Razi University

Cereal Biotechnology and Biochemistry

Online ISSN: 2783-5170

Homepage: https://cbb.razi.ac.ir

Application of association mapping and linkage disequilibrium in plant genomic studies with emphasis on cereal genome

Babak Abdollahi Mandoulakani[1], Hossein Abbasi Holasou[2] ✉ & Salar Shaaf[3]

[1] Professor, Department of Plant Production and Genetics, Faculty of Agriculture, Urmia University, Urmia, Iran.

[2] Post-doctoral student, Department of Plant Breeding and Biotechnology, Faculty of Agricultural Sciences, Tabriz University, Tabriz, Iran.

[3] Department of Agricultural and Environmental Sciences (DISAA), University of Milan, Milan, Italy.

✉ Corresponding author. E-mail: hossein.pdf55@gmail.com

ABSTRACT

Knowledge of genetic diversity and understanding the genetic architecture of complex traits in crop species can play a key role in exploiting genetic diversity and lead to crop cultivation's adaptation, development and expansion in different geographical environments. In this regard, identifying and utilizing informative markers related to quantitative traits is a prerequisite for application in breeding programs and marker-assisted selection (MAS). One of the major uses of these markers is the construction of genome-wide molecular maps and genetic analysis. In genome-wide association study (GWAS) and genomic selection (GS), determining of extent and level of linkage disequilibrium (LD) is critical in sample size and marker density. Moreover, selection for increasing frequency in new mutations advantageous only in a subset of populations leaves some signatures in the genome. Locations of selection signatures are often correlated with genes and QTLs affecting economically important traits. Access to next-generation sequencing technologies, high phenotypic data and a variety of sophisticated statistical tools have enabled LD-based association mapping studies in plants to identify gene loci controlling quantitative traits successfully. Alternatively, the LD-based association mapping (AM) approach can enhance the genetic map resolution to a greater extent due to the representation of a more comprehensive gene pool and more recombination events in history. Association mapping has emerged as a tool to resolve complex trait variation down to the sequence level by exploiting evolutionary recombination events at the population level. Genetic diversity, LD extent in genome and intra-population relationship, determine quality of mapping, marker diversity, statistical methods and power of mapping. LD in population is the foundation of GWAS, whereas it is always affected by genetic drift, population stratification and natural selection. Of the above threats, population stratification is recognized as a major one to the validity of GWAS results. Therefore, knowledge concerning the pattern of LD is essential for performing GWAS and GS. As one of the most important measures in population genetics, LD is the basis of effective population size estimation, genomic selection, GWAS study and QTL mapping. Due to the importance of LD method in mapping studies of the quantitative traits, in this review, we will present LD, its application in population, current status, limitations, and its use in plant breeding, especially in cereals. Finally, this study provides some information about commonly used statistical software and packages in the calculation of LD, and the challenges and perspectives of this approach have been discussed in this study as well.

Keywords: Marker-Trait Association, Population Structure, Cereal, Mixed Linear Model, Association mapping.

Article Type: ReviewArticle

Article history: Received: 27 Mar 2023, **Revised:** 20 Apr 2023, **Accepted:** 16 Jun 2023, **Published online:** 22 Jun 2023

Cite this article: Abdollahi Mandoulakani, B., Abbasi Holasou, H., & Shaaf, S. (2023). Application of association mapping and linkage disequilibrium in plant genomic studies with emphasis on cereal genome. *Cereal Biotechnology and Biochemistry*, *2*(2), 243-268. DOI: 10.22126/cbb.2023.8942.1051

© The Author(s).

Publisher: Razi

10.22126/cbb.2023.8942.1051

بیوتکنولوژی و بیوشیمی غلات

شاپا الکترونیکی: ۵۱۷۰-۲۷۸۳

Homepage: https://cbb.razi.ac.ir

کاربرد مکان‌یابی ارتباطی و عدم تعادل پیوستگی در مطالعات ژنومی گیاهی با تاکید بر ژنوم غلات

بابک عبدالهی مندولکانی[۱]، حسین عباسی هولاسو[۲]* و سالار شعف[۳]

[۱] استاد گروه مهندسی تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران.

[۲] دانشجوی پسا دکتری، گروه به‌نژادی و بیوتکنولوژی گیاهی، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران.

[۳] گروه کشاورزی و علوم محیطی، دانشگاه میلان، میلان، ایتالیا.

*نویسنده مسئول. رایانامه: hossein.pdf55@gmail.com

چکیده

آگاهی از تنوع ژنتیکی و درک صفات پیچیده ژنتیکی در گونه‌های زراعی نقش مهمی در بهره‌برداری از منابع ژنتیکی داشته و باعث سازگاری، توسعه و گسترش کشت ارقام زراعی به ویژه غلات در مناطق مختلف جغرافیایی می‌گردد. در این راستا، شناسایی و استفاده از نشانگرهای آگاهی بخش مرتبط با صفات کمی جهت استفاده در برنامه‌های به‌نژادی و انتخاب به کمک نشانگر از اهمیت زیادی برخوردار می‌باشد. یکی از کاربردهای اصلی نشانگرها تهیه نقشه‌های مولکولی ژنوم و آنالیز پیوستگی جمعیت‌های نقشه‌یابی است. در مطالعات پیوستگی و انتخاب ژنومی تعیین میزان و گستره عدم تعادل پیوستگی (LD) در تشخیص تعداد نشانگر مورد نیاز و اندازه نمونه، اهمیت به‌سزایی دارد. علاوه بر این، گزینش به‌منظور افزایش فراوانی موتاسیون‌های جدیدی که فقط در برخی از زیر جمعیت‌ها سودمند هستند، باعث باقی گذاشتن علائمی در سطح ژنوم می‌شود. اغلب این مناطق با ژن‌ها و QTLهای مرتبط با صفات مهم اقتصادی در ارتباط هستند. دسترسی به فناوری‌های نسل جدید توالی‌یابی، حجم بالای داده‌های فنوتیپی و تنوع زیاد ابزارهای آماری موجب شده است که مطالعات مکان‌یابی ارتباطی مبتنی بر عدم تعادل پیوستگی در گیاهان بتواند موفقیت‌های فراوانی را در شناسایی مکان‌های ژنی کنترل کننده صفات کمی به همراه داشته باشد. نقشه‌یابی ارتباطی مبتنی بر عدم تعادل پیوستگی می‌تواند وضوح نقشه ژنتیکی را با توجه به ارائه مخزن ژنی گسترده‌تر و وقایع نوترکیبی افزایش دهد. برخلاف نقشه‌یابی پیوستگی، روش مکان‌یابی ارتباطی با بهره‌گیری از تنوع موجود در جمعیت‌های طبیعی و لحاظ کردن تمامی وقایعی که در طول تکامل افراد رخ داده است، ارتباط بین تنوع فنوتیپی و چند شکلی موجود در ژنوم را شناسایی می‌کند و روشی امیدوار کننده برای غلبه بر محدودیت‌های نقشه‌یابی پیوستگی است. علی‌رغم اینکه نقشه‌یابی ارتباطی از توان آماری بالایی برخوردار است اما کاربرد این روش در جمعیت‌های دارای ساختار، در گونه‌های با میزان کم عدم تعادل پیوستگی و در صفاتی که توسط آلل‌های نادر کنترل می‌شود، بسیار پیچیده و گاهی ناممکن است. LD پایه و اساس نقشه‌یابی در سطح ژنوم (GWAS) است، که تحت تاثیر عوامل متعددی از جمله رانش ژنتیکی، ساختار جمعیت و گزینش طبیعی است. از عوامل محدود کننده فوق، ساختار جمعیت به عنوان یکی از عوامل اصلی اعتبارسنجی نتایج GWAS شناخته می‌شود. بنابراین آگاهی در مورد الگوی عدم تعادل پیوستگی به منظور مطالعات بیشتر GWAS و گزینش ژنومی اهمیت بسزایی دارد. معیار تعیین کننده موفقیت استفاده از مطالعات GWAS و گزینش ژنومی (GS) به مقدار LD بین نشانگرها و جایگاه‌های ژنومی مسبب صفات کمی در طول کل ژنوم بستگی دارد. با توجه به اهمیت روش عدم تعادل پیوستگی در مطالعات نقشه‌یابی صفات کمی، در این مقاله سعی شده است تا به عمده‌ترین ابعاد عدم تعادل پیوستگی، نحوه کاربرد آن در جمعیت، محدودیت‌های آن و استفاده از آن در به‌نژادی گیاهی به ویژه غلات پرداخته شود. همچنین، در این مقاله برخی از اطلاعات مربوط به نرم‌افزارهای آماری مورد استفاده در عدم تعادل پیوستگی ارایه می‌شود و در نهایت چالش‌ها و چشم‌اندازهای استفاده از این رویکرد بحث خواهد شد.

واژه‌های کلیدی: ارتباط نشانگر- صفت، ساختار جمعیت، غلات، مدل خطی ترکیبی، مکان‌یابی ارتباطی.

نوع مقاله: مقاله مروری

نوع مقاله: دریافت: ۱۴۰۲/۰۱/۰۷ **اصلاح:** ۱۴۰۲/۰۱/۳۱ **پذیرش:** ۱۴۰۲/۰۳/۲۶، **انتشار آنلاین:** ۱۴۰۲/۰۴/۰۱

استناد: عبدالهی مندولکانی، ب.، عباسی هولاسو، ح. و شعف، س. (۱۴۰۲). کاربرد مکان‌یابی ارتباطی و عدم تعادل پیوستگی در مطالعات ژنومی گیاهی با تاکید بر ژنوم غلات. *بیوتکنولوژی و بیوشیمی غلات*، ۲ (۲)، ۲۴۳-۲۶۸. DOI: 10.22126/cbb.2023.8942.1051

ناشر: دانشگاه رازی

© نویسندگان.

مقدمه

توسعه و کاربرد نشانگرهای مولکولی در طی سال‌ها به درک پایه‌های فیزیولوژیکی و ژنتیکی صفات پلی‌ژنتیک همانند مقاومت به تنش‌های زیستی و غیرزیستی در غلات به ویژه گندم کمک شایانی کرده است (Sukumaran *et al.*, 2018). این مسئله منجر به پیشرفت‌های عمده‌ای در تحقیقات ژنومیک گیاهی در طول سی سال اخیر شده است و نشانگرهای مولکولی به بخش مهمی در تحقیقات گیاهی تبدیل شده‌اند (Gupta *et al.*, 2005). دو پدیده عمده درگیر در ایجاد چندشکلی DNA که به‌وسیله نشانگرها شناسایی می‌شوند، جهش و نوترکیبی است. بنابراین شناسایی پیوستگی و روند چندشکلی در DNA محور اصلی کاربرد نشانگرهای مولکولی در طیف وسیعی از مطالعات است. در مطالعات پیوستگی، یکی از ملزومات، طراحی بجا و درست تلاقی‌ها است که منجر به تشکیل جمعیت‌های نقشه و یا لاین‌های شبه ایزوژنیک[1] (NILs)می‌شود. محدودیت‌های جدی در مطالعات پیوستگی وجود دارد. زیرا در همه گیاهان امکان انجام تلاقی‌های مطلوب و تهیه جمعیت‌های در حال تفرق وجود ندارند (مثلا در درختان باغی و جنگلی) و یا جمعیت‌های نقشه که برای این هدف استفاده می‌شوند، اغلب بسیار کوچک هستند و تنها دو آلل در یک مکان بررسی شده دارند.

اگرچه اطلاعات کمی در رابطه با ساختار ژنتیکی صفات مهم و به ویژه صفاتی که بصورت کمی به ارث می‌رسند، وجود دارد (Mackay *et al.*, 2009)، ولی امروزه روش‌های مختلف ژنومیکس توانسته امکان شناسایی دقیق آنها را به وجود آورد (Semagn *et al.*, 2010). شناسایی نواحی ژنومی دخیل در کنترل صفات کمی معمولا به دو روش اصلی نقشه‌یابی پیوستگی و تجزیه ارتباطی یا نقشه‌یابی بر مبنای عدم تعادل پیوستگی انجام می‌گیرد. اگرچه موفقیت نقشه‌یابی پیوستگی در شناسایی QTLها برای صفات متعدد در بسیاری از گونه‌های زراعی به ویژه غلات به اثبات رسیده است (Pasam *et al.*, 2012)، ولی با توجه به اینکه منطقه QTL شناسایی شده بیشتر از چند سانتی‌مورگان و شامل صدها ژن می‌باشد، شناسایی QTLهای کاندید مناسب را با مشکل مواجه می‌سازد.

همچنین ساخت جمعیت‌های نقشه‌یابی مانند لاین‌های اینبرد نوترکیب (RILs) از طریق تلاقی‌های کنترل شده و سپس چندین نسل خودگشنی زمان‌بر بوده و این نیز خود یکی دیگر از محدودیت‌های استفاده از نقشه‌یابی پیوستگی می‌باشد. نقشه‌یابی ارتباطی به عنوان یک روش جایگزین و یا مکمل جهت شناسایی ارتباط بین نشانگر و صفت، دارای مزایای زیادی نسبت به نقشه‌یابی QTL است که از آن جمله می‌توان به افزایش وضوح QTL، استفاده از ژرم‌پلاسم طبیعی و افزایش پوشش آللی اشاره نمود (Yu *et al.*, 2006).

روش نقشه‌یابی ارتباطی علاوه بر نقشه‌یابی دقیق‌تر، با ژرم‌پلاسم‌های دارای تنوع ژنتیکی بیشتر سازگار بوده و اجازه نقشه‌یابی چندین صفت بطور همزمان را می‌دهد. بنابراین، برای هر صفت مورد نظر نیازی به ایجاد

[1] Near-isogenic lines

جمعیت‌های دو والدی که خود باعث هزینه اضافی جهت ارزیابی ژنوتیپی و فنوتیپی می‌شود، نیست. همچنین این روش به طور گسترده‌ای در ژنتیک انسانی و جانوری که در آن‌ها ایجاد جمعیت‌های در حال تفرق بزرگ غیرممکن می‌باشد، استفاده می‌شود (DeWan *et al.*, 2006).

عدم تعادل پیوستگی[۲] (LD)

در برخی موارد بین آلل‌های موجود در جایگاه‌های ژنی مختلف یک ارتباط پیش‌بینی نشده وجود دارد که با قوانین ژنتیک کلاسیک متفاوت است. این ارتباط غیرتصادفی بین آلل‌ها در جایگاه‌های ژنی مختلف را عدم تعادل پیوستگی می‌نامند که باعث می‌شود فراوانی‌های هاپلوتایپی مشاهده شده در جمعیت با فراوانی‌های مورد انتظار آن در قانون هاردی واینبرگ متفاوت باشد (Zhu *et al.*, 2008). عدم تعادل پیوستگی که به آن عدم تعادل فاز گامتی نیز گفته می‌شود، ارتباط غیر تصادفی بین آلل‌های مکان‌های مختلف را نشان می‌دهد. این مکان‌ها ممکن است دارای پیوستگی فیزیکی باشند. عدم تعادل پیوستگی الزاما به مفهوم پیوستگی فیزیکی نیست هرچند در پیوستگی‌های شدید درجه بالایی از عدم تعادل تشخیص داده می‌شود ولی ممکن است بین دو مکان مستقر در کروموزوم‌های مختلف نیز عدم تعادل پیوستگی معنی‌دار باشد(Stich *et al.*, 2007). عدم تعادل پیوستگی می‌تواند برای اهداف متعددی در زمینه ژنومیک گیاهی مورد استفاده قرار گیرد. یکی از کاربردهای اصلی آن در مطالعه ارتباط نشانگر-صفت برای استفاده در گزینش بر مبنای نشانگر است. کاربرد مهم دیگر آن بررسی تنوع ژنتیکی جمعیت‌های طبیعی و ژرم‌پلاسم‌ها و استفاده از آن در مطالعات ژنتیک جمعیت و برنامه‌های اصلاح گیاهی است (Rao Kovi *et al.*, 2015; Yasir *et al.*, 2022).

استفاده از جمعیت‌های مصنوعی همچون F_2 دارای محدودیت‌هایی است که از جمله آن‌ها می‌توان به محدودیت در چرخه‌های میوزی و کم بودن نوترکیبی تا تولید جمعیت و همچنین امکان بررسی فقط دو آلل در هر مکان ژنی اشاره کرد. با توجه به تنوع بالای جمعیت-های طبیعی در مقایسه با جمعیت‌های مصنوعی، استفاده از نقشه‌های عدم تعادل پیوستگی در این جمعیت‌ها، روشی قدرتمند برای شناسایی ژن‌های کنترل کننده صفات کمی شده است. با توجه به اینکه چرخه‌های میوزی فراوانی در فرایند تکامل و توسعه ژرم‌پلاسم‌های طبیعی صورت می‌گیرد، بنابراین امکان شناسایی دقیق جایگاه‌های ژنومی مرتبط با صفات مورد نظر در این روش وجود دارد (Kraakman *et al.*, 2006). در سال‌های اخیر نقشه‌های عدم تعادل پیوستگی در گیاهان مختلف از جمله آرابیدوپسیس، ذرت، برنج، جو، گندم، پنبه و گوجه-فرنگی تهیه شده است (Flint-Garcia *et al.*, 2003).

تفاوت بین هاپلوتایپ مشاهده شده و مورد انتظار مبتنی بر فراوانی آللی تحت عنوان D تعریف می‌شود:

معادله (۱): $D = p_{AB} - p_A p_B$

p_{AB} فراوانی گامت AB است، p_A و p_B فراوانی آلل‌های A و B می‌باشد (Zhu *et al.*, 2008). در غیاب عوامل موثر

[2] Linkage disequilibrium

دیگر، نوترکیبی از طریق جورشدگی‌های تصادفی میزان LD را به‌میزان D_t کاهش می‌دهد.

معادله (۲): $D_t=D_0 (1-r)^t$

D_t میزان LD باقیمانده بین دو مکان پس از t نسل جورشدگی تصادفی است. چندین معیار آماری برای محاسبه LD بکار می‌رود. نحوه تاثیرپذیری این معیارها از فراوانی‌های آللی و اندازه کوچک نمونه متفاوت می‌باشد (Hedrick, 1987). از جمله این معیارها D' و R^2 می‌باشد که در سطح وسیعی برای تعیین کمیت LD بکار می‌روند. در مورد مکان‌های دو آللی، D' و R^2 از فرمول‌های زیر محاسبه می‌شوند:

معادله (۳): $D'= |D| / D_{max}$

معادله (۴): $D_{max}= \min (p_A p_b, p_a p_B)$ *if* $D > 0$

معادله (۵): $D_{max}= \min (p_A p_B, p_a p_b)$ *if* $D < 0$

معادله (۶): $R^2= D^2/ p_A p_a p_B p_b$

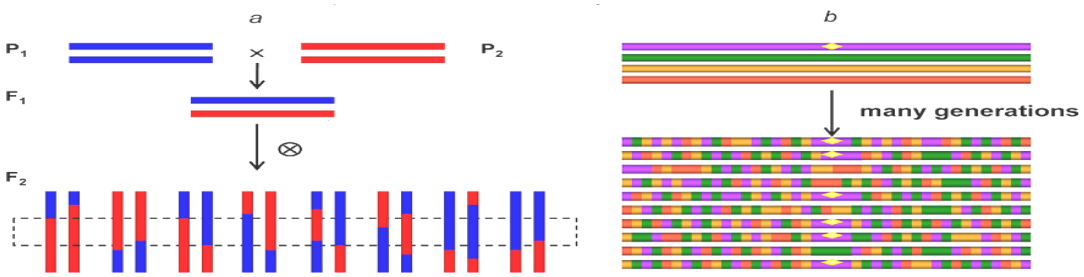
یکی از خصوصیات نامطلوب معیار D این است که دامنه آن بوسیله فراوانی آللی تعیین می‌شود. به همین علت D' برای نرمال کردن میزان D و با توجه به حداکثر مقدار ممکن فراوانی آللی استفاده می‌شود و دامنه آن از صفر تا یک متغیر است (Zhu *et al.*, 2008). R^2 مجذور ضریب همبستگی پیرسون است و میزان مورد انتظار آن از رابطه زیر قابل محاسبه است که N اندازه جمعیت موثر بوده و c ضریب نوترکیبی مورگان می‌باشد (Hill & Robertson, 1968).

معادله (۷): $R^2= 1/ (1+4Nc)$

ساده‌ترین معیار LD که بیشترین استفاده را دارد، D و یا D' (نوع استاندارد شده D) است که تفاوت بین فراوانی گامتی مشاهده شده هاپلوتایپ‌ها و فراوانی گامتی مورد انتظار آنها را تحت شرایط عدم تعادل پیوستگی نشان می‌دهد. انتخاب معیار مناسب LD بستگی به موضوع مطالعه دارد. البته D' و R^2 پر کاربردترین معیارهای LD می‌باشند (Oraguzie *et al.*, 2007). D' برای مقایسه فراوانی‌های آللی مختلف در بین مکان‌ها بکار می‌رود و شدیدا در نمونه‌های کوچک و فراوانی‌های آللی پایین افزایش می‌یابد. بنابراین مقادیر متوسط D' برای آنالیز مقایسه‌ای مطالعات مختلف LD مناسب نمی‌باشد و قبل از استفاده برای کمی کردن دامنه LD باید توسط R^2 تایید شود. R^2 مجذور ضریب همبستگی بین دو مکان بوده و در مواردی که فراوانی آللی پایین می‌باشد قابل اعتمادتر است. R^2 بوسیله جهش و نوترکیبی تحت تاثیر قرار می‌گیرد در حالی‌که D' بیشتر توسط روند جهش تحت تاثیر قرار می‌گیرد. شایان ذکر است که مناسب‌ترین معیار کمی LD مورد نیاز برای مکان‌یابی ارتباطی R^2 است که نشاندهنده همبستگی‌های نشانگر- صفت نیز است. دامنه R^2بین صفر تا یک بوده و وقتی تنها دو هاپلوتایپ وجود داشته باشد برابر با یک خواهد بود (Abdallah *et al.*, 2003). LD می‌تواند برای اهداف مختلفی در تحقیقات ژنومیکس گیاهی استفاده شود. یکی از کاربردهای عمده LD در گیاهان، مطالعه ارتباط نشانگر- صفت (بدون استفاده از یک جمعیت نقشه) و متعاقب آن گزینش به کمک نشانگر است. کاربرد مهم دیگر LD، مطالعه تنوع

ژنتیکی در جمعیت‌های طبیعی و ژرم‌پلاسم‌ها (مکان‌یابی ارتباطی) و استفاده در مطالعه ژنتیک جمعیت و در برنامه‌های بهبود محصولات زراعی است (Gupta *et al.*, 2005).

تاکنون مطالعات متعددی در زمینه برآورد ساختار جمعیت و تعیین عدم تعادل پیوستگی در گیاهان زراعی به ویژه غلات انجام شده است. چن و همکاران (Chen *et al.*, 2012) به منظور تعیین ساختار جمعیت و عدم پیوستگی در بین نمونه‌ای از ارقام و لاین‌های گندم زمستانه چینی، با استفاده از ۲۶۹ نشانگر ریزماهواره که در سراسر ژنوم پراکنده بودند ۹۰ ژنوتیپ را مورد ارزیابی قرار دادند. مشاهدات آن‌ها نشان داد که ساختار جمعیت بدست آمده همبستگی مثبتی با منشا جغرافیایی ژنوتیپ‌ها دارد، به طوری‌که افراد هم گروه از لحاظ منشا جغرافیایی مشابه بودند. در این بررسی حداکثر عدم تعادل پیوستگی ۱۷/۴ سانتی‌مورگان برآورد شد. در مطالعه‌ای بر روی ۹۴ رقم گندم اروپایی از مجموع ۷۰۰۰ نشانگر، ۱۸۴۹ نشانگر چند شکل DArT برای تعیین ساختار جمعیت و میزان عدم تعادل پیوستگی استفاده شد. پس از تعیین ساختار جمعیت ارقام به دو زیر گروه منتسب شدند. به طوری‌که، ارقام مستقر در هر گروه از لحاظ جغرافیایی مشابه بودند. در این بررسی میانگین r2 برای همه جفت نشانگرها ۰/۰۸ و برای جفت نشانگرهای معنی‌دار ۰/۲۷۱ برآورد شد (Nielsen *et al.*, 2014). قاسم و همکاران (Qaseem *et al.*, 2018) به منظور بررسی تغییرات ژنتیکی در یک مجموعه‌ای از ۱۰۸ رقم بهاره گندم طی دو سال تحت تنش خشکی، گرما و ترکیبی از هر دو تنش از ۹۶۴۶ نشانگر SNP استفاده کردند. بررسی عدم تعادل پیوستگی در داخل زیر جمعیت‌ها نشان داد که بیش‌ترین عدم تعادل پیوستگی مربوط به کروموزوم 4D و کم‌ترین مقدار به کروموزوم شماره 5A اختصاص یافت. این محققین بیان کردند گاهی اوقات دو نشانگر که بسیار بهم نزدیک هستند، ممکن است مقدار عدم تعادل پیوستگی پایینی از خود بروز دهند یا نشانگرهایی با فاصله دورتر از هم، مقدار عدم تعادل پیوستگی بالایی را نشان دهند. پنج زیر جمعیت در مجموعه ژنوتیپ‌های مورد مطالعه با استفاده از برنامه Structure شناسایی شد. از آنجا که تعداد زیادی از ژنوتیپ‌ها به گروه ژنوتیپ‌های مخلوط منتسب شده‌اند، نشان دهنده وجود اختلاط در جمعیت است. اختلاط در جمعیت‌های گندم می‌تواند ناشی از تلفیق در نمونه‌های بومی از آلل‌هایی از بیش از یک خزانه ژنی منفرد باشد که این امر به دلیل پراکنش گندم از بیش از یک جمعیت اجدادی تکی است.



شکل ۱- مقایسه شماتیک تجزیه پیوستگی (a) و مکان‌یابی ارتباطی (b) (Zhu *et al.*, 2008).

Figure 1. Schematic comparison of linkage analysis (a) and association mapping (b) (Zhu *et al.*, 2008).

فاکتورهای موثر بر LD

عوامل مختلفی بر میزان LD تاثیر دارند. عواملی که منجر به افزایش LD می‌شوند عبارتند از خویش آمیزی، ساختار جمعیت، اتوگامی، اپیستازی، اندازه کوچک جمعیت، جدایی ژنتیکی بین دودمان‌ها، تفکیک جمعیت به زیر جمعیت‌ها، ضریب نوترکیبی پایین، اختلاط جمعیتی، انتخاب طبیعی و مصنوعی و گزینش متعادل‌کننده (Gupta *et al.*, 2005; Oraguzie *et al.*, 2007). عواملی که منجر به کاهش LD می‌شوند شامل آمیزش با ژنوتیپ‌های دیگر، ضریب نوترکیبی بالا و ضریب جهش بالا می‌باشند (Gupta *et al.*, 2005). عوامل دیگری نیز وجود دارند که منجر به افزایش یا کاهش در LD می‌شوند یا ممکن است LD بین برخی جفت آلل‌ها را افزایش داده و بین سایر جفت آلل‌ها را کاهش دهند. برای مثال، جهش، LD بین جفت آلل‌های وحشی را کاهش می‌دهد در حالیکه باعث افزایش LD بین جفت آلل‌های جهش‌یافته می‌شوند. به‌طور مشابه، نوآرایی‌های ژنومی نیز باعث کاهش LD بین ژن‌های جدا شده در اثر نوآرایی می‌شود. اما LD بین ترکیبات جدید ژنی در مجاورت نقاط انفصال[3] به علت سرکوبی موضعی نوترکیبی افزایش می‌یابد. سایر عوامل موثر بر LD شامل ساختار جمعیت، اپیستازی، تبدیل ژن و اریب تحقیق می‌باشند. مطالعه این عوامل جهت برآورد LD در مطالعات تجزیه ارتباط مبتنی بر پیوستگی مفید و ضروری است و بایستی اثر عواملی که علاوه بر پیوستگی مسبب LD هستند، حذف شوند (شکل ۱) (Rafalski & Morgante, 2004).

ارتباط نشانگر- صفت در گیاهان

ارتباط نشانگر- صفت در گیاهان زراعی عموما از طریق تجزیه پیوستگی، کاربرد روش‌هایی همانند t-test، تجزیه رگرسیون ساده و QTL صورت می‌گیرد (Hackett, 2002). این روش‌ها محدودیت‌های زیادی از جمله امکان مکان‌یابی دقیق ژن‌ها و QTLها مقدور نبوده، شناسایی نواحی کروموزومی مشکل‌تر بوده، احتمال بروز کراسینگ اور نیز در جمعیت در حال تفرق زیاد است، در دسترس نبودن جمعیت‌های نقشه‌یابی و احتیاج به زمان زیاد برای تولید جمعیت‌ها دارند که با استفاده از روش مبتنی بر LD تا حد زیادی بر این محدودیت‌ها غلبه شده است.

[3] Breakpoints

برای انجام مطالعه ارتباط نشانگر- صفت با استفاده از LD، روش‌های محاسبه در مورد صفات کیفی و کمی تفاوت دارند، اگرچه برخی مواقع صفات کمی به عنوان صفات کیفی در نظر گرفته می‌شوند. کاربرد LD برای شناسایی QTLهای کنترل کننده یک صفت کمی بسیار پیچیده‌تر و مشکل‌تر است، اما در عین حال ارزشمندتر نیز می‌باشد چرا که موقعیت دقیق‌تری از وضعیت QTL کنترل کننده صفت مورد نظر را ارائه می‌دهد. با مقایسه روش‌های تجزیه پیوستگی و LD برای تشخیص QTL، مشخص می‌شود که تجزیه مبتنی بر پیوستگی برای شناسایی QTLهای دخیل در صفت در کل ژنوم مناسب‌تر است، در حالی‌که روش LD موقعیت دقیق‌تری از یک QTL ارائه می‌دهد. بنابراین از تجزیه پیوستگی برای تهیه نقشه ژنتیکی و سپس از نقشه ژنتیکی به منظور مطالعات ژنتیکی استفاده می‌شود. وقتی داده‌های صفت ارتباط خطی با ژنوتیپ‌های نشانگر منفرد داشته باشد، LD بین یک نشانگر منفرد و یک QTL به‌وسیله آنالیز رگرسیون قابل اندازه‌گیری است و رگرسیون‌های معنی‌دار، نشانگرهای مرتبط با فنوتیپ را تشخیص می‌دهند. از آنجایی که ارتباط نشانگر و صفت برخی مواقع می‌تواند به‌علت دلایلی غیر از پیوستگی باشد، نیاز به انجام آنالیزهای بیشتر برای انتخاب نشانگرهایی است که واقعا با صفت به علت پیوستگی شدید مرتبط باشند. این رگرسیون صفت و ژنوتیپ یا نشانگر برخی مواقع از طریق آزمون دو نشانگر مجاور به لحاظ ارتباطشان با صفت انجام می‌شود (Remington *et al.*, 2001). در موارد دیگر، اثر هاپلوتایپ‌های نشانگر بر صفت از طریق آنالیز رگرسیون برآورد می‌شود. هاپلوتایپ‌ها، آلل‌های نشانگر مشابه دارند و با اثر فنوتیپی مشابه مرتبط‌اند و شامل یک QTL می‌باشند. قرارگیری چنین موقعیت دقیقی درون یک ناحیه کروموزومی بسیار کوچک از طریق LD امکان‌پذیر است، اما از طریق تجزیه پیوستگی امکان‌پذیر نیست. از طریق تجزیه پیوستگی، نوترکیبی درون چنین ناحیه کوچکی در یک جمعیت معین آزمایشی قابل دسترسی نیست (Meuwissen & Goddard, 2000).

توسعه سریع تکنولوژی‌های آنالیز ژنوتیپی و فنوتیپی در طی سال‌های اخیر به درک پایه‌های فیزیولوژیکی و ژنتیکی صفات کمی در غلات کمک شایانی کرده است (Sukumaran *et al.*, 2018). تا به امروز، از نقشه‌یابی QTL جهت ارزیابی صفات کمی با استفاده از جمعیت‌های حاصل از تلاقی دو والدی در گیاهان زراعی مختلف از جمله گندم (Gahlaut *et al.*, 2017)، ذرت (Zhu *et al.*, 2011)، برنج (Qu *et al.*, 2008) و سورگوم (Kebede *et al.*, 2001) استفاده شده است. با وجود این ارزیابی نقشه‌یابی چنین جمعیت‌هایی بسیار وقت‌گیر می‌باشد. از طرف دیگر، وقتی‌که برای مکان‌یابی QTLها فقط از دو والد استفاده می‌شود، برخی از نشانگرها و QTLهای بالقوه احتمالا مونومورف خواهند بود، حتی اگر لاین‌های والدینی برای صفت مورد نظر به دقت انتخاب شده باشند. از آنجایی که QTL را فقط می‌توان در مکان‌های ژنی پلی‌مورف در داخل ژنوم پیدا کرد، تعداد QTLهایی که انتظار می‌رود از طریق تلاقی دو والدی شناسایی شوند کمتر از

تعدادی خواهد بود که از تجزیه چندین تلاقی بطور همزمان انتظار می‌رود.

مکان‌یابی ارتباطی

صفات پیچیده و مهم زراعی تحت تاثیر QTLهای متعدد، محیط و تعامل بین QTL و محیط هستند. تجزیه پیوستگی و مکان‌یابی ارتباطی دو روش معمول و پرکاربرد برای بررسی صفات پیچیده می‌باشند (شکل ۱). معمولا روش تجزیه پیوستگی در گیاهان، QTLهای مورد بررسی را در فاصله ۱۰ تا ۲۰ سانتی مورگانی مکان‌یابی می‌کند، زیرا افراد مورد مطالعه کم بوده و تعداد موارد نوترکیبی در جمعیت مکان‌یابی محدود می‌باشد. در حالی‌که صدها مطالعه تجزیه پیوستگی در گونه‌های مختلف گیاهی در طی دو دهه اخیر صورت گرفته است، تنها تعداد محدودی از QTLهای شناخته شده، همسانه‌سازی شده یا در سطح ژن بررسی شده‌اند (Holland, 2007). مکان‌یابی ارتباطی با عنوان مکان‌یابی LD نیز شناخته می‌شود و برای تشریح تنوع صفات پیچیده در سطح توالی، با تشریح روند تاریخی و تکاملی نوترکیب در سطح جمعیت بکار می‌رود. مکان‌یابی ارتباطی به عنوان جانشین جدیدی برای تجزیه پیوستگی معمولی، سه مزیت دارد: ۱) مکان‌یابی ارتباطی 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 به دلیل نوترکیبی‌های زیاد نقشه ژنتیکی افزایش می‌دهد. ۲) در این روش به دلیل اینکه از تنوع موجود در جمعیت‌های طبیعی استفاده می‌شود، تا حد زیادی در وقت و هزینه‌ها صرفه‌جویی می‌شود ۳) در این روش به طور بالقوه و هم-زمان تمام آلل‌های یک مکان ژنی (حداقل ۲ آلل)، مطالعه و ارزیابی می‌شود (Zhu *et al.*, 2008). مکان‌یابی ارتباطی شامل شش مرحله می‌باشد: الف) انتخاب گروهی از لاین-ها یا ارقام زراعی با تنوع ژنتیکی گسترده برای تشکیل جمعیت یا پنل نقشه‌یابی ب) ارزیابی فنوتیپی صفات ج) ارزیابی ژنوتیپی جمعیت با استفاده از نشانگرهای پس-زمینه‌ای یا نشانگرهای پیوسته احتمالی به ژن د) تعیین میزان عدم تعادل پیوستگی برای یک کروموزوم و یا ژنوم با استفاده از داده‌های نشانگرهای مولکولی روی جمعیت نقشه‌یابی ر) ارزیابی ساختار و روابط خویشاوندی جمعیت ز) تعیین رابطه بین داده‌های فنوتیپی و ژنوتیپی بر اساس اطلاعات حاصل از LD، ساختار جمعیت و روابط خویشاوندی با استفاده از روش‌های آماری مناسب

مکان‌یابی ارتباطی گامی موثر در به‌نژادی مولکولی است. پیوستگی ژنتیکی بین نشانگرها و مکان‌های ژنی صفات کمی، محتمل‌ترین توجیه برای وجود رابطه بین نشانگرهای مولکولی و نمود صفات کمی است. استفاده از پیوستگی بین نشانگرهای مولکولی و ژن‌های کنترل‌کننده صفات کمی، فرایند به‌نژادی را تسریع می‌کند.

از زمان کاربرد این روش در گیاهان، مکان‌یابی ارتباطی به علت پیشرفت تکنولوژی‌های ژنومیک با توان بالا[4]، علاقه به شناسایی آلل‌های جدید و پیشرفت روش‌های آماری، توجه زیادی را در پژوهش‌های ژنتیکی بخود جلب نموده است (شکل ۲). مطالعات ارتباطی به انواع مختلفی تقسیم

[4] High throughput genomic technologies

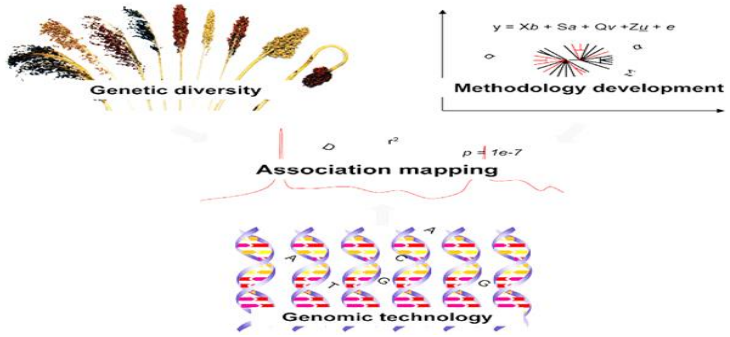
می‌شوند: ۱) چندشکلی کاندید[5]: بر چند شکلی فرد مشکوک به آلودگی توسط عامل بیماری تمرکز دارد، ۲) مکان‌یابی ارتباطی ژن کاندید[6]، ۳) مکان‌یابی دقیق[7]: اغلب به مطالعاتی که در یک ناحیه کاندید ۱ تا ۱۰ میلیون جفت باز شامل صدها SNP صورت می‌گیرد، اتلاق می‌شود. ناحیه کاندید می‌تواند توسط یک مطالعه پیوستگی شناسایی شود و شامل ۵ تا ۵۰ ژن می‌باشد و ۴) مکان‌یابی ارتباطی در سطح ژنوم[8] (Balding, 2006) .

با توجه به مقیاس پژوهش و با تمرکز بر یک مطالعه ویژه، مکان‌یابی ارتباطی معمولا از طریق دو روش اصلی، یعنی مکان‌یابی ارتباطی ژن کاندید و مکان‌یابی ارتباطی در سطح ژنوم انجام می‌گیرد (Zhu *et al.*, 2008). قبل از آغاز مکان‌یابی ارتباطی، پژوهش‌گران باید به دقت تمام جنبه‌های ژنتیکی گونه و ژرم‌پلاسم در دسترس را بررسی کنند. سطح پلوئیدی افراد یک گونه گیاهی که خصوصیات ژنتیکی آنها مشخص نیست باید ارزیابی شود مخصوصا اگر جمعیت شامل اکسشن‌های وحشی موجود در بانک ژرم‌پلاسم باشد. این مسئله از مشکلات موجود در تمایز اثرات چند شکلی‌های عملکردی حاصل از تعداد آلل جلوگیری می‌کند. انتخاب ژرم‌پلاسم در موفقیت مکان‌یابی ارتباطی حیاتی است (Breseghello *et al.*, 2006a). تنوع ژنتیکی، وسعت LD در سطح ژنوم و ارتباط درون جمعیتی، تعیین‌کننده کیفیت مکان‌یابی، تنوع نشانگری، روش‌های آماری و قدرت مکان‌یابی می‌باشند (شکل ۲ و جدول ۱) (Zhu *et al.*, 2008).

خالید و همکاران (Khalid *et al.*, 2018) در مطالعه GWAS ۲۰۹ لاین RIL حاصل از تلاقی لاین سنتتیک W7984 و رقم Opata[9] را با استفاده از تکنولوژی GBS در مرحله گیاهچه‌ای تحت شرایط آبیاری مطلوب و کم آبی مورد ارزیابی قرار دادند. نقشه ژنتیکی جمعیت مشتمل بر ۲۶۳۹ نشانگر SNP بود. در مجموع، ۱۶ QTL در نه کروموزوم برای همه صفات مورد مطالعه به غیر از صفت وزن خشک ریشه شناسایی شد که ۶ و ۱۰ QTL به ترتیب مربوط به آبیاری مطلوب و تنش کم آبی بودند. دامنه تبیین واریانس فنوتیپی QTLهای مکان‌یابی شده از ۴/۷ تا ۵۹ درصد متغیر بود و کروموزوم 7B با شش QTL بیشترین تعداد QTLهای شناسایی شده را در برداشت. از این QTLها، ۳، ۲ و ۱ به ترتیب برای ارتفاع گیاهچه (QSL.nust-7B)، وزن تر ریشه (QFRW.nust-7B) و طول ریشه (QRL.nust-7B) شناسایی شد که QFSW.nust-7B با تبیین ۵۹٪ از واریانس فنوتیپی وزن تر ریشه یک QTL بزرگ اثر محسوب می‌شد. در مطالعه دیگر خالید و همکاران (Khalid *et al.*, 2019) به منظور بررسی تنوع ژنتیکی ۲۱۳ لاین برتر و سنتتیک گندم و همچنین ارزیابی ارتباط نشانگر-صفت تحت تنش خشکی و نرمال در کل ۱۲۴ مارکر KASP مورد استفاده قرار دادند. از این تعداد دو مارکر TaSST-D1 و TaSST-A1 مرتبط با ژن‌های کربوهیدرات محلول در آب، پیوسته با ژن‌های کنترل کننده ارتفاع بوته و وزن هزار دانه بودند.

[5] Candidate polymorphism
[6] Candidate-gene association mapping
[7] Fine mapping
[8] Genome-wide association mapping

تجزیه ارتباط ژن‌های عملکردی با صفات زراعی و بیوشیمیایی در دو شرایط آبیاری مطلوب و تحت تنش خشکی نشان داد که ۲۱ نشانگر ارتباط معنی‌داری با صفات مختلف نشان دادند که این محققین بیان داشتند که استفاده از تجزیه ارتباط برای بررسی ژنوتیپ‌های مختلف گندم در روند برنامه‌های اصلاحی مفید و کارآمد است.



شکل ۲- مراحل مختلف انجام مکان‌یابی ارتباطی (Zhu *et al.*, 2008).

Figure 2. Different stages of association mapping (Zhu *et al.*, 2008).

جدول ۱- نمونه‌هایی از مطالعات انجام گرفته در زمینه مکان‌یابی ارتباطی در گونه‌های گیاهی مختلف.

Table 1. Examples of association mapping studies in various plant species.

گیاه Plant	نوع جمعیت Population	صفات مورد مطالعه Traits studied	گروه پژوهشی Research team
ذرت Maize	لاین‌های اینبرد الیت Elite inbred lines	محتوای اسید اولئیک Oleic acid content	Belo *et al.*, 2008
ملون Melon	توده‌های بومی و ارقام Landraces and varieties	مقاومت به ویروس ZYMV Resistance to ZYMV virus	Abdollahi Mandoulakani *et al.*, 2015
آرابیدوپسیس Arabidopsis	توده بومی Accessions	شاخه دهی Shoot branching	Ehrenreich *et al.*, 2007
گندم Wheat	رقم Cultivar	اندازه مغز، کیفیت آسیاب شدن Kernel size, Milling quality	Breseghello & Sorrells, 2006b
گندم Wheat	لاین Line	صفات مورفولوژیک و فیزیولوژیک Morphological and physiological traits	Khalid *et al.*, 2018

Khalid *et al.*, 2019	صفات مورفولوژیک و فیزیولوژیک Morphological and physiological traits	لاین Line	گندم Wheat
Safdar *et al.*, 2020	صفات مورفولوژیک و فیزیولوژیک Morphological and physiological traits	رقم Cultivar	گندم Whea
Khan *et al.*, 2022	صفات مورفولوژیک Morphological traits	رقم Cultivar	گندم Whea
Abdollahi Mandoulakani *et al.*, 2017	صفات مورفولوژیک Morphological traits	رقم Cultivar	گندم Wheat
Abdollahi Mandoulakani *et al.*, 2015	صفات مورفولوژیک Morphological traits	جمعیت‌های زراعی Cultivated populations	یونجه Alfalfa
Kraakman *et al.*, 2006	صفات زراعی و مقاومت به زنگ برگ Agronomical traits and resistance to leaf rust	رقم Cultivar	جو Barley
Ogrodowicz *et al.*, 2023	صفات مورفولوژیک و صفات مرتبط با ریشه Agronomical and root-related traits	رقم Cultivar	جو Barley
Agrama *et al.*, 2007	عملکرد و ترکیبات Yield and compounds	توده بومی Accessions	برنج Rice
Dolati Baneh *et al.*, 2014	صفات مورفولوژیک Morphological traits	واریته Varieties	انگور Grapevine
Wei *et al.*, 2006	مقاومت به بیماری Resistance to disease	کلون Clone	نیشکر Sugarcane

مکان‌یابی ارتباطی ژن‌های کاندید

مکان‌یابی ارتباطی ژن کاندید، روشی است که برای بررسی صفات پیچیده استفاده می‌شود. ژن‌های کاندید مورد استفاده در این روش معمولا براساس مطالعات ژنتیکی، بیوشیمیایی یا فیزیولوژیکی در گیاهان مدل و غیر مدل انتخاب می‌گردند. در این روش اغلب از نشانگرهای SNP استفاده می‌شود (Rafalski & Morgante, 2004). مکان‌یابی ارتباطی ژن کاندید به چند شکلی‌های موجود در ژن‌های کاندید انتخابی مربوط است که در کنترل تفاوت فنوتیپی صفات ویژه نقش دارند و به شناسایی SNPهای بین لاین‌ها و درون ژن‌های ویژه وابسته است. در این مطالعات اثر ۵ تا ۵۰ عدد SNP درون

یک ژن به عنوان کاندیدهای مرتبط با تنوع صفت استفاده می‌شود، این ژن می‌تواند یک کاندید موقعیتی[9] یا یک کاندید عملکردی[10] باشد.

بنابراین روش مذکور، اصلی‌ترین روش برای شناسایی SNPهای ژن کاندید است که بر توالی‌یابی آمپلیکون‌های چندین فرد در یک جمعیت پیوستگی استوار است که افراد از لحاظ ژنتیکی مشخص و مجزا هستند. افراد با تنوع کمتر در بحث کشف SNP برای شناسایی SNPهای معمولی مورد نیاز هستند در حالی‌که افراد با تنوع بیشتر برای شناسایی SNPهای کمیاب بکار می‌روند. پروموتور، اینترون، اگزون و نواحی ترجمه نشده ′5′3 همگی اهداف معقولی برای شناسایی‌SNPهای ژن کاندید هستند. انتظار می‌رود نواحی غیرکدکننده سطوح بالاتری از تنوع نوکلئوتیدی را نسبت به نواحی کدکننده داشته باشند. ضریب کاهش LD برای یک مکان ژن کاندید ویژه، تعداد SNPها در واحد طول (مثلا کیلو باز) را که برای شناسایی روابط معنی‌دار نیاز است، نشان می‌دهند. بنابراین تعداد و طول جفت باز آمپلیکون‌ها برای نمونه‌گیری مناسب یک مکان ژن کاندید کاملا بستگی به LD و توزیع SNP دارد، در نواحی با LD پایین که تراکم بالاتری از نشانگرهای SNP وجود دارد، تنوع نوکلئوتیدی زیادی نیاز می‌باشد. ژن‌های کاندید بر مبنای دانسته‌های قبلی حاصل از تجزیه جهش، مسیر بیوشیمیایی دخیل در صفت یا تجزیه پیوستگی صفت هدف انتخاب می‌شوند. این روش هزینه کمی دارد و یک سیستم مختص صفت[11] است، اما ژن‌های ناشناخته دخیل در صفت مورد بررسی قرار نمی‌گیرند (Zhu *et al.*, 2008).

مکان‌یابی ارتباطی در سطح ژنوم

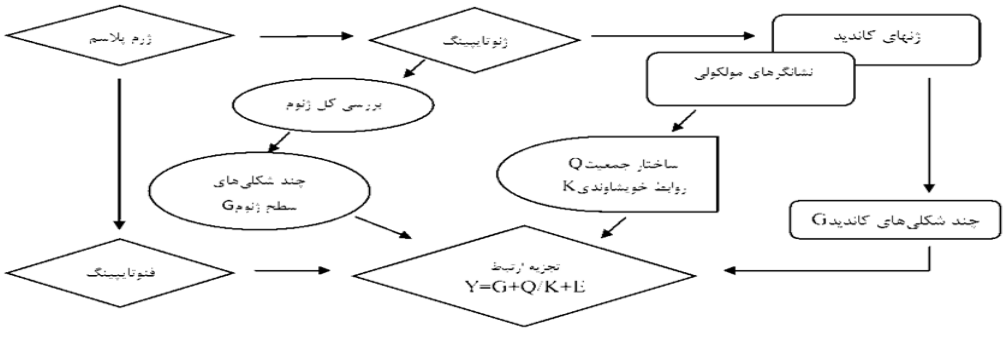
در این روش کل ژنوم برای شناسایی نواحی مرتبط با صفت مورد نظر بررسی می‌شود و معمولا هزاران SNP مورد استفاده قرار می‌گیرد. با ظهور تکنولوژی تعیین ژنوتیپ با توان بالا و نشانگرهای مولکولی کارایی استفاده از این روش گسترش بیشتری یافته است. در اسکن پیوستگی کل ژنوم در محصولات زراعی، گام اولیه و مهم، استفاده از ابزارهای توالی‌یابی DNA با ظرفیت بالا یا آرایه‌های الیگونوکلئوتیدی با تراکم بالا برای شناسایی SNPها در تراکمی که دقیقا ساختار LD کل ژنوم و تنوع هاپلوتایپی را نشان دهد، می‌باشد. تناسب روش توالی‌یابی DNA برای کشف SNP بستگی به تعداد SNPهای مورد نیاز برای یک اسکن ژنومی موثر در یک جمعیت پیوستگی دارد. برای مثال LD بالا در ۹۵ اکسشن آرابیدوپسیس و ۱۰۲ لاین اینبرد جو این امکان را فراهم نمود که آزمون پیوستگی برای تعداد کمی از SNPهای کشف شده، از طریق روش توالی‌یابی سانجر صورت بگیرد و نقشه‌یابی ژنومی در سطح متوسط انجام شود در مقابل صدها هزار نشانگر SNP برای یک اسکن ژنومی قوی در گونه‌هایی با LD پایین و تنوع هاپلوتایپی بالا نظیر ذرت و آفتابگردان مورد نیاز است (Rostoks *et al.*, 2006). مکان‌یابی ارتباطی کل ژنوم یک روش جامع برای بررسی

[9] Positional candidate
[10] Functional candidate
[11] Trait specific system

سیستماتیک ژنوم جهت شناسایی عوامل ژنتیکی دخیل در صفات می‌باشد. در این روش تعداد زیادی نشانگر برای شناسایی ارتباط آنها با صفات پیچیده مختلف تست شده، و اطلاعات قبلی در مورد ژن‌های کاندید مورد نیاز نمی-باشد. این روش پژوهش از لحاظ تکنیک، دقت و هزینه، بهترین است.

مقایسه مکان‌یابی ارتباطی ژن کاندید و مکان‌یابی ارتباطی در سطح ژنوم

مطالعات مکان‌یابی ارتباطی با استفاده از تعداد زیاد SNP، تعداد نمونه بالا و حداقل ساختار جمعیت، در بررسی دقیق صفات پیچیده بسیار مهم و پراهمیت است. مطالعات ارتباطی ژن کاندید تنها بخش کوچکی از ژنوم را بررسی می‌کند. با پیشرفت تکنولوژی ژنومیکس قطعا انتظار می‌رود که مطالعات تجزیه ارتباط در سطح ژنوم در گونه‌های گیاهی به مراتب بیشتر صورت گیرد. تاکنون نتایج موفقیت‌آمیز بسیار کمی از مکان‌یابی ارتباطی ژن کاندید بدست آمده است. اما در برخی تحقیقات، شروع کار با توالی‌های ژن کاندید و نشانگرهای مولکولی، درک بالایی از ساختار جمعیت، روابط خانوادگی، تنوع نوکلئوتیدی، کاهش LD و سایر جنبه‌های مکان‌یابی ارتباطی را فراهم نمود. دلیل دیگر موفقیت‌های محدود روش ژن کاندید، روش انتخاب ژن‌های کاندید است. بدیهی است برخی ژن‌های کاندید از طریق مقایسات بین لاین‌های جهش‌یافته و وحشی شناسایی می‌شوند. درک عمیقی از اثرات طبیعی آلل‌ها در چنین مکان‌هایی وجود ندارد. حتی اگر آلل بدون عملکرد مشخصی باعث تغییر فنوتیپ مشخص شود، تنها می‌توان انتظار داشت که جهش‌های ملایم اثر مختصری بر فنوتیپ داشته باشند. هم فراوانی و هم اثر آلل، بر قابل تشخیص بودن تنوع ایجاد شده توسط یک مکان اثر می‌گذارند. فراوانی اریب‌دار آلل، تشخیص پیوستگی را مشکل می‌سازد حتی اگر چندشکلی ژن کاندید واقعا مسئول تنوع فنوتیپی صفت باشد (شکل ۳) (Zhu *et al.*, 2008).



شکل ۳- دیاگرام مقایسه شماتیک روش‌های مکان‌یابی ارتباطی کل ژنوم و مکان‌یابی ارتباطی ژن کاندید (E: واریانس باقیمانده) (Zhu *et al.*, 2008).

Figure 3. Schematic diagram and contrast of genome-wide association mapping and candidate-gene association mapping (E: stands for residual variance) (Zhu *et al.*, 2008).

روش‌های تجزیه آماری تنوع ژنتیکی در مکان‌یابی ارتباطی

روش‌های اساسی برای مکان‌یابی ارتباطی تحت شرایط ایده‌آل، شامل رگرسیون خطی، تجزیه واریانس ANOVA، t-test و یا آزمون مربع chi (x^2) است. البته از آنجا که ساختار جمعیت می‌تواند روابط ژنوتیپی- فنوتیپی کاذب ایجاد نماید، روش‌های آماری متفاوتی برای مدیریت این عامل مختل کننده طراحی شده‌اند. در مطالعه مبنای ژنتیکی بیماری‌های انسانی که خانواده‌ها به‌عنوان جمعیت محسوب می‌شوند از روش تست عدم تعادل انتقالی[۱۲] (TDT) استفاده می‌شود (Abecasis *et al.*, 2000). در روش‌های مبتنی بر جمعیت برای مدیریت پدیده ساختار جمعیت که باعث اختلال در روند تجزیه و ارائه نتایج مثبت کاذب می‌شود، دو روش کنترل ژنومی (GC[13]) و ارتباط ساختاری (SA[14]) در مطالعات ارتباطی در انسان و گیاهان بکار می‌روند. در روش کنترل ژنومی یک دسته از نشانگرهای تصادفی برای تخمین اثر ساختار جمعیت بر آماره‌های آزمون مکان‌یابی ارتباطی استفاده می‌شود، به طوریکه معنی‌داری آماره ارتباطی تخمین زده شده برای ساختار جمعیت تصحیح می‌شود. در مقابل، تجزیه ارتباط ساختاری در ابتدا از یک سری نشانگر برای ارزیابی ساختار جمعیت (Q) استفاده می‌کند و سپس این برآورد در مراحل بعدی با تجزیه‌های آماری در نظر گرفته می‌شود (Falush *et al.*, 2003) . اخیرا روش مدل ترکیبی[۱۵](MMA) برای مکان‌یابی ارتباطی توسعه یافته است (Yu *et al.*, 2006). در این روش نشانگرهای تصادفی برای محاسبه ماتریس Q و ماتریس خویشاوندی K[16] استفاده می‌شوند. از آنجا که روش‌های MMA مرزهای بین نمونه‌های مبتنی بر خانواده و نمونه‌های مبتنی بر جمعیت را در هم می‌شکنند، مکمل نیرومندی برای روش‌های کنونی در مکان‌یابی ارتباطی محسوب می‌شود (Zhao *et al.*, 2007).

تجزیه به مولفه‌های اصلی[۱۷] مدت‌هاست که در مطالعه تنوع ژنتیکی به‌کار می‌رود و به‌عنوان یک روش سریع و موثر در تشخیص ساختار جمعیت استفاده می‌شود. تجزیه به مولفه‌های اصلی، تنوع مشاهده شده در بین همه نشانگرها را در تعداد کمتری مولفه خلاصه می‌کند. این مولفه‌های اصلی به‌عنوان زیر جمعیت‌های مجزای غیر قابل مشاهده در نظر گرفته می‌شود. قرارگیری هر فرد در هر مولفه اصلی، عضویت در جمعیت یا جد هر فرد را توصیف می‌کند. جایگزینی Q با تجزیه به مولفه‌های اصلی در مدل ترکیبی بسیار مفید است. اما پژوهش‌های بیشتری جهت تثبیت آن در گونه‌های زراعی مورد نیاز است (Weber *et al.*, 2008).

اندازه نمونه در مکان‌یابی ارتباطی کوچک است. در بسیاری از مطالعات مکان‌یابی ارتباطی تنها حدود ۱۰۰ لاین (فرد) بررسی می‌شود. برای توضیح این مسئله در بحث تنوع ژنتیکی یک جمعیت، باید اندازه نمونه تجزیه

[12] Transmission disequilibrium test
[13] Genomic control
[14] Structured analysis

[15] Mixed-model approach
[16] Kinship Matrix
[17] Principal component analysis

پیوستگی و مکان‌یابی ارتباطی مقایسه شود. اندازه نمونه در بسیاری از مطالعات تجزیه پیوستگی در گیاهان حدود ۲۵۰ فرد (F2, BC1, RIL…) با زمینه ژنتیکی دو والدی و یکنواخت است. شناسایی پیوستگی‌های نشانگر- صفت با یک جمعیت کوچک مشکل خواهد بود بجز اینکه مکان عملکردی یک اثر بسیار بزرگ داشته باشد و LD بالایی بین نشانگرهای تست شده با این مکان وجود داشته باشد البته بدون توجه به اینکه روش ژن کاندید برای مکان یابی ارتباطی استفاده می‌شود یا روش اسکن ژنوم. شبیه‌سازی اولیه با استفاده از اطلاعات تجربی حاصل از ذرت نشان داد که یک نمونه با اندازه بزرگ برای داشتن تشخیص ژنی قوی نیاز است (Zhu *et al.*, 2008). تعداد نشانگرهای مورد نیاز برای برآورد دقیق روابط ژنتیکی نیز مهم می‌باشد که به‌وسیله مطالعات مکان‌یابی ارتباطی ژن کاندید به‌دست می‌آید. طول کروموزوم، تنوع گونه، تنوع نمونه، هزینه و در دسترس بودن سیستم‌های نشانگری مختلف بر تعداد نشانگرهای مورد استفاده در یک مطالعه اثر دارد (Zhu *et al.*, 2008).

روش‌شناسی مطالعات ارتباطی در گیاهان جهت بررسی صفات

انواع مختلفی از روش‌ها برای مطالعات مکان‌یابی ارتباطی در انسان در حد وسیعی به‌کار رفته‌اند که اکثر آنها بدون تغییر یا با تغییرات جزئی در سطح وسیعی از ارگانیسم‌ها به‌ویژه گیاهان قابل اجرا هستند. نهایتا برخی از روش‌هایی که مکان‌یابی ارتباطی قوی، دقیق و بدون اریب را در گیاهان انجام می‌دهند انتخاب شده‌اند (Abdurakhmonov & Abdukarimov, 2008).

روش کلاسیک مکان‌یابی ارتباطی، روش CC[18] است که ژن‌های مسئول[19] (موثر) را با مقایسه فراوانی آللی در یک نمونه تحت تاثیر قرار گرفته (به‌عنوان تستر) و یک نمونه سالم به‌عنوان کنترل شناسایی می‌کند. در این روش برای انجام مکان‌یابی دقیق، بایستی اندازه نمونه‌ها در تستر و کنترل برابر باشد و همچنین نمونه‌ها دارای ساختار نباشد و بین افراد نمونه ارتباط خویشاوندی وجود نداشته باشد. آزمون مربع x^2 پیرسون، آزمون دقیق فیشر یا تصحیح پیوستگی یتس برای مقایسه فراوانی‌های آللی و شناسایی پیوستگی بین فنوتیپ بیماری و نشانگر استفاده می‌شوند. اگرچه افراد نمونه به‌طور تصادفی از جمعیت انتخاب می‌شوند، اما این افراد یک نماینده ایده‌آل و متعادل از تسترها و کنترل‌ها نیستند چرا که در جمعیت معمولا تسترها کم هستند. بنابراین نیاز به انتخاب ویژه تسترها می‌باشد. روش CC شدیدا توسط ساختار و طبقه‌بندی جمعیت تحت تاثیر قرار می‌گیرد (Schulze *et al.*, 2002). جهت کاهش میزان این تاثیر روش‌های مختلفی معرفی شده‌اند؛ از جمله روش HRR[20] که میزان این تاثیر را کاهش می‌دهد اما اثر آن را حذف نمی‌کند. در این روش ابتدا یک گروه شبه کنترل[21] ایجاد می‌شود که حامل دو آللی هستند که به نتاج متاثر منتقل نشده‌اند. سپس

[18] Case-control
[19] Causative Gene
[20] Haplotype relative risk
[21] Pseudo-control group

فراوانی آللی نشانگر در گروه‌های تستر و شبه کنترل تجمیع می‌شوند. برای حذف موثر اثرات نامطلوب حاصل از ساختار جمعیت، روش TDT معرفی شده که انتقال و عدم انتقال آلل‌های نشانگر به نتاج متاثر را به‌وسیله آزمون x^2 با فرض وجود پیوستگی بین نشانگر و صفت مقایسه می‌کند. طراحی TDT نیاز به تعیین ژنوتیپ نشانگرهای حاصل از سه فرد دارد: یک والد هتروزیگوس، یک والد هموزیگوس و یک نتاج متاثر. البته روش HRR در نمونه بدون ساختار به علت قدرت زیاد در حذف کامل‌تر پیوستگی کاذب بهتر از TDT عمل می‌کند (Abdurakhmonov & Abdukarimov, 2008).

مدل‌های خطی سازگار با مطالعات ارتباطی کل ژنوم

اگرچه مطالعات مکان‌یابی ارتباطی در سطح کل ژنوم پتانسیل بالایی در شناسایی دقیق چندشکلی‌های ژنتیکی مسئول بیماری‌های انسان و صفات مهم زراعی دارد، اما نتایج اشتباه یک مشکل عمده در این روش بوده و تا حدی پیوستگی‌های نادرستی را به علت ساختار جمعیت و روابط نامتعادل میان افراد در نمونه انتخابی نشان می‌دهد (Abiola, 2003). طبقه‌بندی جمعیت در ابتدا بوسیله روش‌های مبتنی بر GLM نظیر SA و GC و آزمون‌های پیوستگی مبتنی بر خانواده بررسی می‌شد، اما با معرفی روش MLM، روش بهبود یافته‌ای برای محاسبه همزمان ساختار جمعیت و روابط نامتعادل میان افراد حاصل شد (Yu *et al.*, 2006).

در روش‌های مبتنی بر MLM، ساختار جمعیت به عنوان یک اثر ثابت در نظر گرفته می‌شود، در حالی‌که خویشاوندی میان افراد بصورت اثر تصادفی افراد لحاظ می‌شود. صرف نظر از روش آماری مورد استفاده، روش GWAS به نمونه‌هایی با اندازه بزرگ نیاز دارد تا به قدرت آماری قوی دست یابد، به‌ویژه به‌منظور شناسایی چندشکلی‌های کوچک اثر که مسئول صفات پیچیده هستند (Buckler, 2009). در روش MLM، اندازه نمونه‌های بزرگ، یک حجم محاسباتی سنگینی ایجاد می‌نماید زیرا زمان محاسبه برای انجام MLM متناسب با توان سوم تعداد افراد نمونه افزایش می‌یابد. یکی از روش‌های کاهش اندازه اثر تصادفی در MLM روش مدل Sire است که در اصلاح دام استفاده شده است. در این روش اثر ژنتیکی هر فرد با اثر جد خود جایگزین می‌شود، بنابراین این روش الزاما به شجره‌نامه نیاز دارد که در اغلب مطالعات بویژه مطالعات گیاهی همیشه در دسترس نیست، حتی با در دسترس بودن شجره‌نامه نیز استفاده از خویشاوندی مبتنی بر نشانگر به‌علت دقت و صحت بالاتر آن ترجیح داده می‌شود (Zhang *et al.*, 2010). زمان کل محاسبه GWAS با روش MLM استاندارد، برابر mpn^3 است که m تعداد کل نشانگرها، p تعداد تکرارهای لازم برای انجام MLM و n تعداد کل افراد مفروض می‌باشد. روش‌هایی برای کاهش اندازه اثر ژنتیکی تصادفی در غیاب اطلاعات شجره‌نامه و حذف تکرارها برای محاسبه پارامترهای جمعیت برای هر نشانگر، بدون کاهش قدرت آماری وجود دارد. در یکی از این روش‌ها تحت عنوان Compression تعداد افراد n با عدد کوچکتر تعداد گروه‌ها s ($s \leq n$) جایگزین می‌شود و بر مبنای خویشاوندی میان

افراد کلاستربندی می‌شود. متعاقبا خویشاوندی بین جفت گروه‌ها جایگزین خویشاوندی بین جفت افراد برای اثر تصادفی MLM می‌گردد. c=n/s متوسط تعداد افراد در هر گروه بوده و به عنوان سطح تراکم در نظر گرفته می‌شود و در این روش زمان محاسبه به نسبت c^3 کاهش می‌یابد. در این روش اثر تصادفی از فرد به گروه فشرده شده است و بدین علت این نوع MLM را MLM فشرده می‌نامند. MLM فشرده در مرز MLM و GLM قرار دارد، زیرا هر دو می‌توانند به عنوان سطح نهایی MLM فشرده مطرح شوند. MLM زمانی می‌تواند در حکم MLM فشرده بیان شود که هر فرد به عنوان یک گروه در نظر گرفته شود (s=n). روش GLM نیز وقتی همه افراد در یک گروه قرار گیرند (s=1) به عنوان یک MLM فشرده محسوب می‌شود. مورد دوم باعث می‌شود که اثر تصادفی یک سطح داشته باشد و از برآورد مجزای اثر تصادفی و واریانس باقیمانده جلوگیری کند. بعلاوه اثر تصادفی و متوسط کل ارتباط خطی داشته و نمی‌توان بطور مجزا محاسبه نمود (Zhang *et al.*, 2010).

امروزه روش آماری MLM به دلیل امکان گنجاندن ماتریس ساختار جمعیت و ماتریس خویشاوندی به‌عنوان کوواریت در مدل جهت کاهش خطای نوع اول (خطای مربوط به ارتباط دروغین نشانگر-صفت) به‌طور گسترده‌ای برای تجزیه ارتباط در گیاهان استفاده می‌شود. در حالی‌که در مدل GLM تنها ساختار جمعیت در نظر گرفته می‌شود (Alipour & Darvishzadeh, 2019). یو و بوکلر (Yu & Buckler, 2006) در شناسایی نشانگرهای پیوسته با تغییرات فنوتیپی در صفات آگرومورفولوژیک ذرت از مدل MLM به‌منظور بهبود نتایج و کاهش نتایج مثبت دروغین استفاده نمودند و بر اساس اظهار محققین نتایج دقیق‌تری در مقایسه با مدل GLM به‌دست آمد.

احمد و همکاران (Ahmed *et al.*, 2021) ۱۳۸ گندم بهاره را به‌همراه دو ژنوتیپ شاهد به‌نام‌های Wesley (حساس به خشکی و دیگری Harry (متحمل به خشکی) جمع‌آوری شده از ۱۹ کشور مختلف جهان برای چهار صفت فیزیولوژیک تحت تنش خشکی در شرایط گلخانه‌ای ارزیابی کردند. پویش کل ژنوم بر اساس ۴۰۷ نشانگر DArT-seq با استفاده از مدل خطی عمومی و مدل خطی مخلوط منجر به شناسایی ۳۹۸ نشانگر پیوسته با ژن‌های کنترل کننده صفات مورد بررسی گردید. گروه-بندی ژنوتیپ‌ها بر اساس داده‌های مولکولی و با استفاده از تجزیه خوشه‌ای مبتنی بر فاصله و مدل و تجزیه به مولفه-های اصلی، آن‌ها به دو گروه منتسب کرد. نتایج نشان داد که ژنوتیپ‌ها به‌وسیله منشا جغرافیایی‌شان از هم تفکیک نشدند، این نتایج نشان داد که روابط بین ژنوتیپ‌ها به-شدت به‌وسیله تنوع درونی هر منطقه جغرافیایی متاثر می‌شود. در نتیجه کشورهای یا مناطقی که تنوع نسبتا زیادی دارند، به‌آسانی گروه‌بندی نمی‌شوند، چون فاصله ژنتیکی‌شان زیاد است. همچنین نتایج نشان داد که نه ژنوتیپ می‌تواند به‌عنوان ژنوتیپ‌های کاندید جهت بهبود تحمل به خشکی در گندم‌های بهاره مورد استفاده واقع شوند. بیشترین و کمترین تعداد نشانگر و تنوع مولکولی به‌ترتیب مربوط به ژنوم‌های D و A بود. آن‌ها چهار QTL

با تبیین فنوتیپی ۸/۷-۵/۴ درصد بر روی کروموزوم‌های 3D، 2B و 1A شناسایی کردند. برای صفت تعداد روز تا پژمردگی فقط یک QTL با تبیین فنوتیپی ۸/۹ درصد روی کروموزوم 4D، برای طول گیاهچه چهار QTL با تبیین فنوتیپی ۵/۳-۴/۲ درصد بر روی کروموزوم‌های 4A و 7D و برای صفت شاخص تحمل به خشکی ۱۲ QTL با تبیین فنوتیپی ۹/۷-۴/۸ درصد بر روی کروموزوم‌های 4A، 2B، 3D و 7D مکان‌یابی شد. دوازده نشانگر با بیش از یک صفت پیوسته بودند، که نشانگر WPT-2356 در بین تمامی صفات مورد مطالعه مشترک بودند و پتانسیل استفاده در گزینش به‌کمک نشانگر را دارد. به نظر می‌رسد مدل MLM برای مکان‌یابی ارتباطی مدل قابل اطمینانی باشد. در مطالعه مالانا و همکاران (Maulana *et al.*, 2020) با استفاده از مدل خطی مخلوط جهت مکان‌یابی QTL و تشخیص نشانگرهای SNP پیوسته با صفات همبسته با تحمل به تنش خشکی در مرحله گیاهچه‌ای نشان دادند که چندین QTL در ارتباط با صفات مرتبط به تحمل به خشکی در مرحله گیاهچه‌ای بر روی کروموزوم-های 1B، 2A، 2B، 2D، 3A، 3B، 3D، 4B، 5A، 5B، 6B و 7B شناسایی شدند. از این تعداد QTL، ۱۲ QTL پایدار (شامل دو QTL برای محتوای کلروفیل برگ، چهار QTL برای فلورسانس کلروفیل برگ، چهار QTL برای طول ساقه و دو QTL برای تعداد برگ هر گیاهچه) در پاسخ به تنش خشکی برای صفات مورد مطالعه تشخیص داده شد.

نرم‌افزارهای تجزیه اطلاعات در مکان‌یابی ارتباطی

نرم‌افزارهای مختلف برای تجزیه اطلاعات در مکان‌یابی ارتباطی استفاده می‌شود. نرم‌افزار TASSEL (Bradbury *et al.*, 2007) یک نرم‌افزار پرکاربرد برای مکان‌یابی ارتباطی در گیاهان است و براساس روش‌های جدید آماری به‌روز شده است. علاوه بر روش‌های تجزیه ارتباط (مانند رگرسیون توزیعی، مدل خطی و مدل ترکیبی)، TASSEL همچنین برای محاسبه و نمایش گرافیکی LD و جستجوی اطلاعات ژنوتیپی و فنوتیپی بکار می‌رود. ماتریس خویشاوندی K توسط نرم‌افزار TASSEL نیز محاسبه می‌شود. ماتریس K یک ماتریس n×n است. عناصر قطری K برای افراد اینبرد، یک، و برای افراد غیر اینبرد $0.5(1+F_x)$ است که F_x ضریب خویش‌آمیزی می‌باشد. این نرم‌افزار مکان‌یابی ارتباطی را بر مبنای هر دو مدل خطی عمومی (GLM) و مدل خطی مخلوط (MLM) انجام می‌دهد. اگر از مدل خطی عمومی برای تعیین ارتباط نشانگر-صفت استفاده شود، فقط ماتریس Q یا تجزیه به مولفه‌های اصلی (PCA) که نشان‌دهنده ساختار جمعیت است، در محاسبات وارد می‌شوند و روابط خویشاوندی بین افراد در نظر گرفته نمی‌شود. نقشه‌یابی ارتباطی با مدل 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.

نرم‌افزار TASSEL اطلاعات جامعی در مورد مکان‌ها و نشانگرها مانند *P-value*، R^2، F-marker و غیره ارائه

می‌نماید. کلادوگرام‌های افراد را براساس روش‌های Neighbor joining و UPGMA رسم نموده و نمودار دوبعدی LD-plot را بر مبنای داده‌های D'، R^2 و *P-value* رسم می‌نماید.

نرم‌افزار STRUCTURE (Pritchard *et al.*, 2000) معمولا برای محاسبه ماتریس Q بکار می‌رود. ماتریس Q یک ماتریس n×p است که n تعداد افراد و p تعداد زیرجمعیت‌های تعریف شده است. ماتریس Q نشان‌دهنده ساختار جمعیتی موجود در افراد مورد مطالعه و زیرگروه‌های احتمالی می‌باشد. این نرم‌افزار بر مبنای ماتریس Q، ماتریس سهم عضویت را ارائه می‌دهد که مشخص‌کننده درصد قرارگیری جمعیت‌ها در کلاسترها می‌باشند. به‌علاوه برای هر زیرگروه شاخص تثبیت Fst نیز برآورد می‌گردد که بیانگر میزان جریان ژنی بین زیر گروه‌ها و قرابت و فاصله بین زیر گروه‌ها می‌باشد. نرم‌افزار EINGENSTRAT (Price *et al.*, 2006) برای برآورد مولفه‌های اصلی (PCs) اطلاعات نشانگر و تست صحت نتایج آماری حاصل از طبقه‌بندی جمعیت بکار می‌رود. نرم‌افزار R (R Core Team, 2013) شامل مجموعه گسترده‌ای از انواع آنالیزهای آماری بوده و اغلب توسط پژوهشگران آشنا به مهارت‌های برنامه‌نویسی استفاده می‌شود.

یکی دیگر از نرم‌افزارهای مناسب برای گیاهان، GAPIT است که مبتنی بر R می‌باشد (Lipka *et al.*, 2012) و می‌تواند تجزیه ارتباط 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 (جدول ۲).

وضعیت کنونی و چشم انداز مکان‌یابی ارتباطی

عدم تعادل پیوستگی به‌طور گسترده‌ای در مورد اهداف مختلفی نظیر مکان‌یابی QTLهای بیماری در انسان به‌کار گرفته شده است، و این روند در مورد گیاهان زراعی به‌ویژه غلات ادامه‌دار است. با در دسترس قرار گرفتن نقشه‌های با تراکم بالا در برخی گیاهان زراعی، توالی‌های کامل ژنومی در گیاهان مدل مانند آرابیدوپسیس و برنج، و نیز توالی نواحی غنی از ژنوم در محصولاتی نظیر سورگوم، کلزا و گندم، استفاده از LD و مکان‌یابی ارتباطی در دهه‌های اخیر و در جمعیت‌های گیاهی به‌ویژه غلات رو به افزایش است. این روش در ژنوم گیاهان مختلفی برای مطالعه ارتباط نشانگر-صفت و نیز مطالعه ژنتیک و تکامل جمعیت در هر دو حالت طبیعی و اهلی شده مورد استفاده قرار خواهد گرفت. روش سنتی گزینش به کمک نشانگر در غلات با موفقیت برای انتقال صفات عملکردی و مقاومت به بیماری‌ها به مواد اصلاحی به‌کار گرفته شده است. به کمک نشانگرهای DNA می‌توان روش انتخاب به کمک نشانگر (MAS) را برای اصلاح گیاهان زراعی به‌ویژه برای صفات با وراثت‌پذیری پایین (صفات کمی پیچیده مثل عملکرد) به‌کار گرفت (Nadeem *et al.*, 2018) با این حال، اجرای این روش مستلزم تهیه نقشه‌های پیوستگی نشانگرهای مولکولی برای جمعیت‌های مختلف ژنتیکی است. هم‌چنین درک عمل، تنظیم و بیان و همسانه‌سازی ژن‌ها نیز مستلزم تهیه نقشه و توالی‌یابی ژنوم گیاهان

می‌باشد. به کمک روش انتخاب به کمک نشانگر می‌توان با حذف اثرات محیطی، پلیوتروپی یا اپیستاتیک، گیاهان مطلوب را انتخاب و یا صفات آن‌ها را به صورت غیرمستقیم اصلاح کرد. به‌کارگیری این روش در اصلاح گیاهان زراعی مستلزم نشانمند کردن ژن هدف به معنی معرفی یک نشانگر پیوسته به ژن هدف، طراحی آزمایشی مناسب، انجام آنالیزهای نسبتا پیچیده آماری، دسترسی معمول به اطلاعات نشانگر DNA، روش‌های نقشه‌برداری و نیروی انسانی متخصص می‌باشد (Platten *et al.*, 2019).

جدول ۲– فهرست نرم‌افزارهای پرکاربرد در مکان‌یابی ارتباطی (اقتباس از Zhu *et al.*, 2008).

Table 2. List of the software's widely used in association analysis (from Zhu *et al.*, 2008).

نرم افزار Software	عملکرد Yield	وب سایت Web site
TASSEL	مکان‌یابی ارتباطی Association analysis	http://www.maizegenetics.net
R	عمومی General	http://www.r-project.org
GAPIT	عمومی General	http://www.maizegenetics.net/gapit
Structure	ساختار جمعیت Structure analysis	http://pritch.bsd.uchicago.edu/structure.html
SPAGeDi	روابط خویشاوندی Kinship	http://www.ulb.ac.be/sciences/ecoevol/spagedi.html
EINGENSTRAT	PCA، مکان‌یابی ارتباطی Association analysis	http://genepath.med.harvard.edu/~reich/Software.htm
MTDFREML	مدل ترکیبی Mix model	http://aipl.arsusda.gov/curtvt/mtdfreml.html
ASREML	مدل ترکیبی Mix model	http://www.vsni.co.uk/products/asreml

نتیجه‌گیری

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. مطالعات آینده در زمینه LD در گیاهان زراعی ساختارهای ژنومی گیاهان را مشخص‌تر می‌کند و همچنین کاربرد MAS و کلونینگ مبتنی بر نقشه را در مورد ژن‌های مسئول صفات پیچیده تسهیل می‌نماید.

References

Abdallah, J. M., Goffinet, B., Cierco-Ayrolles, C., & Perez-Enciso, M. 2003. Linkage disequilibrium fine mapping of quantitative trait loci: A simulation study. Genetics Selection Evolution, 35, 513–532.

Abdollahi Mandoulakani, B., Rahmanpour, S., Shaaf, S., Gholamzadeh Khoei, S., Rastgou, M., & Rafezi, R. 2015. Towards the identification of retrotransposon-based and ISSR molecular markers associated with populations resistant to ZYMV in melon. South African Journal Botany, 100, 141-147.

Abdollahi Mandoulakani, B., Azizi, H., Piri, Y., Rahmanpour, S., & Hasani, L. 2015. Association analysis for morphological traits in alfalfa (*Medicago sativa* L.) using molecular markers. Journal of Crop Breeding, 8(19), 52-60 (In Persian).

Abdollahi Mandoulakani, B., Nasri, S. H., Dashchi, S., Arzhang, S., Bernousi, I., & Abbasi Holasou, H. 2017. Preliminary evidence for associations between molecular markers and quantitative traits in a set of bread wheat (*Triticum aestivum* L.) cultivars and breeding lines. Comptes rendus biologies, 340, 307-313.

Abdurakhmonov, A. Y., & Abdukarimov, A. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. International Journal of Plant Genomics, https://doi.org/10.1155/2008/574927

Abecasis, G. R., Cardon, L. R., & Cookson, W. O. 2000. A general test of association for quantitative traits in nuclear families. American Journal of Human Genetics, 66, 279-292.

Abiola, O. 2003. The nature and identification of quantitative trait loci: a community's view. Nature Review Genetic, 4, 911-916.

Agrama, H. A., Eizenga, G. C., & Yan, W. 2007. Association mapping of yield and its components in rice cultivars. Molecular Breedings, 19, 341-356.

Ahmed, A. A. M., Mohamed, E. A., Hussein, M. Y., & Sallam, A. 2021. Genomic regions associated with leaf wilting traits under drought stress in spring wheat at the seedling stage revealed by GWAS. Environmental and Experimental Botany, 184, 104393.

Alipour, H., & Darvishzadeh, R. 2019. Association mapping of quantitative traits in molecular cereal breeding. Cereal Research, 9(3), 271-298. https://doi.org/10.22124/cr.2019.14333.1518. (In Persian).

Balding, D. J. 2006. A tutorial on statistical methods for population association studies. Nature Publishing Group, 7, 781-791.

Belo, A., Zheng, P., Luck, S., Shen, B., Meyer, D. J., Li, B., Tingey, S., & Rafalski, A. 2008. Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. Molecular Genetics and Genomics, 279, 1-10.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23, 2633-2635.

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

Breseghello, F., & Sorrells, M. E. 2006a. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Science, 46, 1323–1330.

Breseghello, F., & Sorrells, M. E. 2006b. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics, 172, 1165-1177.

Buckler, E. S. 2009. The genetic architecture of maize flowering time. Science, 325, 714-718.

Chen, X., Min, D., Yasir, T. A., & Hu, Y. G. 2012. Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. PLOS One, 7, e44510. https://doi.org/10.1371/journal.pone.0044510.

DeWan, A., Liu, M., Hartman, S., & Zhang, S. S. M. 2006. HTRA1 promoter polymorphism in wet age-related macular degeneration. Science, 314, 989-992.

Dolati Baneh, H., mohammadi, S. A., Abdollahi Mandoulakani, B., & Rahmanpour, S. 2014. Association analysis for morphological traits in grapevine using SSR and AFLP markers. Iranian Journal of Agricultural Biotechnology, 1, 46-60 (In Persian).

Ehrenreich, I. M., Stafford, P. A., & Purugganan, M. D. 2007. The genetic architecture of shoot branching in Arabidopsis thaliana: A comparative assessment of candidate gene associations vs. quantitative trait locus mapping. Genetics, 176, 1223-1236.

Falush, D., Stephens, M., & Pritchard, J. K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics, 164, 1567-1587.

Flint-Garcia, S. A., Thrnsberry, J. M., & Buckler, E. S. 2003. Structure of linkage disequilibrium in plants. Annal Review of Plant Biology, 54, 357-374.

Gahlaut, V., Jaiswal, V., Tyagi, B. S., Singh, G., Sareen, S., & Balyan, H. S. 2017. QTL mapping for nine drought-responsive agronomic traits in bread wheat under irrigated and rain-fed environments. PLoS One, 12, e0182857. https://doi.org/10.1371/journal.pone.0182857.

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

Hackett, C. A. 2002. Statistical methods of QTL mapping in cereals. Plant Molecular Biology, 48, 585-599.

Hedrick, P. W. 1987. Gametic disequilibrium measures: proceed with caution. Genetics, 117, 331-341.

Hill, W. G., & Robertson, A. 1968. Linkage disequilibrium in populations. Theorical and Applied Genetics, 38, 226-231.

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

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

Kebede, H., Subudhi, P. K., Rosenow, D. T., & Nguyen, H. T. 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theoretical and Applied Genetics, 103, 266-276. https://doi.org/10.1007/s001220100541.

Khalid, M., Afzal, F., Gul, A., Amir, R., Ahmed, Z., Mahmood, Z., Rasheed, A., & He, Z. 2019. Molecular characterization of 87 functional genes in wheat diversity panel and their association with phenotypes under well-watered and water-limited conditions. Frontiers in Plant Science, 10, 717. https://doi.org/10.3389/fpls.2019.00717.

Khalid, M., Gul, A., Amir, R., Ali, M., Afzal, F., & Quraishi, U. 2018. QTL mapping for seedling morphology under drought stress in wheat cross synthetic (W7984)/Opata. Plant Genetic Resources, 16, 359-366.

Khan, M. I., Kainat, Z., Maqbool, S., Mehwish, A., Ahmad, S., Suleman, H. M., Mahmood, Z., Ali, M., Aziz, A., Rasheed, A., & Li, H. 2022. Genome-wide association for heat tolerance at seedling stage in historical spring wheat cultivars. Frontiers in Plant Science, https://doi.org/10.3389/fpls.2022.972481.

Kraakman, A. T. W., Martínez, F., Mussiraliev, B., Eeuwijk, F. A., & Niks, R. E. 2006. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Molecular Breeding, 17, 41-58.

Mackay, T. F., Stone, E. A., & Ayroles, J. F. 2009. The genetics of quantitative traits: Challenges and prospects. Nature Reviews Genetics, 10, 565-577.

Maulana, F., Huang, W., Anderson, J. D., & Ma, X-F. 2020. Genome-wide association mapping of seedling drought tolerance in winter wheat. Frontiers in Plant Science, 11,573786. https://doi.org/10.3389/fpls.2020.57378.

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., & Stein, N. 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature, 491(7426), 711-716.

Meuwissen, T.H.E., & Goddard, M.E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics, 155, 421-430.

Nadeem, M. Z., Nawaz, M. N., Shahid, M. Q., Dogan, Y., Comertpay, G., Yıldız, M., Hatipoglu, R., Ahmad, F., Alsaleh, A., Labhane, N. O., Zkan, H., Chung, G., & Baloch, F. S. 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology and Biotechnological Equipment, 32(2), 261-285.

Nielsen, N. H., Backes, G., Stougaard, J., & Jahoor, A. 2014. Genetic diversity and population structure analysis of European hexaploid bread wheat (*Triticum aestivum* L.) varieties. PLos One, https://doi.org/10.1371/journal.pone.0094000.

Ogrodowicz, P., Mikolajczak, K., Kempa, M., Mokrzycka, M., Krajewski, P., & Kuczynska, A. 2023. Genome-wide association study of agronomical and root-related traits in spring barley collection grown under field conditions. Frontiers in Plant Science, https://doi.org/10.3389/fpls.2023.1077631.

Oraguzie, N. C., Wilcox, P. L., Rikkerink, E. H. A., & De Silva, H. N. 2007. Linkage disequilibrium, Association mapping in plants. Springer New York USA, 11-39.

Pasam, R. K., Sharma, R., Malosetti, M., & Van Eeuwijk, F. A. 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biology, 12, 16.

Platten, J. D., Cobb, J. N., & Zantua, R. E. 2019. Criteria for evaluating molecular markers: comprehensive quality metrics to improve marker -assisted selection. PLoS One, 14(1), e0210529.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38, 904-909.

Pritchard, J. K., Stephens, M., & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Qaseem, M. F., Qureshi, R., Muqaddasi, Q. H., Kousar, R., & Roder, M. S. 2018. Genome-wide association mapping in bread wheat subjected to independent and combined high temperature and drought stress. Plos One, https://doi.org/10.1371/journal.pone.0199121.

Qu, Y. Y., Mu, P., Li, X. Q., Tian, Y. X., Wen, F., & Zhang, H. L. 2008. QTL mapping and correlations between leaf water potential and drought resistance in rice under upland and lowland environments. Acta Agronomica Sinica, 34, 198-206. https://doi.org/10.1016/S1875-2780(08)60008-5.

R Core Team 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.r-project.org/

Rao Kovi, R., Fjellheim, S., Larsen, A., Rudi, H., Asp, T., Peter Kent, M., & Rognli, O. A. 2015. Population structure, genetic variation, and linkage disequilibriumin perennial ryegrass populations divergently selected for freezing tolerance. Frontiers in Plant Science, https://doi.org/10.3389/fpls.2015.00929

Rafalski, A., & Morgante, M. 2004. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends in Genetic, 20, 103-111.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. National Academic Science USA, 98, 11479-11484.

Rostoks, N., Ramsay, L., Mackenzie, K., Cardle, L., Bhat, P. R., Roose, M. L., Svensson, J. T., Stein, N., Varshney, R. K., Marshall, D. F., Graner, A., Close, T. J., & Waugh, R. 2006. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc National Academic Science, USA.

Safdar, L., Bin, A., Ndleeb, T., Latif, S., Umer, M. J., Tang, M., Li, X., Liu, Sh., & Quraishi, U. M. 2020. Genome-wide association study and QTL meta-analysis identified novel genomic loci controlling potassium use efficiency and agronomic traits in bread wheat. Frontiers in Plant Science, 11, 1-14 https://doi.org/ 10.3389/fpls.2020.00070

Schulze, T. G., & 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.

Semagn, K., Bjornstad, A., & Xu, Y. 2010. The genetic dissection of quantitative traits in crops. Electronic Journal of Biotechnology, 13, 1-45.

Stich, B., Yu, J., Melchinger, A. E., Piepho, H. P., Utz, H. F., Maurer, H. P., & Buckler, E. S. 2007. Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. Genetics, 176, 563-570.

Sukumaran, S., Reynolds, M. P., & Sansaloni, C. 2018. Genome-wide association analyses identify QTL hotspots for yield and component traits in durum wheat grown under yield potential, drought, and heat stress environments. Frontiers in Plant Science, https://doi.org/10.3389/fpls.2018.00081.

Weber, A., Clark, R. M., Vaughn, L., Sánchez-Gonzalez, J. D. J., Yu, J., Yandell, B. S., Bradbury, P., & Doebley, J. F. 2008. Major regulatory genes in maize contribute to standing variation in Teosinte (*Zea mays* ssp. *parviglumis*). Genetics, 177, 2349-2359.

Wei, X. M., Jackson, P. A., McIntyre, C. L., Aitken, K. S., & Crot, B. 2006. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. Theatrical and Applied Genetics, 114, 155-164.

Yasir, M., Kanwal, H. H., Hussain, Q., Riaz, M. W., & Sajjad, M. 2022. Status and prospects of genome-wide association studies in cotton. Frontiers in Plant Science, https://doi.org/10.3389/fpls.2022.1019347.

Yu, J., & 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., Vroh, B. I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38, 203-208.

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., & Buckler, E. S. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genetics, 24, 355-362.

Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., & Nordborg, M. 2007. An arabidopsis example of association mapping in

structured samples. PLoS Genetics, 3(1), 71-82.

Zhu, C., Gore, M., Bucker, E. S., & Yu, J. 2008. Status and Prospects of Association Mapping in Plants. The plant genome, 1(1), 5-20.

Zhu, J. J., Wang, X. P., Sun, C. X., Zhu, X. M., Li, M., & Zhang, G. D. 2011. Mapping of QTL associated with drought tolerance in a semi-automobile rain shelter in maize (*Zea mays* L.). Agricultural Science in China, 10, 987-996. https://doi.org/10.1016/ S1671-2927(11)60085-0.