

افزایش صحت پیش‌بینی مدل‌های انتخاب ژنومی با بهینه‌سازی مدل‌ها توسط متغیرهای کمکی در گندم نان (*Triticum aestivum* (L.

Prediction accuracy improvement of genomic selection models through the optimization of models using covariates in bread wheat (*Triticum aestivum* L.)

مجید جهرمی^۱، علیرضا عباسی^{۱*}، محمدرضا بی‌همتا^{۱*}، محمد آرمیون^۲

۱- به‌ترتیب، دانشجوی دکتری ژنتیک و به‌نژادی گیاهی، استاد گروه زراعت و اصلاح نباتات، دانشکده‌گان

کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران

۲- استادیار، بخش تحقیقات علوم زراعی و باغی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی ایلام،

سازمان تحقیقات، آموزش و ترویج کشاورزی، ایلام، ایران

Jahromi M¹, Abbasi A^{*1}, Bihamta MR^{*1}, Armion M²

1- Respectively, PhD Student in Genetics and Plant Breeding, Professor, Professor, Department of Agronomy and Plant Breeding, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

2- Assistant Professor of Crop and Horticultural Science Research Department, Ilam Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Ilam, Iran

* نویسنده مسئول مکاتبات، پست الکترونیکی: mrghanad@ut.ac.ir ; rezabbasi@ut.ac.ir

(تاریخ دریافت: ۱۴۰۳/۰۹/۱۹ - تاریخ پذیرش: ۱۴۰۴/۱۱/۱۲)

چکیده

پیش‌بینی ژنومی یک رویکرد امیدوارکننده برای تسریع پیشرفت ژنتیکی صفات پیچیده در به‌نژادی گندم است. با این حال، افزایش صحت پیش‌بینی مدل‌های انتخاب ژنومی یک چالش جدی در اجرای موفقیت‌آمیز این رویکرد است. در این خصوص، استفاده از متغیرهای کمکی زمانی که شرایط لازم را داشته باشند، نوید بخش بوده است. هدف از این مطالعه بررسی نقش دو متغیر کمکی، یعنی تعداد روز تا سنبله‌دهی (DH) و تعداد روز تا گل‌دهی (DF) در بهینه‌سازی و افزایش صحت پیش‌بینی هفت مدل مختلف انتخاب ژنومی (GS)، شامل GBLUP، RRBLUP و مدل‌های بیزی (BayesA، BayesB، BayesC، BayesRR و BayesLASSO) بود. برای این منظور از دو راهبرد یعنی استفاده از مدل‌های انتخاب ژنومی بدون متغیرهای کمکی (راهبرد اول) و همراه با متغیرهای کمکی (راهبرد دوم)، برای صفات عملکرد دانه (GY)، وزن هزاردانه (TKW)، عملکرد بیولوژیک (BY)، تعداد روز تا رسیدگی کامل (DFM) و طول دوره پرشدن دانه (GFP) استفاده شد. پژوهش در مزرعه تحقیقاتی دانشگاه تهران با استفاده از ۱۰۰ ژنوتیپ در دو سال زراعی ۱۴۰۰-۱۴۰۱ و ۱۳۹۹-۱۴۰۰ انجام شد. متغیرهای کمکی علاوه بر همبستگی‌های فنوتیپی و ژنوتیپی بالا با یکدیگر، همبستگی‌های مثبت و منفی معنی‌داری با صفات مورد مطالعه داشتند. همچنین وراثت‌پذیری صفات از ۶۹٪ (متعلق به صفت GFP) تا ۹۷٪ (برای صفت DH) متغیر بود. پیش‌بینی ژنومی توسط مدل‌ها، با استفاده از ۶۵۹۴۸ نشانگر SNP همراه با داده‌های فنوتیپی انجام شد. با افزودن متغیرهای کمکی DH و DF به مدل‌ها، بیشترین صحت پیش‌بینی GY در راهبرد اول، از ۰/۰۸۵ به ۰/۲۷۹ در راهبرد دوم و به همین ترتیب برای TKW از ۰/۲۳۸ به ۰/۴۱۱، BY از ۰/۲۳ به ۰/۳۹۹، GFP از ۰/۳۱۴ به ۰/۷۱۴ و DFM از ۰/۲۵۶ به ۰/۸۸۹ افزایش یافت. در راهبرد دوم، مدل‌های بیزی به‌ویژه BayesRR و BayesLASSO، بالاترین صحت‌های پیش‌بینی را نشان دادند. این یافته‌ها حاکی از آن است که افزودن متغیرهای کمکی مناسب به مدل‌های GS، می‌تواند ابزاری مؤثر برای بهینه‌سازی مدل‌ها و ارتقاء صحت پیش‌بینی ژنومی برخی صفات فنولوژیک و زراعی گندم باشد.

واژه‌های کلیدی

انتخاب ژنومی
پیش‌بینی ژنومی
صحت پیش‌بینی
متغیر کمکی
مدل‌های بیزی

ژنومیک و انتخاب به‌کمک نشانگر می‌توانند در برنامه‌های اصلاحی گنجانده شوند و این امر اغلب موجب افزایش دو برابری نرخ پیشرفت ژنتیکی در مقایسه با انتخاب فنوتیپی استاندارد می‌شود (Eathington et al. 2007; Battenfield et al. 2016).

به‌طور معمول، MAS ارزش صفات را تنها بر اساس تعداد محدودی از نشانگرهای مرتبط با جایگاه‌های صفت کمی^۲ (QTL) با اثرات بزرگ پیش‌بینی می‌کند (Collard and Mackill 2008). اگرچه MAS در پیش‌بینی فنوتیپی صفات مندلی مفید است، اما بسیاری از صفات زراعی مهم پیچیده‌اند؛ به این معنا که این صفات تحت تاثیر بسیاری از مؤلفه‌های ژنتیکی با اندازه اثرهای متفاوت (اغلب کوچک) قرار دارند (Barton et al. 2017). هنگامی که منابع ژنومی موثر بر یک صفت شامل هزاران ژن با اثر کوچک باشد، انتخاب بر اساس یک یا چند نشانگر ژنتیکی از نظر تئوری بی‌اثر خواهد بود (Xu and Crouch 2008). از این رو، کاربرد عملی MAS با محدودیت‌های زیادی روبرو بوده است (Bernardo, 2008)، عمدتاً به این دلیل که QTL‌های شناسایی شده تنها بخش محدودی از واریانس ژنتیکی را به خود اختصاص می‌دهند (Tessem et al. 2020). علاوه بر این، پیشرفت ژنتیکی نسبتاً کمی با استفاده از تعداد اندکی نشانگر قابل دستیابی است (Abdollahi Arpanahi 2012). بنابراین، برای دستیابی به پیشرفت‌های قابل توجه در انتخاب، روشی پیچیده‌تر و کارآمدتر از MAS مورد نیاز است.

علاوه بر MAS، استراتژی برجسته دیگری به نام انتخاب ژنومی^۳ (GS) وجود دارد. اگرچه هر دو روش بر تجزیه و تحلیل‌های آماری نشانگرهای ژنتیکی برای برآورد سهم هر نشانگر در تنوع فنوتیپی تکیه دارند (Bernardo 2010)، اما GS مبتنی بر فرضیه و مدل بی‌نهایت ژنگاه با اثرات جزئی^۴ (IFM) می‌باشد. این فرضیه بیان می‌کند که ارزش یک صفت نتیجه ترکیب خطی منابع ژنتیکی افزایشی و غیرژنتیکی می‌باشد (Fisher 1919). GS که اولین بار توسط Meuwissen et al. (2001) پیشنهاد شد، اثر تمامی

طبق برآوردهای مرکز بین‌المللی تحقیقات ذرت و گندم، نیاز به گندم و محصولات آن تا سال ۲۰۵۰ حداقل ۵۰ درصد افزایش خواهد یافت (CIMMYT 2021). در حالی که عملکرد گندم از دهه ۱۹۵۰ به دلیل بهبود و تکامل شیوه‌های به‌نژادی و به‌زراعی به سرعت افزایش یافته است، اما با این وجود نرخ فعلی در تامین نیازهای کالری بشر در آینده کافی نیست و در برخی موارد این نرخ ثابت مانده است (Graybosch and Peterson 2010). برای افزایش دو برابری تولید گندم تا سال ۲۰۵۰، بهبود عملکرد در حدود ۲/۴ درصد در سال، در مقایسه با ۰/۹ درصد فعلی، مورد نیاز است (Ray et al. 2013). افزایش سطح زیر کشت نیز به دلیل محدودیت اراضی زراعی و چالش‌های اقلیمی راهکار پذیرفته شده‌ای برای تحقق نیازهای آینده محسوب نمی‌شود (Jaggard et al. 2010). گزارش شده است که برنامه‌های کلاسیک و معمول به‌نژادی گندم که از ارزش فنوتیپی برای انتخاب بهترین افراد استفاده می‌کنند، پیشرفت ژنتیکی سالانه کمتر از یک درصد دارند (Godin et al. 2012; Tessem et al. 2020). وراثت‌پذیری پایین صفات کمی، اثرات محیطی، اثرات متقابل ژنتیکی و همچنین مدت زمان طولانی روش‌های اصلاحی کلاسیک، انتخاب بر اساس فنوتیپ را با چالش‌های جدی مواجه می‌سازد. بنابراین، برای افزایش پیشرفت ژنتیکی و تضمین پایداری تامین منابع غذایی، به‌نژادگران و متخصصان ژنتیک باید از روش‌های نوین به‌منظور تقویت استراتژی‌های اصلاحی موجود استفاده کنند (Crain et al. 2018). پیشرفت‌های اخیر در فناوری‌های تعیین توالی ژنوم، این امکان را برای به‌نژادگران فراهم کرد تا معماری ژنتیکی غلات مهم اقتصادی را مشخص کنند. لذا با ظهور و تکامل نشانگرهای مولکولی، تصمیم‌گیری برای گزینش با ادغام اطلاعات نشانگرهای مولکولی و داده‌های فنوتیپی، از طریق انتخاب به‌کمک نشانگر^۱ (MAS)، امکان‌پذیر شد (Fujino et al. 2019). برآورد ژنومی ارزش‌های اصلاحی، مزایای بیشتری نسبت به انتخاب فنوتیپی دارد، زیرا صحت برآورد ارزش‌های اصلاحی زمانی بالاتر است که اطلاعات ژنومی در تصمیم‌گیری در نظر گرفته شوند (Daetwyler et al. 2013). مطالعات نشان داده‌اند که

^۱ Marker Assisted Selection

^۲ Quantitative Trait Loci

^۳ Genomic Selection

^۴ Infinitesimal Model

درست‌نمایی برای این اثرات به‌دست می‌آید (Gianola 2013). با توجه به اینکه یک خطر مهم در توسعه مدل‌های پیش‌بینی، مسئله برآورد اغراق‌آمیز است و این می‌تواند توانایی پیش‌بینی را تضعیف کند، برای چیره شدن بر این مسئله، طیفی از روش‌های آماری برای مدل‌های GS توسعه یافته است. این روش‌ها عمدتاً در مفروضات توزیع اثرات نشانگر متفاوت هستند و از رویکردهای انتخاب متغیر⁶، جریمه‌ای⁷ یا انقباضی⁸ به‌صورت جداگانه یا ترکیبی استفاده می‌کنند (Wang et al. 2018; Plavšín et al. 2021).

بیشتر مدل‌های GS، مدل‌های تک صفتی هستند و اطلاعات فنوتیپی را فقط در مورد صفت هدف در خود جای می‌دهند (Jia and Jannink 2012). اگرچه مفید بودن مدل‌های تک‌صفتی ثابت شده است (Heffner et al. 2009, 2011)، این مدل‌ها از همبستگی بین صفات بهره نمی‌برند (Jia and Jannink 2012). گاهی اوقات یک صفت، که به‌عنوان متغیر پاسخ اصلی⁹ (MRV) شناخته می‌شود، ممکن است تحت تأثیر یک صفت یا مجموعه‌ای از صفات دیگر به‌عنوان متغیر(های) کمکی¹⁰ قرار گیرد. از نظر آماری، این امر دلالت بر وجود همبستگی یا ارتباط بالا بین MRV و متغیر(های) کمکی دارد. بسیاری از صفات با یکدیگر همبستگی مثبت یا منفی دارند، زیرا توسط ژن‌های یکسان کنترل می‌شوند. اهمیت متغیرهای کمکی در این است که می‌توانند تخمین عملکرد افراد را بهبود بخشند و خطای آزمایشی را کاهش دهند (Alvarado et al. 2020). در این صورت، این متغیرها می‌توانند به مدل در شناسایی اثرات پیچیده ژنتیکی و در نتیجه افزایش صحت پیش‌بینی کمک کنند. با این وجود، اغلب عملکرد دانه توسط مدل‌های پایه GS پیش‌بینی می‌شود، اما فرآیندهای متعددی وجود دارند که به عملکرد دانه ختم می‌شوند (Pask et al. 2012). بهبود صحت پیش‌بینی روش‌های انتخاب ژنومی یکی از اهداف و دغدغه‌های محققین بوده است. در گندم، انتخاب با افزودن صفاتی به‌عنوان صفات ثانویه یا متغیرهای کمکی در مدل‌های

نشانگرهای ژنتیکی موجود در سراسر ژنوم را برای پیش‌بینی ارزش‌های اصلاحی ژنومی¹ (GEBVs) در نظر می‌گیرد، به جای آنکه تنها به نشانگرهایی که از یک آستانه معنی‌داری عبور می‌کنند، توجه شود. این روش با تکیه بر مدل IFM، پایه‌های ژنومی یک صفت پیچیده را تخمین می‌زند. بنابراین انجام GS نیاز به تخمین اثر هر نشانگر دارد، که در عمل معمولاً به‌صورت افزایشی در نظر گرفته می‌شود، اگرچه روش‌هایی برای گنجاندن اثرات غالبیت (Technow et al. 2012)، اپیستاتیک (Jiang and Reif, 2015) و ژنوتیپ × محیط (Cuevas et al. 2016; Crossa et al. 2018; Montesinos-López et al. 2016) در دسترس جامعه تحقیقاتی قرار دارد. GS از یک جمعیت مرجع یا آموزش² (TP) که ژنوتیپ و فنوتیپ‌سنجی شده است، استفاده می‌کند تا مدلی را ایجاد کند که بتواند برای محاسبه GEBVs افراد جمعیت اصلاحی³ (BP)، فقط از داده‌های ژنوتیپی آن‌ها استفاده کند (Jannink et al. 2010). بنابراین، از آنجایی که روش GS باعث کاهش هزینه‌های ژنوتیپ‌سنجی، فنوتیپ‌سنجی و زمان چرخه اصلاحی می‌شود و با توجه به اینکه پیشرفت ژنتیکی قابل توجهی نسبت به انتخاب فنوتیپی و MAS ارائه می‌دهد، این رویکرد احتمالاً در حال ایجاد انقلابی در برنامه‌های به‌نژادی است و پایه و بنیان محکمی برای تحقیقات بیشتر در زمینه انتخاب ژنومی گندم فراهم می‌کند (Heffner et al. 2009; Heslot et al. 2015; Plavšín et al. 2021).

ژنوتیپ‌سنجی با توان بالا⁴ تعداد زیادی داده نشانگر تولید می‌کند که در GS مورد استفاده قرار می‌گیرد. بنابراین، به دلیل ابعاد بالای داده‌های ژنتیکی، برآزش مدل‌های GS نیازمند توجه به این واقعیت است که تعداد نشانگرها یا متغیرهای پیش‌بینی کننده⁵ (p) موجود در یک مطالعه، از تعداد افراد (n) بیشتر است (de los Campos et al. 2013). در نتیجه، هنگامی که مدل GS که اثرات افزایشی نشانگرها را در نظر می‌گیرد، روی چنین داده‌هایی ($p > n$) برآزش داده می‌شود، تعداد بی‌نهایت برآورد حداکثر

⁶ Variable Selection

⁷ Penalized approach

⁸ Shrinkage

⁹ Main Response Variable

¹⁰ Covariate

¹ Genomic Estimated Breeding Values

² Training Population

³ Breeding Population

⁴ High-throughput genotyping

⁵ Predictor variables

شد. آزمایش در ۳۰۰ کرت، هر یک دارای چهار خط، با فاصله ۰/۲۵ متر و با طول یک متر اجرا شد.

جمع‌آوری داده‌های فنوتیپی

در این آزمایش صفات تعداد روز تا سنبله‌دهی^۳، تعداد روز تا گلدهی^۴، تعداد روز تا رسیدگی کامل^۵ و طول دوره پرشدن دانه^۶ (از اینجا به بعد به ترتیب به اختصار DH، DF، DFM و GFP) و همچنین صفات عملکرد دانه^۷، عملکرد بیولوژیک^۸ و وزن هزار دانه^۹ (از اینجا به بعد به ترتیب به اختصار BY، GY و TKW) مورد ارزیابی قرار گرفتند. فاصله تعداد روز از زمان کاشت تا رسیدن به مرحله فنولوژیکی مدنظر (Zadoks et al. 1974) با در نظر گرفتن ورود ۵۰ درصد گیاهان هر کرت آزمایشی به هر یک از مراحل فنولوژیکی و با حذف اثر حاشیه، به‌عنوان زمان وقوع آن مرحله نمودی برای هر ژنوتیپ ثبت شد. همچنین عملکرد دانه، عملکرد بیولوژیک و وزن هزار دانه بعد از رسیدگی و برداشت کرت‌ها به‌وسیله کمباین، برای هر دو سال اندازه‌گیری شد.

تجزیه و تحلیل داده‌ها

بهترین تخمین‌های ناریب خطی^{۱۰} (BLUEs)، بهترین پیش‌بینی‌های ناریب خطی^{۱۱} (BLUPs) و مؤلفه‌های واریانس^{۱۲} برای تجزیه مرکب (معادله ۱) با استفاده از نرم‌افزار META-R (Alvarado et al. 2020) و بسته LME4 در نرم‌افزار R (Bates et al. 2015) برآورد شدند. فرض تصادفی برای اثر ژنوتیپ، امکان محاسبه وراثت‌پذیری عمومی را فراهم می‌کند. با این حال ژنوتیپ‌ها در هنگام محاسبه BLUEها به‌عنوان اثر ثابت در نظر گرفته می‌شوند.

$$Y_{ijk} = \mu + Yr_i + Rep_j(Yr_i) + Block_k(Yr_i Rep_j) + Gen_l + Yr_i \times Gen_l + \varepsilon_{ijk} \quad (1)$$

³ Days to Heading

⁴ Days to Flowering

⁵ Days to Full Maturity

⁶ Grain Filling Period

⁷ Grain Yield

⁸ Biological Yield

⁹ Thousand-Kernel Weight

¹⁰ Best Linear Unbiased Estimates

¹¹ Best Linear Unbiased Predictions

¹² Variance Components

انتخاب ژنومی تک متغیره و چند متغیره تسریع می‌شود (Haghighattalab et al. Rutkoski et al. 2016; Sun et al. 2017); (2016). طبق یافته‌های (Rutkoski et al. 2016)، افزودن صفات فنوتیپی با استفاده از مدل بهترین پیش‌بینی ناریب خطی ژنومی^۱ (GBLUP) چندمتغیره، می‌تواند صحت پیش‌بینی را در مقایسه با استفاده صرف از مدل‌های تک متغیره، بهبود بخشد. دمای کانوپی و شاخص نرمال شده تفاوت پوشش گیاهی^۲ (NDVI) به‌عنوان صفات ثانویه در گندم، صحت پیش‌بینی عملکرد دانه را تا ۷۰ درصد بهبود داده‌اند (Sun et al. 2017). همچنین، استفاده از یک QTL اصلی به‌عنوان اثر ثابت در مدل GP، می‌تواند صحت پیش‌بینی را افزایش دهد (Zhang et al. 2017). برای صفت تعداد روز تا رسیدگی، با استفاده از نسبت‌های تحمل و تنش به‌عنوان متغیرهای کمکی در مدل‌های انتخاب ژنومی تک متغیره، بهبود سه تا چهار برابری در صحت پیش‌بینی‌های ژنومی توسط (Shabannejad et al. 2020) گزارش شده‌است. بر اساس این یافته‌ها، هدف از این مطالعه بررسی و مقایسه صحت پیش‌بینی مدل‌های انتخاب ژنومی، با استفاده از مدل‌های GS بدون متغیر کمکی (راهبرد اول) و همراه با متغیر کمکی (راهبرد دوم) و همچنین ارزیابی نقش متغیرهای کمکی در بهینه‌سازی و بهبود صحت پیش‌بینی مدل‌ها است.

مواد و روش‌ها

مواد گیاهی و شرایط آزمایش

مواد گیاهی مورد استفاده شامل ۹۶ رقم گندم تجاری به‌همراه چهار توده بومی از بانک ژن ایران، واقع در پردیس کشاورزی و منابع طبیعی دانشگاه تهران بود. آزمایش‌ها در مزرعه تحقیقاتی گروه زراعت و اصلاح نباتات دانشگاه تهران واقع در محمدشهر کرج (طول جغرافیایی ۵۸°۵۱' شرقی، عرض جغرافیایی ۴۸°۳۵' شمالی و ارتفاع ۱۳۲۱ متر از سطح دریا) به مدت دو سال متوالی، در فصول زراعی ۱۴۰۰-۱۳۹۹ و ۱۴۰۱-۱۴۰۰ با استفاده از طرح لاتیس سه‌گانه (۱۰×۱۰) در شرایط معمولی با سه تکرار انجام

¹ Best Linear Unbiased Prediction

² Normalized difference vegetation index

σ_{g_i} و σ_{g_j} جذر واریانس‌های ژنوتیپی صفات i و j است. ژنوتیپ‌سنجی از طریق توالی‌یابی^۲ (GBS) و فراخوانی SNPها^۳

تهیه و توالی‌یابی کتابخانه‌های GBS برای گندم ایرانی پیش‌تر توسط (Alipour et al. 2017)، شرح داده شده‌است. به‌طور خلاصه در این فرآیند پس از برش خوانش‌های توالی‌یابی شده به‌طول ۶۴ جفت باز و گروه‌بندی آن‌ها در تگ‌های توالی^۴، SNPها بر اساس هم‌ترازسازی داخلی^۵ که اجازه حداکثر سه جفت باز عدم تطابق^۶ را می‌داد، شناسایی شدند. برای فراخوانی SNPها از پایپ‌لاین GBS که با نام UNEAK^۷ شناخته می‌شود، استفاده شد و خوانش‌هایی که امتیاز کیفیت پایینی (کمتر از ۱۵) داشتند، حذف شدند. جانهی داده‌ها^۸ با استفاده از BEAGLE نسخه ۳،۳،۲ (Browning et al. 2009) و ژنوم مرجع w7984 انجام شد (Alipour et al. 2019). در نهایت SNPهایی با هتروزیگوسیتی کمتر از ۱۰٪ و فراوانی آلل جزئی بیشتر از ۱٪ برای تجزیه و تحلیل‌های بعدی در نظر گرفته شدند.

انتخاب ژنومی و راهبردها

برای پیش‌بینی ژنومی، هفت مدل شامل بهترین پیش‌بینی ناریب خطی ژنومی (GBLUP)، بهترین پیش‌بینی ناریب خطی رگرسیون ریج^۹ (RRBLUP) و مدل‌های بیزی شامل BayesB، BayesA، BayesC، BayesRR^{۱۰} و BayesLASSO^{۱۱} استفاده شدند. مروری کوتاه بر این مدل‌ها، توسط (Plavšinet al. 2021) and (Juliana et al. 2017) ارائه شده است. دو راهبرد انتخاب ژنومی شامل ۱- اجرای مدل‌های انتخاب ژنومی بدون متغیر کمکی ($S_1=GS$) و ۲- اضافه کردن متغیرهای کمکی DF و DH به مدل‌ها ($S_2DH=GS+DH$ و $S_2DF=GS+DF$)، برای پنج صفت (DFM)

که در این معادله Y_{ijk} مقدار صفت مورد نظر، μ اثر میانگین کل، Y_{ri} اثر سال i ام، $Rep_i(Y_{ri})$ اثر تکرار i ام درون سال i ام، $Block_k$ اثر بلوک ناقص k ام درون سال i ام، تکرار i ام، اثر Gen_i ژنوتیپ i ام، $Y_{ri} \times Gen_i$ اثر متقابل سال i × ژنوتیپ هستند. ϵ_{ijkl} اثر خطای مربوط به سال i ام، تکرار i ام، بلوک ناقص k ام و ژنوتیپ i ام که فرض می‌شود به‌صورت مستقل و هم توزیع^۱ (i.i.d) از توزیع نرمال با میانگین صفر و واریانس σ^2_ϵ پیروی کند. وراثت‌پذیری عمومی هر صفت در تجزیه مرکب مطابق معادله (۲) محاسبه شد:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ge}^2 / nYr + \sigma_e^2 / (nYr \times nRep)} \quad (2)$$

در این معادله، σ_g^2 ، σ_{ge}^2 و σ_e^2 به ترتیب، مؤلفه‌های واریانس خطا، ژنوتیپ و اثر متقابل ژنوتیپ × سال، $nRep$ و nYr به ترتیب، تعداد سال‌ها و تکرارها هستند. نحوه توزیع صفات، محاسبه ضرایب همبستگی پیرسون (r_p) بین صفات مورد مطالعه و همچنین نمایش نمودارهای مربوطه با استفاده از بسته‌های `ggpubr`، `psych` و `ggplot2` در محیط R (R Core Team 2018) انجام شد. همبستگی ژنتیکی (ρ_g) بین صفات، طبق آنچه توسط (Alvarado et al. 2020) گزارش شده است، برآورد شد. به‌طوری که، مدل به‌صورت جداگانه برای صفت i ، برای صفت j و برای مجموع هر دو صفت i و j برآزش داده می‌شود. سپس برآوردی از کوواریانس ژنوتیپی بین صفات i و j به‌دست می‌آید، بصورت زیر:

$$\sigma_{g_{ij}}^2 = \frac{1}{2} (\sigma_{g_{i+j}}^2 - \sigma_{g_i}^2 - \sigma_{g_j}^2) \quad (3)$$

که $\sigma_{g_{i+j}}^2$ کوواریانس ژنوتیپی مجموع صفات i و j است؛ و $\sigma_{g_i}^2$ و $\sigma_{g_j}^2$ به ترتیب، واریانس ژنوتیپی صفات i و j هستند. در نهایت همبستگی ژنتیکی بین صفات i و j به صورت زیر محاسبه می‌شود:

$$\rho_{g_{ij}} = \frac{\sigma_{g_{ij}}^2}{\sigma_{g_i} \sigma_{g_j}} \quad (4)$$

¹ independent and identically distributed

² Genotyping by sequencing

³ Single Nucleotide Polymorphisms

⁴ Sequence tags

⁵ Internal alignment

⁶ Mismatch

⁷ Universal Network-Enabled Analysis Kit

⁸ Imputation

⁹ Ridge-Regression Best Linear Unbiased Prediction

¹⁰ Bayesian Ridge Regression

¹¹ Bayesian Least Absolute Shrinkage and Selection Operator

نتایج

ارزیابی داده های فنوتیپی

تجزیه واریانس مرکب تفاوت معنی داری را در بین ژنوتیپ ها برای تمامی صفات مورد مطالعه شناسایی کرد ($p < 0.001$). اثر متقابل ژنوتیپ \times سال نیز برای تمامی صفات به طور قابل توجهی معنی دار بود ($p < 0.001$). همچنین اثر سال در سطوح معنی داری مختلف برای همه صفات معنی دار بود (جدول ۱). وراثت پذیری صفات از ۶۹٪ (متعلق به صفت GFP) تا ۹۷٪ (صفت DH) متغیر بود (جدول ۱). بر اساس نتایج همبستگی فنوتیپی (r_p)، صفات DH و DF به عنوان متغیر کمکی در این مطالعه، علاوه بر همبستگی بسیار بالا با یکدیگر، با صفات TKW (به ترتیب $r_p = -0.31^{**}$ و $r_p = -0.30^{**}$)، GFP ($r_p = -0.69^{***}$) و $r_p = -0.69^{***}$ و $r_p = -0.29^{**}$ و $r_p = -0.29^{**}$) GY و همبستگی منفی معنی دار و با صفات DFM ($r_p = 0.88^{***}$) و $r_p = 0.89^{***}$) و $r_p = 0.41^{***}$ و $r_p = 0.41^{***}$) BY همبستگی مثبت معنی دار داشتند. در این خصوص الگوی همبستگی های ژنتیکی (ρ_g) نیز مشابهت زیادی با همبستگی های فنوتیپی داشت (شکل ۱). همچنین همبستگی فنوتیپی و ژنتیکی مثبت و معنی دار، بین GY با صفات GFP ($r_p = 0.45^{***}$ و $\rho_g = 0.57^{***}$) و BY ($r_p = 0.59^{***}$ و $\rho_g = 0.55^{***}$) وجود داشت. پایین ترین همبستگی بین صفات TKW و BY ($r_p = -0.11$ و $\rho_g = -0.14$) مشاهده شد (شکل ۱).

GFP، BY، GY و TKW)، به طور جداگانه مورد بررسی قرار گرفتند. متغیرهای کمکی به عنوان اثر ثابت در مدل های راهبرد دوم گنجانده شدند. برای اجرا و برازش مدل های بیزی، فرایند نمونه گیری زنجیره مارکوف مونت کارلو^۱ (MCMC) با ۱۰۰۰۰ تکرار^۲ انجام شد که طی آن، ۱۰۰۰ تکرار ابتدایی به عنوان تکرارهای حذف شده یا سوخته (burn-ins) در نظر گرفته شدند تا مدل به حالت تعادل برسد. برای ارزیابی صحت پیش بینی مدل ها، از فرایند اعتبارسنجی متقابل پنج بخشی^۳، با ۱۰ تکرار استفاده شد. در این فرایند جمعیت به طور مساوی به پنج زیرجمعیت تقسیم شد، به طوری که چهار زیرجمعیت (۸۰٪) به عنوان مجموعه آموزشی (شامل داده های فنوتیپی و نشانگری) برای آموزش مدل و یک زیرجمعیت (۲۰٪) به عنوان مجموعه تایید یا اعتبارسنجی (تنها شامل داده های نشانگرها) برای پیش بینی استفاده شد. میانگین ضریب همبستگی پیرسون بین ارزش های اصلاحی ژنومی پیش بینی شده (GEBVs) و مقادیر فنوتیپی مشاهده شده در فرایند اعتبارسنجی متقابل پنج بخشی و تکرارها، به عنوان صحت پیش بینی مدل ها (معیار ارزیابی عملکرد مدل ها) برای صفت مد نظر گزارش شد. برای اجرای GS از نرم افزار iPat (Chen and Zhang 2018) و بسته های "rrBLUP" (Endelman 2011) و "BGLR" (Pérez and de los campos 2014) در R استفاده شد.

¹ Markov Chain Monte Carlo

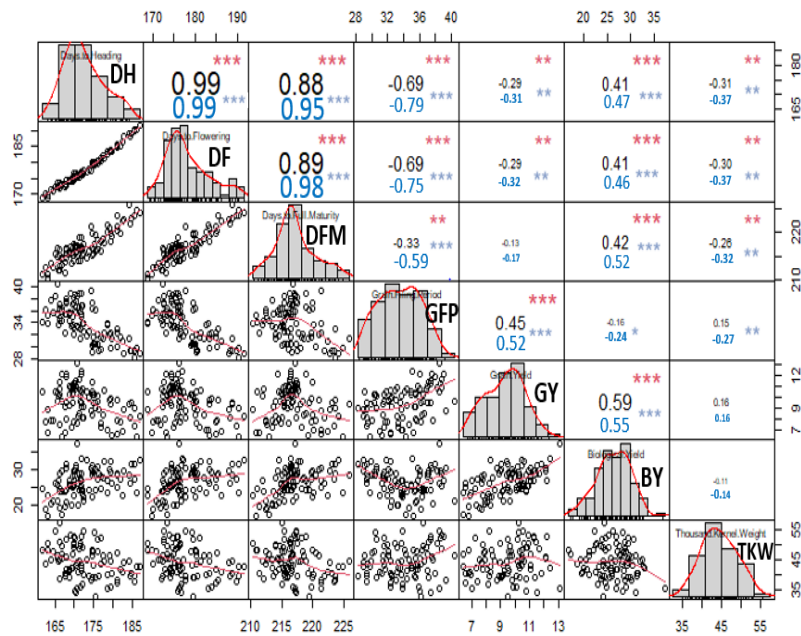
² Iterations

³ Five-Fold Cross-Validation

جدول ۱- تجزیه واریانس مرکب، وراثت پذیری عمومی و ضریب تغییرات صفات کمی گندم نان در قالب طرح لائیس جزئی متعادل سه گانه 10×10 ، در دوسال زراعی (۱۳۹۹-۱۴۰۱) و سه تکرار

منبع تغییرات	میانگین مربعات						
	تعداد روز تا سنبله دهی	تعداد روز تا گلدهی	تعداد روز تا رسیدگی کامل	طول دوره پر شدن دانه (روز)	وزن هزار دانه (g)	عملکرد دانه (t/ha)	عملکرد بیولوژیک (t/ha)
سال	۰/۶۶ ^{**}	۳/۱۱ ^{***}	۸/۷۱ ^{**}	۴/۷۲ [*]	۱/۱۷ [*]	۲/۶۷ ^{**}	۱۴/۹۲ ^{**}
ژنوتیپ	۳۶/۴۳ ^{***}	۲۸/۰۱ ^{***}	۹/۷۵ ^{***}	۶/۴۵ ^{***}	۲۳/۶۸ ^{***}	۲/۱۰ ^{***}	۱۲/۱۲ ^{***}
ژنوتیپ \times سال	۱/۴۶ ^{***}	۲/۳۳ ^{***}	۱/۶۰ ^{***}	۴/۴۸ ^{***}	۴/۷۹ ^{***}	۰/۳۸ ^{***}	۲/۲۶ ^{***}
باقیمانده	۱/۹۳	۱/۸۱	۳/۰۷	۳/۵۸	۱۶/۱۰	۰/۷۳۵	۷/۶۵
میانگین کل	۱۷۲/۹۶	۱۸۷/۰۱	۲۱۷/۰۳	۳۳/۸۴	۴۴/۱۵	۹/۳۷	۲۶/۱۷
وراثت پذیری %	۹۷	۹۵	۸۸	۶۹	۸۲	۸۷	۸۳
ضریب تغییرات %	۰/۸۰	۰/۷۵	۰/۸۱	۵/۵۹	۹/۰۸	۹/۱۵	۱۰/۵۷

***، **، *، به ترتیب اختلاف معنی دار در سطح ۰.۵، ۱ و ۰.۱ درصد



شکل ۱ - ماتریس نمودار پراکنده‌گی و توزیع‌های فنوتیپی به همراه همبستگی‌های فنوتیپی (مثلث بالایی، فونت مشکی) و ژنتیکی (مثلث بالایی، فونت آبی) بین صفات مورد مطالعه گندم نان (DH): تعداد روز تا سنبله‌دهی، DF: تعداد روز تا گلدهی، DFM: تعداد روز تا رسیدگی کامل، GFP: طول دوره پرشدن دانه، GY: عملکرد دانه، BY: عملکرد بیولوژیک و TKW: وزن هزار دانه. * $P < 0.05$ ، ** $P < 0.01$ ، *** $P < 0.001$

جدول ۲ - تعداد و پراکنش SNPها بر روی کروموزوم‌ها و ژنوم‌ها

ژنوم	کروموزوم		
D	B	A	
۱۱۴۰	۴۶۰۳	۲۸۲۶	۱
۱۳۹۹	۵۷۲۵	۴۳۰۲	۲
۸۲۲	۵۹۴۵	۲۷۴۵	۳
۴۵۲	۲۲۹۳	۴۳۳۰	۴
۷۸۱	۴۹۷۱	۳۰۴۸	۵
۸۵۹	۵۶۳۴	۳۰۷۵	۶
۹۹۵	۴۹۹۸	۴۴۴۹	۷
۶۴۴۸	۳۴۱۶۹	۲۴۷۷۵	جمع
	۶۵۹۴۸		جمع کل*

* با احتساب ۵۵۶ نشانگر ناشناخته

ارزیابی نشانگرهای SNP

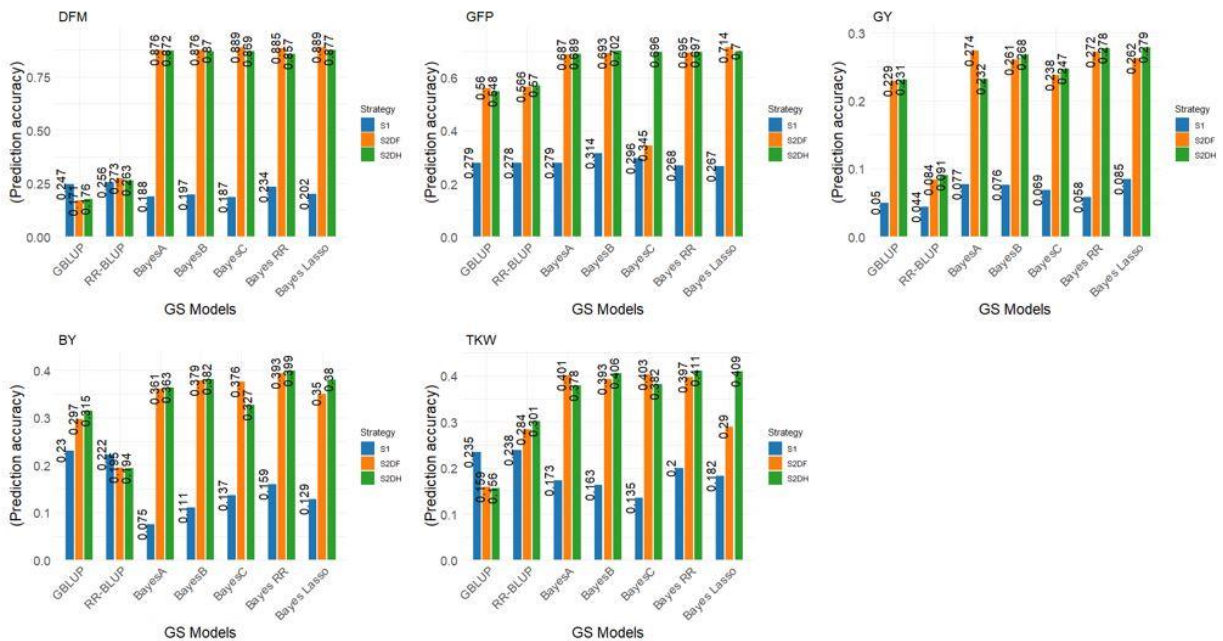
بر اساس پالایش داده‌ها مطابق با آنچه در مواد و روش‌ها ذکر شد، تعداد کل SNPهای صحیح فراخوانی شده با در نظر گرفتن ۲۰ درصد داده‌های گم شده برابر با ۶۵۹۴۸ نشانگر بود که حدوداً ۳۸ درصد آن‌ها (۲۴۷۷۵ نشانگر) بر روی ژنوم A، ۵۲ درصد (۳۴۱۶۹ نشانگر) بر روی ژنوم B و ۱۰ درصد (۶۴۴۸ نشانگر) بر روی ژنوم D قرار داشتند (جدول ۲). بیشترین تعداد نشانگرهای SNP بر روی کروموزوم دوم و سوم ژنوم B و کمترین تعداد بر روی کروموزوم چهارم ژنوم D قرار داشتند.

عملکرد مدل‌های انتخاب ژنومی و مقایسه راهبردها

بر اساس نتایج مدل‌های انتخاب ژنومی، صحت پیش‌بینی مدل‌ها در S_1 از ۰/۰۴۴ (برای GY) تا ۰/۳۱۴ (برای GFP)، در S_2 DF از ۰/۰۸۴ (برای GY) تا ۰/۸۸۹ (برای DFM) و در S_2 DH از ۰/۰۹۱ (برای GY) تا ۰/۸۷۷ (برای DFM) متغیر بود (شکل ۲). بنابراین، صحت‌های پیش‌بینی قابل توجهی در راهبرد اول مشاهده نشد. برای صفت GY، در S_1 بیشترین صحت پیش‌بینی، ۰/۰۸۵ (متعلق به مدل BayesLASSO) بود، اما صحت پیش‌بینی این مدل در S_2 DH (۰/۲۷۹) و در S_2 DF (۰/۲۶۲) به ترتیب حدود ۳/۲ و ۳

۲). برای صفت DFM، مقدار بیشترین صحت پیش‌بینی در S_1 ($0/256$) (مربوط به RRBLUP)، در مقایسه با همین مدل در S_2DH و S_2DF ، به ترتیب به $0/273$ و $0/263$ افزایش یافت. اما بیشترین صحت پیش‌بینی در S_2DF (حدود $0/88$) متعلق به مدل‌های BayesLASSO، BayesC و BayesRR (نسبت به بیشترین $0/256$) و کمترین صحت پیش‌بینی در S_1 (حدود $0/18$) متعلق به مدل‌های BayesA و BayesC، به ترتیب حدود $3/4$ و $4/7$ برابر ارتقاء داشت (شکل ۲). برای صفت GFP، بالاترین صحت در S_1 $0/314$ (متعلق به BayesB) بود، اما صحت پیش‌بینی این مدل در S_2DF ($0/693$) و S_2DH ($0/702$) حدود $2/2$ برابر افزایش یافت. مقدار ارتقاء بیشترین صحت پیش‌بینی در S_2DF ($0/714$) متعلق به BayesLASSO (در مقایسه با بیشترین $0/314$) و کمترین صحت پیش‌بینی (حدود $0/26$) متعلق به مدل‌های BayesRR و BayesLASSO (در S_1)، به ترتیب $2/2$ و $2/6$ برابر بود. شایان ذکر است که S_2DH و S_2DF در اکثر موارد، تقریباً تأثیر مشابهی در بهبود صحت پیش‌بینی مدل‌ها داشتند (شکل ۲).

برابر ارتقاء یافت. همچنین مقدار ارتقاء نسبت به کمترین صحت پیش‌بینی در S_1 ($0/44$) (مربوط به مدل RRBLUP) برابر بود (شکل ۲). در خصوص صفت TKW، مقدار بیشترین صحت پیش‌بینی در S_1 ، از $0/238$ (مربوط به RRBLUP)، به $0/284$ در S_2DF و $0/301$ در S_2DH افزایش یافت (نسبت به مدل مشابه در S_1). همچنین مقدار بیشترین صحت پیش‌بینی در بین همه مدل‌ها ($0/411$) که متعلق به BayesRR در S_2DH بود، نسبت به بیشترین صحت پیش‌بینی ($0/238$) (مربوط به RRBLUP) و کمترین صحت پیش‌بینی ($0/135$) (مربوط به BayesC) در S_1 ، به ترتیب حدود $1/7$ و 3 برابر ارتقاء یافت (شکل ۲). برای صفت BY، در S_1 بیشترین صحت پیش‌بینی ($0/230$) مربوط به مدل GBLUP بود، اما بیشترین صحت پیش‌بینی برای همین مدل در S_2DH و S_2DF به ترتیب، به $0/297$ و $0/315$ افزایش یافت. افزون بر این، مقدار بیشترین صحت پیش‌بینی در S_2DH ($0/399$) متعلق به BayesRR، نسبت به بیشترین صحت پیش‌بینی ($0/230$) مربوط به مدل GBLUP و کمترین صحت پیش‌بینی در S_1 ($0/075$) (متعلق به BayesA)، به ترتیب $1/7$ و $5/3$ برابر ارتقاء یافت (شکل



شکل ۲- نمودار مقایسه صحت پیش‌بینی مدل‌های انتخاب ژنومی برای صفات مورد مطالعه گندم نان (DFM: تعداد روز تا رسیدگی کامل، GFP: طول دوره پرشدن دانه، GY: عملکرد دانه، BY: عملکرد بیولوژیک و TKW: وزن هزار دانه، S_1 : مدل پیش‌بینی ژنومی بدون متغیر کمکی، S_2DH : مدل انتخاب ژنومی همراه با متغیر کمکی تعداد روز تا سنبله دهی (DH)، S_2DF : مدل انتخاب ژنومی همراه با متغیر کمکی تعداد روز تا گلدهی (DF)).

فوتوتیپ سنجی، وراثت‌پذیری و همبستگی صفات

بر اساس نتایج تجزیه مرکب، اثر مؤلفه‌های ژنوتیپ و سال بر روی صفات از نظر آماری به‌طور معنی‌داری تایید شد (جدول ۱). این موضوع نشان‌دهنده تأثیر تنوع موجود در ژنوتیپ‌ها بر صفات مورد مطالعه و همچنین احتمال نقش عوامل محیطی مانند دما، بارندگی و سایر شرایط محیطی در تغییرات فنوتیپی مشاهده شده بین سال‌ها است (Lobell et al. 2011). معنی‌دار بودن اثر متقابل ژنوتیپ در سال برای تمام صفات، بیانگر تفاوت در پاسخ و واکنش ژنوتیپ‌های مختلف به تغییرات شرایط محیطی در سال‌های مختلف است. این نتایج با نتایج مطالعات پیشین (Tomar et al. 2021; Alipour et al. 2021; Rahimi et al. 2019a) همخوانی داشت.

به غیر از صفت GFP با وراثت‌پذیری متوسط (۶۹٪)، بقیه صفات مورد مطالعه و همچنین متغیرهای کمکی، وراثت‌پذیری بالایی (< ۸۰٪) نشان دادند (Singh 2001). این نتیجه، نشان‌دهنده کیفیت بالای اندازه‌گیری‌های فنوتیپی و پتانسیل پیش‌بینی قابل توجه صفات است (Tomar et al. 2021). افزون بر این، انتخاب بر اساس چنین صفاتی به دلیل مطابقت نزدیک فنوتیپ و ژنوتیپ نسبتاً آسان است (Singh 2001; Bernardo 2002). این امر فرایند اصلاح را تسهیل کرده و پیش‌بینی دقیق‌تر را ممکن می‌سازد. وراثت‌پذیری و همبستگی بین صفات از عوامل مهم برای دستیابی به صحت پیش‌بینی بالاتر محسوب می‌شوند (Rutkoski et al. 2016; Tomar et al. 2021).

در این مطالعه، متغیرهای کمکی DH و DF با تمام صفات همبستگی فنوتیپی و ژنتیکی مثبت یا منفی معنی‌داری داشتند. بالاترین همبستگی آن‌ها، با صفات GFP و DFM بود (شکل ۱). تصور بر این است صفت زودرسی در گندم ممکن است اثر منفی بر عملکرد و اندازه دانه داشته باشد. در صورتی که سنبله‌دهی اگر زود هنگام باشد، با توجه به اینکه زمان بیشتری برای دوره پرشدن دانه فراهم می‌کند در افزایش وزن دانه و عملکرد مؤثر است (Armion et al. 2020). علاوه بر این سنبله‌دهی زود هنگام یک سازوکار مهم برای فرار از شرایط تنش انتهایی در زمان دوره پرشدن دانه است (Tewolde et al. 2006; Sharma and Smith,)

(1986). همبستگی منفی بین صفت DH و GY در دیگر مطالعات، (مانند Ullah et al. 2018; Rahimi et al. 2019b; Gill et al. 2021) گزارش شده است. اطلاع از زمان رسیدگی نیز اهمیت زیادی دارد، زیرا محصولاتمانند گندم در معرض مشکلاتی نظیر جوانه‌زنی پیش از برداشت، ورس و تنش‌های زیستی قرار می‌گیرند که می‌تواند در طی فرایند رسیدگی موجب کاهش عملکرد شود. همان‌طور که انتظار می‌رفت همبستگی مثبت معنی‌دار بین GFP و GY ($r_p = 0.45^{***}$ و $\rho_g = 0.52^{***}$) وجود داشت، که نتایج تقریباً مشابهی در دیگر مطالعات مانند Rahimiet al. (2019a) and al. (2016) Mohammadi Gonbad گزارش شده است. طولانی‌تر شدن دوره پر شدن دانه منجر به افزایش زمان انباشت مواد فتوسنتزی در دانه و فراهم شدن فرصت بیشتر برای بهره‌برداری گیاه از منابع موجود و در نتیجه افزایش عملکرد خواهد شد (Amjadi et al. 2023).

ژنوتیپ‌سنجی و نشانگرهای SNP

مطابق با گزارش‌های قبلی، ژنوم D دارای تعداد کمی SNP است، در حالی‌که بیشتر SNP‌ها در ژنوم‌های B و A قرار دارند (Javid et al. 2025; Armion et al. 2020; Rahimi et al. 2019a). نتایج ما نیز با این یافته‌ها مطابقت داشت. به نظر می‌رسد ارتباط مستقیمی بین اندازه کروموزوم و چگالی SNP وجود دارد (Berkman et al. 2013). با افزایش اندازه کروموزوم و توالی بازهای نوکلئوتیدی روی آن، احتمال جهش و ایجاد SNP‌های جدید روی ژنوم A و B بیشتر خواهد شد. از سوی دیگر، فراوانی بیشتر SNP‌ها در ژنوم B احتمالاً ناشی از فرآیندهای تکاملی متعددی (Rabieyan et al. 2020; Mourad et al. 2020; Alipour et al. 2023) از جمله منشاء گرفتن این ژنوم از جمعیت‌های وحشی با تنوع ژنتیکی بالاتر و گلوگاه ژنتیکی کمتر در فرایند انتقال ژنوم B به گندم نان است (Pont et al. 2019; Avni et al. 2017). افزون بر این تنوع بالاتر کشف شده در ژنوم‌های B و A می‌تواند به دو عامل عمده نسبت داده شود: ۱- جریان ژنی از *T. turgidum* در مقابل عدم وجود آن از *Ae. tauschii* به *T. aestivum*؛ ۲- تاریخچه تکاملی قدیمی‌تر ژنوم‌های B و A نسبت به ژنوم D (Jordan et al. 2006; Dvorak et al. 2015).

نقش راهبردهای انتخابی در بهینه‌سازی مدل‌های GS

افزایش صحت پیش‌بینی مدل‌های انتخاب ژنومی...
 اضافه کردن صفات ثانویه یا متغیرهای کمکی به مدل‌های تک متغیره و چند متغیره، می‌تواند صحت پیش‌بینی مدل‌ها را افزایش دهد (Rutkoski et al. 2016; Sun et al. 2017; Crain et al. 2018; Lozada and Carter 2020). بر اساس اطلاعات ما، صفات ارتفاع، تعداد روز تا سنبله‌دهی، اطلاعات فیزیولوژیکی، متغیرهای محیطی، نسبت تحمل، نسبت تنش و شاخص نرمال شده تفاوت پوشش گیاهی به‌عنوان اثرات ثابت جهت افزایش صحت پیش‌بینی و بهینه‌سازی مدل‌های GS، در برخی از مطالعات مورد استفاده قرار گرفتند (Jarquin et al. 2014; Haghigattalab et al. 2016; Pierre et al. 2016; Rutkoski et al. 2016; Sun et al. 2017; Crain et al. 2018; Shabannejad et al. 2020; Guo et al. 2020). در این تحقیق، افزودن متغیر کمکی DF و DH به مدل‌ها در راهبرد دوم، باعث ارتقاء قابل توجه صحت پیش‌بینی اکثر مدل‌های GS در همه صفات مورد مطالعه نسبت به راهبرد اول (بدون متغیر کمکی) شد (شکل ۲). همچنین، افزایش‌های ۵ تا ۶ برابری صحت پیش‌بینی در راهبرد دوم نسبت به راهبرد اول، به‌ترتیب برای BY و GY قابل توجه بود، چراکه در راهبرد اول مدل‌ها عملکرد مناسبی نداشتند و در نتیجه صحت‌های پیش‌بینی مناسبی به‌دست نیامد. جالب توجه اینکه، برای صفات GY و GFP در راهبرد دوم بهبود صحت پیش‌بینی به گونه‌ای بود که مقدار پایین‌ترین صحت‌های پیش‌بینی مدل‌ها در راهبرد دوم، از بالاترین صحت در راهبرد اول بیشتر بود، این نشان‌دهنده نقش مؤثر متغیرهای کمکی استفاده شده در این تحقیق، در بهبود صحت پیش‌بینی مدل‌ها بود (شکل ۲). ترکیب متغیرهای کمکی و صفات ثانویه همبسته و با وراثت‌پذیری بالا برای بهبود توانایی پیش‌بینی عملکرد دانه ایده‌آل است (Crain et al. 2017; Lozada and Carter, 2020). هر دو متغیر کمکی و اکثر صفات هدف این مطالعه وراثت‌پذیری بالایی داشتند. همان‌طور که انتظار می‌رفت، در صفاتی که همبستگی بالاتری با متغیرهای کمکی داشتند، یعنی GFP و DFM صحت پیش‌بینی بالاتری مشاهده شد (شکل ۱ و ۲). بهبود ۷۰ درصدی صحت پیش‌بینی عملکرد گندم، با اضافه کردن اثرات ثابت به مدل‌های GS گزارش شده است (Sun et al. 2017). صحت پیش‌بینی با افزایش اندازه جمعیت آموزشی می‌تواند افزایش یابد (Asoro et al. 2011). در برخی جمعیت‌های

کوچک، احتمالاً به دلیل تنوع و تعداد کمتر داده‌ها، مدل به خوبی آموزش نمی‌بیند و این می‌تواند بر صحت پیش‌بینی تاثیر منفی داشته باشد. گنجاندن صفات و محیط‌های همبسته در مدل‌های پیش‌بینی می‌تواند محدودیت یک جمعیت آموزشی کوچک را جبران کند (Gill et al. 2021) و صحت پیش‌بینی پایین در برخی جمعیت‌های کوچک را افزایش دهد. متغیرهای کمکی می‌توانند بخشی از واریانس باقی‌مانده را توضیح دهند و خطای آزمایش را کاهش دهند (Alvarado et al. 2020). این ویژگی می‌تواند توانایی مدل‌های انتخاب ژنومی را در جداسازی سیگنال ژنتیکی از نویز محیطی ارتقاء دهد. در نتیجه، برآورد اثرات ژنتیکی واقعی نشانگرها با دقت بیشتری انجام می‌شود و صحت پیش‌بینی ارزش‌های اصلاحی ژنومی افزایش می‌یابد. همچنین، برخی متغیرها ممکن است با ژن‌هایی که در کنترل یک صفت یا صفات دیگر دخیل هستند، همبستگی داشته باشند. همبستگی ژنتیکی تخمینی از اثر ژنتیکی افزایشی است که بین یک جفت صفت مشترک است. بسیاری از صفات مهم، همبستگی مثبت یا منفی دارند زیرا توسط ژن‌های یکسانی کنترل می‌شوند (Alvarado et al. 2020). برای مثال، ژن‌های مرتبط با صفت DF بر روی تمام کروموزوم‌های گندم توزیع شده‌اند و به خوبی ثابت شده است، آن‌ها نقش‌های اضافی و فراتر از زمان گلدهی دارند و با صفات مختلف از جمله صفات مرتبط با عملکرد در ارتباط هستند (Worland, 1996; Snape et al. 2001; Cockram et al. 2007,). (Kamran et al. 2014). بنابراین، متغیر کمکی اگر به درستی انتخاب شود، می‌تواند به مدل کمک کند تا اثرات پیچیده ژنتیکی را با دقت بیشتری شناسایی کند.

عملکرد مدل‌های GS و مقایسه آن‌ها

انتخاب بهترین مدل GS به حداکثر توانایی پیش‌بینی و بازده محاسباتی یک مدل در طیف گسترده‌ای از صفات و مجموعه داده‌ها بستگی دارد (Ali et al. 2020). در مطالعه حاضر، نه تنها مدل‌های بیزی در راهبرد دوم برای تمام صفات بهتر عمل نمودند، بلکه در بیشتر موارد نسبت به راهبرد اول ارتقاء چشمگیری داشتند. با این حال در مجموع، مدل‌های BayesLASSO و BayesRR برای بیشتر صفات مورد مطالعه، بالاترین یا نزدیک به بالاترین صحت‌های پیش‌بینی را ارائه کردند (شکل ۲). مدل‌های

نتیجه‌گیری کلی

تولید پایدار غذا در آینده با توجه به پیش‌بینی افزایش جمعیت و محدودیت منابع و زمین، یک چالش اساسی خواهد بود. بنابراین، برای پاسخ به این نیاز رو به رشد، افزایش قابل توجه پیشرفت ژنتیکی در برنامه‌های به‌نژادی ضروری است. ما دو راهبرد انتخاب ژنومی (GS) را مورد ارزیابی قرار دادیم و نقش متغیرهای کمکی را در بهینه‌سازی و بهبود صحت پیش‌بینی هفت مدل GS بررسی کردیم. نتایج نشان داد افزودن متغیرهای کمکی، شامل تعداد روز تا گلدهی (DF) و تعداد روز تا سنبله‌دهی (DH) به مدل‌های GS، می‌تواند عملکرد مدل‌ها و در نتیجه صحت پیش‌بینی آن‌ها را نسبت به استفاده صرف از این مدل‌ها ارتقاء دهد. افزون بر این، گنجاندن صفات همبسته در مدل‌های پیش‌بینی می‌تواند محدودیت یک جمعیت آموزشی کوچک را جبران کند. این یافته‌ها نشان می‌دهد که GS پتانسیل قابل توجهی برای بهبود دقت انتخاب صفات پیچیده در برنامه‌های اصلاحی گندم دارد، چرا که ارزش‌های اصلاحی ژنومی را با صحت بیشتری پیش‌بینی می‌کند، به‌ویژه در صورت بهینه‌سازی مدل‌ها. در نتیجه، راهبردهای به‌نژادی آینده باید بر افزایش کارایی برنامه‌های اصلاحی از طریق به حداکثر رساندن پیشرفت ژنتیکی تمرکز کنند.

بیزی به دلیل استفاده از توزیع‌های احتمال مختلف، انعطاف‌پذیری بیشتری در مدل‌سازی تاثیرات ژنومی دارند و می‌توانند اثرات متفاوت نشانگرها را با دقت بیشتر شبیه‌سازی کنند. بنابراین، این مدل‌ها فرض واقع‌گرایانه‌تری از نشانگرها دارند. هنگامی که تعداد QTLها کاهش و اثر آن‌ها افزایش می‌یابد، روش‌های بیزی می‌توانند نتایج بهتری ارائه دهند (Daetwyler et al. 2010). در بعضی از صفات در راهبرد دوم، عملکرد برخی مدل‌های بیزی نسبت به مدل‌های GBLUP یا RRBLUP تقریباً مشابه بود و یا با یکدیگر در رقابت بودند (شکل ۲). مانند صفات عملکرد دانه (مدل‌های GBLUP و BayesA)، عملکرد بیولوژیک (مدل‌های GBLUP و BayesC) و وزن هزار دانه (مدل‌های RRBLUP و BayesLASSO). زمانی که تعداد زیادی از ژن‌ها با اثرات کم وجود داشته باشند، روش RRBLUP برتر است و نتایج قابل اعتمادتری را برای QTLهای با اثرات کوچک نشان می‌دهد (Tibshirani 1996; Luan et al. 2009). همچنین GBLUP از نظر ریاضی معادل هستند (Meuwissen et al. 2001; Juliana et al. 2017). وقتی تعداد QTLهای مؤثر بر صفت زیاد باشد، روش GBLUP مشابه و یا حتی بهتر از روش‌های بیزی عمل می‌کند (2007; Meuwissen, 2009; Daetwyler et al. 2010). (Habier et al.

منابع

- Abdollahi Arpanahi R, Pakdel A, Zandi-Baghchehmaryam MB (2012) From Infinitesimal Model to Genomic Selection. *Modern Genetics Journal* 7:105-114. (In Farsi).
- Ali M, Zhang Y, Rasheed A, Wang J, Zhang L (2020) Genomic prediction for grain yield and yield-related traits in chinese winter wheat. *International journal of molecular sciences* 21:1342.
- Alipour H, Abdi H, Rahimi Y, Bihamta MR (2021) Dissection of the genetic basis of genotype-by-environment interactions for grain yield and main agronomic traits in Iranian bread wheat landraces and cultivars. *Scientific reports* 11:17742.
- Alipour H, Bai G, Zhang G, Bihamta MR, Mohammadi V, Peyghambari SA (2019) Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. *PLoS One* 14: e0208614.
- Alipour H, Bihamta MR, Mohammadi V, Peyghambari SA, Bai G, Zhang G (2017) Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Frontiers in Plant Science* 8:1293.
- Alvarado G, Rodríguez FM, Pacheco A, Burgueño J, Crossa J, Vargas M, Pérez-Rodríguez P, Lopez-Cruz MA (2020) META-R: A software to analyze data from multi-environment plant breeding trials. *The Crop Journal* 8: 745-756.
- Amjadi H, Hosseinpanahi F, Siosemardeh A (2023) Evaluation of grain filling rate and duration in wheat cultivars under different supplementary irrigation regimes. *Plant Production and Genetics* 4:59-70. (In Farsi).
- Armion M (2020) Evaluation of Genomic Selection Model in Predicting Genomic Estimated Breeding Values of Quantitative traits and bread-making qualities in bread wheat (*Triticum aestivum* L.). Dissertation, University of Tehran, Iran.

- Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink JL (2011) Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *The Plant Genome* 4:132-144.
- Barton NH, Etheridge AM, Veber A (2017) The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* 118:50-73.
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1-48.
- Battenfield SD, Guzmán C, Gaynor RC, Singh RP, Peña RJ, Dreisigacker S, Fritz AK, Poland JA (2016) Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* 9:1-12.
- Berkman PJ, Visendi P, Lee HC, Stiller J, Manoli S, Lorenc MT, Edwards D (2013) Dispersion and domestication shaped the genome of bread wheat. *Plant biotechnology journal* 11:564-71.
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Science* 48: 1649-1664.
- Bernardo R (2010) Breeding for quantitative traits in plants. Stemma Press, Woodbury, Minnesota, 1-411.
- Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics* 84:210-223.
- Chen CJ, Zhang Z (2018) iPat: intelligent prediction and association tool for genomic research. *Bioinformatics* 34:1925-1927.
- CIMMYT (2021). International Maize and Wheat Improvement Center. Available at <https://www.cimmyt.org/work/wheat-research>. CIMMYT, Mexico City, Mexico.
- Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *Journal of experimental botany* 58:1231-44.
- Collard BC, Mackill DJ (2008) Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society* 363:557-572.
- Crain J, Mondal S, Rutkoski J, Singh RP, Poland J (2018) Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *The Plant Genome* 11:170043.
- Crossa J, Campos GDI, Maccaferri M, Tuberosa R, Burgueño J, Pérez-Rodríguez P (2016) Extending the Marker × Environment Interaction Model for Genomic-Enabled Prediction and Genome-Wide Association Analysis in Durum Wheat. *Crop Science* 56: 2193-2209.
- Cuevas J, Crossa J, Soberanis V, Perez-Elizalde S, Perez-Rodríguez P, Campos GL, et al. (2016) Genomic prediction of genotype × environment interaction kernel regression models. *Plant Genome* 9.
- Daetwyler HD, Calus MPL, Pong-Wong R, Campos GDL, Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193:347-365.
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021-1031.
- de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327-345.
- Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC (2006) Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Molecular Biology and Evolution* 23:1386-1396.
- Eathington, SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. *Crop Science* 47(S3): S154-S163.
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* 4:.
- Fisher RA (1919) XV.-The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh* 52:399-433.
- Fujino K, Hirayama Y, Kaji R (2019) Marker-assisted selection in rice breeding programs in Hokkaido. *Breeding Science* 69:383-392.
- Gill HS, Halder J, Zhang J, Brar NK, Rai TS, Hall C, Sehgal SK (2021) Multi-trait multi-environment genomic prediction of agronomic traits in advanced breeding lines of winter wheat. *Front Plant Sci* 12:709545.
- Godin C, Mailliar A, Chassin A, Gardet O, Giraud A, Heumez E, et al. (2012) A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. *European Journal of Agronomy* 40:28-38.
- Graybosch RA, Peterson CJ (2010) Genetic Improvement in Winter Wheat Yields in the Great Plains of North America. *Crop Science* 50: 1882-1890.
- Guo J, Pradhan S, Shahi D, Khan J, Mcbreen J, Bai G, Murphy JP, Babar MA (2020) Increased Prediction Accuracy Using Combined Genomic Information and Physiological Traits in A Soft Wheat Panel Evaluated in Multi-Environments. *Scientific reports* 10:7023.
- Habier D, Fernando RL, Dekkers JC (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389-2397.
- Haghighattalab A, Gonzalez Perez L, Mondal S, Singh D, Schinstock D, Rutkoski J, Ortiz-Monasterio I, Singh RP, Goodin D, Poland J (2016) Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods* 12:35.
- Heffner EL, Sorrells ME, Jannink JL (2009) Genomic Selection for Crop Improvement. *Crop Science* 49: 1-12.
- Heffner EL, Jannink J, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4:65-75.

- Heslot, N., Jannink, J.-L. and Sorrells, M.E. (2015) Perspectives for Genomic Selection Applications and Research in Plants. *Crop Science* 55:1-12.
- Jaggard KW, Qi A, Ober ES (2010) Possible changes to arable crop yields by 2050. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:2835–2851.
- Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics* 9:166-177.
- Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, Piroux F, Guerreiro L, Perez P, Calus M, Burgueno J (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet* 127:595-607.
- Javid S, Bihamta MR, Omid M, Abbasi AR, Alipour H, Ingvarsson PK, Pocza P (2025) Genome-wide association study (GWAS) uncovers candidate genes linked to the germination performance of bread wheat (*Triticum aestivum* L.) under salt stress. *BMC Genomics* 26:5.
- Jia Y, Jannink JL (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192:1513-1522.
- Jiang Y, Reif JC (2015) Modeling epistasis in genomic selection. *Genetics* 201:759-768.
- Jordan KW, Wang S, Lun Y, Gardiner LJ, MacLachlan R, Hucl P, Wiebe K, Wong D, Forrest KL, IWGS Consortium, Sharpe AG, Sidebottom CH, Hall N, Toomajian C, Close T, Dubcovsky J, Akhunova A, Talbert L, Bansal UK, Bariana HS, Akhunov E (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biology* 16:48.
- Juliana P, Singh RP, Singh PK, Crossa J, Rutkoski JE, Poland JA, Bergstrom GC, Sorrells ME (2017) Comparison of Models and Whole-Genome Profiling Approaches for Genomic-Enabled Prediction of Septoria Tritici Blotch, Stagonospora Nodorum Blotch, and Tan Spot Resistance in Wheat. *The Plant Genome* 10:10.3835.
- Kamran A, Iqbal M, Spaner D (2014) Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability. *Euphytica* 197:1-26
- Lobell DB, Schlenker W, Costa-Roberts J (2011) Climate Trends and Global Crop Production since 1980. *Science* 333:616-620.
- Lozada DN, Carter AH (2020) Genomic Selection in Winter Wheat Breeding Using a Recommender Approach. *Genes* 11:779.
- Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen TH (2009) The accuracy of Genomic Selection in Norwegian red cattle assessed by cross-validation. *Genetics* 183:1119-26.
- Meuwissen, THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- Meuwissen TH (2009) Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics Selection Evolution* 41:35.
- Mohammadi Gonbad R, Esfahani, M, Roustaii M, Sabouri H (2016) Effect of planting date on grain filling trend and grain yield of bread wheat genotypes using multivariate analysis method under dryland conditions. *Iranian Journal of Field Crop Science* 47:329-340. (In Farsi).
- Montesinos-López A, Montesinos-López OA, Gianola D, Crossa J, Hernández-Suárez CM (2018) Multi-environment Genomic Prediction of Plant Traits Using Deep Learners With Dense Architecture. *G3 Genes Genomes Genetics* 8:3813-3828.
- Mourad AMI, Belamkar V, Baenziger PS (2020) Molecular genetic analysis of spring wheat core collection using genetic diversity, population structure, and linkage disequilibrium. *BMC Genomics* 21:1-12.
- Pask AJD, Pietragalla J, Mullan D, Reynolds MP (2012) Physiological breeding II: A field guide to wheat phenotyping. *The International Maize and Wheat Improvement Center, CIMMYT, Mexico City, Mexico.*
- Pérez P, de los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483-495.
- Pierre SC, Burgueño J, Crossa J, Dávila GF, López PF, Moya ES, Moreno JI, Muela VH, Villa VZ, Vikram P, Mathews K (2016) Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Scientific reports* 6:1-11.
- Plavšín I, Gunjača J, Šatović Z, Šarčević H, Ivić M, Dvojković K, Novoselović D (2021) An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits. *Plants* 10:745.
- Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5:92-102.
- Pont C, Leroy T, Seidel M, Tondelli A, Duchemin W, Armisen D, Lang D, Bustos-Korts D, Goué N, Balfourier F, Molnár-Láng M, Lage J, Kilian B, Özkan H, Waite D, Dyer S, Letellier T, Alaux M, Wheat and Barley Legacy for Breeding Improvement (WHEALBI) consortium, Russell J, Keller B, van Eeuwijk F, Spannagl M, Mayer KFX, Waugh R, Stein N, Cattivelli L, Haberer G, Charmet G, Salse J (2019) Tracing the ancestry of modern bread wheats. *Nat Genet* 51:905-911.
- R Core Team (2018) R: A Language and Environment for Statistical Computing. Available at <https://www.scirp.org/reference/ReferencesPapers?ReferenceID=2342186>
- Rabieyan E, Bihamta MR, Moghaddam ME, Alipour H, Mohammadi V, Azizyan K, Javid S (2023) Analysis of genetic diversity and genome-wide association study for drought tolerance related traits in Iranian bread wheat. *BMC Plant Biology* 23:431.
- Rahimi Y, Bihamta MR, Taleei A, Alipour H, Ingvarsson PK (2019a) Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. *BMC Plant Biology* 19:541.
- Rahimi Y, Bihamta MR, Taleei A, Alipour H, Ingvarsson PK (2019b) Applying an artificial neural network approach for drought tolerance screening among Iranian wheat landraces and cultivars grown under well-watered and rain-fed conditions. *Acta Physiologiae Plantarum* 41:156.
- Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8(6):e66428.

- Rutkoski J, Poland J, Mondal S, Autrique E, Perez LG, Crossa J, Reynolds M, Singh R (2016) Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes|Genomes|Genetics* 6:2799-808.
- Shabannejad M, Bihamta MR, Majidi-Hervan E, Alipour H, Ebrahimi A (2020) A simple, cost-effective high-throughput image analysis pipeline improves genomic prediction accuracy for days to maturity in wheat. *Plant Methods* 16:146.
- Sharma RC, Smith EL (1986) Selection for high and low harvest index in three winter wheat populations. *Crop Science* 26:1147-1150.
- Singh BD (2001) *Plant Breeding: Principles and Methods*. Kalyani Publishers, New Delhi, India, 1-896.
- Snape J, Butterworth K, Whitechurch E, et al. (2001) Waiting for fine times: genetics of flowering time in wheat. *Euphytica* 119:185-190.
- Sun J, Rutkoski JE, Poland JA, Crossa J, Jannink JL, Sorrells ME (2017) Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome* 10:10.3835.
- Technow F, Riedelsheimer C, Schrag TA, Melchinger AE (2012) Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. *Theoretical and Applied Genetics* 125:1181-1194.
- Tessema BB, Liu H, Sørensen AC, Andersen JR, Jensen J (2020) Strategies using genomic selection to increase genetic gain in breeding programs for wheat. *Frontiers in Genetics* 11:578123.
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, University of Toronto, Canada, Series B*, 267-288.
- Tomar V, Singh D, Dhillon GS, Chung YS, Poland J, Singh RP, Joshi AK, Gautam Y, Tiwari BS, Kumar U (2021) Increased predictive accuracy of multi-environment genomic prediction model for yield and related traits in spring wheat (*Triticum aestivum* L.). *Front Plant Sci* 12:720123.
- Tewolde H, Fernandez CJ, Erickson CA (2006) Wheat cultivars adapted to post-heading high temperature stress. *J Agron Crop Sci* 192:111-120.
- Ullah N, Ullah H, Afridi K, Alam M, Jadoon A, Khan W, Ahmad M, Uddin H (2018) Genetic variability, heritability and correlation analysis among morphological and yield traits in wheat advanced lines. *Biol Divers Conserv* 11:166-180.
- Wang X, Xu Y, Hu Z, Xu C (2018) Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal* 6:330-340.
- Worland AJ (1996) The influence of flowering time genes on environmental adaptability in European wheats. *Euphytica* 89: 49-57.
- Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: From publications to practice. *Crop science* 48:391-407.
- Zhang J, Naik HS, Assefa T, Sarkar S, Reddy RV, Singh A, Ganapathysubramanian B, Singh AK (2017) Computer vision and machine learning for robust phenotyping in genome-wide studies. *Scientific reports* 7:44048.