احتمالی افراد به هر گروه را می‌دهد و با به حداکثر رساندن تنوع ژنتیکی بین گروهی، در توصیف

specific PCR) برای تولید نشانگرهای SNP استفاده می‌شوند (Hussain *et al.*, 2022). در میان این روش‌ها، روش GBS می‌تواند با تولید SNPهای جدید در گیاهانی با ژنوم پیچیده بدون نیاز به اطلاعات توالی‌یابی قبلی استفاده گردد (Getachew *et al.*, 2019). روش GBS فراوانی داده گمشده بالایی دارد اما به‌خوبی می‌توان با روش‌های جانمایی (Imputation) این مشکل را رفع کرد (Alipour *et al.*, 2019). کتابخانه‌های GBS در ژنوتیپ‌های مورد مطالعه با روش معرفی شده توسط پالند و همکاران (Poland *et al.*, 2012) ساخته شدند که در آن از آنزیم‌های محدودکننده *PstI* و *MspI* استفاده شد. فراخوانی SNPها با استفاده از UNEAK (Universal Network Enabled Analysis Kit) GBS pipeline در برنامه TASSEL (Bradbury *et al.*, 2007) صورت گرفت. نشانگرهای SNP با هتروزیگوسیتی بیشتر از ۱۰ درصد، فراوانی آلی جزئی کمتر از پنج درصد و داده گمشده بیشتر از ۲۰ درصد برای حصول نتایج دقیق‌تر فیلتر و حذف شدند. در نهایت، ۱۶۲۷۰ نشانگر SNP برای تجزیه و تحلیل ساختار جمعیت باقی‌مانده و مورد استفاده قرار گرفتند. لازم به توضیح است که جزئیات مربوط به روش GBS و همچنین نتایج تنوع ژنتیکی با این داده‌ها در مطالعه علی‌پور و همکاران (Alipour *et al.*, 2017) ارائه شده است.

تجزیه و تحلیل آماری: تجزیه واریانس مولکولی (AMOVA: Analysis of molecular variance)، شاخص تمایز ژنتیکی (PhiPT) و شاخص تثبیت F_{st} (Fixation index) بین گروه‌های مورد مطالعه با استفاده از نرم‌افزار GenAlex محاسبه شد (Peakall and Smouse, 2006). همچنین تجزیه PCA و DAPC در بسته adegenet برنامه R-4.3.2 انجام گرفت (Jombart, 2008). در نهایت آماره‌های مختلف نظیر هتروزیگوسیتی مشاهده شده (Ho)، هتروزیگوسیتی مورد انتظار (Hs)، هتروزیگوسیتی مورد انتظار در آمیزش تصادفی (Ht)، Ht تصحیح شده (Htp)، تنوع ژنی در بین نمونه‌ها (Dst)، Dst تصحیح شده (Dstp)، شاخص تثبیت (F_{st}) تصحیح شده (Fstp) و ضریب خویش‌آمیزی (Fis) در بسته adegenet برآورد شدند.

زیرجمعیت‌ها بهتر از STRUCTURE عمل می‌کند (Jombart *et al.*, 2010). نتایجی مشابه توسط تهسین و همکاران (Tehseen *et al.*, 2022) گزارش گردیده است که طی آن به برتری روش‌های مبتنی بر فاصله (PCA و DAPC) نسبت به STRUCTURE اذعان شده است. تاکنون در طیفی از تحقیقات از روش DAPC برای ارزیابی ساختار جمعیت ژنوتیپ‌های گندم بر اساس اطلاعات نشانگرهای چندشکلی تک نوکلئوتیدی (SNP: Single nucleotide polymorphism) استفاده شده است (Cavanagh *et al.*, 2013; Kabbaj *et al.*, 2020; Alemu *et al.*, 2020; Mahboubi *et al.*, 2020).

گرچه پیش از این، ساختار جمعیت نمونه‌های گندم هگزاپلوئید ایرانی با استفاده از نشانگرهای SNP از طریق ماتریس Q حاصل از نرم‌افزار STRUCTURE و روش PCA مورد مطالعه قرار گرفته است (Alipour *et al.*, 2017; Rahimi *et al.*, 2022; Rabieyan *et al.*, 2019)، اما در تحقیق حاضر برای نخستین بار، رویکردهای PCA و DAPC در ارزیابی ساختار جمعیت این نمونه‌های همراه با نمونه‌های تتراپلوئید گندم مورد مقایسه قرار گرفتند.

مواد و روش‌ها

مواد گیاهی: در تحقیق حاضر تعداد ۳۸۳ ژنوتیپ گندم ایرانی شامل ۳۷۰ نمونه هگزاپلوئید (۱۰۰ رقم زراعی معرفی شده در طی یک قرن اخیر و ۲۷۰ توده بومی جمع‌آوری شده از مناطق مختلف کشور) و ۱۳ نمونه تتراپلوئید مورد مطالعه قرار گرفتند (جدول ۱).

ارزیابی ژنوتیپی: ژنوتیپ‌های مورد مطالعه قبلاً با استفاده از روش ژنوتایپینگ به‌وسیله از توالی‌یابی (GBS: Genotyping by sequencing) در دانشگاه ایالتی کانزاس آمریکا ژنوتیپ‌سنجی شدند (Alipour, 2016). ژنوتایپینگ به‌وسیله توالی‌یابی، یکی از روش‌های توالی‌یابی نسل بعد (NGS: Next-generation sequencing) می‌باشد که در کنار سایر روش‌ها نظیر توالی‌یابی آمپلیکون هدفمند (TAS: Targeted amplicon sequencing)، آرایه‌های ایلومینا (Illumina SNP arrays)، فناوری تنوع آرایه‌ها (DArT: Diversity arrays technology) و PCR آل اختصاصی رقابتی (KASP: Kompetitive allele-

جدول ۱- عادت رشد، کد ژنوتیپ و سال معرفی تعداد ۳۸۳ ژنوتیپ گندم ایرانی

Table 1. Growth habit, genotype code and released year of 383 Iranian wheat genotypes

شماره	نام رقم	عادت رشدی	سال معرفی	شماره	نام رقم	عادت رشدی	سال معرفی
No	Variety name	Growth Habit	Released year	No	Variety name	Growth Habit	Released year
ارقام زراعی هگزاپلوئید (Hexaploid cultivars)							
1	4820	Spring	1951	51	Koohdasht	Spring	2002
2	Adl	Spring	1976	52	Mahdavi	Spring	1995
3	Aflak	Spring	2010	53	Maroon	Spring	1991
4	Akbari	Spring	2006	54	Marvdasht	Spring	1999
5	Akova	Winter	1958	55	Mihan	Winter	2010
6	Alborz	Spring	1978	56	Misr 1	Facultative	2009
7	Alvand	Facultative	1995	57	Moghan 1	Spring	1974
8	Argelato	No information	No information	58	Moghan 2	Spring	1974
9	Aroona	No information	No information	59	Moghan 3	Spring	2006
10	Arta	Spring	2006	60	Morvarid	Spring	2009
11	Arvand 1	Spring	1974	61	Munal #1	No information	No information
12	Atrak	Spring	1995	62	Mv-17	Winter	1993
13	Azadi	Facultative	1979	63	Navid 1990	Facultative	1990
14	Azar	Winter	1957	64	Naz	Spring	1978
15	Azar 2	Winter	1997	65	Neishabour	Spring	2006
16	Bahar	Spring	2007	66	Nicknejad	Spring	1995
17	Bajio	Spring	1966	67	Ofogh	Spring	2012
18	Bam	Spring	2006	68	Ohadi	Winter	2010
19	Barta	Facultative	1970	69	Panjamo 62	Spring	1968
20	Bayat	Spring	1976	70	Parsi	Spring	2009
21	Bezostaya	Winter	1969	71	Pishgam	Facultative	2008
22	Biston	Spring	1980	72	Pishtaz	Spring	2002
23	Chamran	Spring	1997	73	Qaboos	Spring	2014
24	Chamran2	Spring	2013	74	Rashid	Facultative	1968
25	Darab1	Spring	1980	75	Rask	Spring	2005
26	Darab2	Spring	1995	76	Rayhani	Spring	1942
27	Darya	Spring	2006	77	Rijaw	Facultative	2011
28	Dastjerdi	Spring	1960	78	Roshan	Spring	1960
29	Dayhim	Spring	1968	79	Sabalan	Facultative	1980
30	Dena	Spring	2007	80	Sefidak	Facultative	1946
31	Dez	Spring	2002	81	Sepahan	Spring	2006
32	Dn-11	No information	No information	82	Shahi	Winter	1967
33	Falat	Spring	1990	83	Shahpassand	Winter	1942
34	Fong	No information	No information	84	Shahryar	Winter	2002
35	Fontana	No information	No information	85	Shanghai #7	Spring	No information
36	Gahar	Spring	1996	86	Shiraz	Spring	2002
37	Gascogne	No information	1994	87	Shiroodi	Spring	1997
38	Ghods	Spring	1988	88	Siosson	No information	1994
39	Golestan	Spring	1986	89	Sirvan	Spring	2012
40	Hamoon	Spring	2002	90	Sistan	Spring	2006
41	Homa	Winter	2009	91	Sivand	Spring	2009
42	Inia 66	Spring	1969	92	Tajan	Spring	1995
43	Karaj 1	Facultative	1974	93	Tak-Ab	Winter	2013
44	Karaj 2	Winter	1974	94	Tobari 66	Spring	1969
45	Karaj 3	Winter	1974	95	Tous	Facultative	2002
46	Karim	Spring	2011	96	Uroum	Winter	2009
47	Karkheh	Spring	2003	97	Vee/Nac	Spring	1997
48	Kaveh	Spring	1980	98	Zagros	Spring	1996
49	Kavir	Spring	1997	99	Zare	Facultative	2010
50	Khazar 1	Spring	1974	100	Zarin	Facultative	1995

نتایج و بحث

به منظور بررسی سطح تنوع در درون و بین سه گروه مورد مطالعه (ارقام زراعی هگزاپلوئید، توده‌های بومی هگزاپلوئید و ژنوتیپ‌های تتراپلوئید) تجزیه واریانس مولکولی انجام گرفت. نتایج نشان داد که ۲۴/۱۲ درصد از تنوع محاسبه شده مربوط به بین گروه‌ها و ۷۵/۸۸ درصد از تنوع مربوط به درون گروه‌ها بود (جدول ۲). وجود تنوع ژنتیکی بالا در بین و درون جمعیت‌ها از ویژگی‌های گندم نان (Alipour *et al.*, 2017) و گندم دوروم (Fayaz *et al.*, 2019) ایرانی هستند. پایین بودن تنوع بین جمعیتی، احتمالاً ناشی از جریان ژنی بالا در بین آن‌ها می‌باشد. از طرفی دیگر، تنوع بالا در درون جمعیت را هم می‌توان با ماهیت چندپلوئیدی گندم، ناهمگنی افراد درون گروه‌ها و تلاقی‌های مختلف بین ژنوتیپ‌ها به‌ویژه میان ارقام زراعی و توده‌های بومی توجیه کرد. سهم اندک تنوع بین گروهی از تنوع کل در مطالعات متعددی گزارش شده است (Rasheed *et al.*, 2016; Miazzi *et al.*, 2022; Sthapit *et al.*, 2022).

آماره Fst امکان تجزیه و تحلیل ساختار جمعیت‌های تقسیم‌شده را فراهم می‌کند. همچنین جهت اندازه‌گیری فاصله ژنتیکی بین زیرجمعیت‌ها نیز استفاده می‌گردد. همان‌طور که انتظار می‌رفت بین دو زیرجمعیت هگزاپلوئید آماره Fst پایین (۰/۱۵۴۴) بود، در حالی که این شاخص بین نمونه‌های هگزاپلوئید توده بومی و رقم زراعی با نمونه‌های تتراپلوئید به شکل قابل توجهی بالا و به ترتیب برابر با ۰/۴۳۶۲ و ۰/۳۵۶۲ به دست آمد (جدول ۳). آماره Fst می‌تواند مقداری بین صفر (عدم تمایز) تا یک (تمایز کامل) داشته باشد و اساساً مقادیر بالاتر از ۰/۲۵ نشان از تمایز ژنتیکی بسیار بالا است و مقادیر بین ۰/۱۵-۰/۲۵ نیز حاکی از تمایز بالا در زیرجمعیت‌ها دارد. یافته‌ها، منحصر به فرد بودن گندم‌های تتراپلوئید نسبت به گندم‌های هگزاپلوئید را ثابت می‌نماید. چنین نتیجه‌ای با توجه به عدم حضور ژنوم D در گندم‌های تتراپلوئید کاملاً قابل توجیه است. تمایز گسترده و مقادیر بالای آماره Fst بین گونه‌های گندم پیش از این نیز گزارش شده است (Fiore *et al.*, 2022).

جدول ۲- تجزیه واریانس مولکولی در سه گروه از ژنوتیپ‌های گندم مورد مطالعه بر اساس نشانگرهای SNP

Table 2. Analysis of molecular variance (AMOVA) in three groups of studied wheat genotypes based on SNP markers

منابع تغییرات Sources of variation	درجه آزادی Degrees of freedom	مجموع مربعات Sum of squares	میانگین مربعات Mean squares	تنوع Variation	درصد %	معیار Stat	مقدار Value
بین گروه Between group	2	12929.8	6464.9	75.0	24.12	PhiPT	0.241
درون گروه Within group	380	89668.5	236.0	236.0	75.88		
کل Total	382	102598.3	268.6	311.0	100		

جدول ۳- مقادیر شاخص تثبیت زوجی بین سه گروه گندم مورد مطالعه

Table 3. Pairwise fixation index (Fst) values among three studied wheat groups

زمینه ژنتیکی Genetic background	هگزاپلوئید (ارقام زراعی) Hexaploid (Cultivars)	هگزاپلوئید (توده‌های بومی) Hexaploid (Landraces)	تتراپلوئید Tetraploid
Hexaploid (cultivars)	-	-	-
Hexaploid (Landraces)	0.1544	-	-
Tetraploid	0.3562	0.4362	-

موضوع ناشی از تأثیر متفاوت گزینش بر هر ژنوم گندم است که پیش از این ثابت شده است (Chao *et al.*, 2010). درک ساختار جمعیت برای تجزیه و تحلیل‌های پایین‌دستی مانند تجزیه GWAS بسیار مهم است (Tehseen *et al.*, 2022). در مطالعات متعددی از روش PCA برای ارزیابی ساختار جمعیت و تأیید ساختار ژنتیکی استفاده کرده‌اند (Nielsen *et al.*, 2014; Fayaz *et al.*, 2019; Khadka *et al.*, 2020; Rabieyan *et al.*, 2022). شکل ۱، درصد واریانس جمعی توجیه‌شده بر اساس تعداد مؤلفه‌های اصلی متفاوت را نشان می‌دهد که در آن دو مؤلفه اول ۲۲/۸۸ درصد از تغییرات موجود در داده‌های ژنوتایپینگ را توجیه کردند. بر اساس بای‌پلات دو مؤلفه اول، توده‌های بومی هگزاپلوئید به‌وضوح پراکنش متفاوتی از ارقام زراعی هگزاپلوئید داشتند. با این حال چنین تمایزی در بین ژنوتیپ‌های تتراپلوئید مشاهده نشد، به‌طوری که توزیعی مشابه با ارقام زراعی هگزاپلوئید داشتند (شکل ۲). ارقام زراعی هگزاپلوئید و نمونه‌های تتراپلوئید مقادیر مثبت اولین و دومین مؤلفه اصلی را به خود اختصاص دادند ولی توده‌های بومی هگزاپلوئید از نظر هر دو مؤلفه عمدتاً دارای مقادیر منفی بودند. چنین اختلافی در فضای بای‌پلات حاکی از ساختار ژنتیکی متفاوت آن‌ها می‌باشد. به‌دلیل پراکنش گسترده‌تر در فضای بای‌پلات، به‌نظر می‌رسد تنوع ژنتیکی در توده‌های بومی بیشتر از ارقام زراعی است و تمایز این دو گروه می‌تواند بیان‌گر این موضوع باشد که بخش بزرگی از تنوع ژنتیکی موجود در توده‌های بومی در برنامه‌های اصلاح گندم‌های مدرن گنجانده نشده است. تفکیک ارقام زراعی می‌تواند مربوط به فرآیندهای گزینشی مختلف، تلاقی‌های مکرر و وجود منبع خارجی در شجره آن‌ها باشد. همچنین مشاهده شد که برخی از ارقام زراعی پراکنشی مشابه با توده‌های بومی داشتند. با نگاهی دقیق‌تر به شجره این ارقام متوجه شدیم که برخی از آن‌ها مانند ارقام اوحدی، روشن و هما در اصل توده بومی بوده‌اند که مورد خالص‌سازی قرار گرفته و به‌عنوان رقم زراعی معرفی شده‌اند. سایر ارقامی که مشابه توده‌های بومی بودند، جزء ارقام قدیمی هستند که اطلاعات دقیقی از شجره آن‌ها در دسترس نیست.

برآورد آماره‌های مختلف به تفکیک سه ژنوم گندم نشان می‌دهد که ژنوم D نسبت به ژنوم‌های A و B دارای پایین‌ترین تعداد نشانگر SNP و کمترین مقادیر هتروزیگوسیتی مشاهده شده، هتروزیگوسیتی مورد انتظار، تنوع ژنی و شاخص تثبیت بود، در حالی که ضریب خویش‌آمیزی بالایی داشت (جدول ۴). در میان کروموزوم‌ها، 3B دارای بیشترین تعداد نشانگر (۱۱۵۴) و هتروزیگوسیتی مشاهده شده (۰/۰۴۲۷) بود. کروموزوم‌های 5B، 7B و 6A به‌ترتیب با مقادیر ۰/۲۲۱۷، ۰/۲۲۱۶ و ۰/۲۲۱۰ بیشترین هتروزیگوسیتی مورد انتظار را داشتند. در اکثر کروموزوم‌ها مقدار هتروزیگوسیتی مورد انتظار (Hs) مشابه با هتروزیگوسیتی مورد انتظار در آمیزش تصادفی (Ht) بود؛ اما اختلاف بین این دو شاخص برای برخی از کروموزوم‌ها نظیر 4B بالا بود. از نظر شاخص تنوع ژنی (Dst) و همچنین شاخص تنوع ژنی تصحیح شده (Dstp) به‌ترتیب کروموزوم‌های 4B (۰/۱۰۲۶)، 7B (۰/۰۸۸۳) و 3A (۰/۰۸۶۸) بیشترین و کروموزوم 6D (۰/۰۲۳۸) کمترین مقادیر را داشتند. به لحاظ شاخص Fst، کروموزوم‌های 4B (۰/۲۵۱۴)، 4A (۰/۲۲۰۰)، 3A (۰/۲۱۹۲) و 7B (۰/۲۱۶۵) به‌ترتیب بیشترین و کروموزوم‌های 6D (۰/۰۲۵۹)، 2D (۰/۰۲۵۹) و 1D (۰/۰۲۷۹) و 3D (۰/۰۲۹۳) به‌ترتیب کمترین میزان را به خود اختصاص دادند. ترتیب ژنوتیپ‌ها از نظر Fst تصحیح شده مشابه با Fst بود. در نهایت، کروموزوم‌های 4D (۰/۸۷۳۰) و 1D (۰/۸۷۲۳) بیشترین ضریب خویش‌آمیزی و کروموزوم‌های 4B (۰/۷۵۸۶) و 7B (۰/۷۶۳۲) کمترین ضریب مذکور را داشتند (جدول ۴). تنوع ژنی بالا و ضریب خویش‌آمیزی پایین، نشان‌دهنده هیبریداسیون گسترده و تبادل ژرم‌پلاسم می‌باشد (Ayalew *et al.*, 2020)؛ بنابراین کروموزوم 4B در مطالعه حاضر ناحیه‌ای از ژنوم گندم بود که حاوی بیشترین تنوع ژنی و تمایز ژنتیکی است. پایین بودن ضریب تنوع و تعداد نشانگر SNP و همچنین بالا بودن ضریب خویش‌آمیزی در ژنوم D با تاریخچه تکاملی گندم سازگار است و این موضوع به جوان بودن ژنوم D و هیبریداسیون کمتر این ژنوم برمی‌گردد (Alipour *et al.*, 2017; Rimbart *et al.*, 2018). چنین شاخص‌هایی می‌توانند مناطقی از ژنوم گندم که در طی سالیان طولانی توسط کشاورزان محلی و به‌نژادگران در معرض گزینش بوده‌اند را آشکار کنند (Cavanagh *et al.*, 2013). این

جدول ۴- آماره‌های برآورد شده در ژنوتیپ‌های گندم مورد مطالعه با استفاده از ۱۶۲۷۰ نشانگر SNP

Table 4. Statistics estimated for studied wheat genotypes using 16270 SNP markers

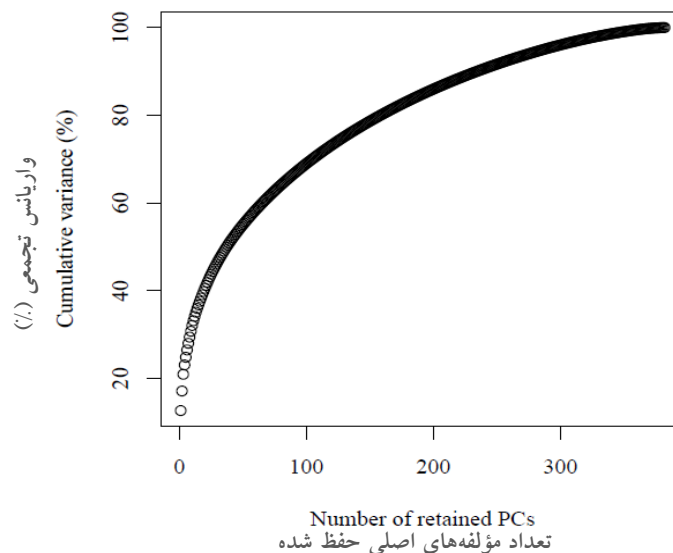
Chr	NS	Ho	Hs	Ht	Htp	Dst	Dstp	Fst	Fstp	Fis
1A	749	0.0398	0.1981	0.2770	0.3165	0.0789	0.1184	0.1960	0.2407	0.7936
1B	903	0.0445	0.2127	0.2841	0.3198	0.0714	0.1071	0.1790	0.2219	0.7844
1D	439	0.0223	0.1905	0.2158	0.2286	0.0253	0.0381	0.0279	0.0155	0.8723
2A	778	0.0427	0.2119	0.2770	0.3095	0.0651	0.0976	0.1618	0.2036	0.8009
2B	1136	0.0417	0.2146	0.2896	0.3270	0.0750	0.1125	0.1835	0.2296	0.8009
2D	617	0.0284	0.1678	0.1918	0.2039	0.0240	0.0361	0.0259	0.0122	0.8511
3A	889	0.0413	0.1952	0.2820	0.3253	0.0868	0.1301	0.2192	0.2689	0.7771
3B	1154	0.0419	0.2022	0.2768	0.3141	0.0746	0.1120	0.1859	0.2309	0.7931
3D	388	0.0227	0.1444	0.1756	0.1914	0.0312	0.0470	0.0293	0.0078	0.8411
4A	687	0.0399	0.1926	0.2774	0.3198	0.0848	0.1271	0.2200	0.2682	0.7738
4B	523	0.0441	0.1810	0.2836	0.3349	0.1026	0.1539	0.2514	0.3008	0.7586
4D	145	0.0202	0.1818	0.2063	0.2187	0.0245	0.0369	0.0318	0.0282	0.8730
5A	622	0.0403	0.1956	0.2714	0.3093	0.0758	0.1137	0.1925	0.2353	0.7754
5B	882	0.0425	0.2217	0.2889	0.3226	0.0673	0.1009	0.1685	0.2148	0.8047
5D	331	0.0257	0.1487	0.1946	0.2176	0.0459	0.0689	0.0714	0.0640	0.8307
6A	552	0.0446	0.2210	0.2844	0.3161	0.0634	0.0950	0.1569	0.1987	0.8027
6B	850	0.0397	0.2118	0.2621	0.2873	0.0503	0.0755	0.1341	0.1747	0.8112
6D	325	0.0259	0.1893	0.2131	0.2251	0.0238	0.0358	0.0259	0.0154	0.8491
7A	1098	0.0392	0.1963	0.2712	0.3087	0.0749	0.1124	0.1879	0.2340	0.7946
7B	914	0.0494	0.2216	0.3099	0.3540	0.0883	0.1324	0.2165	0.2649	0.7632
7D	527	0.0231	0.1515	0.1870	0.2048	0.0356	0.0533	0.0460	0.0299	0.8300
NA	1761	0.0427	0.2013	0.2598	0.2890	0.0584	0.0876	0.1418	0.1748	0.7899
A genome	5375	0.0409	0.2006	0.2768	0.3149	0.0762	0.1143	0.1919	0.2372	0.7884
B genome	6362	0.0433	0.2109	0.2851	0.3222	0.0742	0.1113	0.1850	0.2303	0.7901
D genome	2772	0.0246	0.1660	0.1960	0.2111	0.0300	0.0451	0.0362	0.0229	0.8475
Whole genomes	16270	0.0392	0.1988	0.2645	0.2973	0.0656	0.0985	0.1572	0.1912	0.7993

Chr: کروموزوم، NS: تعداد SNPها، Ho: هتروزیگوسیتی مشاهده شده، Hs: هتروزیگوسیتی مورد انتظار، Ht: هتروزیگوسیتی مورد انتظار در آمیزش تصادفی، Htp: تصحیح شده، Dst: تنوع ژنی در بین نمونه‌ها، Dstp: تصحیح شده، Fst: شاخص تثبیت، Fstp: تصحیح شده و Fis: ضریب خویش‌آمیزی

Chr: Chromosome, NS: Number of SNP, Ho: Observed heterozygosity, Hs: Expected genetic diversity (heterozygosity), Ht: Expected heterozygosity in the random-mating, Htp: Corrected Ht, Dst: Gene diversity among samples, Dstp: Corrected Dst, Fst: Fixation index, Fstp: Corrected Fst, and Fis: Inbreeding coefficient

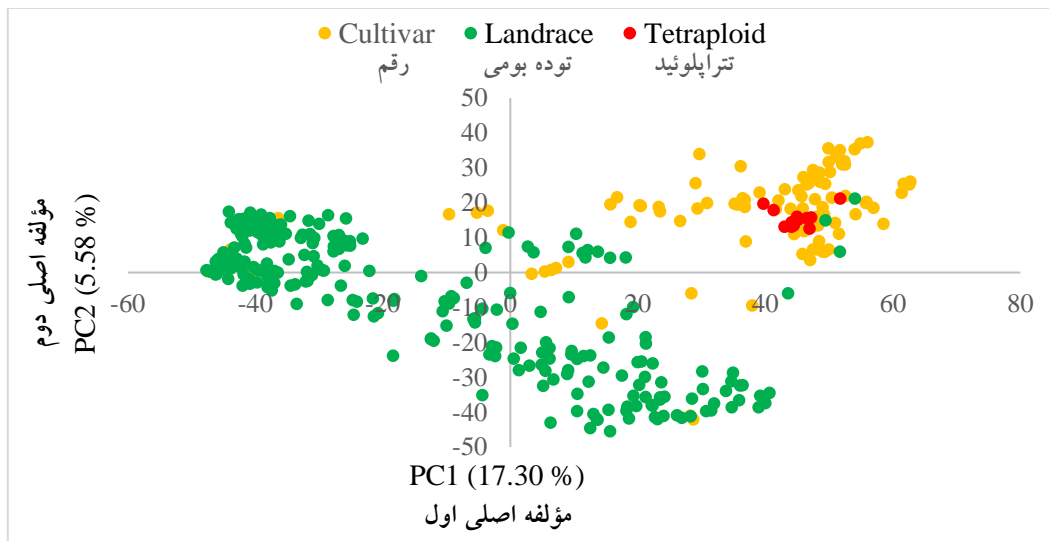
مجموعه داده‌های بزرگ ژنومی در زمان محاسباتی اندک و بدون هیچ فرضی درباره زمینه مدل ژنتیکی جمعیت است، با این حال این روش فاقد برخی از ویژگی‌های ضروری برای بررسی ساختار جمعیت می‌باشد (Jombart *et al.*, 2010). عدم تمایز نمونه‌های تتراپلوئید توسط بای‌پلات PCA می‌تواند ناشی از آن باشد که دو مؤلفه اصلی اول تنها حدود یک‌چهارم از تنوع موجود در کل داده‌ها را توجیه می‌نمایند و به نظر می‌رسد اطلاعات موجود در مؤلفه‌های باقی‌مانده می‌توانند منجر به تمایز مورد انتظار بین نمونه‌های تتراپلوئید با سایر نمونه‌ها گردد. از طرفی دیگر، تقسیم جمعیت به دو زیرگروه گاهی اوقات بدین معنی می‌باشد که ساختار ارائه شده نمی‌تواند به‌درستی ساختار ژنتیکی جمعیت را شناسایی کند (Janes *et al.*, 2017). از این‌رو در مطالعه حاضر، جهت تشریح دقیق‌تر و بیشتر ساختار واقعی جمعیت از رویکرد دیگر یعنی DAPC نیز استفاده شد.

یک احتمال دیگر که در این باره وجود دارد، می‌تواند مربوط به طبقه‌بندی اشتباه نمونه‌های گندم باشد؛ به‌طوری که بانک‌های ژن مراکز بین‌المللی CIMMYT و ICARDA تأیید کرده‌اند که ۴/۴ درصد از نمونه‌های تتراپلوئید و ۲/۴ درصد از نمونه‌های هگزاپلوئید در اطلاعات پاسپورت خود به اشتباه طبقه‌بندی شده‌اند (Sansaloni *et al.*, 2020). گزارشی مشابه توسط خدکا و همکاران (Khadka *et al.*, 2020) در رابطه با احتمال جمع‌آوری اشتباه توده‌های بومی گندم در خلال سال‌های ۱۹۷۰ الی ۱۹۹۰ در کشور نپال وجود دارد. حتی قرارگیری برخی از توده‌های بومی در زیرگروه ارقام زراعی را به برچسب‌گذاری اشتباه نمونه‌ها توسط بانک‌های ژن ارتباط داده‌اند (Kabbaj *et al.*, 2017). در روش PCA، گروه‌بندی بر اساس مؤلفه‌ها می‌باشد و هر مؤلفه خود ترکیب خطی از تمام متغیر دیگر است. مشخصه اصلی PCA توانایی آن برای شناسایی ساختارهای ژنتیکی در



شکل ۱- درصد واریانس تجمعی توجیه‌شده بر اساس تعداد مؤلفه‌های اصلی متفاوت

Figure 1. Cumulative variance percentage explained by different number of principal components



شکل ۲- ساختار جمعیت ۳۸۳ ژنوتیپ گندم نان و دوروم ایرانی بر اساس تجزیه به مؤلفه‌های اصلی

Figure 2. Population structure of 383 Iranian bread and durum wheat genotypes based on principal component analysis

بلکه ناشی از واقعیت تبادل ژنی بین این دو گروه است. همان-طور که پیش‌تر ملاحظه گردید، بای‌پلات روش PCA با دو مؤلفه نخست نتوانست تمایز ایجاد شده توسط روش DAPC را ایجاد نماید؛ بنابراین، برتری روش DAPC نسبت به PCA در به تصویر کشیدن تفاوت بین گروهی است. رویکرد DAPC ابتدا داده‌ها را با استفاده از PCA به مؤلفه‌هایی تبدیل می‌کند و سپس یک تجزیه تابع تشخیص بر روی مؤلفه‌ها اصلی حفظ‌شده انجام می‌دهد. جویمبارت و همکاران

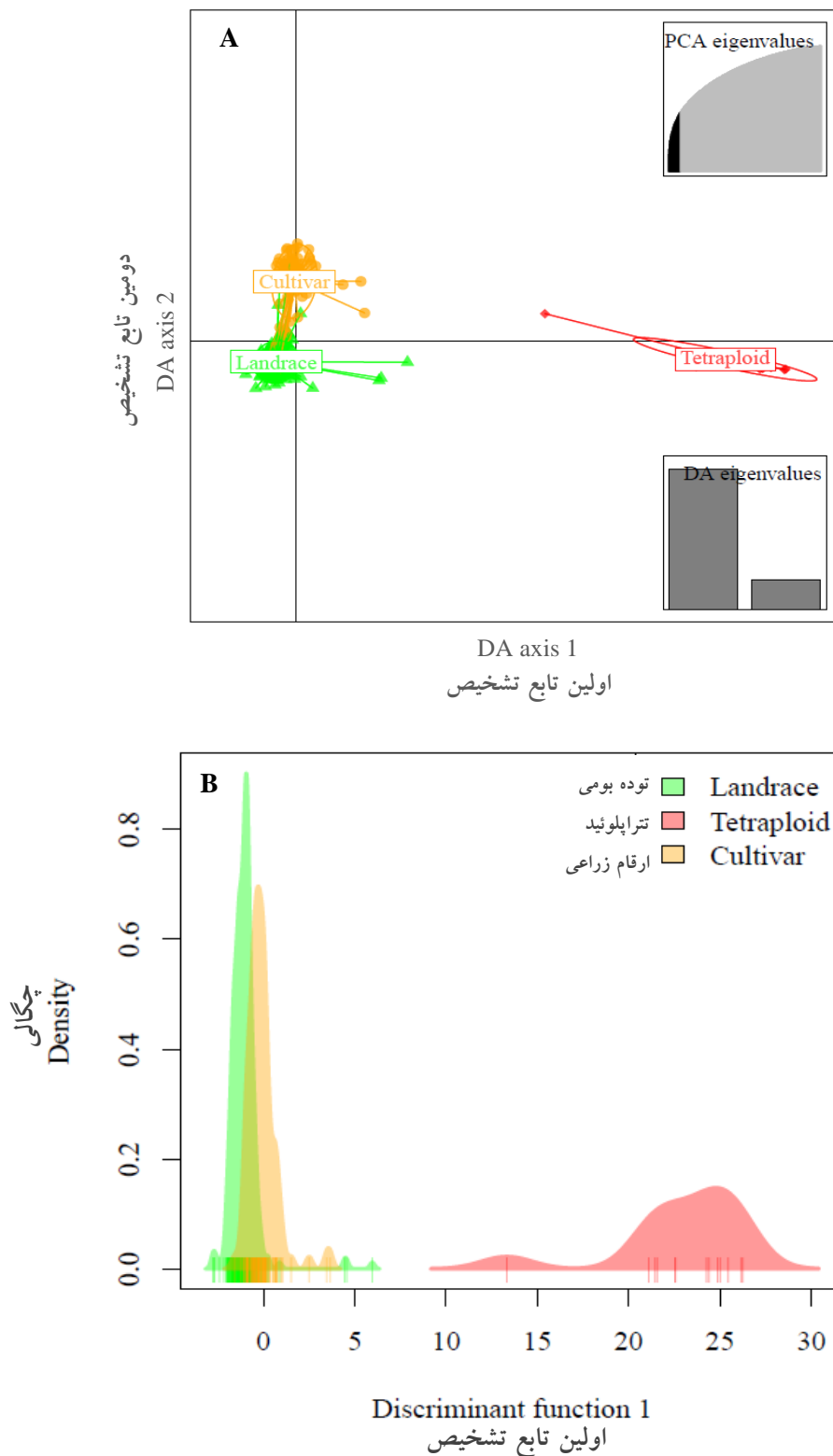
بر اساس نتایج روش DAPC، اولین تابع تشخیص تمایز قابل‌توجهی بین نمونه‌های تتراپلوئید و نمونه‌های هگزاپلوئید ایجاد کرد؛ در حالی‌که دومین تابع تشخیص منجر به تمایز ارقام زراعی از توده‌های بومی گردید؛ بنابراین در یک فضای دوبعدی از این دو تابع با وجود این‌که اختلاطی جزئی بین توده‌های بومی و ارقام زراعی وجود دارد اما به‌طور کلی شاهد تمایز گروه‌های مورد مطالعه هستیم (شکل ۳). عدم تمایز کامل توده‌های بومی از ارقام زراعی نه ناشی از ضعف روش DAPC

جدید است (Kabbaj *et al.*, 2017). تمایز بین ارقام زراعی و توده‌های بومی را می‌توان به تغییر ترکیبات آللی در نتیجه پروسه‌های گزینشی مختلف ارتباط داد. در یک مطالعه مشابه، با تجزیه و تحلیل تمایز ژنتیکی بین جمعیت‌ها و میزان اشتراک هاپلوتایپ‌ها در ژرم‌پلاسم جهانی متشکل از ارقام زراعی و توده‌های بومی، انواعی از آلل‌ها (به‌ویژه ژن‌های دخیل در تنظیم مراحل فنولوژیک) شناسایی شدند که برای بهبود صفات زراعی در ارقام زراعی مورد گزینش قرار گرفته‌اند (Cavanagh *et al.*, 2013). بر اساس نتایج ما، روش DAPC با استفاده از نشانگرهای SNP به شکل بارزی گونه‌های *Triticum aestivum* L. و *Triticum turgidum* L. را شناسایی نمود. در توافق با این یافته‌ها، فیوره و همکاران (Fiore *et al.*, 2022) با مطالعه ساختار جمعیت توسط روش‌های مختلف از جمله PCA و DAPC نتیجه گرفتند که نشانگرهای SNP برای تشخیص آسان گونه‌ها و زیرگونه‌های *Triticum* توانایی قابل ملاحظه‌ای دارند. در یک تحقیقی سانسالونی و همکاران (Sansaloni *et al.*, 2022) حدود ۸۰ هزار نمونه گندم از سه گروه بیولوژیکی (گونه‌های وحشی، تتراپلوئید اهلی شده و هگزاپلوئید اهلی شده) را مورد بررسی قرار دادند و اظهار داشتند که نمودار مقیاس‌بندی چندبعدی (MSD: Multidimensional scaling) بر اساس نشانگرهای SNP توانایی تفکیک ژنوتیپ‌ها به توده‌های بومی، ارقام زراعی را دارند.

با وجود مزایای DAPC، این روش ایراداتی دارد، از جمله این که DAPC از تمام اطلاعات موجود استفاده نمی‌کند. همچنین DAPC اجازه نمی‌دهد زیرجمعیت‌های درون گروه-ها ایجاد شود و ساختار ژنتیکی ارائه شده توسط این روش به درستی منشأ جغرافیایی افراد را پیش‌بینی نمی‌کند (Qin *et al.*, 2022). برخلاف این گزارش، کامپوی و همکاران (Campoy *et al.*, 2016) ضمن مشاهده هماهنگی بالا بین نتایج روش‌های STRUCTURE و DAPC نتیجه گرفتند که ساختار جمعیت ارائه شده توسط DAPC در توافق با شجره ارقام و منشأ جغرافیایی افراد است. چنین یافته‌های متفاوت در رابطه با خروجی DAPC تا حد زیادی به ماهیت ژرم‌پلاسم مورد مطالعه و میزان اختلاط ژنتیکی افراد بستگی دارد.

(Jombart *et al.*, 2010) در مقایسه روش‌های PCA با DAPC اظهار داشتند که PCA جهتی را جستجو می‌کند که بیشترین واریانس کل را نشان دهد، در حالی که DAPC تمایز بین گروه‌ها را به حداکثر و تغییرات درون گروه را به حداقل می‌رساند. همچنین، عدم نیاز به مفروضاتی درباره مدل ژنتیکی جمعیت به‌ویژه فرضیاتی در مورد تعادل هاردی واینبرگ یا تعادل پیوستگی (لینکازی) از ویژگی‌های بارز روش DAPC می‌باشد. کباج و همکاران (Kabbaj *et al.*, 2017) در مقایسه ساختار جمعیت بر اساس روش‌های PCA، STRUCTURE، DAPC و خوشه‌بندی بر مبنای UPGMA اظهار داشتند که نتایج این روش‌ها می‌تواند متفاوت باشد. برخلاف این گزارش، محبوبی و همکاران (Mahboubi *et al.*, 2020) نشان دادند که PCA و DAPC مطابقت زیادی با نتایج تجزیه خوشه‌ای و STRUCTURE دارد. در مطالعه‌ای نتیجه‌گیری شد که وجود نمونه‌های پرت می‌تواند نتایج PCA در ارزیابی ساختار جمعیت گندم را تحت تأثیر قرار دهد و بهتر است از روش‌های جایگزین دیگر استفاده شود (Abdi *et al.*, 2023). علی‌رغم این که مولر و همکاران (Müller *et al.*, 2018) مشاهده کردند که شاخص Fst و تجزیه DAPC نتایج PCA را تأیید می‌کنند و تفکیک واضحی بین گونه‌های *T. spelta* و *T. aestivum* ایجاد می‌کنند، اما رویکرد مبتنی بر DAPC و Fst در جدایی ژنوتیپ‌های بهاره و پاییزه *T. spelta* مؤثرتر از PCA بود. در بررسی ۲۴۰۳ توده بومی گندم ایرانی گزارش شد که روش PCA به شکل قابل قبولی نمونه‌های تتراپلوئید را از نمونه‌های هگزاپلوئید جدا می‌کند (Vikram *et al.*, 2020)؛ بنابراین، این روش‌ها تکمیل‌کننده یکدیگر می‌باشند و در ارزیابی ساختار جمعیت‌ها بهتر است هم‌زمان از چندین روش استفاده کرد.

خوشه‌بندی توده‌های بومی با شباهت‌های آللی عملاً تلاشی برای ردیابی آن دسته از آلل‌هایی است که از نظر اجدادی یکسان هستند و از رویداد اهلی‌سازی اولیه یا فشارهای محیطی مشترک حفظ شده‌اند. در عوض، در مورد ارقام زراعی، شباهت ژنتیکی به شدت تحت تأثیر انتخاب به‌نژادگران از هیبریداسیون منابع ژرم‌پلاسم خاص برای ایجاد لاین‌های



شکل ۳- تجزیه تابع تشخیص مؤلفه‌های اصلی (DAPC) حاصل از ۳۸۳ ژنوتیپ گندم. (A) پراکنش ژنوتیپ‌ها بر مبنای دو تابع تشخیص اول و (B) چگالی گروه‌ها بر اساس اولین تابع تشخیص.

Figure 3. Discriminant analysis of principal component (DAPC) obtained from 383 wheat genotypes. (A) The distribution of genotypes based on the first two discriminant function, and (B) the density of groups based on the first discriminant function.

دستیابی به نتایج جامع‌تر لازم است تحقیقات آتی با تعداد نمونه بیشتر صورت گیرد. همچنین نظر به تمایز توده‌های بومی از ارقام زراعی می‌توان توصیه کرد که در برنامه‌های به‌نژادی آتی جهت افزایش تنوع ژنتیکی در خزانه ژنی ارقام گندم از توده‌های بومی استفاده بیشتری شود. اطلاعات ارائه شده در این مقاله می‌تواند در ترکیب با داده‌های فنوتیپی برای شناسایی مکان‌های ژنی از طریق GWAS به‌کار آیند. در نهایت، پیشنهاد می‌شود که در مطالعات آتی، برای تجزیه دقیق ساختار ژنتیکی جمعیت از روش‌های مختلف آماری در کنار یکدیگر استفاده شود.

به‌طور کلی، خصوصیات ژنتیکی ۳۸۳ ژنوتیپ گندم به ما امکان داد تا ساختار آن‌ها را ارزیابی کنیم و روش DAPC گروه‌های از پیش تعیین‌شده یعنی نمونه‌های تتراپلوئید، ارقام زراعی و توده‌های بومی هگزاپلوئید را به‌خوبی از هم تفکیک نمود. از آنجایی که عوامل متعددی می‌توانند بر تنوع ژنتیکی و ساختار جمعیت در مجموعه ژرم‌پلاسم تأثیر بگذارند، بنابراین می‌توان اظهار داشت که تجزیه و تحلیل‌های ارائه شده در مطالعه حاضر تنها می‌تواند کسری از تنوع و ساختار ژنتیکی موجود در نمونه‌های گندم هگزاپلوئید و تتراپلوئید ایرانی را توضیح دهد و برای

References

- Abdi, H., Alipour, H., Bernousi, I., Jafarzadeh, J. and Rodrigues, P.C.** (2023). Identification of novel putative alleles related to important agronomic traits of wheat using robust strategies in GWAS. *Scientific Reports*, **13**(1): 9927.
- Alemu, A., Feyissa, T., Letta, T. and Abeyo, B.** (2020). Genetic diversity and population structure analysis based on the high-density SNP markers in Ethiopian durum wheat (*Triticum turgidum* ssp. durum). *BMC Genetics*, **21**(1): 1-12.
- Alipour, H.** (2016). Association mapping of the main agronomic traits in bread wheat. PhD Thesis. University of Tehran, Tehran, Iran (In Persian).
- Alipour, H., Bai, G., Zhang, G., Bihamta, M.R., Mohammadi, V. and Peyghambari, S.A.** (2019). Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. *PLoS One*, **14**(1): e0208614.
- Alipour, H., Bihamta, M.R., Mohammadi, V., Peyghambari, S.A., Bai, G. and Zhang, G.** (2017). Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Frontiers in Plant Science*, **8**: 1293.
- Ayalew, H., Sorrells, M.E., Carver, B.F., Baenziger, P.S. and Ma, X.F.** (2020). Selection signatures across seven decades of hard winter wheat breeding in the Great Plains of the United States. *The Plant Genome*, **13**(3): e20032.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S.** (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**(19): 2633-2635.
- Campoy, J.A., Lerigoleur-Balsemin, E., Christmann, H., Beauvieux, R., Girollet, N., Quero-García, J., Dirlewanger, E. and Barreneche, T.** (2016). Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC Plant Biology*, **16**(1): 1-15.
- Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G.L., Akhunova, A. and See, D.** (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences*, **110**(20): 8057-8062.
- Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.C., Baenziger, S.P., Matnyazov, R., Clark, D.R., Talbert, L.E., Anderson, J.A., Dreisigacker, S. and Glover, K.** (2010). Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genomics*, **11**(1): 1-17.
- Deperi, S.I., Tagliotti, M.E., Bedogni, M.C., Manrique-Carpintero, N.C., Coombs, J., Zhang, R., Douches, D. and Huarte, M.A.** (2018). Discriminant analysis of principal components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *PLoS One*, **13**(3): e0194398.

- Elhaik, E.** (2022). Principal component analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, **12(1)**: 14683.
- Fayaz, F., Aghaee Sarbarzeh, M., Talebi, R. and Azadi, A.** (2019). Genetic diversity and molecular characterization of Iranian durum wheat landraces (*Triticum turgidum durum* (Desf.) Husn.) using DArT markers. *Biochemical Genetics*, **57**: 98-116.
- Fiore, M.C., Blangiforti, S., Preiti, G., Spina, A., Bosi, S., Marotti, I., Mauceri, A., Puccio, G., Sunseri, F. and Mercati, F.** (2022). Elucidating the genetic relationships on the original old Sicilian *Triticum* Spp. collection by SNP genotyping. *International Journal of Molecular Sciences*, **23(21)**: 13378.
- Getachew, S.E., Ngalle, H.B., Joseph, M.B. and Wosene, G.** (2019). Genotyping by sequencing for plant breeding- A review. *Advances in Microbiology and Biotechnology*, **14**: 555891.
- Ghorbani, R., Chasemzadeh, R. and Alipour, H.** (2023). Genome-wide association study of seedling characteristics in bread wheat cultivars under normal and salt stress conditions. *Plant Genetic Researches*, **9(1)**: 13-26 (In Persian).
- Hussain, S., Habib, M., Ahmed, Z., Sadia, B., Bernardo, A., Amand, P.S., Bai, G., Ghori, N., Khan, A.I., Awan, F.S. and Maqbool, R.** (2022). Genotyping-by-sequencing based molecular genetic diversity of Pakistani bread wheat (*Triticum aestivum* L.) accessions. *Frontiers in Genetics*, **13**: 772517.
- Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I. and Andrew, R.L.** (2017). The K= 2 conundrum. *Molecular Ecology*, **26(14)**: 3594-3602.
- Jombart, T.** (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24(11)**: 1403-1405.
- Jombart, T., Devillard, S. and Balloux, F.** (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11(1)**: 1-15.
- Kabbaj, H., Sall, A.T., Al-Abdallat, A., Geleta, M., Amri, A., Filali-Maltouf, A., Belkadi, B., Ortiz, R. and Bassi, F.M.** (2017). Genetic diversity within a global panel of durum wheat (*Triticum durum*) landraces and modern germplasm reveals the history of alleles exchange. *Frontiers in Plant Science*, **8**: 1277.
- Khadka, K., Torkamaneh, D., Kaviani, M., Belzile, F., Raizada, M.N. and Navabi, A.** (2020). Population structure of Nepali spring wheat (*Triticum aestivum* L.) germplasm. *BMC Plant Biology*, **20(1)**: 1-12.
- Liu, L., Zhang, D., Liu, H. and Arendt, C.** (2013). Robust methods for population stratification in genome wide association studies. *BMC Bioinformatics*, **14**: 1-12.
- Mahboubi, M., Mehrabi, R., Naji, A.M. and Talebi, R.** (2020). Whole-genome diversity, population structure and linkage disequilibrium analysis of globally diverse wheat genotypes using genotyping-by-sequencing DArTseq platform. *3 Biotech*, **10(2)**: 48.
- Miazi, M.M., Babay, E., De Vita, P., Montemurro, C., Chaabane, R., Taranto, F. and Mangini, G.** (2022). Comparative genetic analysis of durum wheat landraces and cultivars widespread in Tunisia. *Frontiers in Plant Science*, **13**: 939609.
- MirMohammadi Maibody, S.A.M. and Golkar, P.** (2019). Application of DNA molecular markers in plant breeding. *Plant Genetic Researches*, **6(1)**: 1-30 (In Persian).
- Müller, T., Schierscher-Viret, B., Fossati, D., Brabant, C., Schori, A., Keller, B. and Krattinger, S.G.** (2018). Unlocking the diversity of genebanks: whole-genome marker analysis of Swiss bread wheat and spelt. *Theoretical and Applied Genetics*, **131**: 407-416.
- Nielsen, N.H., Backes, G., Stougaard, J., Andersen, S.U. and Jahoor, A.** (2014). Genetic diversity and population structure analysis of European hexaploid bread wheat (*Triticum aestivum* L.) varieties. *PLoS One*, **9(4)**: e94000.
- Patterson, N., Price, A.L. and Reich, D.** (2006). Population structure and eigenanalysis. *PLoS Genetics*, **2(12)**: e190.
- Peakall, R.O.D. and Smouse, P.E.** (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6(1)**: 288-295.
- Peng, J.H., Sun, D. and Nevo, E.** (2011). Domestication evolution, genetics and genomics in wheat. *Molecular Breeding*, **28**: 281-301.

- Poland, J.A., Brown, P.J., Sorrells, M.E. and Jannink, J.L.** (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *Plos One*, **7(2)**: e32253.
- Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D.** (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38(8)**: 904-909.
- Pritchard, J.K., Stephens, M. and Donnelly, P.** (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155(2)**: 945-959.
- Pritchard, J.K., Wen, X. and Falush, D.** (2010). *Documentation for Structure Software*. University of Chicago, Chicago, IL, USA.
- Privé, F., Luu, K., Blum, M. G., McGrath, J.J. and Vilhjálmsson, B.J.** (2020). Efficient toolkit implementing best practices for principal component analysis of population genetic data. *Bioinformatics*, **36(16)**: 4449-4457.
- Qin, X., Chiang, C.W. and Gaggiotti, O.E.** (2022). KLFDA: a supervised machine learning approach for spatial genetic structure analysis. *Briefings in Bioinformatics*, **23(4)**: bbac202.
- Rabieyan, E., Bihamta, M. R., Moghaddam, M.E., Mohammadi, V. and Alipour, H.** (2022). Genome-wide association mapping and genomic prediction for pre-harvest sprouting resistance, low α -amylase and seed color in Iranian bread wheat. *BMC Plant Biology*, **22(1)**: 1-23.
- Rahimi, Y., Bihamta, M.R., Taleei, A., Alipour, H. and Ingvarsson, P.K.** (2019). Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. *BMC Plant Biology*, **19(1)**: 1-19.
- Rasheed, A., Xia, X., Mahmood, T., Quraishi, U.M., Aziz, A., Bux, H., Mahmood, Z., Mirza, J.I., Mujeeb-Kazi, A. and He, Z.** (2016). Comparison of economically important loci in landraces and improved wheat cultivars from Pakistan. *Crop Science*, **56(1)**: 287-301.
- Rimbert, H., Darrier, B., Navarro, J., Kitt, J., Choulet, F., Leveugle, M., Duarte, J., Rivière, N., Eversole, K., International Wheat Genome Sequencing Consortium and Le Gouis, J.** (2018). High throughput SNP discovery and genotyping in hexaploid wheat. *PloS One*, **13(1)**: e0186329.
- Sansaloni, C., Franco, J., Santos, B., Percival-Alwyn, L., Singh, S., Petroli, C., Campos, J., Dreher, K., Payne, T., Marshall, D. and Kilian, B.** (2020). Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nature Communications*, **11(1)**: 4572.
- Sharifi-Tehrani, M.** (2018). A review on the analysis of population genetic structure using dominant molecular markers and introducing the new program STRUCTUREEasy. *Taxonomy and Biosystematics*, **10(36)**: 35-47 (In Persian).
- Sthapit, S.R., Ruff, T.M., Hooker, M.A. and See, D.R.** (2022). Population structure and genetic diversity of US wheat varieties. *The Plant Genome*, **15(2)**: e20196.
- Tehseen, M.M., Tonk, F.A., Tosun, M., Istipliler, D., Amri, A., Sansaloni, C.P., Kurtulus, E., Mubarik, M.S. and Nazari, K.** (2022). Exploring the genetic diversity and population structure of wheat landrace population conserved at ICARDA GenBank. *Frontiers in Genetics*, **13**: 900572.
- Vikram, P., Franco, J., Burgueño, J., Li, H., Sehgal, D., Saint-Pierre, C., Ortiz, C., Singh, V.K., Sneller, C., Sharma, A. and Tattaris, M.** (2021). Strategic use of Iranian bread wheat landrace accessions for genetic improvement: Core set formulation and validation. *Plant Breeding*, **140(1)**: 87-99.