doi: 10.22124/cr.2019.14333.1518

(مقاله مروری)

تحقیقات غلات

دوره نهم/ شماره سوم/ پاییز ۱۳۹۸ (۲۹۸-۲۷۱)

تجزیه ارتباطی صفات کمی در به‌نژادی مولکولی غلات

هادی علی‌پور[۱] و رضا درویش‌زاده[۲*]

تاریخ دریافت: ۱۳۹۸/۰۵/۱۵ تاریخ پذیرش: ۱۳۹۸/۰۷/۱۷

چکیده

زمینه مطالعاتی نقشه‌یابی ارتباطی اخیراً توجه زیادی را برای مطالعه ژنتیکی صفات کمی در بسیاری از گیاهان مهم به خود جلب کرده است. دسترسی به فناوری‌های نسل جدید توالی‌یابی، حجم بالای داده‌های فنوتیپی و تنوع زیاد ابزارهای آماری موجب شده است که مطالعات نقشه‌یابی ارتباطی در گیاهان بتواند موفقیت‌های فراوانی را در شناسایی مکان‌های ژنی کنترل کننده صفات کمی به‌همراه داشته باشد. با توجه به اهمیت این روش در مطالعات نقشه‌یابی صفات کمی، مقاله حاضر تهیه شد که هدف از آن، توضیح روش نقشه‌یابی ارتباطی و استفاده از آن در به‌نژادی گیاهی به‌ویژه غلات بود. همچنین، در این مقاله برخی از اطلاعات مربوط به نرم‌افزارهای آماری مورد استفاده در نقشه‌یابی ارتباطی ارایه می‌شود و پس از آن در مورد فرصت‌ها و چالش‌های نقشه‌یابی ارتباطی و مطالعات پسا نقشه‌یابی در سطح کل ژنوم بحث خواهد شد. در انتها نیز با ارایه مثالی ساده، مقدار عددی نامتعادلی پیوستگی و نقشه‌یابی ارتباطی بر اساس مدل خطی عمومی (GLM=General Linear Model) و مدل خطی مخلوط (MLM=Mixed Linear Model) برآورد خواهد شد.

واژه‌های کلیدی: روابط خویشاوندی، ساختار جمعیت، مدل خطی مخلوط، نقشه‌یابی در سطح ژنوم

۱- استادیار، گروه تولید و ژنتیک گیاهی، دانشکده کشاورزی و منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران

۲- استاد، گروه تولید و ژنتیک گیاهی، دانشکده کشاورزی و منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران

* نویسنده مسئول: r.darvishzadeh@urmia.ac.ir

مقدمه

تنوع فنوتیپی موجود در صفات کمّی تحت تأثیر چندین جایگاه ژنی با آثار کوچک، عوامل محیطی و برهمکنش این دو قرار دارد. برای مطالعه ژنتیکی این گونه صفات، دو روش کلی وجود دارد:

نقشه‌یابی پیوستگی: هدف از این روش، شناسایی نشانگرهای پیوسته با یک صفت کمّی در یک جمعیت خویشاوند است. این نقشه‌یابی بر پایه ایجاد یک رابطه آماری بین آلل‌های نشانگر و آلل‌های کنترل‌کننده صفت کمّی است و به‌طور وسیعی در جمعیت‌های گیاهی (به‌دلیل سهولت ایجاد جمعیت‌های خویشاوند) استفاده می‌شود.

نقشه‌یابی یا تجزیه ارتباطی: در این روش، مجموعه بزرگی از افراد یک جمعیت به‌طور تصادفی جمع‌آوری و نقشه‌یابی بر اساس عدم تعادل پیوستگی (LD=Linkage Disequilibrium) انجام می‌شود که در ژنتیک انسانی به‌وفور از آن استفاده شده است (Schulze and McMahon, 2002).

با وجود توان آماری بالا و قابلیت تکرارپذیری نقشه‌یابی پیوستگی، این روش دارای محدودیت‌های اساسی است که در زیر به برخی از آن‌ها اشاره می‌شود (Gupta *et al.*, 2005; Holland, 2007):

۱- چون جمعیت نقشه‌یابی از تلاقی دو یا چند والد به‌دست می‌آید، تنوع ژنتیکی جمعیت محدود به تنوع والدین مورد استفاده است.

۲- در جمعیت‌های نقشه‌یابی مانند F_2، دابل‌هاپلوئیدها و حتی رگه‌های خویش‌آمیخته نوترکیب، به‌دلیل آنکه افراد جمعیت تعداد نسل نوترکیبی کمی را پشت سر گذاشته‌اند، نشانگرهایی که بسیار دور از QTL قرار گرفته‌اند (۱۰-۲۰ سانتی‌مورگان) هنوز با آن ارتباط دارند. چنین ارتباطات کاذبی نقشه‌یابی QTLها را با مشکل مواجه می‌سازد.

۳- برای تکثیر لاین‌ها (رگه‌ها) به‌منظور دستیابی به تعداد کافی کراسینگ‌اور، علاوه بر هزینه بالا، زمان قابل توجهی نیز لازم است صرف شود.

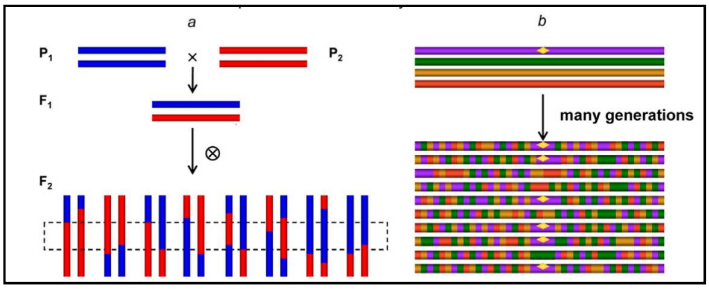
۴- جمعیت ایجادشده فقط برای مطالعات و صفات محدودی کارایی دارد.

۵- علی‌رغم مطالعات نقشه‌یابی وسیعی که در طول دو دهه اخیر روی گیاهان صورت گرفته است، فقط تعداد محدودی از QTLها در سطح ژن همسانه‌سازی شده‌اند (Li *et al.*, 2005; Sameri *et al.*, 2006; Roy *et al.*, 2010; Varshney *et al.*, 2012).

بر خلاف نقشه‌یابی پیوستگی، نقشه‌یابی ارتباطی با استفاده از تنوع موجود در جمعیت‌های طبیعی، ارتباط بین تنوع فنوتیپی و چندشکلی موجود در ژنوم (تنوع ژنوتیپی) را بررسی و نشانگرهای پیوسته با ژن‌های کنترل کننده صفات را شناسایی می‌کند. این روش به‌عنوان جایگزین یا مکمل نقشه‌یابی پیوستگی با بهره‌گیری از تنوع موجود در جمعیت‌های طبیعی و لحاظ کردن تمامی وقایع میوزی که در طول تکامل جمعیت رخ داده است، وضوح نقشه‌یابی را افزایش می‌دهد (شکل ۱). در این روش، به‌دلیل اینکه از تنوع موجود در جمعیت‌های طبیعی استفاده می‌شود، تا حد زیادی در وقت و هزینه‌ها صرفه‌جویی می‌شود. از طرف دیگر، در مطالعات نقشه‌یابی پیوستگی برای هر مکان ژنی حداکثر دو آلل مورد مطالعه قرار می‌گیرد، در صورتی‌که در نقشه‌یابی ارتباطی به‌طور بالقوه و هم‌زمان تمام آلل‌های یک مکان ژنی (حداقل ۲ آلل)، مطالعه و ارزیابی می‌شود (شکل ۲). نقشه‌یابی ارتباطی در مقایسه با نقشه‌یابی پیوستگی، به‌مراتب دقیق‌تر است و به‌راحتی می‌توان از آن در فرآیند گزینش به‌کمک نشانگر استفاده کرد (Moose and Mumm, 2008).

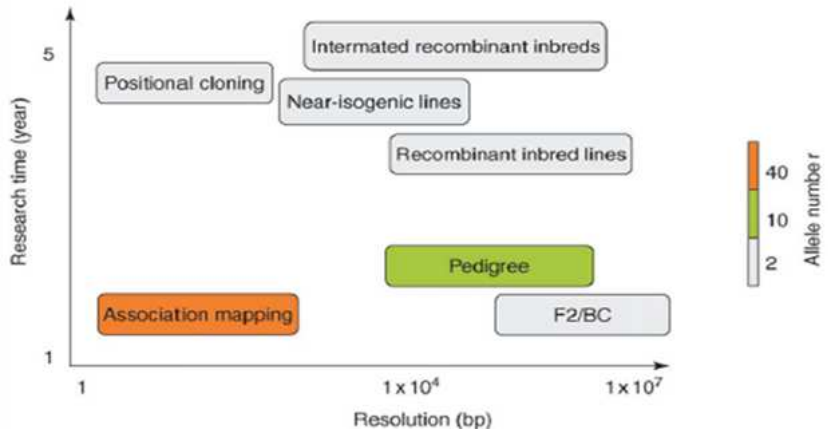
از زمانی که نقشه‌یابی ارتباطی در گیاهان مطرح شد، علاقه‌مندی برای شناسایی ژن‌های جدید و بهبود روش‌های آماری، به‌ویژه با پیشرفت‌های چشم‌گیر در فن‌آوری توالی‌یابی نسل جدید، بیش‌تر شده است.

اصطلاح‌های نقشه‌یابی ارتباطی و عدم تعادل پیوستگی در منابع مکرراً به‌جای یکدیگر مورد استفاده قرار می‌گیرند، در حالی‌که منظور از نقشه‌یابی ارتباطی، ارتباط معنی‌دار بین یک نشانگر و یک صفت کمی است و عدم تعادل پیوستگی به ارتباط معنی‌دار دو نشانگر، دو ژن و یا یک نشانگر و یک ژن اشاره دارد. به‌عبارت بهتر، نقشه‌یابی ارتباطی، کواریانس معنی‌دار نشانگر و صفت کمی است و عدم تعادل پیوستگی کواریانس بین نشانگرها (ژن‌ها) است. نقشه‌یابی ارتباطی روشی امیدوار کننده برای غلبه بر محدودیت‌های نقشه‌یابی پیوستگی است. موفقیت نقشه‌یابی ارتباطی به امکان وجود عدم تعادل بین نشانگرها و آلل‌های کنترل‌کننده صفت کمی بستگی دارد (Zhu *et al.*, 2008). این روش برای اولین بار در ژنتیک انسانی و برای صفات کیفی (مانند بیماری‌های ژنتیکی) مورد استفاده قرار گرفت، اما امروزه در جمعیت‌های گیاهی و جانوری نیز به‌کار می‌رود (Reich *et al.*, 2001).



شکل ۱- مقایسه شماتیک (a) نقشه‌یابی پیوستگی با جمعیت دو والدی و (b) نقشه‌یابی ارتباطی با بهره‌گیری از جمعیت‌های طبیعی (اقتباس از Zhu *et al.*, 2008).

Figure 1. Schematic comparison of (a) linkage analysis with designed mapping populations and (b) association mapping with diverse collections (see Zhu *et al.*, 2008)



شکل ۲- مقایسه شماتیک روش‌های مختلف شناسایی تنوع نوکلئوتیدی مرتبط با صفت از نظر وضوح، زمان تحقیق و تعداد آلل (اقتباس از Yu and Buckler, 2006).

Figure 2. Schematic comparison of various methods for identifying trait-associated nucleotide polymorphism in terms of resolution, research time and allele number (see Yu and Buckler, 2006).

مراحل انجام یک آزمایش نقشه‌یابی ارتباطی

یک آزمایش نقشه‌یابی ارتباطی، شامل شش مرحله زیر است (Abdurakhmonov and Abdukarimov, 2008; Zhu *et al.*, 2008; Myles *et al.*, 2009):

۱- انتخاب گروهی از لاین‌ها یا ارقام زراعی با تنوع ژنتیکی گسترده برای تشکیل جمعیت یا پَنِل نقشه‌یابی

قبل از شروع نقشه‌یابی ارتباطی می‌بایست از تمام جنبه‌های ژنتیکی گونه‌ها و ژرم‌پلاسم در دسترس و نیز سطح پلوئیدی افراد اطلاع کافی داشت. بررسی این عوامل باعث می‌شود که از مشکلاتی که در اثر تفسیر چندشکلی به‌وجود می‌آید، جلوگیری شود. انتخاب ژرم‌پلاسم یکی از موضوعات مهم برای موفقیت در نقشه‌یابی ارتباطی است. تنوع ژنتیکی، وسعت و میزان عدم تعادل پیوستگی (LD) در ژنوم و ارتباطات داخل جمعیت، تراکم نشانگرها و روش‌های آماری که باعث تخمین وضوح نقشه می‌شوند را تعیین می‌کند. ویژگی‌های جمعیت مورد استفاده در نقشه‌یابی ارتباطی مانند نقشه‌یابی پیوستگی است، با این تفاوت که دیگر از جمعیت‌های پیشرفته حاصل از تلاقی دو ژنوتیپ استفاده نمی‌شود، بلکه از جمعیت‌های طبیعی و یا لاین‌های از پیش ساخته شده‌ای استفاده می‌شود که برای صفات مورد بررسی دارای تنوع وسیعی هستند.

نقشه‌یابی بر مبنای جمعیت طبیعی ابزار قدرتمندی برای شناسایی تنوع طبیعی صفات کمّی در بسیاری از گیاهان زراعی از سال ۲۰۰۱ فراهم کرده است. البته توان آماری نقشه‌یابی بر مبنای جمعیت طبیعی به‌شدت به میزان LD، ساختار جمعیت، روابط خویشاندی، اندازه جمعیت و

فراوانی آلل جزئی بستگی دارد (Yu and Buckler, 2006). وجود ساختار در جمعیت‌های نقشه‌یابی ارتباطی به‌عنوان یک محدودیت مهم ممکن است باعث ایجاد ارتباطات مثبت کاذب شوند، زیرا بسیاری از نشانگرهای خنثی به‌طور معنی‌داری با تفاوت‌های صفات مورد ارزیابی در بین زیرجمعیت‌ها همبستگی دارند. فراوانی آلل جزئی به‌عنوان محدودیت مهم دیگر در نقشه‌یابی بر مبنای جمعیت‌های طبیعی است و آلل‌های عملکردی با فراوانی کم‌تر از پنج درصد به‌سختی می‌توانند شناسایی شوند، مگر اینکه اثر بسیار بزرگی داشته باشند. در این رابطه، نقشه‌یابی بر مبنای خانواده، گزینه خوبی برای شناسایی چنین آلل‌هایی می‌باشد، زیرا فراوانی آلل‌ها را می‌توان به‌طور مصنوعی با ساخت جمعیت نقشه‌یابی افزایش داد. علاوه بر نقشه‌یابی بر مبنای جمعیت طبیعی، می‌توان از جمعیت‌های نقشه‌یابی چند والدی مانند نقشه‌یابی ارتباطی آشیانه‌ای و تلاقی‌های نسل پیشرفته چندوالدی نیز برای نقشه‌یابی استفاده کرد. این جمعیت‌ها برای غلبه بر محدودیت‌های جمعیت‌های دو والدی تولید می‌شوند که تنوع ژنتیکی چند والد باعث ایجاد تنوع فنوتیپی بالا و نیز وضوح مناسب برای نقشه‌یابی پیوستگی می‌شود (Xu *et al*., 2017).

۲- ثبت ارزیابی‌های فنوتیپی صفات

این مرحله همانند ارزیابی فنوتیپی در روش نقشه‌یابی پیوستگی است. اهمیت ارزیابی فنوتیپی به‌علت اثرگذاری بر ارزیابی‌های ژنوتیپی است و دقت ارزیابی ژنوتیپی به‌کمک اطلاعات و داده‌های دقیق فنوتیپی و با انتخاب نوع طرح و تکرار مناسب، روش آماری مناسب و کارآمد و بررسی برهمکنش محیط و QTL افزایش می‌یابد و باعث قدرتمندتر شدن نقشه‌یابی می‌شود. در انجام آزمایش برای نقشه‌یابی ارتباطی در محیط‌های مختلف، در نظر گرفتن برخی صفات مهم که روی نمود صفات دیگر اثر می‌گذارند، ضروری است. به‌عنوان مثال، زمان گلدهی، حساسیت به فتوپریود، خوابیدگی و واکنش به بیماری‌های شایع ژنوتیپ‌ها بهتر است بررسی شود، زیرا این صفات، اندازه مورفولوژی یا زراعی صفات دیگر به‌ویژه در شرایط مزرعه را تحت تأثیر قرار می‌دهند، به این مفهوم که در نقشه‌یابی ارتباطی صفتی مانند عملکرد دانه، ممکن است برخی ژنوتیپ‌ها تحت تأثیر حساسیت به خوابیدگی و یا بیماری‌ها نتوانند عملکرد پتانسیل خود را نشان دهند و نتایج نقشه‌یابی ارتباطی برای این صفت با اریب همراه شود. بنابراین، در صورتی که نقشه‌یابی ارتباطی صفتی تحت تأثیر صفات مختل‌کننده دیگری قرار گیرد، بهتر است قبل از انجام نقشه‌یابی ارتباطی، اثر صفات مختل‌کننده روی صفت مورد نقشه‌یابی ارتباطی با استفاده از تجزیه کوواریانس تصحیح شود و سپس برای داده‌های تصحیح‌شده، نقشه‌یابی ارتباطی انجام شود.

۳- ارزیابی ژنوتیپی جمعیت با استفاده از نشانگرهای پس‌زمینه‌ای یا نشانگرهای پیوسته احتمالی به ژن

بر این اساس، نقشه‌یابی ارتباطی به دو گروه اصلی شامل نقشه‌یابی ارتباطی در سطح کل ژنوم و نقشه‌یابی ارتباطی ژن کاندید تقسیم می‌شود.

الف- نقشه‌یابی ارتباطی در سطح کل ژنوم

در این روش، تنوع ژنتیکی کل ژنوم با استفاده از تعداد زیادی نشانگر مورد بررسی قرار می‌گیرد تا محل ژن مرتبط با فنوتیپ شناسایی شود (Zhu *et al*., 2008). ارزیابی ژنوتیپی به‌کمک چند دسته از نشانگرها انجام می‌شود. نشانگرهای پس‌زمینه‌ای که اگر جمعیت دارای ساختار باشد و وابستگی بین آن‌ها برقرار باشد، از ارتباطات کاذب بین آن‌ها جلوگیری می‌کند و خویشاوندی را تخمین می‌زند. نشانگرهای RAPD و AFLP می‌توانند به‌عنوان نشانگرهای پس‌زمینه‌ای به‌کار روند. البته امروزه بیش‌تر از نشانگرهای SSR و SNP استفاده می‌شود. نشانگرها می‌بایست کل ژنوم را پوشش دهند و به‌صورت تصادفی روی ژنوم قرار گیرند. نشانگرهای SSR به‌دلیل دارا بودن آلل‌های بیش‌تر، حاوی اطلاعات بیش‌تری نسبت به نشانگرهای SNP است و به‌همین دلیل اگر از نشانگرهای SNP استفاده شود، به تعداد بیش‌تری نشانگر احتیاج است. تقریباً قدرت تشخیص هر ۱۰۰ نشانگر SNP برابر ۲۰-۱۰ نشانگر SSR است. در ارزیابی ژنوتیپی با استفاده از نشانگرهای SNP که امروزه گسترش یافته است (Gupta *et al*., 2005; Jain and Brar, 2010)، روش‌های مختلفی برای توسعه نشانگر در دسترس است که دقیق‌ترین آن روش توالی‌یابی است.

ب- نقشه‌یابی ارتباطی ژن کاندید

در روش نقشه‌یابی ژن کاندید، علاوه بر نشانگرهای پس‌زمینه‌ای، از نشانگرهای اختصاصی پیوسته به ژن نیز برای شناسایی نشانگرهای پیوسته به فنوتیپ استفاده می‌شود. در این روش، رابطه بین چندشکلی داخل ژن کاندید با تنوع فنوتیپی صفت بررسی می‌شود. انتخاب نوع نقشه‌یابی ارتباطی در سطح کل ژنوم و یا ژن کاندید، به

میزان عدم تعادل پیوستگی و نشانگرهای موجود بستگی دارد. هر چند نقشه‌یابی ارتباطی در سطح ژنوم، روش امیدوارکننده‌ای برای غربال کل ژنوم با استفاده از تعداد زیادی نشانگر برای شناسایی رابطه بین نشانگر و صفت است، اما روش نقشه‌یابی ژن کاندید نیز برای نقشه‌یابی ژن‌های هدف با عملکردِ شناخته شده مفید است.

۴- تعیین میزان عدم تعادل پیوستگی (LD) برای یک کروموزوم و یا ژنوم با استفاده از داده‌های نشانگرهای مولکولی روی جمعیت نقشه‌یابی

عدم تعادل پیوستگی وضعیتی را نشان می‌دهد که در آن حضور هم‌زمان دو آلل مربوط به مکان‌های ژنی متفاوت در گروهی از افراد بیش از مقدار مورد انتظار تفرق مندلی است. عدم تعادل پیوستگی نشان‌دهنده همبستگی غیرتصادفی آلل‌ها در مکان‌های ژنی مختلف روی یک کروموزوم (گروه پیوستگی) یا کروموزوم‌های مختلف است.

گستره عدم تعادل پیوستگی در گیاهان، بسته به گونه و نوع جمعیت مورد بررسی، از صدها جفت باز تا صدها جفت کیلوباز مشاهده شده است. در جمعیت های مصنوعی اصلاحی، تنها عامل ایجادکننده عدم تعادل، پیوستگی ژنی (به‌علت نزدیکی فیزیکی ژن‌ها بر روی کروموزوم‌ها) است، ولی در جمعیت‌های طبیعی، عدم تعادل پیوستگی می‌تواند توسط عوامل دیگری غیر از پیوستگی ژنی مانند جهش، مهاجرت، گزینش و رانده‌شدگی ژنتیکی طی دوره تکامل یک جمعیت به‌وجود آید (Gupta *et al.*, 2005; Stich *et al.*, 2005, 2006, 2007; Oraguzie *et al.*, 2007). پیوستگی زیاد بین دو ژن (آلل) باعث ایجاد سطح بالایی از عدم تعادل پیوستگی می‌شود. برای مثال، اگر دو جهش در نزدیکی یکدیگر اتفاق بیفتند، به‌دلیل آن‌که این‌دو طی زمان دست‌خوش نیروهای گزینشی و رانشی مشابهی قرار می‌گیرند و به‌دلیل آن‌که احتمال وقوع نوترکیبی بین این‌دو جهش بسیار کم است، از این‌رو سطح بالایی از همبستگی و عدم تعادل پیوستگی بین این دو وجود خواهد داشت.

عواملی مانند جهش جدید، خودگشنی، ساختار جمعیت، روابط خویشاوندی افراد، رانش ژنتیکی و انتخاب (طبیعی، مصنوعی و متعادل کننده) باعث افزایش و عواملی مانند نوترکیبی بالا، نرخ جهش، جهش‌های تکراری و دگرگشنی باعث کاهش میزان عدم تعادل پیوستگی می‌شوند (Gupta *et al.*, 2005; Oraguzie *et al.*, 2007). انتخاب هم‌زمان بین جایگاه‌های ژنی در طول برنامه‌های به‌نژادی برای چندین صفت به‌طور معمول باعث ایجاد عدم تعادل پیوستگی بین جفت نشانگرهای مستقل می‌شود. از نقطه نظر تئوریک، روابط خویشاوندی در جمعیت‌ها باعث ایجاد عدم تعادل پیوستگی بین مکان‌های ژنی پیوسته می‌شود. با این‌حال، وجود یک والد غالب در جمعیت (به‌طوری که بسیاری از افراد جمعیت به‌نحوی از این والد مشتق شده‌اند)، ممکن است باعث ایجاد عدم تعادل پیوستگی بین نشانگرهای غیرپیوسته شود. علاوه بر این عوامل، ساختار جمعیت و رانش ژنتیکی، نیروهایی هستند که باعث ایجاد عدم تعادل پیوستگی در بین نشانگرهای غیرپیوسته در جمعیت می‌شوند (Abdurakhmonov *et al.*, 2008). عدم تعادل پیوستگی ایجاد شده توسط انتخاب، ساختار و روابط خویشاوندی جمعیت و رانش ژنتیکی، با توجه به این‌که منجر به کاهش تعداد نشانگرهای مورد نیاز برای مطالعه نقشه‌یابی ارتباطی می‌شوند، ممکن است در برخی موقعیت‌ها و جمعیت‌ها سودمند باشد (Stich *et al.*, 2005, 2006)، اما برای به‌دست آوردن نتایج نااریب از نقشه‌یابی ارتباطی لازم است عوامل مؤثر بر عدم تعادل پیوستگی مانند ساختار و روابط خویشاوندی جمعیت کنترل شوند (Liu and Muse, 2005; Pritchard *et al.*, 2000b).

آماره‌های مختلفی برای اندازه‌گیری میزان عدم تعادل پیوستگی وجود دارند (Hedrick, 1987) که هدف همه آن‌ها، تخمین ارزش پیشگویی یک مکان ژنی بر اساس مکان ژنی دیگر است. اگر آماره عدم تعادل پیوستگی (LD) صفر باشد، دو مکان ژنی مورد بررسی ارزش پیشگویی صفر برای یک‌دیگر خواهند داشت. بسیاری از اندازه‌گیری‌های LD بر پایه انحراف فراوانی هاپلوتیپ مشاهده شده از مقدار مورد انتظار آن‌ها است. یکی از آماره‌هایی که برای محاسبه LD استفاده می‌شود، آماره D لوونتین (Lewontin, 1964) است که از تفاوت بین فراوانی گامت‌های جفتی و غیرجفتی در دو مکان ژنی محاسبه می‌شود. با فرض وجود دو مکان ژنی A و B هر یک با دوآلل، فراوانی‌ها و هاپلوتیپ‌ها به‌صورت جدول ۱ خواهند بود.

جدول ۱- فراونی‌های آللی و هاپلوتیپی دو مکان ژنی دوآللی
Table 1. Haplotype and allele frequencies of two biallelic loci

Allele	B_1	B_2	Total
A_1	P_{11}	P_{12}	p_1
A_2	P_{21}	P_{22}	p_2
Total	q_1	q_2	1

آماره D بر اساس اطلاعات جدول ۱ به‌صورت رابطه (۱) محاسبه می‌شود:

$$\begin{aligned} D &= Pr(A_1,B_1) - Pr(A_1)Pr(B_1) \\ &= P_{11} - p_1q_1 = P_{22} - p_2q_2 \\ &= P_{11}P_{22} - P_{12}P_{21} \end{aligned} \quad (۱)$$

که در آن، $Pr(A_1,B_1)$ فراوانی گامت یا هاپلوتیپ A_1B_1 و $Pr(A_1)$ و $Pr(B_1)$ به‌ترتیب فراوانی آلل‌های A_1 و B_1 هستند.

ترتیب آلل‌ها در ردیف‌ها و ستون‌های جدول ۱ به‌صورت اختیاری است و مقدار آماره D اغلب بدون علامت و به‌صورت $|D|$ گزارش می‌شود. اما مقدار آماره D وابسته به فراوانی آللی است و برای مقایسه مکان‌های ژنی مختلف با فراوانی‌های آللی متفاوت، مقدار استانداردشده این آماره می‌تواند تخمین بهتری باشد. لوونتین (Lewontin, 1964) مقدار استاندارد شده D را تحت عنوان 'D به‌صورت رابطه (۲) معرفی کرد:

$$D' = \begin{cases} \dfrac{D}{Min(p_1q_2, p_2q_1)} & D\rangle 0 \\ \dfrac{D}{Min(p_1q_1, p_2q_2)} & D\langle 0 \end{cases} \quad (۲)$$

که در آن، Min نشان‌دهنده حداقل فراوانی است.

زمانی که 'D=۱ باشد، نشان‌دهنده LD کامل بین دو ژن است. این حالت زمانی اتفاق می‌افتد که دو مکان ژنی LD کاملی داشته باشند و یا از بین چهار هاپلوتیپ، فقط سه هاپلوتیپ در جمعیت وجود داشته باشد. مقدار 'D<۱ نشان دهنده مختل شدن احتمالی LD کامل اجدادی در نتیجه نوترکیبی است، البته زمانی که تمامی چهار هاپلوتیپ ممکن مشاهده شوند.

هیل و ویر (Hill and Weir, 1994) آماره دیگری تحت عنوان r^2 معرفی کردند که یکی از معمول‌ترین آماره‌ها برای تخمین عدم تعادل پیوستگی در مطالعات نقشه‌یابی ارتباطی است. از نظر مفهومی و آماری r ضریب همبستگی پیرسون است که همبستگی بین حالت آللی یک مکان ژنی چندشکل را با حالت آللی مکان ژنی چندشکل دیگر نشان می‌دهد و r^2 همان توان دوم ضریب همبستگی است که ضریب تبیین (تشخیص) نیز نامیده می‌شود. از روابط (۳) و (۴) به‌ترتیب برای محاسبه r و r^2 در مطالعات نقشه‌یابی ارتباطی استفاده می‌شود:

$$\begin{aligned} r_{AB} &= \frac{Cov(X_A, X_B)}{\sqrt{Var(X_A)Var(X_B)}} \\ &= \frac{D_{AB}}{\sqrt{p_A(1-p_A)p_B(1-p_B)}} \end{aligned} \quad (۳)$$

$$r^2_{AB} = \frac{D^2_{AB}}{p_A(1-p_A)p_B(1-p_B)} \quad (۴)$$

'D برای مقایسه مکان‌های ژنی با فراوانی‌های متفاوت مناسب است. با این حال، برای مکان‌های ژنی با فراوانی آللی پایین، r^2 نتایج قابل اعتمادتری را نسبت به $|D|$ ارایه می‌دهد. تفاوت اساسی بین r^2 و 'D در این است که 'D تاریخچه نوترکیبی را در جمعیت نشان می‌دهد، در حالی‌که r^2 تحت تأثیر ترکیبی از وقایع نوترکیبی و جهش قرار می‌گیرد. بنابراین، تفاوت بین 'D و r^2 از وقایعی مانند ترکیب دو یا چند جمعیت و یا وقوع گلوگاه جمعیت در نسل‌های اخیر که جهش‌های جدید اتفاق افتاده‌اند، اما زمان کافی برای نوترکیبی کامل وجود نداشته است، می‌باشد.

در کنار 'D و r^2، آماره‌های دیگری به‌دلیل خطر انتساب جمعیت، برای محاسبه LD در علم اپیدمیولوژی ارایه شده است. یکی از این آماره‌ها که با δ نشان داده می‌شود (رابطه ۵)، برابر است با (Levin and Bertell, 1978):

$$\delta = \frac{D}{q_1P_{22}} \quad (۵)$$

که در آن، P_{22} نشان‌دهندهٔ فراوانی گامت یا هاپلوتیپ A_2B_2 است.

آماره دیگری (d) نیز برای محاسبه LD در اپیدمیولوژی پیشنهاد شده است (رابطه ۶)، که بر اساس تفاوت فراوانی‌های شرطی می‌باشد (Kaplan and Weir, 1992):

$$d = \frac{P_{11}}{q_1} = \frac{P_{12}}{q_2} = \frac{D}{q_1q_2} \quad (۶)$$

آخرین آماره (Q) برای برآورد LD که در اپیدیمولوژی استفاده می‌شود، توسط دیولین و ریچ (Devlin and Risch, 1995) معرفی شد (رابطه ۸)، که بر اساس نسبت شانس (Odds Ratio) است. نسبت شانس (OR) برای هر واقعه‌ای (رابطه ۷) برابر با احتمال وقوع آن واقعه تقسیم بر احتمال عدم وقوع آن است:

$$OR\ \ (Odds\ Ratio) = \frac{P_{11}P_{22}}{P_{12}P_{21}} \quad (۷)$$

$$Q = \frac{OR - 1}{OR + 1} = \frac{P_{11}P_{22} - P_{12}P_{21}}{P_{11}P_{22} + P_{12}P_{21}} = \frac{D}{P_{11}P_{22} + P_{12}P_{21}} \quad (۸)$$

که در آن، P_{11}، P_{22}، P_{12} و P_{21} به‌ترتیب نشان‌دهنده فراوانی هاپلوتیپ A_1B_1، A_2B_2، A_1B_2 و A_2B_1 هستند. بر خلاف نسبت شانس که دامنه‌ای بین صفر تا بی‌نهایت دارد، مقدار Q بین ۱- تا ۱+ متغییر است.

جدا از آثار جمعیت که خود ممکن است باعث پیچیده کردن مطالعات ارتباطی شود، آماره‌هایی نظیر r^2 قادر به تعریف و اندازه‌گیری سطح وقوع هم‌زمان آلل‌ها در دو مکان ژنی هستند (Hill and Robertson, 1968). هنگامی‌که این آماره برابر با صفر است، به این معنی است که دو آلل نسبت به هم در تعادل کامل هاردی-واینبرگ قرار دارند. هنگامی که مقدار این آماره به یک برسد، به این مفهوم است که این دو آلل همیشه در کنار یکدیگر قرار دارند و همیشه با هم به ارث می‌رسند. نقشه‌یابی ارتباطی از این ویژگی آماره‌ها در جهت پیش‌بینی یک فنوتیپ بر اساس مکان ژنی که در مجاورت ناحیه کروموزمی کنترل کننده صفت مورد نظر قرار دارد، استفاده می‌کند. توان آماری نقشه‌یابی ارتباطی به‌وسیله مقدار LD بین چندشکلی‌های علّی (نواحی ژنومی موثر بر صفت) و اندازه جمعیت مورد استفاده تعیین می‌شود (Long and Langley, 1999; Wang and Rannala, 2005). میزان LD در نتیجه اتفاقات نوترکیبی در طول زمان کاهش می‌یابد. در صورتی‌که تعداد اتفاقات نوترکیبی بیش‌تری بین دو مکان ژنی اتفاق بیافتد، LD سریع‌تر کاهش می‌یابد. اگر LD در یک ناحیه ژنومی خیلی سریع کاهش یابد، برای پویش آن منطقه ژنومی به تعداد بیش‌تری نشانگر نیاز خواهد بود. در مقابل، اگر LD خیلی آهسته کاهش یابد، اندازه بلوک‌های هاپلوتایپی بسیار بزرگ خواهد بود و بنابراین به‌طور واضح نمی‌توان مکان‌های ژنی علّی (نواحی موثر بر صفت) را شناسایی کرد. به‌عبارت دیگر، کاهش LD روی فاصله فیزیکی در جمعیت‌های مورد مطالعه، تراکم نشانگری مورد نیاز و سطح وضوح مطالعات نقشه‌یابی را تعیین می‌کند. بنابراین در کنار آگاهی از میزان عدم تعادل پیوستگی، آگاهی از ساختار بلوک‌های هاپلوتیپی می‌تواند اطلاعات مهمی را در مطالعات نقشه‌یابی ارتباطی، انتخاب ژنومی و مناطق ژنومی تحت انتخاب طی تکامل فراهم کند (Sabeti *et al.*, 2002). بلوک‌های هاپلوتایپی، گروهی از آلل‌ها در بخشی از توالی DNA هستند که پیوستگی بسیار بالایی با هم دارند و به احتمال زیاد با هم به ارث می‌رسند (Gabriel *et al.*, 2002).

۵- ارزیابی ساختار و روابط خویشاوندی جمعیت (ضریب رابطه بین دو به دوی افراد)

به‌طور ساده می‌توان گفت که نقشه‌یابی ارتباطی عبارت از یک همبستگی ساده بین صفت و نشانگر است، اما پیچیدگی‌های موجود در ساختار جمعیت آنالیز را پیچیده می‌کند، به‌طوری‌که مهم‌ترین مشکل استفاده از روش نقشه‌یابی ارتباطی برای گیاهان زراعی به‌دلیل ساختار تعیین نشده جمعیت و اختلاط جمعیت به‌دلیل عامل‌هایی چون سازگاری و اهلی شدن است (Thornsberry *et al.*, 2001; Wright and Gaut, 2004). نقشه‌یابی ارتباطی تا دهه‌های اخیر به‌دلیل ارتباط دروغین که توسط ساختار جمعیت ایجاد می‌شد، با تردید در ژنتیک و اصلاح گیاهان مورد استفاده قرار می‌گرفت. ساختار جمعیت باعث ایجاد عدم تعادل پیوستگی بین مکان‌های ژنی غیرپیوسته می‌شود. هنگامی که فراوانی‌های آللی بین زیرجمعیت‌های یک گونه به‌دلیل عواملی نظیر رانش ژنتیکی، اهلی شدن یا گزینش پس‌زمینه‌ای، به‌طور معنی‌داری متفاوت باشد، در این حالت مکان‌های خنثی (که هیچ ارتباطی با صفت مورد نظر ندارند) ممکن است ارتباط معنی‌داری با صفت نشان دهند. در صورتی‌که تعداد زیادی از نشانگرهای خنثی برای تخمین اثر ساختار جمعیت فراهم باشد، در این‌صورت می‌توان به‌طور آماری اثر ساختار روی تجزیه ارتباطی را تخمین زد و از آن جدا کرد (Yu *et al.*, 2006).

توسعه مدل‌های آماری محاسبه ساختار جمعیت برای مطالعات نقشه‌یابی ارتباطی، استفاده از نقشه‌یابی ارتباطی برای شناسایی مکان‌های ژنی کنترل کننده صفات کمی را برای گیاهان زراعی بهبود بخشید (Pritchard, 2001; Bamshad *et al.*, 2004). در جمعیت‌هایی که دارای ساختار هستند، اغلب از سه روش به‌عنوان روش‌های مناسب برای کنترل آماری اثر ساختار جمعیت در آزمون‌های ارتباطی استفاده می‌شود:

- کنترل ژنومی (Genomic Control) (Devlin and Roeder, 1999; Devlin *et al.*, 2001, 2004).
- روش ارتباط ساختاری (Structural Association)، که بر اساس نوع مطالعه ارتباطی به دو نوع کنترل موردی (Pritchard *et al.*, 2000b) و مطالعات ارتباطی صفت کمّی (Thornsberry *et al.*, 2001; Camus-Kulandaivelu *et al.*, 2006) تقسیم می‌شود.

- مدل مخلوط متحدشده (Mixed Model) (ماتریس‌های Q+K) (Yu *et al.*, 2006).

روش کنترل ژنومی که برای کنترل آماری ساختار جمعیت استفاده می‌شود، فرض می‌کند که ساختار جمعیت دارای آثار یکسان روی تمامی مکان‌های ژنی است. در روش کنترل ژنومی، یک دسته از نشانگرهای تصادفی (مانند چندشکلی‌هایی که احتمال نمی‌رود با صفت مورد نظر مرتبط باشند) برای تخمین اثر ساختار جمعیت بر آماره‌های آزمون ارتباطی استفاده می‌شود، به‌طوری‌که معنی‌داری آماره ارتباطی (P-value) تخمین زده‌شده برای ساختار جمعیت تصحیح می‌شود. اصول کلی کنترل ژنومی، استفاده از ژنوم افراد دخیل در نمونه برای تخمین سطوح پیچیدگی‌های ایجاد شده توسط زیرساختارهای جمعیت و خویشاوندی‌های مستقیم‌تر و نزدیک‌تر شامل خویشاوندی‌های فامیلی است و مطابق با آن سطح نهایی معنی‌داری ارتباطی گزارش‌شده تعیین می‌شود (Devlin *et al.*, 2001).

در روش ارتباط ساختاری، از نشانگرهای غیرپیوسته به ژن‌های کاندید برای استنباط تعداد زیرجمعیت‌ها و افراد در هر زیرجمعیت استفاده می‌شود. کاربرد ارتباط ساختاری در مورد صفات کمی و کیفی با توجه به نوع جمعیت و صفت مورد مطالعه، با مدل مناسب، به‌ترتیب مدل ارتباطی صفت کمّی و کنترل موردی، تجزیه می‌شود. در استفاده از ارتباط ساختاری برای نقشه‌یابی صفات کمّی دو مرحله وجود دارد. در مرحله اول، احتمال عضویت هر فرد به یک زیرجمعیت محاسبه می‌شود (Pritchard *et al.*, 2000a, b) و سپس در مرحله بعد، احتمال عضویت محاسبه‌شده هر فرد به هر زیرجمعیت به‌عنوان یک متغیر در مدل تجزیه ارتباطی استفاده می‌شود (Pritchard *et al.*, 2000b).

در مطالعات کنترل موردی، احتمال توزیع فراوانی نشانگرها بر پایه ساختار جمعیت بین نمونه‌های شاهد و مورد مطالعه مقایسه می‌شود. برای صفات کمّی، ساختار تخمین زده شده جمعیت به‌عنوان یک متغیر کمکی در مدل رگرسیون استفاده می‌شود که در این مدل ارتباط بین ژنوتیپ و فنوتیپ محاسبه می‌شود (Thornsberry *et al.*, 2001; Camus-Kulandaivelu *et al.*, 2006).

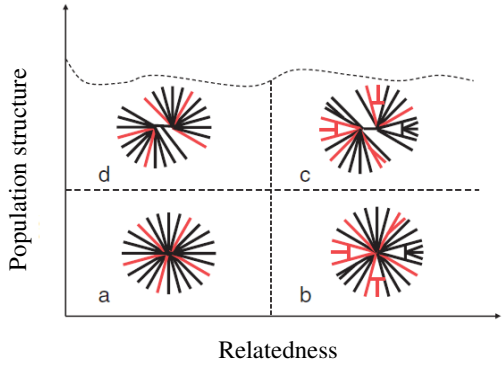
در مدل مخلوط متحدشده که توسط Yu و همکاران (Yu *et al.*, 2006) پیشنهاد شده است، تعداد زیادی از نشانگرهای تصادفی که کل ژنوم را می‌پوشانند، برای تخمین ساختار جمعیت (Q) و ماتریس خویشاوندی نسبی (K) استفاده می‌شود. در این روش، هر یک از عامل‌های ارتباط فامیلی بین افراد (K) و ارتباط بر اساس ساختار جمعیت (Q) که باعث پیچیدگی تجزیه ارتباطی می‌شوند را به‌عنوان یک متغیر مستقل در نظر می‌گیرند. برای محاسبه آثار مشترک عامل‌های خویشاوندی در محاسبات ارتباطی، این عامل‌ها به‌صورت متغیرهای کمکی در مدل رگرسیونی که ارتباط بین صفت و ژنوتیپ را نشان می‌دهند، به‌کار می‌روند. آرایش ژنتیکی جمعیت استفاده شده در مطالعه ارتباطی، مدل و نوع آماره‌های ارتباطی به‌کار رفته برای آزمون ارتباطی را مشخص می‌کند.

مدل مخلوط Q+K با رابطه (۹) نشان داده می‌شود:

$$Y = X\beta + S\alpha + Qv + Zu + e \qquad (۹)$$

در این رابطه، Y بردار مشاهدات فنوتیپی، β بردار آثار ثابت به‌غیر از ساختار جمعیت و داده‌های نشانگر، α بردار آثار نشانگر، u بردار آثار خویشاوندی پس‌زمینه‌ای، e بردار باقیمانده‌ها، Q ماتریس ساختار مربوط به v و X، S و Z به‌ترتیب نشان‌دهنده ماتریس‌های صفر و یک مرتبط با β، α و u هستند.

سطوح مختلفی از ساختار جمعیت از مقدار صفر تا جمعیت‌های با ساختار بالا در گندم شناسایی شده است و برخلاف برنج و ذرت، گروه‌های هتروتیک یا ساختار مشخص در گندم شناخته نشده است (Zhao *et al.*, 2015). لازم به توضیح است که گروه‌های هتروتیک به گروهی از ژنوتیپ‌ها اطلاق می‌شود که ترکیب‌پذیری و واکنش مشابهی در تلاقی با ژنوتیپ‌های متمایز از نظر ژنتیکی نشان می‌دهند (Melchinger and Gumber, 1998). الگوی هتروتیک نیز به یک جفت خاص یعنی دو گروه هتروتیک که هتروزیس بالایی در تلاقی با یکدیگر نشان می‌دهند، گفته می‌شود. الگوی هتروتیک و متوسط هتروزیس در گندم در مطالعات مختلف، متفاوت گزارش شده است. ساختار شماتیک جمعیت‌های مختلف مورد استفاده در مطالعات ارتباطی به‌همراه مدل مناسب برای تجزیه هر یک در شکل ۳ ارایه شده است.



شکل ۳- ساختار و روابط خویشاوندی انواع جمعیت‌های مورد استفاده در مطالعات نقشه‌یابی ارتباطی. a) یک جمعیت ایده‌آل با ساختار ظریف و روابط فامیلی (همانند جمعیت F_2 و یا جمعیت‌های ساختگی). روش مناسب تجزیه: روش رگرسیون و کنترل ژنومی، b) خانواده مبتنی بر نمونه (همانند شجره گسترش‌یافته). روش مناسب تجزیه: آزمون عدم تعادل انتقال، آزمون انتقال کمی عدم تعادل، کنترل ژنومی و مدل مخلوط، c) نمونه با ساختار جمعیت (همانند توده‌های ذرت). روش مناسب تجزیه: روش ارتباط ساختاری و کنترل ژنومی، d) نمونه با ساختار جمعیت و ارتباطات فامیلی (همانند پنل ارتباطی ذرت). روش مناسب تجزیه: روش ارتباط ساختاری، کنترل ژنومی و مدل مخلوط (ماتریس ساختار جمعیت (Q) + ماتریس ضریب خویشاوندی نسبی (K)).

Figure 3. Different types of population structure and relatedness in association mapping studies. a) An ideal population with delicate structure and family relationships (such as F_2 population or artificial populations). Appropriate method of analysis: Regression and genomic control; b) Sample-based family (such as extended pedigree). Appropriate analysis method: Transmission unequilibrium test, quantitative unequilibrium transfer test, genomic control and mixed model; c) Sample with population structure (such as maize landraces). Appropriate method of analysis: Structural association and genomic control; d) Sample with population structure and family relationships (such as maize communication panel). Appropriate method of analysis: Structural association method, genomic control and mixed model (Q, population structure matrix + K, relative kinship coefficient matrix).

۶- تعیین رابطه بین داده‌های فنوتیپی و ژنوتیپی بر اساس اطلاعات حاصل از LD، ساختار جمعیت و روابط خویشاوندی با استفاده از روش‌های آماری مناسب

روش آماری پایه برای نقشه‌یابی ارتباطی شامل رگرسیون خطی، تجزیه واریانس، آزمون t و X^2 می باشد. در این روش نقشه‌یابی، ارتباط یک نشانگر با صفت در شرایط مختلف مورد بررسی قرار می‌گیرد.

این روش برای اولین بار در ژنتیک انسانی و برای تجزیه و تحلیل ژنتیکی صفات کیفی مانند بیماری‌ها مورد استفاده قرار گرفت، اما در حال حاضر استفاده از آن در جمعیت‌های گیاهی و جانوری رو به افزایش است. اولین گزارش تجزیه ارتباطی در گیاهان در سال ۲۰۰۱ در گیاه ذرت منتشر شد (Thornsberry *et al.*, 2001). پس از آن، تعیین نشانگرهای مرتبط با صفات مهم از طریق تجزیه ارتباط در گونه‌های گیاهی مختلفی از جمله توتون (Darvishzadeh *et al.*, 2014, 2016; Abedi *et al.*, 2014; Basirnia *et al.*, 2014; Dadras *et al.*, 2014; Darvishzadeh, 2016c)، پنبه (Abdurakhmonov and Abdukarimov, 2008)، سورگوم (Shehzad *et al.*, 2009; El Mannai *et al.*, 2011)، آفتابگردان (Darvishzadeh *et al.*, 2008; Darvishzadeh, 2016a; Fusari *et al.*, 2012; Davar and Darvishzadeh, 2012; Ahmadpour *et al.*, 2018; Soleimani *et al.*, 2018, Najafzadeh *et al.*, 2018)، گیاه دارویی گرچک (Goodarzi *et al.*, 2015; Darvishzadeh, 2016b)، نخود (Saeed *et al.*, 2013; Saeed and Darvishzadeh, 2017)، انگور (Razi *et al.*, 2018) و به‌ویژه در غلات (جدول ۲) انجام شده است. با کاهش هزینه‌های ارزیابی ژنوتیپی با تراکم بالا و گسترش داده‌های ژنوم مرجع برای بیش‌تر گیاهان، پیش‌بینی می‌شود که استفاده از نقشه‌یابی ارتباطی برای شناسایی مکان‌های ژنی کنترل کننده صفات کمی بیش از این نیز گسترش یابد.

جدول ۲- برخی مطالعات نقشه‌یابی ارتباطی در غلات

Table 2. Some of the association mapping studies in cereal crops

Plant	Population	Sample size	Marker	Trait	Reference
Rice	Cultivars and landraces	220	4929 SNP markers	Salinity stress tolerance	Kumar *et al.*, 2015
	Cultivars	328	30000 SNP markers	Ozone tolerance	Ueda *et al.*, 2015
	Elite breeding lines	363	71170 SNP markers	Nineteen agronomic traits	Begum *et al.*, 2015
	Cultivars and landraces	95	263 SSR markers	Grain filling rate	Liu *et al.*, 2015
	Japanese rice collection	175	3168 SNP markers	Metabolites	Matsuda *et al.*, 2015
	Elite hybrid cultivars	1495	1654030 SNP markers	38 agronomic traits	Huang *et al.*, 2015
	Iranian rice varieties	132	12 Closely linked SSR markers to *SalTol*	Salinity stress	Kordrostami *et al.*, 2016
	Japanese *japonica* cultivars	176	43323 SNP markers	Agronomic traits	Yano *et al.*, 2016
	Cultivars and landraces	270	1019883 SNP markers	Drought stress tolerance	Ma *et al.*, 2016
	Diverse cultivars	315	44100 SNP markers	Agronomic traits	Liu *et al.*, 2016
	Diverse indica samples	432	5291 SNP markers	Flood tolerance	Zhang *et al.*, 2017
	Diverse indica population	203	16232 SNP markers	Spike sterility	Dingkuhn *et al.*, 2017
	Elite breeding lines	217	43394 SNP markers	Hybrid vigor	Guo *et al.*, 2017
	Diverse population	258	22488 SNP markers	Grain quality	Wang *et al.*, 2017
	Diverse population from IRRI	222	700000 SNP markers	Zinc and iron toxicity	Zhang *et al.*, 2017
	Diverse indica population	203	16232 SNP markers	Flowering time	Dingkuhn *et al.*, 2017
	Diverse indica population	453	5291 SNP markers	Seed dormancy	Lu *et al.*, 2018
	Diverse population	478	162529 SNP markers	Salinity stress tolerance	Cui et al., 2018
	Cultivars	190	112565 SNP markers	Salinity stress tolerance	Lekkar et al., 2019
	Diverse cultivars	83	44 SSR and miRNA-SSR markers	Drought stress	Tabkhkar *et al.*, 2020
	Diverse cultivars	121	42 Linked SSR markers to major blast resistance genes	Leaf blast disease	Zarbafi *et al.*, 2020
Wheat	Spring genotypes	246	17937 SNP markers	Grain protein and quality	Kumar *et al.*, 2018
	Winter and facultative cultivars	150	18085 SNP markers	Salinity stress tolerance	Oyiga *et al.*, 2018
	Soft winter lines and cultivars	238	3919 SNP markers	Fusarium head blight	Tessmann and Sanford, 2018
	Elite breeding lines	66	15555 SNP markers	Yield and its components	Ma *et al.*, 2018

جدول ۲- ادامه

Table 2. Continued

Plant	Population	Sample size	Marker	Trait	Reference
Wheat	Winter lines	215	20881 SNP markers	Root traits	Beyer *et al.*, 2018
	Elite germplasm	213	11461 SNP markers	Fusarium head blight	Wu *et al.*, 2019
	Diverse population	404	35143 SNP markers	38 agronomic traits	Sheoran *et al.*, 2019
	Cultivars and landraces	1292	5011 SNP markers	Powdery mildew	Li *et al.*, 2019
Barley	Spring cultivars	154	14 AFLP and 32 SSR markers	Drought stress tolerance	Jabbari *et al.*, 2018
	Double haploid lines	122	9680 SNP markers	Agronomic traits	Hu *et al.*, 2018
	Cultivars	379	6810 SNP markers	Grain yield	Xu *et al.*, 2018
	Diverse population	282	12239 DArT markers	Leaf rust	Singh *et al.*, 2018
	Diverse population	221	7864 SNP markers	Root system architecture	Jai *et al.*, 2019
	NAM population	1420	5333 SNP markers	Drought stress tolerance	Pham *et al.*, 2019
	Diverse population	449	33818 SNP markers	Net blotch resistance	Novakazi *et al.*, 2019
Maize	Inbred lines	257	48193 SNP markers	Stem lodging	Zhang *et al.*, 2018
	Inbred lines	923	347765 SNP markers	Grain zinc and iron	Hindu *et al.*, 2018
	Inbred lines	126	46046 SNP markers	Grain yield	Li *et al.*, 2018
	Inbred lines	292	56110 SNP markers	Ear quantitative traits	Zhu *et al.*, 2018
	Elite inbred lines	253	2824 SNP markers	Grain size and weight	Hao *et al.*, 2019
	Natural diverse population	445	1242565 SNP markers	Salinity stress tolerance	Luo *et al.*, 2019
	Inbred lines	100	81 ISSR markers	Agro-morphologic traits	Ghaffari Azar *et al.*, 2018

نرم‌افزارهای مورد استفاده در تجزیه ارتباطی

در حال حاضر چندین بسته نرم‌افزاری برای انجام تجزیه ارتباطی در دسترس هستند (جدول ۳). از آنجایی‌که بسیاری از این بسته‌ها در اصل برای سیستم‌های جانوری طراحی شده‌اند، قبل از استفاده از این برنامه‌ها، مناسب بودن آن‌ها برای انجام نقشه‌یابی ارتباطی در سیستم‌های گیاهی باید مورد بررسی قرار گیرد (Zhang *et al.*, 2009).

نرم‌افزار TASSEL یکی از مهم‌ترین نرم‌افزارهایی است که برای نقشه‌یابی ارتباطی در گیاهان استفاده (Bradbury *et al.*, 2007) و به‌طور مداوم به‌روزرسانی می‌شود. برای انجام نقشه‌یابی ارتباطی در این نرم‌افزار، باید داده‌های فنوتیپی و ژنوتیپی و ماتریس Q به نرم‌افزار معرفی شوند. البته می‌توان با استفاده از این نرم‌افزار، تجزیه به مولفه‌های اصلی (PCA) انجام داد و نتیجه PCA را به‌عنوان متغیر کمکی (کوواریت) به‌جای ماتریس Q در تجزیه ارتباطی در نظر گرفت. تجزیه ارتباطی در این نرم‌افزار با استفاده از مدل خطی عمومی (GLM) و مدل خطی مخلوط (MLM) انجام می‌شود. اگر از مدل خطی عمومی برای تعیین ارتباط نشانگر-صفت استفاده شود، روابط خویشاوندی بین افراد در نظر گرفته نمی‌شود و فقط ماتریس Q یا PCA که نشان دهنده ساختار جمعیت است، در محاسبات وارد می‌شوند. نقشه‌یابی ارتباطی با مدل خطی مخلوط، نیازمند محاسبه ماتریس K (ماتریس روابط خویشاوندی بین افراد جمعیت) است که با استفاده از داده‌های ژنوتیپی در این نرم‌افزار محاسبه و سپس ارتباط بین نشانگر-صفت به‌همراه احتمال اشتباه نوع اول محاسبه می‌شود.

یکی دیگر از نرم‌افزارهای مناسب برای گیاهان، GAPIT است که مبتنی بر R می‌باشد (Lipka *et al.*, 2012) و می‌تواند تجزیه ارتباط و پیش‌بینی ژنومی را با روش‌های دیگری مانند ECMLM و CMLM انجام دهد. از ویژگی‌های مهم این بسته نرم‌افزاری اداره مجموعه بسیار بزرگی از داده‌ها در زمان محاسبه کم‌تر است.

روش‌های پسا-نقشه‌یابی ارتباطی (Post-GWAS) برای به‌نژادی گیاهی

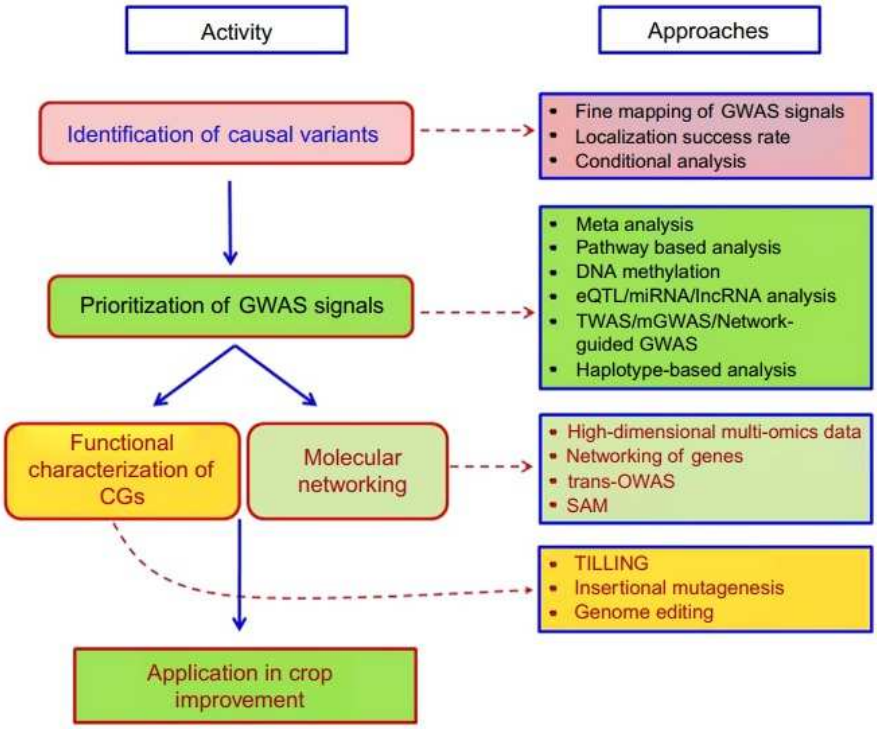
نقشه‌یابی فاصله‌ای (IM=Interval Mapping) و مطالعات نقشه‌یابی در سطح ژنوم (GWAS=Genome Wide Association Studies) دو رویکرد مهمی هستند که به‌طور گسترده برای تعیین ارتباط بین نشانگر-صفت در بسیاری از گیاهان زراعی مهم استفاده می‌شوند. تا به حال، میلیون‌ها SNP در مطالعات مختلف نقشه‌یابی، توسعه یافته و هزاران ارتباط نشانگر-صفت شناسایی شده است. از نشانگرهای شناسایی شده در مطالعات ارتباط نشانگر-صفت، بدون اینکه اطلاعاتی از ارتباط سببی بین نشانگرهای پیوسته و ژن‌های کنترل کننده صفت یا صفات وجود داشته باشد، برای گزینش به‌کمک نشانگر استفاده می‌شود. در این رابطه، نقشه‌یابی فاصله‌ای با وجود نداشتن اطلاعات در مورد روابط پیوستگی، در مقایسه با مطالعات نقشه‌یابی در سطح کل ژنوم بسیار مفیدتر است، هر چند که مطالعات نقشه‌یابی در سطح کل ژنوم وضوح بهتری دارد. به هر حال، اخیراً در پسا-نقشه‌یابی ارتباطی، تلاش‌های زیادی برای پیدا کردن SNPهای سببی از بین نشانگرهایی که در نقشه‌یابی ارتباطی با صفت همبستگی نشان داده‌اند، انجام شده است. استفاده از این SNPهای سببی، ژن‌های کاندید و مسیرهای منتهی به فنوتیپ می‌تواند اطلاعات ارزشمندی در بهبود گیاهان زراعی فراهم آورد (شکل ۴). چنین مطالعاتی نه‌تنها برای درک معماری ژنتیکی صفت، بلکه برای دستکاری ژن‌های صفات هدف در گیاهان زراعی ویژه نیز سودمند هستند. بنابراین، اولین قدم در مطالعه پسا- نقشه‌یابی ارتباطی، تعیین SNPهای سببی و ژن‌های کاندید مرتبط شامل ژن‌های رمزکننده فاکتورهای رونویسی هستند. البته این SNPها ممکن است در نواحی غیررمزکننده ژنوم نیز باشند که اغلب به‌عنوان نواحی تنظیم‌کننده شامل توالی‌های پروموتری و توالی‌هایی که به miRNAها، IncRNAها و ncRNAهای دیگر ترجمه می‌شوند، هستند (Gupta *et al.*, 2019).

نقشه‌یابی ارتباطی برای مطالعات نقشه‌یابی در سطح کل فنوم نیز استفاده شده است. به‌عنوان مثال، در برنج بر اساس پروفیل متابولیت (شامل ۸۴۰ متابولیت) به‌عنوان صفت فنوتیپی و ۶/۴ میلیون SNP به‌عنوان داده‌های ژنوتیپی، برای نقشه‌یابی ارتباطی در سطح کل متابولیکی ژنوم انجام شده است (Chen *et al.*, 2014). رویکرد تغییریافته دیگری نیز در برنج تحت عنوان نقشه‌یابی در سطح کل فنوم بر مبنای مسیرهای متابولیتی انجام شده است (Lu *et al.*, 2015). مطالعه مشابه دیگری برای بیوسنتز روغن در ذرت توسط لی و همکاران (Li *et al.*, 2013) انجام شده است.

جدول ۳- برخی از بسته‌های نرم‌افزاری مورد استفاده در مطالعات نقشه‌یابی ارتباطی (اقتباس از Zhu *et al.,* 2008)

Table 3. Some of the software packages used for association mapping studies (from Zhu *et al.,* 2008)

Software	Comment	Website
TASSEL	LD statistics, GLM, MLM, CMLM, P3D, genomic selection; graphical interphase, PCA and kinship; free	http://www.maizegenetics.net/tassel/
GAPIT	R-based, CMLM, fast computation, free	http://www.maizegenetics.net/gapit
R	Generic, commonly used for programing; free	http://www.r- ptoject.org/
PLINK	Handles virtually unlimited numbers of SNPs; MDS to visualize substructure; free	http://pngu.mgh.harvard.edu/~purcell/plink/
EMMA	Mixed model, corrects for the confounding from population structure and genetic relatedness; free	http://mouse.cs.ucla.edu/emma/
EMMAX	Large-scale association mapping, corrects for the confounding from population structure and genetic relatedness, increased computational speed; free	http://genetics.cs.ucla.edu/emmax/
EIGENSOFT	Uses principal components analysis to explicitly model ancestry differences between cases and controls; free	http://www.hsph.harvard.edu/alkes-price/ software/
GGT 2.0	Graphical genotypes; LD statistics; FDR calculation, does not control for population stratification of its own; free	http://www.wageningenur.nl/en/show/ Graphical-GenoTypes-transformmolecular-data-to-colorful-chromosomedrawings. htm
GenAMap	Performs automatic structured association mapping (SAM) using different algorithms; good graphical presentation; free	http://sailing.cs.cmu.edu/genamap/
Matapax	GWAS is performed in R environment with EMMA and GAPIT libraries; performs all essential steps for basic GWAS, population structure, fast computation; free	http://matapax.mpimp-golm.mpg.de
Merlin	Includes an integrated genotype inference feature for improved analysis when some genotypes are missing, does not control for population stratification of its own; free	http://www.sph.umich.edu/csg/abecasis/ merlin/tour/assoc.html
ASReml	Handle large data set, calculates population structure and pedigree-based kinship; commercial	http://www.vsni.co.uk/software/asreml
SAS	Generic program commonly used in data analysis; commercial	http://www.sas.com
JMP Genomics	Calculates population structure and marker-based kinship; commercial	http://www.jmp.com/software/genomics/
SVS	Comprehensive package with better visualization of the results; offers different options; commercial	http://www.goldenhelix.com/SNP_Variation/
GenStat	Performs GLM and MLM, takes care of population structure;commercial	http://www.vsni.co.uk/software/genstat
FaST-LMM	For analysis of large data sets (up to 120,000 individuals); free	http://fastlmm.codeplex.com/
GenABEL	Performs GWAS for quantitative as well as binary traits; free	http://www.genabel.org/packages/GenABEL



شکل ۴- نمودار روند مطالعات و فعالیت‌هایی که در زمینه پسا-نقشه‌یابی ارتباطی می‌تواند انجام شود (اقتباس از Gupta *et al.*, 2019)
Figure 4. A flow chart of the studies and activities which can be carried out in the post-GWAS era (see Gupta *et al.*, 2019)

چالش‌ها و فرصت‌های استفاده از نقشه‌یابی ارتباطی

فناوری‌های توالی‌یابی نسل بعد، فرصت‌ها و چالش‌های جدیدی را در مطالعات نقشه‌یابی ارتباطی ایجاد کرده‌اند. رویکردهای جدید برای ارزیابی‌های فنوتیپی با توان بالا در سطح گسترده نیازمند توسعه ظرفیت‌های ارزیابی‌های ژنوتیپی در همان سطح است. از طرف دیگر، داده‌های تولیدشده در مقیاس بالا با استفاده از توالی‌یابی‌ها، مزراع، اتاقک‌های رشد و گلخانه‌ها، نیازمند توسعه پایگاه‌های داده‌ای، روش‌های آماری، بیوانفورماتیکی و طرح‌های ژنتیکی هستند. در حال حاضر، راهکارهای توالی‌یابی مجدد با استفاده از توالی‌یابی RNA (Wang *et al.*, 2009b)، توالی‌یابی اگزوم (Ng *et al.*, 2009) و ازیابی ژنوتیپی با استفاده از توالی‌یابی (Huang *et al.*, 2009; Elshire *et al.*, 2011) در مطالعات نقشه‌یابی ارتباطی، بهینه‌سازی شده و در ادامه، بسیاری از زمینه‌های مطالعاتی جدید مانند شناسایی تنوع تعداد کپی (Springer *et al.*, 2009; Rogers and Bendich 1987) و تنوع حضور و عدم (Springer *et al.*, 2009) نیز می‌توانند به‌عنوان راه‌کارهای ارزیابی ژنوتیپی در نقشه‌یابی ارتباطی استفاده شوند. مطالعات توالی‌یابی ژنوم در برخی گیاهان زراعی مهم مانند گندم (Brenchley *et al.*, 2012) و جو (Mayer *et al.*, 2012)، منابع و فرصت‌های ارزشمندی برای تسریع شناسایی ژن‌های کلیدی در کنار ارزیابی‌های فنوتیپی در سطح گسترده فراهم کرده‌اند.

ارزیابی فنوتیپی در سطح گسترده یکی از بخش‌های پرهزینه در مطالعات ژنتیک گیاهی است. در حال حاضر، فنوتیپ‌های مورد استفاده در نقشه‌یابی ارتباطی، از روش‌های سنتی و پرهزینه ارزیابی‌های فنوتیپی به سطح بیان ژن، متابولیت‌ها و پروتئین‌ها توسعه داده می‌شوند. از طرف دیگر، فناوری‌های جدید مانند CT-SCAN، اسپکتوسکوپی مادون قرمز (NIR)، سیستم ارزیابی تک‌دانه برای کیفیت بذر، سیستم موقعیت‌یابی جهانی و تجزیه معکوس، امکان ارزیابی فنوتیپی دقیق و سریع را فراهم کرده‌اند. در ادمه لازم است مطالعات روی آلل‌های نادر، آلل‌های چندگانه، افزایش سرعت محاسبات، ذخیره داده‌ها و توان محاسبات متمرکز شوند. همچنین، اعتبارسنجی نتایج حاصل از نقشه‌یابی با استفاده از توالی‌یابی RNA و

روش‌های دیگر ارزشمند خواهد بود. مسلماً لازم خواهد بود که تحقیقات نقشه‌یابی ارتباطی با زمینه‌های دیگری مانند MutMap (Abe *et al.*, 2012)، نقشه‌یابی نسل بعد (NGM) (Schneeberger *et al.*, 2009)، انتخاب ژنومی (Bernardo and Yu 2007; Meuwissen *et al.*, 2001) و ژنومیک تطبیقی ادغام شود.

مثال‌های عددی در زمینه مطالعات نقشه‌یابی ارتباطی

۱- محاسبه LD در یک جمعیت نقشه‌یابی

به‌عنوان مثال وضعیتی را در نظر بگیرید که فراوانی‌های آللی برابر هستند، یعنی:

$f(A_1)=f(A_2)=f(B_1)=f(B_2)=۰/۵$

و اگر فراوانی هاپلوتایپ‌ها به‌صورت جدول ۴ باشند:

جدول ۴- فراوانی هاپلوتایپ‌ها در یک جمعیت فرضی

Table 4. Haplotypes frequency in a hypothetical population

		Marker A		
		A1	A2	Frequency
Marker B	B1	0.4	0.1	0.5
	B2	0.1	0.4	0.5
Frequency		0.5	0.5	

بنابراین:

$f(A1B1)=0.4 \quad , \quad f(A1B2)=0.1$
$f(A2B1)=0.1 \quad , \quad f(A2B2)=0.4$

در این‌صورت، مقادیر آماره‌های D و r^2 به‌صورت زیر برآورد خواهند شد:

$D=[f(A1B1)\times f(A2B2)]-f(A1B2)\times f(A2B1)]=0.15$

$$\Rightarrow D^2 = 0.0225$$

$$r^2 = \frac{D^2}{f(A_1)\times f(A_2)\times f(B_1)\times f(B_2)} = \frac{0.0225}{0.5\times 0.5\times 0.5\times 0.5} = 0.36$$

۲- محاسبه ماتریس خویشاوندی

ضریب خویشاوندی احتمال تشابه آلل‌های یک مکان ژنی در دو فرد، به‌دلیل وجود حداقل یک جد مشترک بین آن‌دو فرد و از رابطه (۱۰) محاسبه می‌شود:

$$Kf = \sum_k \sum_a \frac{(f_{ai} \times f_{aj})_k}{n} \qquad (۱۰)$$

که در آن، f_{ai} فراوانی آلل a در جمعیت یا فرد i ام، f_{aj} فراوانی آلل a در جمعیت یا فرد j ام و n تعداد مکان‌های ژنی (k=1, …, n) را نشان می‌دهد. به‌عنوان مثال، ارزیابی ژنوتیپی دو فرد I1 و I2 برای سه مکان ژنی (A، B و C) در جدول ۵ ارایه شده است:

جدول ۵- ژنوتیپ افراد I1 و I2 در سه مکان ژنی

Table 5. The genotypes of I1 and I2 in three loci

Individual	Locus A	Locus B	Locus C
I1	A	b	c
	A	b	c
I2	A	B	c
	A	B	c

بنابراین، ضریب خویشاوندی این‌دو فرد برابر است با:

$$Kf = \frac{(f_{A1}.f_{A2})_1 + (f_{a1}.f_{a2})_1 + ... + (f_{C1}.f_{C2}) + (f_{c1}\times f_{c2})_3}{n}$$
$$= \frac{(1\times1)+(0\times0)+...+(0\times0)+(1\times1)}{3} = \frac{2}{3}$$

۳- تجزیه ارتباطی

فرض کنید برای تجزیه ارتباطی، پنج فرد انتخاب شده و ضمن ارزیابی فنوتیپی در قالب یک طرح آزمایشی تکراردار (جدول ۶)، با یک نشانگر RAPD تعیین ژنوتیپ شده باشند (جدول ۷). تجزیه ساختار جمعیت نیز نشان دهد که افراد اول و دوم متعلق به زیرجمعیت اول و افراد سوم، چهارم و پنجم متعلق به زیرجمعیت دوم هستند. بررسی خویشاوندی نیز نشان دهد که افراد ۱، ۲، ۳ و ۴ غیرخویشاوند هستند، اما فرد ۵ از نواده‌های افراد ۳ و ۴ است. برآورد اثر نشانگر با روش مدل خطی عمومی (GLM) و مدل خطی مخلوط (MLM) به‌صورت زیر خواهد بود:

جدول ۶- ارزش‌های فنوتیپی صفت مورد بررسی

Table 6. Phenotypic values of investigated traits

Individual	1	2	3	4	5
Average values	9.3	7.6	5.3	4.4	6.1

جدول ۷- باندهای مشاهده شده در پنج فرد با نشانگر RAPD

Table 7. Observed bands with a RAPD markers

Individual	1	2	3	4	5
Band 1	-	-	-	-	-
Band 2		-	-		-

الف- مدل خطی عمومی (GLM)

مدل خطی عمومی به‌صورت $y = 1\mu + Xm_i + e$ می‌باشد. در این مدل y بردار فنوتیپ‌های مشاهده شده است، 1 بردار عددی با مقادیر یک، μ میانگین کل، m_i اثر

ثابت نشانگر و ساختار جمعیت، e بردار خطای تصادفی مدل و X ماتریس وقوع (Incidence Matrix) برای m_i است. این مدل فرض می‌کند که نشانگر صفت را فقط زمانی تحت تأثیر قرار می‌دهد که نشانگر و QTL در حالت LD باشند. برآورد هم‌زمان میانگین کل ($\hat{\mu}$)، اثر نشانگر ($\hat{m}$) و ساختار جمعیت ($\hat{v}_1$، $\hat{v}_2$ و ...) با حل معادله (۱۱) انجام می‌شود. این ماتریس‌ها در جدول ۸ تعریف شده‌اند:

$$\begin{bmatrix}\hat{\mu}\\ \hat{m}_i\end{bmatrix}=\begin{bmatrix}\{\hat{\mu}\\ \begin{cases}\hat{m}\\ \hat{v}_1\\ \hat{v}_2\end{cases}\end{bmatrix}=\begin{bmatrix}1'1 & 1'X\\ X'1 & X'X\end{bmatrix}^{-1}\begin{bmatrix}1'y\\ X'y\end{bmatrix} \qquad (۱۱)$$

جدول ۸- بردارها و ماتریس‌های رابطه (۱۱) و ماتریس‌های حاصل از حاصل‌ضرب آن‌ها

Table 8. Vectors and matrices of the equation (11) and their multiplicative matrices

$y=\begin{bmatrix}9.3\\7.6\\5.3\\4.4\\6.1\end{bmatrix}_{5\times1}$	$1'1=[5]_{1\times1}$	$X'X=\begin{bmatrix}3&1&2\\1&2&0\\2&0&3\end{bmatrix}_{3\times3}$
$1=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}_{5\times1}$	$1'X=[3\quad 2\quad 3]_{1\times3}$	$1'y=[32.7]_{1\times1}$
$X=\begin{bmatrix}0&1&0\\1&1&0\\1&0&1\\0&0&1\\1&0&1\end{bmatrix}_{5\times3}$	$X'1=\begin{bmatrix}3\\2\\3\end{bmatrix}_{3\times1}$	$X'y=\begin{bmatrix}19.0\\16.9\\15.8\end{bmatrix}_{3\times1}$

تصویر ژل پنج فرد با یک نشانگر RAPD در جدول ۷ ارایه شده است که در آن، ردیف اول مربوط به یک باند یک‌شکل و ردیف دوم مربوط به یک باند چندشکل است. ستون اول ماتریس X مربوط به نشانگر مولکولی است که در آن امتیاز افراد به‌صورت اعداد یک و صفر به‌ترتیب برای وجود یا عدم وجود باند نوشته می‌شود. ستون دوم و سوم ماتریس X نیز مربوط به ساختار جمعیت است که با توجه به فرضیات مساله، اعداد صفر و یک اختصاص یافته‌اند. در صورتی‌که رابطه (۱۱) بر اساس اطلاعات این مثال حل شود، مقادیر زیر به‌دست می‌آید:

$\hat{\mu}$ برآوردی از μ یا میانگین کل است که در این مثال مقدار ۴/۵۶۶۶۶۶۷ برآورد شده است. $\hat{m}$ برآورد اثر ثابت نشانگر مولکولی است و در این مثال مقدار آن ۰/۰۱۴۲۸۵۷ برآورد شده است. $\hat{v}_1$ و $\hat{v}_2$ نیز برآورد آثار زیرجمعیت‌ها هستند که در این مثال که دو زیرجمعیت در جمعیت نقشه‌یابی شناسایی شده است، مقدار ۳/۸۷۶۱۹۰۵ و ۰/۶۹۰۴۷۶۲ به‌ترتیب برای زیرجمعیت اول و دوم برآورد شده است.

$$\begin{bmatrix}\hat{\mu}\\ \hat{m}_i\end{bmatrix}=\begin{bmatrix}\hat{\mu}\\ \hat{m}\\ \hat{v}_1\\ \hat{v}_2\end{bmatrix}=\begin{bmatrix}5&3&2&3\\3&3&1&2\\2&1&2&0\\3&2&0&3\end{bmatrix}^{-1}$$

$$=\begin{bmatrix}4.5666667\\0.0142857\\3.87661905\\0.6904762\end{bmatrix}$$

مقدار آماره F برای آزمون معنی‌داری اثر نشانگر نیز به‌صورت رابطه (۱۲) محاسبه می‌شود:

$$F = \frac{MS_{REgression}}{\hat{\sigma}_e^2} = \frac{\hat{m}x'y + \hat{\mu}\ 1'y - (1/n)(1'y)^2}{\frac{y'y - \hat{m}x'y - \hat{\mu}1'y}{n-2}} \qquad (۱۲)$$

در این رابطه، n تعداد افراد است. عدد محاسبه‌شده آماره F در سطح احتمال مورد نظر با F جدول با درجه آزادی یک و n-۲ به‌ترتیب برای صورت و مخرج کسر F مقایسه می‌شود تا در رابطه با معنی‌دار بودن یا نبودن اثر نشانگر تصمیم‌گیری شود. بدیهی است که اگر مقدار F محاسبه شده بزرگ‌تر یا مساوی F جدول باشد، به مفهوم معنی‌دار بودن اثر نشانگر خواهد بود.

اگر در این مثال، نشانگرهای مولکولی دیگری نیز استفاده شده باشند، ماتریس X با یا بدون در نظر گرفتن ماتریس Q (یا همان ماتریس ساختار جمعیت) به‌صورت جدول ۹ خواهد بود.

جدول ۹- ماتریس X برای نشانگرهای مولکولی دیگر در مطالعات نقشه‌یابی ارتباطی

Table 9. The X matrix for other molecular markers in association mapping studies

Marker	Type	Gel image	Genotyping code	Comment	X matrix
RAPD	Dominant	1 2 3 4 5 - - - - - - - -	0 and 1	Without considering population structure (Q matrix)	$X = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}_{5\times1}$
RAPD	Dominant	1 2 3 4 5 - - - - - - - -	0 and 1	Considering population structure (Q matrix)	$X = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}_{5\times3}$
SNP	Codominant	1 2 3 4 5 AA Aa aa Aa AA	Genotype Code AA 2 Aa 1 aa 0	Without considering population structure (Q matrix)	$X = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}_{5\times1}$
SNP	Codominant	1 2 3 4 5 AA Aa aa Aa AA	Genotype Code AA 2 Aa 1 aa 0	Considering population structure (Q matrix)	$X = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 2 & 0 & 1 \end{bmatrix}_{5\times3}$
SSR	Codominant	1 2 3 4 5 - - - -	Genotype Code AA 1 AB 2 AC 3 BB 4 BC 5 CC 6	Without considering population structure (Q matrix)	$X = \begin{bmatrix} 1 \\ 4 \\ 6 \\ 1 \\ 6 \end{bmatrix}_{5\times1}$
SSR	Codominant	1 2 3 4 5 -[a] - - -	Genotype Code AA 1[b] AB 2 AC 3 BB 4 BC 5 CC 6	Considering population structure (Q matrix)	$X = \begin{bmatrix} 1^c & 1 & 0 \\ 4 & 1 & 0 \\ 6 & 0 & 1 \\ 1 & 0 & 1 \\ 6 & 0 & 1 \end{bmatrix}_{5\times3}$

a) با سه آلل، ژنوتیپ‌های AA، AB، AC، BB، BC و CC می‌توانند تولید شوند.

b) همانند اختصاص کد به ترکیب تیمارها در آزمایش‌های فاکتوریل، به ژنوتیپ افراد نیز کد اختصاص می‌یابد.

c) کدهای مربوط به ژنوتیپ افراد به‌ترتیب ۱، ۴، ۶، ۱ و ۶ خواهند بود.

a) The genotypes AA, AB, AC, BB, BC, and CC can be created using three alleles.

b) The genotypic codes for individuals is assigned similar to code assignment to treatment combinations in the factorial experiments.

c) The genotypic codes for individuals will be 1, 4, 6, 1, and 6, respectively.

برنامه آماری توسعه‌یافته برای انجام تجزیه ارتباطی با روش GLM

برنامه آماری توسعه‌یافته برای برآورد میانگین، اثر نشانگر و ساختار جمعیت بر اساس روش GLM با استفاده از نرم‌افزارهای R و SAS به‌صورت زیر می‌باشد. در برنامه SAS، دستور t() برای محاسبه برگردان یک بردار یا ماتریس، ginv() و inv() به‌ترتیب برای محاسبه عکس تعمیم‌یافته و عکس یک ماتریس، || برای در کنار هم قرار دادن دو ماتریس مستقل در داخل یک ماتریس و // برای در زیر هم قرار دادن دو ماتریس مستقل در داخل یک ماتریس استفاده می‌شود. در خروجی برنامه، ۴/۵۶۶۶۶۶۷ برآورد میانگین، ۰/۰۱۴۲۸۵۷ برآورد اثر نشانگر و ۳/۸۷۶۱۹۰۵ و ۰/۶۹۰۴۷۶۲ به‌ترتیب برآورد آثار ساختار دو زیرجمعیت اول و دوم هستند.

GLM in SAS software:

```
proc iml;
a={1,1,1,1,1};
x={0 1 0,1 1 0,1 0 1, 0 0 1,1 0 1};
y={9.3,7.6,5.3,4.4,6.1};
a1=t(a)*a;
a2=t(a)*x;
a3=t(x)*a;
a4=t(x)*X;
a5=a1||a2;
a6=a3||a4;
a7=a5//a6;
a8=ginv(a7);
a9=t(a)*y;
a10=t(x)*y;
a11=a9//a10;
a12=a8*a11;
print a12;
```

```
a12
4.5666667
0.0142857
3.8761905
0.6904762
```

GLM in R software:

```
a<-matrix(c(1,1,1,1,1),5,1)
x<-matrix(c(0,1,1,0,1,1,1,0,0,0,0,0,1,1,1),5,3)
y<-matrix(c(9.3,7.6,5.3,4.4,6.1),5,1)
a1<-t(a)%*%a
a2<-t(a)%*%x
a3<-t(x)%*%a
a4<-t(x)%*%x
a5<-cbind(a1,a2)
a6<-cbind(a3,a4)
a7<-rbind(a5,a6)
library(MASS)
a8<-ginv(a7)
a9<-t(a)%*%y
a10<-t(x)%*%y
a11<-rbind(a9,a10)
a12<-a8%*%a11

a12
     [,1]
[1,] 4.5666667
[2,] 0.0142857
[3,] 3.8761905
[4,] 0.6904762
```

ب) مدل خطی مخلوط (MLM)

مدل خطی مخلوط به‌صورت روابط (۱۳) و (۱۴) است:

$$Y = \underbrace{X\beta}_{\text{آثار ثابت}} + \underbrace{Zu}_{\text{آثار تصادفی}} + \underbrace{e}_{\text{اشتباه}} \qquad (۱۳)$$

$$Y = \underbrace{X\beta + S\alpha + Qv}_{\text{آثار ثابت}} + \underbrace{Zu + e}_{\text{آثار تصادفی}} \qquad (۱۴)$$

در این رابطه، Y بردار فنوتیپ‌های مشاهده شده، β بردار آثار محیطی (غیر از آثار نشانگر و ساختار جمعیت) است که اینجا معادل جزء اول مدل خطی عمومی (یا همان "μ1" است، α و v به‌ترتیب بردار آثار ثابت نشانگر و ساختار جمعیت، Q ماتریسی است که با نرم‌افزار Structure برآورد می‌شود و Y را به v ارتباط می‌دهد، X، S و Z ماتریس‌های تلاقی یک و صفرها هستند که به‌ترتیب β، α و u را به Y ارتباط می‌دهند، u اثر تصادفی زمینه ژنتیکی و e بردار خطاهای تصادفی است. بردارهای e و u (آثار تصادفی) دارای توزیع نرمال با میانگین صفر و

واریانس $\mathrm{Var}\begin{bmatrix}u\\e\end{bmatrix}=\begin{bmatrix}G & 0\\0 & R\end{bmatrix}$ می‌باشند که در آن:

$$G=\mathrm{Var}(u)=K\sigma_u^2 \qquad (۱۵)$$

$$R=\mathrm{Var}(e)=RV_R \qquad (۱۶)$$

K یک ماتریس n×n ضرایب خویشاندی (Kinship Coefficients) است که درجه کوواریانس ژنتیکی بین زوج افراد را بیان می‌کند، R یک ماتریس n×n با عناصر خارج قطر صفر و عناصر روی قطر برابر با عکس تعداد مشاهدات برای اندازه‌گیری فنوتیپ هر فرد است، σ_u^2 واریانس ژنتیکی افزایشی و V_R واریانس مقادیر باقی‌مانده است. مقادیر σ_u^2 و σ_e^2 با استفاده از الگوریتم EM (Expectation-Maximization) برآورد می‌شوند. بهترین برآورد نااریب خطی (BLUE) بردارهای α، β و v (آثار ثابت) و بهترین پیشگویی نااریب خطی (BLUP) بردار u (آثار تصادفی) با حل معادلات مدل مخلوط زیر (رابطه ۱۷) حاصل می‌شوند:

$$\begin{bmatrix}\begin{Bmatrix}\hat{\mu}\\ \hat{\alpha}\\ \hat{v}_1\\ \hat{v}_2\end{Bmatrix}\\ \begin{Bmatrix}u_1\\ u_2\\ u_3\\ u_4\\ u_5\end{Bmatrix}\end{bmatrix}=\begin{bmatrix}X'X & X'Z\\ Z'X & Z'Z+Q_1\end{bmatrix}^{-1}\begin{bmatrix}X'y\\ Z'y\end{bmatrix} \qquad (۱۷)$$

که در آن، Q_1 برابر است با:

$$Q_1=K^{-1}\left(\frac{\sigma_e^2}{\sigma_u^2}\right) \qquad (۱۸)$$

ماتریس‌های رابطه (۱۷) در جدول ۱۰ ارایه شده است.

جدول ۱۰- بردارها و ماتریس‌های رابطه (۱۷) و ماتریس‌های حاصل از حاصل‌ضرب آن‌ها

Table 10. Vectors and matrices of the equation (17) and their multiplicative matrices

$$y=\begin{bmatrix}9.3\\7.6\\5.3\\4.4\\6.1\end{bmatrix}_{5\times1} \qquad X'X=\begin{bmatrix}5&3&2&3\\3&3&1&2\\2&1&2&0\\3&2&0&3\end{bmatrix}_{4\times4} \qquad Z'Z+K^{-1}=\begin{bmatrix}2&0&0&0&0\\0&2&0&0&0\\0&0&2.5&0.5&-1\\0&0&0.5&2.5&-1\\0&0&-1&-1&3\end{bmatrix}_{5\times5}$$

$$X=\begin{bmatrix}1&0&1&0\\1&1&1&0\\1&1&0&1\\1&0&0&1\\1&1&0&1\end{bmatrix}_{5\times4} \qquad X'Z=\begin{bmatrix}1&1&1&1&1\\0&1&1&0&1\\1&1&0&0&0\\0&0&1&1&1\end{bmatrix}_{4\times5} \qquad X'y=\begin{bmatrix}32.7\\19.0\\16.9\\15.8\end{bmatrix}_{4\times1}$$

$$Z=Z'Z=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}_{5\times5} \qquad Z'X=\begin{bmatrix}1&0&1&0\\1&1&1&0\\1&1&0&1\\1&0&0&1\\1&1&0&1\end{bmatrix}_{5\times4} \qquad Z'y=\begin{bmatrix}9.3\\7.6\\5.3\\4.4\\6.1\end{bmatrix}_{5\times1}$$

ماتریس X مربوط به تمام آثار ثابت است. با توجه به فرضیات مثال، ماتریس خویشاوندی به‌صورت زیر تشکیل می‌شود:

$$K=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0.5\\0&0&0&1&0.5\\0&0&0.5&0.5&1\end{bmatrix}_{5\times5}$$

با فرض $\left(\frac{\sigma_e^2}{\sigma_u^2}\right)=1$ حل مثال به صورت زیر خواهد بود:

$$\begin{bmatrix}\begin{cases}\hat{\beta}\\ \hat{\alpha}\\ \hat{v}_1\\ \hat{v}_2\end{cases}\\ \begin{cases}u_1\\ u_2\\ u_3\\ u_4\\ u_5\end{cases}\end{bmatrix}=\begin{bmatrix}5&3&2&3&1&1&1&1&1\\3&3&1&2&0&1&1&0&1\\2&1&2&0&1&1&0&0&0\\3&2&0&3&0&0&1&1&1\\1&0&1&0&2&0&0&0&0\\1&1&1&0&0&2&0&0&0\\1&1&0&1&0&0&2.5&0.5&-1\\1&0&0&1&0&0&0.5&2.5&-1\\1&1&0&1&0&0&-1&-1&3\end{bmatrix}^{-1}\begin{bmatrix}32.7\\19\\16.9\\15.8\\9.3\\7.6\\5.3\\4.4\\6.1\end{bmatrix}=\begin{bmatrix}4.47\\0.18\\3.89\\0.58\\0.47\\-0.47\\0.18\\-0.18\\0.29\end{bmatrix}$$

در سمت چپ مساوی، β برداری شامل میانگین و آثار محیطی (غیر از آثار ثابت نشانگر و ساختار جمعیت) است که در این مثال برای میانگین ($\hat{\beta}$) مقدار ۴/۴۷ برآورد شده است. در صورت گنجاندن سایر آثار محیطی می‌توان مقدار آن‌ها را نیز برآورد کرد. α بردار آثار ثابت نشانگر است که در این مثال مقدار آن ($\hat{\alpha}$) برابر با ۰/۱۸ برآورد شده است. $\hat{v}_1$ و $\hat{v}_2$ برآورد آثار زیرجمعیت‌ها هستند که در این مثال برای دو زیرجمعیت شناسایی شده در جمعیت نقشه‌یابی به‌ترتیب ۳/۸۹ و ۰/۵۸ برآورد شده‌اند. $\hat{u}_1$ تا $\hat{u}_5$ نیز آثار تصادفی زمینه ژنتیکی هستند که در این مثال به‌ترتیب ۰/۴۷، ۰/۴۷-، ۰/۱۸، ۰/۱۸- و ۰/۲۹ برآورد شده‌اند.

برنامه آماری برای انجام تجزیه ارتباطی با روش MLM

برنامه آماری برای برآورد میانگین، اثر نشانگر، ساختار جمعیت و روابط خویشاندی بر اساس روش MLM با استفاده از نرم‌افزارهای R و SAS به‌صورت زیر است:

MLM in R software:

```
x<-matrix(c(1,1,1,1,1,0,1,1,0,1,1,1,0,0,0,0,0,1,1,1),5,4)
y<-matrix(c(9.3,7.6,5.3,4.4,6.1),5,1)
z<-matrix(c(1,0,0,0,0,0,1,0,0,0,0,0,1,0,0, 0,0,0,1,0,0,0,0,0,1),5,5)
k<-matrix(c(1,0,0,0,0,0,1,0,0,0,0,0,1,0,0.5,0,0,0,1,0.5,0,0,0.5,0.5,1),5,5)
a1<-t(x)%*%x
a2<-t(x)%*%z
a3<-t(z)%*%x
a4<-t(z)%*%z
a5<-a4+solve(k)
a6<-rbind(a1,a3)
a7<-rbind(a2,a5)
a8<-cbind(a6,a7)
a9<-t(x)%*%y
a10<-t(z)%*%y
a11<-rbind(a9,a10)
a12<-ginv(a8)%*%a11

a12
     [,1]
[1,]  4.47
[2,]  0.18
[3,]  3.89
[4,]  0.58
[5,]  0.47
[6,] -0. 47
[7,]  0.18
[8,] -0.18
[9,]  0.29
```

MLM in SAS software:

```
proc iml;
x={1 0 1 0,1 1 1 0,1 1 0 1,1 0 0 1,1 1 0 1};
y={9.3,7.6,5.3,4.4,6.1};
z={1 0 0 0 0, 0 1 0 0 0, 0 0 1 0 0, 0 0 0 1 0, 0 0 0 0 1};
k={1 0 0 0 0, 0 1 0 0 0, 0 0 1 0 0.5, 0 0 0 1 0.5, 0 0 0.5 0.5 1};
a1=t(x)*x;
a2=t(x)*z;
a3=t(z)*x;
a4=t(z)*Z;
a5=a4+inv(k);
a6=a1//a3;
a7=a2//a5;
a8=a6||a7;
a9=t(x)*y;
a10=t(z)*y;
a11=a9//a10;
a12=ginv(a8)*a11;
print A8, A11, a12;
```

```
a12
4.47
0.18
3.89
0.58
0.47
-0.47
0.18
-0.18
0.29
```

References

Abdurakhmonov, I. Y. and Abdukarimov, A. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. **International Journal of Plant Genomics** 2008 (2): 574927.

Abdurakhmonov, I. Y., Kohel, R. J., Yu, J. Z., Pepper, A. E., Abdullaev, A. A., Kushanov, F. N., Salakhutdinov, I. B., Buriev, Z. T., Saha, S., Scheffler, B. E. and Jenkins, J. N. 2008. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. **Genomics** 92 (6): 478-487.

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C. and Tamiru, M. 2012. Genome sequencing reveals agronomically important loci in rice using MutMap. **Nature Biotechnology** 30 (2): 174.

Abedi, S., Darvishzadeh, R., Bernousi, I., Abdollahi Mandoulakani, B., Hatami Maleki, H. and Shah D. 2014. Genetic variability of *Orobanche aegyptiaca* infesting tobacco in Iran by Bayesian analysis. **Biologia** 69 (12): 1652-1659.

Ahmadpour, S., Sofalian, O., Darvishzadeh, R. and Abbaspour, N. 2018. Preliminary evidence of the associations between DNA markers and morphological characters in sunflower under natural and salt stress conditions. **Zemdirbyste-Agriculture** 105 (3): 279-286.

Bamshad, M., Wooding, S., Salisbury, B. A. and Stephens, J. C. 2004. Deconstructing the relationship between genetics and race. **Nature Reviews Genetics** 5 (8): 598.

Basirnia, A., Hatami Maleki, H., Darvishzadeh, R. and Ghavami F. 2014. Mixed linear model association mapping for low chloride accumulation rate in oriental-type tobacco (*Nicotiana tabacum* L.) germplasm. **Journal of Plant Interactions** 9 (1): 666-672.

Begum, H., Spindel, J. E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., Virk, P., Collard, B. and McCouch, S. R. 2015. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa* L.). **PloS One** 10 (3): e0119873.

Bernardo, R. and Yu, J. 2007. Prospects for genomewide selection for quantitative traits in maize. **Crop Science** 47 (3): 1082-1090.

Beyer, S., Daba, S., Tyagi, P., Bockelman, H., Brown-Guedira, G. and Mohammadi, M. 2019. Loci and candidate genes controlling root traits in wheat seedlings-a wheat root GWAS. **Functional and Integrative Genomics** 19 (1): 91-107.

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.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'Amore, R., Allen, A. M., McKenzie, N., Kramer, M., Kerhornou, A. and Bolser, D. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. **Nature** 491 (7426): 705.

Camus-Kulandaivelu, L., Veyrieras, J. B., Madur, D., Combes, V., Fourmann, M., Barraud, S., Dubreuil, P., Gouesnard, B., Manicacci, D. and Charcosset, A. 2006. Maize adaptation to temperate climate: relationship between population structure and polymorphism in the *Dwarf8* gene. **Genetics** 172 (4): 2449-2463.

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., Li, Y., Liu, X., Zhang, H. and Dong, H. 2014. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. **Nature Genetics** 46 (7): 714.

Cui, Y., Zhang, F. and Zhou, Y. 2018. The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. **Frontiers in Plant Science** 9: 1464. doi:10.3389/fpls.2018.01464.

Dadras, A. R., Sabouri, H., Nejad, G. M., Sabouri, A. and Shoai-Deylami, M. 2014. Association analysis, genetic diversity and structure analysis of tobacco based on AFLP markers. **Molecular Biology Reports** 41 (5): 3317-3329.

Darvishzadeh, R. 2016a. Population structure, linkage disequilibrium and association mapping for morphological traits in sunflower (*Helianthus annuus* L.), **Biotechnology and Biotechnological Equipment** 30 (2): 236-246.

Darvishzadeh, R. 2016b. Detection of ISSR markers linked to seed oil biochemical characteristics in castor (*Ricinus communis* L.) through association analysis. **Genetika** 48 (3): 807-817.

Darvishzadeh, R. 2016c. Genetic variability, structure analysis, and association mapping of resistance to broomrape (*Orobanche aegyptiaca*) in tobacco. **Journal of Agricultural Science and Technology** 18: 1407-1418.

Darvishzadeh, R., Basirnia A., Hatami Maleki H. and Jafari M. 2014. Association mapping for resistance to powdery mildew in oriental tobacco (*Nicotiana tabacum* L.) germplasm. **Iranian Journal of Genetic and Breeding** 3 (1): 21-30.

Darvishzadeh, R., Heidari, A. and Hatami Maleki, H. 2016. Identification of SSR markers associated with resistance to potato virus Y in tobacco germplasm. **Genetics in the 3rd Millennium** 14 (2): 4262-4269.

Darvishzadeh, R., Poormohammad Kiani, S., Huguet, T. and Sarrafi, A. 2008. Genetic variation and identification of molecular markers associated with partial resistance to black stem in gamma-irradiation induced mutants in sunflower (*Helianthus annuus* L.). **Canadian Journal of Plant Pathology** 30: 106-114.

Davar, R., Darvishzadeh, R., Rezaee Danesh, Y., Kholghi, M., Azizi, M and Shah, D. A. 2012. Single sequence repeat markers associated with partial resistance in sunflower to *Phoma macdonaldii*. **Phytopathologia Mediterrania** 51 (3): 541-548.

Devlin, B., Bacanu, S. A. and Roeder, K. 2004. Genomic control to the extreme. **Nature Genetics** 36 (11): 1129.

Devlin, B. and Risch, N. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. **Genomics** 29 (2): 311-322.

Devlin, B. and Roeder, K. 1999. Genomic control for association studies. **Biometrics** 55 (4): 997-1004.

Devlin, B., Roeder, K. and Wasserman, L. 2001. Genomic control, a new approach to genetic-based association studies. **Theoretical Population Biology** 60 (3): 155-166.

Dingkuhn, M., Pasco, R., Pasuquin, J. M., Damo, J., Soulié, J. C., Raboin, L. M., Dusserre, J., Sow, A., Manneh, B. and Shrestha, S. 2017. Crop-model assisted phenomics and genome-wide association study for climate adaptation of indica rice. 1. Phenology. **Journal of Experimental Botany** 68 (15): 4369-4388.

El Mannai, Y., Shehzad, T. and Okuno, K. 2011. Variation in flowering time in sorghum core collection and mapping of QTLs controlling flowering time by association analysis. **Genetic Resources and Crop Evolution** 58 (7): 983.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. and Mitchell, S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS One** 6 (5): e19379. doi:10.1371/journal.pone.0019379.

Fusari, C. M., Di Rienzo, J. A., Troglia, C., Nishinakamasu, V., Moreno, M. V., Maringolo, C., Quiroz, F., Álvarez, D., Escande, A. and Hopp, E. 2012. Association mapping in sunflower for sclerotinia head rot resistance. **BMC Plant Biology** 12 (1): 93.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. and Liu-Cordero, S. N. 2002. The structure of haplotype blocks in the human genome. **Science** 296 (5576): 2225-2229.

Ghaffari Azar A., Darvishzadeh, R., Hatami Maleki, H., Kahrizi, D., Darvishi, B. and Bernoosi, I. 2018. Identification of inter simple sequence repeat regions associated with agro-morphological traits in maize genome. **Cereal Research** 8 (1): 97-109. (In Persian with English Abstract).

Goodarzi, F., Hassani, A., Darvishzadeh, R. and Hatami Maleki, H. 2015. Genetic variability and traits association in castor bean (*Ricinus communis* L.). **Genetika** 47 (1): 265-274.

Gupta, P. K., Kulwal, P. L. and Jaiswal, V. 2019. Association mapping in plants in the post-GWAS genomics era. **Advances in Genetics** 104: 75-154.

Gupta, P. K., Rustgi, S. and Kulwal, P. L. 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Molecular Biology** 57 (4): 461-485.

Hao, D., Xue, L., Zhang, Z., Cheng, Y., Chen, G., Zhou, G., Li, P., Yang, Z. and Xu, C. 2019. Combined linkage and association mapping reveal candidate loci for kernel size and weight in maize. **Breeding Science**: 18185.

Hayes, B. and Goddard, M. 2001. Prediction of total genetic value using genome-wide dense marker maps. **Genetics** 157 (4): 1819-1829.

Hedrick, P. W. 1987. Gametic disequilibrium measures: Proceed with caution. **Genetics** 117 (2): 331-341.

Hill, W. and Robertson, A. 1968. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics** 38 (6): 226-231.

Hill, W. and Weir, B. 1994. Maximum-likelihood estimation of gene location by linkage disequilibrium. **American Journal of Human Genetics** 54 (4): 705.

Hindu, V., Palacios-Rojas, N., Babu, R., Suwarno, W. B., Rashid, Z., Usha, R., Saykhedkar, G. R. and Nair, S. K. 2018. Identification and validation of genomic regions influencing kernel zinc and iron in maize. **Theoretical and Applied Genetics** 131 (7): 1443-1457.

Holland, J. B. 2007. Genetic architecture of complex traits in plants. **Current Opinion in Plant Biology** 10 (2): 156-161.

Hu, X., Zuo, J., Wang, J., Liu, L., Sun, G., Li, C., Ren, X. and Sun, D. 2018. Multi-locus genome-wide association studies for 14 main agronomic traits in barley. **Frontiers in Plant Science** 9: 1683.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q. and Huang, T. 2009. High-throughput genotyping by whole-genome resequencing. **Genome Research** 19 (6): 1068-1076.

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B. and Xia, J. 2015. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. **Nature Communications** 6: 6258.

Jabbari, M., Fakheri, B. A., Aghnoum, R., Mahdi Nezhad, N. and Ataei, R. 2018. GWAS analysis in spring barley (*Hordeum vulgare* L.) for morphological traits exposed to drought. **PLoS One** 13 (9): e0204952. doi:10.1371/journal.pone.0204952.

Jain, S. M., Brar, D. S. 2010. Molecular techniques in crop improvement. 2nd Edition. Springer.

Jia, Z., Liu, Y., Gruber, B. D., Neumann, K., Kilian, B., Graner, A. and von Wiren, N. 2019. Genetic dissection of root system architectural traits in spring barley. **Frontiers in Plant Science** 10: 400. doi:10.3389/fpls.2019.00400.

Kaplan, N. and Weir, B. 1992. Expected behavior of conditional linkage disequilibrium. **American Journal of Human Genetics** 51 (2): 333.

Kordrostami, M., Rabiei, B. and Hassani Kumleh, H. 2016. Association analysis, genetic diversity and haplotyping of rice plants under salt stress using SSR markers linked to SalTol and morpho-physiological characteristics. **Plant Systematics and Evolution** 302 (7): 871-890.

Kumar, J., Saripalli, G., Gahlaut, V., Goel, N., Meher, P. K., Mishra, K. K., Mishra, P. C., Sehgal, D., Vikram, P. and Sansaloni, C. 2018. Genetics of Fe, Zn, β-carotene, GPC and yield traits in bread wheat (*Triticum aestivum* L.) using multi-locus and multi-traits GWAS. **Euphytica** 214 (11): 219.

Kumar, V., Singh, A., Mithra, S. A., Krishnamurthy, S., Parida, S. K., Jain, S., Tiwari, K. K., Kumar, P., Rao, A. R. and Sharma, S. 2015. Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa* L.). **DNA Research** 22 (2): 133-145.

Lekklar, C., Pongpanich, M., Suriya-Arunroj, D., Chinpongpanich, A., Tsai, H., Comai, L., Chadchawan, S. and Buaboocha, T. 2019. Genome-wide association study for salinity tolerance at the flowering stage in a panel of rice accessions from Thailand. **BMC Genomics** 20 (1): 76.

Levin, M. L. and Bertell, S. R. 1978. RE:simple estimation of population attributable risk from case-control studies. **American Journal of Epidemiology** 108 (1): 78-79.

Lewontin, R. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. **Genetics** 49 (1): 49.

Li, G., Xu, X., Tan, C., Carver, B. F., Bai, G., Wang, X., Bonman, J. M., Wu, Y., Hunger, R. and Cowger, C. 2019. Identification of powdery mildew resistance loci in wheat by integrating genome-wide association study (GWAS) and linkage mapping. **The Crop Journal** 7 (3): 294-306.

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T. and Yang, N. 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. **Nature Genetics** 45 (1): 43.

Li, J., Huang, X., Heinrichs, F., Ganal, M. and Röder, M. 2005. Analysis of QTLs for yield, yield components, and malting quality in a BC 3-DH population of spring barley. **Theoretical and Applied Genetics** 110 (2): 356-363.

Li, T., Qu, J., Wang, Y., Chang, L., He, K., Guo, D., Zhang, X., Xu, S. and Xue, J. 2018. Genetic characterization of inbred lines from Shaan A and B groups for identifying loci associated with maize grain yield. **BMC Genetics** 19 (1): 63. doi:10.1186/s12863-018-0669-9.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S. and Zhang, Z. 2012. GAPIT: genome association and prediction integrated tool. **Bioinformatics** 28 (18): 2397-2399.

Liu, E., Liu, X., Zeng, S., Zhao, K., Zhu, C., Liu, Y., Breria, M. C., Zhang, B. and Hong, D. 2015. Time-course association mapping of the grain-filling rate in rice (*Oryza sativa* L.). **PLoS One** 10 (3): e0119959.

Liu, E., Liu, Y., Wu, G., Zeng, S., Thi, T., Thu, G., Liang, L., Liang, Y., Dong, Z. and She, D. 2016. Identification of a candidate gene for panicle length in rice (*Oryza sativa* L.) via association and linkage analysis. **Frontiers in Plant Science** 7: 596.

Liu, K. and Muse, S. V. 2005. PowerMarker: integrated analysis environment for genetic marker data. **Bioinformatics** 21 (9): 2128-2129.

Long, A. D. and Langley, C. H. 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. **Genome Research** 9 (8): 720-731.

Lu, Q., Niu, X., Zhang, M., Wang, C., Xu, Q., Feng, Y., Yang, Y., Wang, S., Yuan, X. and Yu, H. 2018. Genome-wide association study of seed dormancy and the genomic consequences of improvement footprints in rice (*Oryza sativa* L.). **Frontiers in Plant Science** 8: 2213.

Lu, Y., Liu, Y., Niu, X., Yang, Q., Hu, X., Zhang, H. Y. and Xia, J. 2015. Systems genetic validation of the SNP-metabolite association in rice via metabolite-pathway-based phenome-wide association scans. **Frontiers in Plant Science** 6: 1027.

Luo, X., Wang, B., Gao, S., Zhang, F., Terzaghi, W. and Dai, M. 2019. Genome-wide association study dissects the genetic bases of salt tolerance in maize seedlings. **Journal of Integrative Plant Biology** 61 (6): 658-674.

Ma, F., Xu, Y., Ma, Z., Li, L. and An, D. 2018. Genome-wide association and validation of key loci for yield-related traits in wheat founder parent Xiaoyan 6. **Molecular Breeding** 38 (7): 91.

Ma, X., Feng, F., Wei, H., Mei, H., Xu, K., Chen, S., Li, T., Liang, X., Liu, H. and Luo, L. 2016. Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. **Frontiers in Plant Science** 7: 1801.

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J. I., Ebana, K., Yano, M. and Saito, K. 2015. Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. **The Plant Journal** 81 (1): 13-23.

Mayer, K. F., Waugh, R., Brown, J. W., Schulman, A., Langridge, P., Platzer, M., Fincher, G. B., Muehlbauer, G. J., Sato, K., Close, T. J., Wise, R. P. and Stein, N. 2012. A physical, genetic and functional sequence assembly of the barley genome. **Nature** 491 (7426): 711-716.

Melchinger, A. E. and Gumber, R. K. 1998. Overview of heterosis and heterotic groups in agronomic crops. In: Larnkey, K. R. and Staub, J. E. (Eds.). Concepts and breeding of heterosis in crop plants. CSSA Special Publications. pp: 29-44.

Moose, S. P. and Mumm, R. H. 2008. Molecular plant breeding as the foundation for 21st century crop improvement. **Plant Physiology** 147 (3): 969-977.

Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E. and Buckler, E. S. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. **The Plant Cell** 21 (8): 2194-2202.

Najafzadeh, R., Darvishzadeh, R., Musa-Khalifani, Kh., Abrinbana, M. and Alipour, H. 2018. Retrotransposonable regions of sunflower genome having relevance with resistance to *Sclerotinia* species: *S. sclerotiorum* and *S. minor*. **Australasian Plant Pathology** 47: 511-519.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A. and Eichler, E. E. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. **Nature** 461 (7261): 272.

Novakazi, F., Afanasenko, O., Anisimova, A., Platz, G. J., Snowdon, R., Kovaleva, O., Zubkovich, A. and Ordon, F. 2019. Genetic analysis of a worldwide barley collection for resistance to net form of net blotch disease (*Pyrenophora teres* f. teres). **Theoretical and Applied Genetics**: 132 (9): 2633-2650.

Oraguzie, N. C., Wilcox, P. L., Rikkerink, E. H. and de Silva, H. N. 2007. Linkage disequilibrium. In: Oraguzie, N. C., Rikkerink, E. H. A., Gardiner, S. E. and De Silva, H. N. (Eds.). Association mapping in plants. Springer, New York. pp: 11-39.

Oyiga, B. C., Sharma, R. C., Baum, M., Ogbonnaya, F. C., Leon, J. and Ballvora, A. 2018. Allelic variations and differential expressions detected at quantitative trait loci for salt stress tolerance in wheat. **Plant, Cell and Environment** 41 (5): 919-935. doi:10.1111/pce.12898.

Pham, A. T., Maurer, A., Pillen, K., Brien, C., Dowling, K., Berger, B., Eglinton, J. K. and March, T. J. 2019. Genome-wide association of barley plant growth under drought stress using a nested association mapping population. **BMC Plant Biology** 19 (1): 134. doi:10.1186/s12870-019-1723-0.

Pritchard, J. K. 2001. Deconstructing maize population structure. **Nature Genetics** 28 (3): 203.

Pritchard, J. K., Stephens, M. and Donnelly, P. 2000a. Inference of population structure using multilocus genotype data. **Genetics** 155 (2): 945-959.

Pritchard, J. K., Stephens, M., Rosenberg, N. A. and Donnelly, P. 2000b. Association mapping in structured populations. **The American Journal of Human Genetics** 67 (1): 170-181.

Razi, M., Darvishzadeh R., Amiri, M. E., Doulati-Baneh, H. and Martínez-Gómez P. 2018. Molecular characterization of a diverse Iranian table grapevine germplasm using REMAP markers: Population structure, linkage disequilibrium and association mapping of berry yield and quality traits. **Biologia** https://doi.org/10.2478/s11756-018-0158-7.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F. and Ward, R. 2001. Linkage disequilibrium in the human genome. **Nature** 411 (6834): 199.

Rogers, S. O. and Bendich, A. J. 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. **Plant Molecular Biology** 9 (5): 509-520.

Roy, J. K., Smith, K. P., Muehlbauer, G. J., Chao, S., Close, T. J. and Steffenson, B. J. 2010. Association mapping of spot blotch resistance in wild barley. **Molecular Breeding** 26 (2): 243-256.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J. and Ackerman, H. C. 2002. Detecting recent positive selection in the human genome from haplotype structure. **Nature** 419 (6909): 832.

Saeed A. and Darvishzadeh R. 2017. Association analysis of biotic and abiotic stresses resistance in chickpea (*Cicer* spp.) using AFLP markers. **Biotechnology and Biotechnological Equipment** https://doi.org/10.1080/13102818.2017.1333455.

Saeed, A., Darvishzadeh, R. and Basirnia, A. 2013. Simple sequence repeat markers associated withagro-morphological traits in chickpea (*Cicer arietinum* L.). **Zemdirbyste-Agriculture** 100 (4): 433-440.

Sameri, M., Takeda, K. and Komatsuda, T. 2006. Quantitative trait loci controlling agronomic traits in recombinant inbred lines from a cross of oriental-and occidental-type barley cultivars. **Breeding Science** 56 (3): 243-252.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., Jørgensen, J. E., Weigel, D. and Andersen, S. U. 2009. SHOREmap: Simultaneous mapping and mutation identification by deep sequencing. **Nature Methods** 6 (8): 550.

Schulze, T. G. and McMahon, F. J. 2002. Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines. **American Journal of Medical Genetics** 114 (1):1-11.

Shehzad, T., Iwata, H. and Okuno, K. 2009. Genome-wide association mapping of quantitative traits in sorghum (*Sorghum bicolor* L. Moench) by using multiple models. **Breeding Science** 59 (3): 217-227.

Sheoran, S., Jaiswal, S., Kumar, D., Raghav, N., Sharma, R., Pawar, S., Paul, S., Iquebal, M. A., Jaiswar, A., Sharma, P., Singh, R., Singh, C. P., Gupta, A., Kumar, N., Angadi, U. B., Rai, A., Singh, G. P., Kumar, D. and Tiwari, R. 2019. Uncovering genomic regions associated with 36 agro-morphological traits in indian spring wheat using GWAS. **Frontiers in Plant Science** 10: 527.

Singh, D., Ziems, L.A., Dracatos, P. M., Pourkheirandish, M., Tshewang, S., Czembor, P., German, S., Fowler, R. A., Snyman, L., Platz, G. J. and Park, R. F. 2018. Genome-wide association studies provide insights on genetic architecture of resistance to leaf rust in a worldwide barley collection. **Molecular Breeding** 38(4). doi:10.1007/s11032-018-0803-4.

Soleimani Gezeljeh, A., Darvishzadeh, R., Ebrahimi, A. and Bihamta, M. R. 2018. Identification of SSR and retrotransposon-based molecular markers linked to morphological characters in oily sunflower (*Helianthus annuus* L.) under natural and water-limited states. **Journal of Genetics** https://doi.org/10.1007/s12041-018-0901-4.

Springer, N. M., Ying, K., Fu, Y., Ji, T., Yeh, C.T., Jia, Y., Wu, W., Richmond, T., Kitzman, J. and Rosenbaum, H. 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. **PLoS Genetics** 5 (11): e1000734.

Stich, B., Maurer, H. P., Melchinger, A. E., Frisch, M., Heckenberger, M., van der Voort, J. R., Peleman, J., Sørensen, A. P. and Reif, J. C. 2006. Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. **Molecular Breeding** 17 (3): 217-226.

Stich, B., Melchinger, A. E., Frisch, M., Maurer, H. P., Heckenberger, M. and Reif, J. C. 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. **Theoretical and Applied Genetics** 111 (4): 723-730.

Stich, B., Melchinger, A. E., Piepho, H. P., Hamrit, S., Schipprack, W., Maurer, H. P. and Reif, J. C. 2007. Potential causes of linkage disequilibrium in a European maize breeding program investigated with computer simulations. **Theoretical and Applied Genetics** 115 (4): 529-536.

Tabkhkar, N., Rabiei, B., Samizadeh Lahiji, H. and Hosseini Chaleshtori, M. 2020. Identification of a new set of drought-related miRNA-SSR markers and association analysis under drought stress in rice (*Oryza sativa* L.). **Plant Gene** 100220.

Tessmann, E. and Van Sanford, D. 2018. GWAS for fusarium head blight related traits in winter wheat (*Triticum aestivum* L.) in an artificially warmed treatment. **Agronomy** 8 (5): 68.

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E. S. 2001. Dwarf8 polymorphisms associate with variation in flowering time. **Nature Genetics** 28 (3): 286.

Ueda, Y., Frimpong, F., Qi, Y., Matthus, E., Wu, L., Höller, S., Kraska, T. and Frei, M. 2014. Genetic dissection of ozone tolerance in rice (*Oryza sativa* L.) by a genome-wide association study. **Journal of Experimental Botany** 66 (1): 293-306.

Varshney, R., Paulo, M., Grando, S., Van Eeuwijk, F., Keizer, L., Guo, P., Ceccarelli, S., Kilian, A., Baum, M. and Graner, A. 2012. Genome wide association analyses for drought tolerance related traits in barley (*Hordeum vulgare* L.). **Field Crops Research** 126: 171-180.

Wang, Y. and Rannala, B. 2005. In silico analysis of disease-association mapping strategies using the coalescent process and incorporating ascertainment and selection. **The American Journal of Human Genetics** 76 (6): 1066-1073.

Wang, Z., Gerstein, M. and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics** 10 (1): 57.

Wright, S. I. and Gaut, B. S. 2004. Molecular population genetics and the search for adaptive evolution in plants. **Molecular Biology and Evolution** 22 (3): 506-519.

Wu, L., Zhang, Y., He, Y., Jiang, P., Zhang, X. and Ma, H. 2019. Genome-wide association mapping of resistance to fusarium head blight spread and deoxynivalenol accumulation in chinese elite wheat germplasm. **Phytopathology** 109 (7): 1208-1216.

Xu, X., Sharma, R., Tondelli, A., Russell, J., Comadran, J., Schnaithmann, F., Pillen, K., Kilian, B., Cattivelli, L., Thomas, W. T. B. and Flavell, A. J. 2018. Genome-wide association analysis of grain yield-associated traits in a pan-european barley cultivar collection. **Plant Genome** 11 (1). doi:10.3835/plantgenome2017.

Xu, Y., Li, P., Yang, Z. and Xu, C. 2017. Genetic mapping of quantitative trait loci in crops. **The Crop Journal** 5 (2):175-184.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H. and Hirano, K. 2016. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. **Nature Genetics** 48 (8): 927.

Yu, J. and Buckler, E. S. 2006. Genetic association mapping and genome organization of maize. **Current Opinion in Biotechnology** 17 (2): 155-160.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M. and Holland, J. B. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics** 38 (2): 203.

Zarbafi, S. S., Rabiei, B., Ebadi, A. A. and Ham, J. H. 2020. Association mapping of traits related to leaf blast disease in rice (*Oryza sativa* L.). **Australasian Plant Pathology** 49 (1): 31-43.

Zhang, M., Lu, Q., Wu, W., Niu, X., Wang, C., Feng, Y., Xu, Q., Wang, S., Yuan, X. and Yu, H. 2017. Association mapping reveals novel genetic loci contributing to flooding tolerance during germination in indica rice. **Frontiers in Plant Science** 8: 678.

Zhang, Y., Liu, P., Zhang, X., Zheng, Q., Chen, M., Ge, F., Li, Z., Sun, W., Guan, Z. and Liang, T. 2018. Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. **Frontiers in Plant Science** 9: 611.

Zhang, Z., Buckler, E. S., Casstevens, T. M. and Bradbury, P. J. 2009. Software engineering the mixed model for genome-wide association studies on large samples. **Briefings in Bioinformatics** 10 (6): 664-675.

Zhao, Y., Li, Z., Liu, G., Jiang, Y., Maurer, H.P., Würschum, T., Mock, H.-P., Matros, A., Ebmeyer, E. and Schachschneider, R. 2015. Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. **Proceedings of the National Academy of Sciences** 112 (51): 15624-15629.

Zhu, C., Gore, M., Buckler, E. S. and Yu, J. 2008. Status and prospects of association mapping in plants. **The Plant Genome** 1 (1): 5-20.

Zhu, X., Shao, X., Pei, Y., Guo, X., Li, J., Song, X. and Zhao, M. 2018. Genetic diversity and genome-wide association study of major ear quantitative traits using high-density SNPs in maize. **Frontiers in Plant Science** 9: 966.

doi: 10.22124/cr.2019.14333.1518

University of Guilan
Faculty of Agricultural
Sciences

(Review Paper)
Cereal Research
Vol. 9, No. 3, Autumn 2019 (271-298)

Association mapping of quantitative traits in molecular cereal breeding

Hadi Alipour[1] and Reza Darvishzadeh[2*]

Received: August 6, 2019 Accepted: October 9, 2019

Abstract

The field of association mapping studies has recently received major attention for genetic studies of quantitative traits in many important plants. Access to next generation sequencing technologies, high phenotypic data and a variety of sophisticated statistical tools have enabled association mapping studies in plants to be successful in identifying gene loci controlling quantitative traits. Due to the importance of association mapping method in mapping studies of the quantitative traits, the present paper was prepared to explain the association mapping method and its use in plant breeding especially cereals. This paper, also provides some information about statistical softaware packages used in association mapping and then the opportunities and challenges of association mapping and post-genome wide association studies at the whole genome level will be discussed. Finally, linkage disequilibrium value and association mapping analysis will be evaluated based on general linear model (GLM) and mixed linear method (MLM) using a simple example.

Keywords: Genome-wide association mapping, Kinship, Mixed linear model, Population structure

1. Assist. Prof., Dept. of Plant Production and Genetics, Faculty of Agriculture and Natural Resources, Urmia University, Urmia, Iran
2. Prof., Dept. of Plant Production and Genetics, Faculty of Agriculture and Natural Resources, Urmia University, Urmia, Iran
[*] Corresponding author: r.darvishzadeh@urmia.ac.ir