

نقشه‌یابی ارتباطی در گیاهان

رضا عطایی^{۱*}، مجید غلامحسینی^۱، ولی الله محمدی^۲

۱. استادیار، موسسه تحقیقات اصلاح و تهیه نهال و بذر، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران

۲. دانشیار، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران

(تاریخ دریافت: ۱۳۹۶/۲/۱۰ - تاریخ تصویب: ۱۳۹۶/۷/۲۹)

Association mapping in plants

Reza Ataei^{1*}, Majid Gholamhoseini¹, Valiollah Mohammadi³

1. Assistant Professor, Seed and Plant Improvement Institute (SPII), Agriculture Research, Education and Extension (AREEO), Karaj, Iran

2. Associate Professor, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

(Received: Apr. 30, 2017 - Accepted: Oct. 19, 2017)

Abstract

The phenotypic diversity of many important traits in plants is influenced by several loci, environmental factors and their interactions. Association mapping is one of the methods proposed in recent decades for genetic study and detection of quantitative trait loci (QTLs). Association mapping was used in human genetics and qualitative traits (such as diseases), but recently its use is increasing in the plant science because of advances in high throughput genomic technologies, interests in identifying novel and superior alleles, and improvements in statistical methods. Association mapping through linkage disequilibrium analysis is a purposeful method for identifying marker alleles and quantitative traits association. Unlike linkage mapping, this method identifies the association between phenotypic and polymorphic diversity in the genome by exploiting the diversity of natural populations and taking into account all the events that occurred during the evolution and is a promising approach for overcoming the limitations of linkage mapping. Despite association mapping has high statistical power, the application of this method in structured populations, species with low level of linkage disequilibrium and in traits controlled by rare alleles is complicated and difficult. In this review, we will present a comprehensive view, its application in population, current status and limitation of association mapping in plant science

Keywords: Population structure, Linkage disequilibrium, Association mapping, QTL.

چکیده

تنوع فنوتیپی موجود در بسیاری از صفات مهم در گیاهان تحت تأثیر چندین جایگاه ژنی، عوامل محیطی و اثرات متقابل این دو می‌باشد. نقشه‌یابی ارتباطی یکی از روش‌هایی است که در دهه‌های اخیر برای مطالعه ژنتیکی و تعیین تعداد مکانهای ژنی کنترل کننده صفات کمی پیشنهاد شده است. این روش برای اولین بار در ژنتیک انسانی و برای صفاتی کیفی (مانند بیماری‌های ژنتیکی) مورد استفاده قرار گرفت اما امروزه به دلیل پیشرفت‌های چشمگیر در تکنولوژی توالی‌یابی DNA، علاقمندی برای شناسایی ژن‌های جدید و بهبود روش‌های آماری استفاده از آن در جمعیت‌های گیاهی رو به افزایش است. نقشه‌یابی ارتباطی روشی هدفمند برای شناسایی ارتباط آماری بین آلل‌های نشانگری و صفات کمی بر اساس عدم تعادل لینکازی است. برخلاف نقشه‌یابی لینکازی، این روش 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 چندشکلی موجود در ژنوم را شناسایی می‌کند و روشی امیدوارکننده برای غلبه بر محدودیت‌های نقشه‌یابی لینکازی 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 لینکازی و در صفاتی که توسط آلل‌های نادر کنترل می‌شود بسیار پیچیده و دشوار است. در این مقاله وضعیت کلی نقشه‌یابی ارتباطی، نحوه کاربرد آن در جمعیت و محدودیت‌های آن در گیاهان مورد بررسی قرار خواهد گرفت.

واژه‌های کلیدی: ساختار جمعیت، عدم تعادل لینکازی، نقشه‌یابی ارتباطی، QTL.

مقدمه

مطالعه ژنتیکی صفات کمی یکی از اهداف اصلی بهنژادی گیاهی است که امروزه با استفاده از مکان‌یابی ژن‌های کنترل‌کننده صفات کمی انجام می‌شود. اغلب صفاتی که ارزش اقتصادی دارند و با شایستگی افراد مرتبط است به صورت کمی و توسط چندین جایگاه ژنی کنترل می‌شوند (Stich *et al.*, 2010). تا دهه ۱۹۸۰ بررسی چنین صفاتی (بدون اینکه اطلاعاتی در باره جایگاه ژنومی ژن‌های کنترل‌کننده این صفات در دسترس باشد) با استفاده از روش‌های بیومتری مانند تجزیه میانگین نسل‌ها و واریانس جمعیت انجام می‌گرفت. توسعه و گسترش نشانگرهای مولکولی نقطه عطفی در مطالعات صفات کمی به وجود آورد (Kearsey *et al.*, 1998). نقشه‌یابی لینکاژی^۱ یکی از روش‌هایی است که در دهه اخیر برای مطالعه ژنتیکی صفات کمی مورد استفاده قرار گرفته است. مقایسه دو روش نقشه‌یابی صفات کمی (نقشه‌یابی لینکاژی و ارتباطی) نشان می‌دهد نقشه‌یابی ارتباطی در مقایسه با نقشه‌یابی لینکاژی به مراتب دقیق‌تر است زیرا به دلیل نوترکیبی‌های زیاد نقشه ژنتیکی وضوح بالایی دارد و به راحتی می‌توان از آن در فرآیند گزینش به کمک نشانگر^۲ استفاده کرد (Moose *et al.*, 2008). علاوه بر این در نقشه‌یابی ارتباطی به دلیل اینکه از تنوع موجود در جمعیت‌های طبیعی استفاده می‌شود لذا تا حد زیادی در وقت و هزینه صرفه‌جویی می‌شود (جدول ۱).

از زمانی که نقشه‌یابی ارتباطی در گیاهان مطرح شد (Thornsberry *et al.*, 2001)، این روش به دلیل پیشرفت‌های چشمگیر در تکنولوژی توالی‌یابی DNA، علاقمندی برای شناسایی ژن‌های جدید و بهبود روش‌های آماری امروزه طرفداران زیادی به

خود اختصاص داده است. با وجود اینکه استفاده از این روش در دهه‌های اخیر و در جمعیت‌های گیاهی رو به افزایش است (Reich *et al.*, 2001) اما متأسفانه کاربرد آن محدود به گونه‌های مدل با ژنوم شناخته شده مانند آراییدوپسیس (Aranzana *et al.*, 2005)، برنج (Agrama *et al.*, 2007)، ذرت، (Chandler *et al.*, 2002, Lawrence *et al.*, 2008) و جو (Ataei *et al.*, 2016, Ferreira *et al.*, 2016, Laurentin, 2009) مورد استفاده قرار گرفته است و در گیاهان دیگر کاربرد محدودی داشته است (جدول ۲).

جدول ۱. مقایسه دو روش نقشه‌یابی لینکاژی و نقشه‌یابی ارتباطی (Stich *et al.*, 2010)

نقشه‌یابی لینکاژی	نقشه‌یابی ارتباطی
وضوح نقشه ژنتیکی	کم
تعداد نشانگر مورد نیاز برای بررسی ژنوم	کم
میزان نوترکیبی در جمعیت نقشه‌یابی	کم
وجود ساختار و زیرجمعیت	کم
زمان و هزینه مورد نیاز برای تهیه جمعیت	متوسط تا زیاد
قدرت شناسایی QTL در ژنوم	کم
تعداد آلل‌های مورد ارزیابی در جمعیت	بیش از ۴۰
دیپلوئید	۲ آلل

عدم تعادل لینکاژی

در یک جمعیت بزرگ که به صورت تصادفی با یکدیگر آمیزش می‌کنند و ژن‌ها به صورت مستقل از هم تفرق می‌یابند، در صورت عدم وجود گزینش، جهش و مهاجرت، جایگاه‌های ژنی که از چندشکلی برخوردارند در حالت تعادل لینکاژی (تعادل هاردی-واینبرگ) قرار خواهند گرفت. در این صورت فراوانی‌های ژنی و ژنوتیپی از نسلی به نسل دیگر ثابت باقی می‌مانند (Falconer *et al.*, 1996). عدم تعادل لینکاژی (LD) مفهومی اساسی در ژنتیک

1. Linkage Mapping
2. Marker-assisted selection

تفاوت قابل توجهی بین این دو وجود دارد. منظور از لینکاژ توارث همبسته جایگاه‌های ژنی است که به علت نزدیکی فیزیکی ژن‌ها بر روی کروموزوم اتفاق می‌افتد در حالی که عدم تعادل لینکاژی به معنی همبستگی بین آلل‌ها در یک جمعیت است. روش‌های آماری متفاوتی برای اندازه‌گیری عدم تعادل لینکاژی پیشنهاد شده است که در اینجا به دو مورد از پرکاربردترین آن‌ها اشاره خواهد شد. معایب و مزیت‌های این روش‌های آماری در مطالعات پیشین به طور کامل بحث شده است (Jorde, 2000).

جمعیت است و به معنی عدم تعادل مربوط به دو یا چند مکان ژنی، صرف‌نظر از پیوسته یا مستقل بودن آن‌ها، در یک جمعیت است (Falconer, 1981). به عبارت دیگر عدم تعادل لینکاژی همبستگی بین چندشکلی‌هاست (مانند SNPها). علت ایجاد چنین همبستگی‌هایی، تأثیر مشترک نیروهای تکاملی مانند جهش، گزینش و نوترکیبی است (Flint-Garcia *et al.*, 2003). دو اصطلاح لینکاژ و عدم تعادل لینکاژی اغلب گمراه کننده است. اگرچه این دو اصطلاح ارتباط نزدیکی با یکدیگر دارند، با این وجود

جدول ۲. برخی مطالعات نقشه‌یابی ارتباطی در گیاهان مختلف (Zhu *et al.*, 2008)

گونه گیاهی	جمعیت	اندازه نمونه	صفات
	اینبرد لاین‌ها	۹۲	زمان گلدهی
	اینبرد لاین‌ها	۷۱	زمان گلدهی
	اینبرد لاین‌ها و ارقام بومی	۳۷۵+۲۷۵	زمان گلدهی
	اینبرد لاین‌ها	۹۵	زمان گلدهی
ذرت	اینبرد لاین‌ها	۱۰۲	ترکیبات دانه و خصوصیات خمیری نشاسته
	اینبرد لاین‌ها	۸۶	ساخت Maysin
	اینبرد لاین‌ها	۷۵	رنگ دانه
	اینبرد لاین‌ها	۵۷	طعم مطبوع
	اینبرد لاین‌ها	۵۵۳	میزان اولئیک اسید
	اینبرد لاین‌ها	۲۸۲	میزان کارتنوئید
	اکوتیپ‌ها	۹۵	زمان گلدهی
آراییدوپسیس	اکوتیپ‌ها	۹۵	مقاومت به بیماری و زمان گلدهی
	اکوتیپ‌ها	۹۶	انشعابات اندام‌های هوایی
سورگوم	اینبرد لاین‌ها	۳۷۷	ارتفاع، زمان گلدهی، طول و عرض برگ پرچمی و طول پانیکول
گندم	ارقام مختلف	۹۵	اندازه دانه و کیفیت آسیاب شدن
جو	ارقام مختلف	۱۴۸	تعداد روز تا گلدهی، بیماری زنگ، ویروس کوتولگی زرد، اندازه لودیپکول، مقاومت به سرما، عملکرد و صفات مرتبط با ریشه
سیب‌زمینی	ارقام مختلف	۱۲۳	مقاومت به آلودگی فیتوفترایی
	ارقام بومی	۱۰۵	گلوتنی شدن
برنج	ارقام بومی	۵۷۷	کیفیت نشاسته
	ژنوتیپ‌های مختلف	۱۰۳	عملکرد و اجزای عملکرد

آنگاه می‌توان نشان داد تفاوت بین فراوانی‌های مشاهده شده و مورد انتظار $D_{ab} = P_{AB} - P_A P_B$ می‌باشد. یکی از روش‌های بسیار متداول برای

اگر دو ژن با آلل‌های A و a در جایگاه ژنی اول و B و b در جایگاه ژنی دوم وجود داشته باشد و فراوانی هر آلل به ترتیب P_A, P_a, P_B و P_b باشد

کوچک، رانش تصادفی باعث از بین رفتن ترکیبات آل‌های نادر (با فراوانی کم) شده و از این طریق باعث افزایش عدم تعادل لینکاژی در جمعیت می‌شود. هنگامی که رانش تصادفی و نوترکیبی در حالت تعادل ژنتیکی قرار دارند

$$2 = \frac{1}{1 + 4Nc}$$

می‌باشد که N اندازه مؤثر جمعیت و c میزان نوترکیبی بین ژن‌هاست (Weir, 1996).

عدم تعادل لینکاژی در گونه‌های دگرگشن نسبت به گونه‌های خودگشن سریعتر کاهش می‌یابد زیرا نوترکیبی در گونه‌های خودگشن که افراد اغلب از خلوص بالایی برخوردارند کمتر مؤثر است (Nordborg, 2000). مهاجرت باعث ورود کروموزوم‌هایی از اجداد مختلف و فراوانی‌های آلی متفاوت به درون یک جمعیت شده و باعث افزایش عدم تعادل لینکاژی می‌شود با این وجود آمیزش تصادفی بین افراد بومی جمعیت و افراد مهاجر، عدم تعادل لینکاژی را در مدت زمان کوتاهی کاهش می‌دهد (Pritchard et al., 1999). پدیده عدم تعادل لینکاژی می‌تواند در جمعیت‌هایی که کاهش اندازه (دهانه بطری) را به همراه رانش تصادفی تجربه کرده‌اند نیز بوجود آید. طی فرآیند کاهش اندازه جمعیت، فقط ترکیبات آلی کمی به نسل بعد منتقل شده که این امر موجب افزایش عدم تعادل لینکاژی می‌شود (Dunning et al., 2000).

گزینش اعم از گزینش طبیعی و یا مصنوعی باعث ایجاد عدم تعادل لینکاژی بین ژن‌های تحت گزینش و ژن‌های پیوسته با آن می‌شود. به‌علاوه گزینش صعودی (و یا نزولی) صفتی که با دو یا چند ژن ناپیوسته کنترل می‌شود، علی‌رغم وجود فاصله ژنتیکی (فیزیکی) زیاد بین ژن‌ها باعث افزایش عدم تعادل لینکاژی می‌شود.

اندازه‌گیری عدم تعادل لینکاژی استفاده از آماره r^2 است که عبارت است از:

$$r^2 = \frac{()^2}{2}$$

و به صورت مربع ضریب همبستگی بین دو جایگاه ژنی تعریف می‌شود (Hill et al., 1968). معنی‌دار بودن آماری عدم تعادل لینکاژی معمولاً با آزمون دقیق فیشر برای ژن‌های دوآلی و تجزیه جایگشت چند عاملی^۱ برای ژن‌هایی با بیش از دو آلل آزمون می‌شود.

یکی دیگر از روش‌های اندازه‌گیری عدم تعادل لینکاژی آماره D' است که به صورت زیر تعریف می‌شود.

$$D' = \frac{(D_{ab})^2}{\min(P_A P_b; P_a P_b)} \text{ For } D_{ab} < 0$$

$$D' = \frac{(D_{ab})^2}{\min(P_A P_B; P_a P_b)} \text{ For } D_{ab} > 0$$

مخرج کسر کمترین مقدار دو حاصلضرب فراوانی‌های آلی را نشان می‌دهد. این آماره فارغ از ارزش فراوانی‌های آلی در دو جایگاه ژنی، دامنه آن همواره بین -1 و 1 می‌باشد. دو آماره r^2 و D' جنبه‌های مختلفی از عدم تعادل لینکاژی را در نظر می‌گیرند. آماره D' فقط وقایع نوترکیبی را در نظر می‌گیرد در حالی که r^2 علاوه بر نیروی تکاملی حاصل از نوترکیبی، وقایع حاصل از جهش را نیز در بر می‌گیرد (Jorde, 2000). به دلیل اینکه فراوانی‌های آلی و نوترکیبی بین جایگاه‌های ژنی، عدم تعادل لینکاژی را تحت تأثیر قرار می‌دهند لذا تأثیر اغلب نیروهای تکاملی را می‌توان در الگوی عدم تعادل لینکاژی مشاهده نمود. علاوه بر تأثیر دو پدیده ژنتیکی جهش و نوترکیبی بر روی میزان عدم تعادل لینکاژی، اندازه جمعیت نقش مهمی در تغییر عدم تعادل لینکاژی ایفا می‌کند. در جمعیت‌های

بررسی و ارزیابی قرار گیرد. مهم‌ترین عاملی که در نقشه‌یابی ارتباطی باعث ایجاد ارتباط کاذب نشانگر-صفت می‌شود ساختار موجود در جمعیت است. در مکان‌یابی ژن‌های کنترل‌کننده صفتی که باعث سازگاری به مناطق جغرافیایی می‌شود (مانند زمان گلدهی) این مشکل به وضوح خود را نشان می‌دهد. در چنین مواردی تنوع فنوتیپی موجود در جمعیت با تفاوت فراوانی آللی بین زیرجمعیت‌ها همبسته است (Aranzana *et al.*, 2005; Flint-Garcia *et al.*, 2005). معمولاً در نقشه‌یابی ارتباطی مجموعه‌ای از نشانگرهای ناپیوسته^۲ که از لحاظ گزینشی خنثی هستند برای تعیین ترکیب ژنتیکی افراد جمعیت مورد استفاده قرار می‌گیرند و به‌عنوان نشانگرهای پس‌زمینه شناخته می‌شوند. نشانگرهای پس‌زمینه ابزاری مفید برای اختصاص افراد به زیرجمعیت‌ها، جلوگیری از ایجاد ارتباطات کاذب نشانگر-صفت در جوامع دارای ساختار و روابط خویشاوندی، برآورد میزان روابط خویشاوندی افراد و درون زادآور است (Yu *et al.*, 2006).

نشانگرهای RAPD و AFLP به دلیل اینکه نشانگرهایی تصادفی هستند به خوبی می‌توانند به عنوان نشانگرهای پس‌زمینه مورد استفاده قرار گیرند اما به دلیل اینکه هر دو این نشانگرها توارث غالب/مغلوب دارند لذا به روش‌های آماری خاصی نیاز است تا از آن‌ها بتوان در تعیین پارامترهای ژنتیکی جمعیت استفاده کرد (Falush *et al.*, 2007; Ritland, 2005). برخلاف نشانگرهای RAPD و AFLP، نشانگرهای SSR و SNP توارث هم‌بارز دارند و برای برآورد ساختار و روابط خویشاوندی افراد از قدرت بیشتری برخوردارند. نشانگرهای SSR به دلیل چندآللی بودن، امکان استفاده از PCR، قابلیت تکرارپذیری بالا و نیز به دلیل اینکه عموماً از لحاظ

عدم تعادل لینکاژی کاربرد مهمی در مطالعات نقشه‌یابی ارتباطی دارد. در صورت وجود عدم تعادل لینکاژی بین نشانگر مولکولی و ژن‌های کنترل‌کننده یک صفت خاص می‌توان ارتباط معنی‌دار نشانگر-صفت را تعیین کرد و از آن در برنامه‌های به‌نژادی استفاده کرد. امروزه مطالعه عدم تعادل لینکاژی در بسیاری از گیاهان چه در سطح ژن و چه در سطح کل ژنوم صورت گرفته است (جدول ۳) (Gupta *et al.*, 2005).

انتخاب ژرم‌پلاسم مناسب

قبل از آغاز تجزیه ارتباطی باید تمامی جنبه‌های ژنتیکی ژرم‌پلاسم قابل دسترس به دقت بررسی شود. سطح پلویدی افراد یک گونه گیاهی که از نظر ژنتیکی به خوبی شناخته شده نیست باید مورد ارزیابی قرار گیرد. بویژه هنگامی که جمعیت مورد مطالعه شامل گونه‌های وحشی می‌باشد. به دلیل اینکه جمع‌آوری و مطالعه ژنتیکی جمعیت نیاز به همکاری دراز مدت افراد مختلف دارد لذا بررسی دقیق ژنتیکی جمعیت از اهمیت خاصی برخوردار است.

انتخاب ژرم‌پلاسم گام مهمی در موفقیت تجزیه ارتباطی به‌شمار می‌رود (Breseghello *et al.*, 2006). تنوع ژنتیکی، میزان عدم تعادل لینکاژی و میزان خویشاوندی افراد جمعیت عواملی هستند که وضوح نقشه، تراکم نشانگرها، روش‌های آماری و توان نقشه‌یابی را تعیین می‌کنند. جمعیت‌ها بر اساس منبع مواد گیاهی به سه دسته مواد ژنتیکی حاصل از بانک‌های ژن، جمعیت‌های مصنوعی^۱ و ژرم‌پلاسم الیت تقسیم‌بندی می‌شوند (Breseghello *et al.*, 2006).

تجزیه ژنتیکی ژرم‌پلاسم

پس از گزینش ژرم‌پلاسم مناسب باید تنوع ژنتیکی، ساختار جمعیت و میزان عدم تعادل لینکاژی مورد

یک نشانگر SNP مورد نیاز است در حالی که در آرآیدوپسیس تقریباً هر ۵۰kb یک نشانگر SNP مورد نیاز است (Nordborg, 2000, Tenaillon *et al.*, 2001). مطالعه عدم تعادل چهار ژن در ۳ جمعیت جو (۷۴ رقم زراعی، ۲۳ واریته بومی و ۳۴ ژنوتیپ جو وحشی) نشان داد بیشترین میزان عدم تعادل به جمعیت متشکل از ارقام زراعی اختصاص دارد و جمعیت بومی و ژنوتیپ‌های وحشی به ترتیب بعد از آن قرار داشتند. الگوی متفاوت عدم تعادل در این مطالعه به تفاوت در سازماندهی ژنوم، وجود عناصر ترانسپوزون و یا عناصر کوچکی که در ژنوم مانع نوترکیبی می‌شوند و تأثیر متفاوت نیروهای تکاملی مانند گزینش و یا رانش نسبت داده شد (Caldwell *et al.*, 2006). بررسی عدم تعادل لینکاژی در جمعیتی متشکل از ۲۵ ژنوتیپ جو وحشی نشان داد سرعت کاهش عدم تعادل لینکاژی در این جمعیت مشابه سرعت کاهش عدم تعادل لینکاژی در ذرت است (Morrell *et al.*, 2005).

گزینشی خنثی هستند برای تعیین روابط خویشاوندی کاربرد بیشتری دارند. به دلیل اینکه هتروزیگوسیتی مورد انتظار SNP نسبت به SSR کمتر است لذا برای تعیین ساختار و خویشاوندی افراد جمعیت در اغلب گیاهان تعداد بیشتری نشانگر SNP مورد نیاز است. این امر نباید به عنوان یک نقص در نظر گرفته شود زیرا SNPها به تعداد نامحدودی در ژنوم وجود دارند و با استفاده از سیستم‌های خودکار و با هزینه‌ای به مراتب کمتر از SSR می‌توانند مورد استفاده قرار گیرند. میزان عدم تعادل لینکاژی از گونه‌ای به گونه دیگر و بسته به نوع سیستم تولید مثل (خودگشن یا دگرگشن) تفاوت دارد (جدول ۳). در گونه‌های خودگشن مانند جو تأثیر نوترکیبی کمتر از گونه‌های دگرگشنی مانند ذرت است لذا اندازه بلوک‌های عدم تعادل در جو بسیار بزرگتر از ذرت است (Flint-Garcia *et al.*, 2003). تفاوت در میزان عدم تعادل بین گونه‌های گیاهی و بین جمعیت‌های مختلف از یک گونه بر تعداد نشانگر مورد نیاز برای پوشش ژنوم نیز تأثیر می‌گذارد. در ذرت هر ۲۰۰-۱۰۰ جفت‌باز

جدول ۳. دامنه عدم تعادل لینکاژی در گیاهان مختلف (Gupta *et al.*, 2005).

دامنه عدم تعادل لینکاژی	سیستم آمیزشی	گونه گیاهی
۷-۰/۵ کیلوباز (kb)	دگرگشن	ذرت
۲۰-۱۰ سانتی‌مورگان	خودگشن	جو
۲۰-۱۰ سانتی‌مورگان	خودگشن	گندم تتراپلوئید
۱۰۰ کیلوباز (kb)	خودگشن	برنج
۴ < سانتی‌مورگان یا ۱۰ ≤ کیلوباز (kb)	خودگشن	سورگوم
۱۰ سانتی‌مورگان	دگرگشن / تکثیر غیر جنسی	نیشکر
۲۵۰ کیلوباز (kb)	خودگشن	آرآیدوپسیس
۵۰ > کیلوباز (kb)	خودگشن	سویا
۳ < سانتی‌مورگان	دگرگشن	چغندر قند
۱-۰/۳ سانتی‌مورگان	خودگشن	سیب‌زمینی
۲۰۰ ~ کیلوباز (kb)	خودگشن	کاهو
۵۰۰ جفت‌باز (bp)	تکثیر غیر جنسی	انگور
۲۰۰-۱۰۰ ~ جفت‌باز (bp)	دگرگشن	کاج نروژی
۱۵۰-۱۰۰ جفت‌باز (bp)	دگرگشن	کاج استخری
۱۵۰۰ ~ جفت‌باز (bp)	دگرگشن	کاج استخری

ارزیابی فنوتیپی

علی‌رغم تأکید زیاد بر اهمیت ارزیابی ژنتیکی در گزارشات، تأثیر ارزیابی فنوتیپی بر نقشه‌یابی ارتباطی کمتر مورد بحث قرار گرفته است. به دلیل اینکه نقشه‌یابی ارتباطی نیازمند ارزیابی تعداد زیادی ژنوتیپ است لذا ارزیابی فنوتیپی دقیق و صحیح در مقایسه با ارزیابی ژنوتیپی می‌تواند گران‌تر و دشوارتر باشد. همچنین حصول داده‌های فنوتیپی قابل اعتماد از جمعیتی که افراد آن به شرایط رشدی متفاوت سازگاری دارند بسیار دشوار است و در برخی از موارد کاربرد چنین ژرم‌پلاسمی محدود می‌شود (Myles *et al.*, 2009). با این وجود با تکرار ژنوتیپ‌ها در دو یا چند محیط مختلف می‌توان تأثیر عوامل محیطی و اثرات متقابل ژنوتیپ و محیط را بر روی تنوع فنوتیپی ژنوتیپ‌ها برآورد کرد. طرح‌های مزرعه‌ای کارآمد (مانند α -لاتیس)، روش‌های آماری مناسب به همراه بررسی اثر متقابل QTL و محیط تا حد زیادی می‌تواند صحت و دقت شناسایی QTL‌ها را افزایش دهد به خصوص هنگامی که شرایط مزرعه از یکنواختی خوبی برخوردار نیست (Zhu *et al.*, 2008). امروزه افزایش توان نقشه‌یابی بوسیله داده‌های تکرار شده به خوبی شناخته شده است و همچنین توسط مطالعات شبیه‌سازی در ژرم‌پلاسم-های شجره‌دار این امر ثابت شده است (Arbelbide *et al.*, 2006). با توجه به تنوع بالای جمعیت‌های مورد استفاده در نقشه‌یابی ارتباطی، بررسی تأثیر زمان گلدهی بر روی تظاهر دیگر صفات همبسته مهم و ضروری است لذا در صورتی که صفت مورد نظر وابسته به مراحل رشدی باشد، بلوک‌بندی مزرعه بر اساس زمان گلدهی می‌تواند مفید واقع شود. دیگر مسائلی که نیاز به توجه و بررسی بیشتر دارد ارزیابی فنوتیپی صفاتی مانند حساسیت به طول روز، ورس و مقاومت به بیماری‌هاست زیرا این صفات تا حد زیادی تحت تأثیر دیگر صفات مورفولوژیکی و زراعی تحت شرایط مزرعه می‌باشند (Zhu *et al.*, 2008).

تجزیه ارتباطی

نقشه‌یابی ارتباطی اولین بار با روش تیمار-شاهد^۱ در ژنتیک انسانی مورد استفاده قرار گرفت. یکی از بررسی‌های اولیه با استفاده از این روش با هدف یافتن نشانگر پیوسته با بیماری دیابت بود (Knowler *et al.*, 1988). در این بررسی نمونه‌ای متشکل از ۴۹۲۰ فرد بومی از دو قبیله پیما و پاپاگو جمع‌آوری شد و نشان داد که بیماری دیابت نوع ۲ به طور معنی‌داری با آلل G (ایمونوگلوبین G) ارتباط دارد. هنگامی که این بررسی فقط معطوف به افراد بالای ۳۵ سال شد چنین پیوستگی وجود نداشت. مطالعات بعدی نشان داد که جمعیت مورد بررسی کاملاً بومی نبوده و مخلوطی از شجره اروپایی و آمریکایی هستند به عبارت دیگر ترکیبی از دو جمعیت هستند که هر یک ساختار ژنتیکی منحصر به فردی دارد. این اولین مطالعه‌ای بود که اهمیت وجود ساختار جمعیت و اشتباه نوع اول در چنین جمعیت‌هایی را گوشزد کرد. مطالعات بعدی نشان داد وجود ساختار در جمعیت باعث ایجاد اشتباه در این روش می‌شود. برای حل چنین مشکلی در دهه ۹۰ آزمون‌هایی بر اساس خانواده گسترش پیدا کرد. برخلاف روش تیمار-شاهد قدرت چنین آزمون‌هایی وابسته به وجود و یا عدم وجود ساختار در جمعیت نمی‌باشد و شناخته‌شده‌ترین این آزمون‌ها آزمون عدم تعادل انتقال^۲ است که در سال ۱۹۹۳ پیشنهاد شد (Spielman *et al.*, 1993). در سال ۲۰۰۰ روش ارتباط ساختارمند^۳ را معرفی کردند (Pritchard *et al.*, 2000). در این روش ابتدا جمعیت با استفاده از روش بیس^۴ به زیرجمعیت‌ها تقسیم می‌شود و ماتریس Q با استفاده از رگرسیون لگاریتمی برای تصحیح اشتباه نوع اول در نقشه‌یابی وارد می‌شود.

1. Case-Control

2. Transmission disequilibrium test (TDT)

3. Structured association

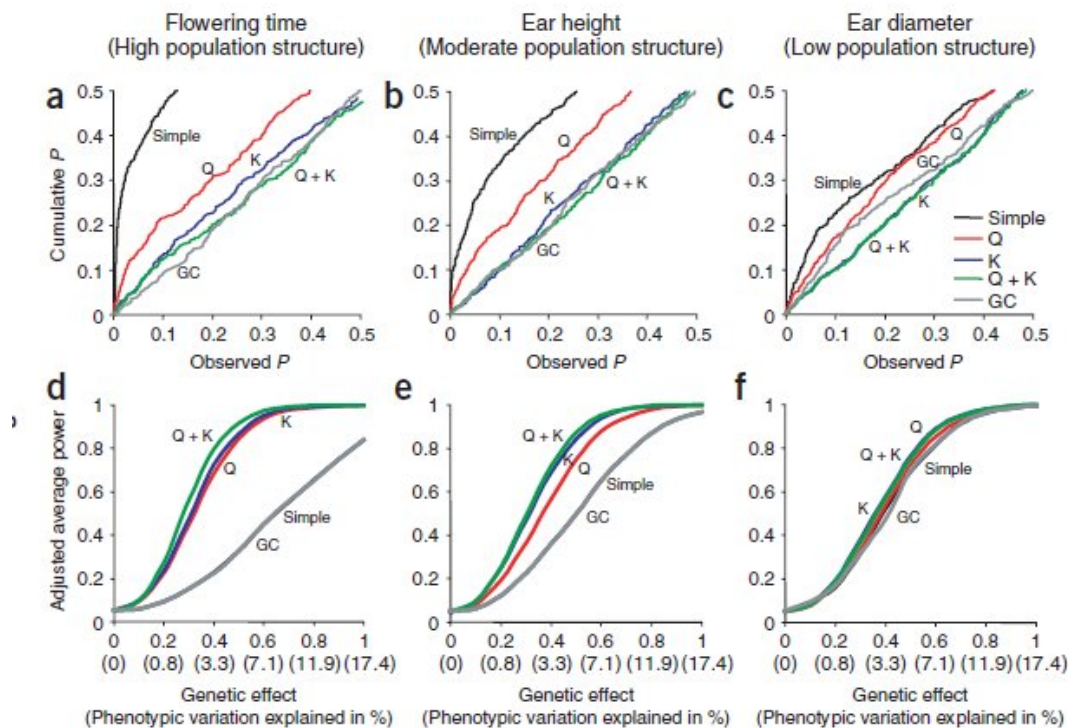
4. Bayesian method

ژنوتیپی فرد i در زیرجمعیت K ، α اثر نشانگر و $\varepsilon_{i(k)}$ اثرات باقیمانده فرد i در زیرجمعیت K است. همانگونه که در مدل آماری دیده می‌شود مدل خطی مخلوط تنوع فنوتیپی جمعیت را به اثرات مربوط به ساختار جمعیت و نشانگر تجزیه می‌کند. به دلیل توان آماری متفاوت روش‌های نقشه‌یابی، در مطالعه‌ای کارایی روش‌های مختلف نقشه‌یابی ارتباطی مورد بررسی قرار گرفت و گزارش شد که مدل خطی مخلوط به همراه ماتریس‌های K و Q برای تصحیح ساختار و روابط خویشاوندی موجود در جمعیت، وضوح نقشه‌یابی را نسبت به دیگر روش‌ها تا حد زیادی افزایش می‌دهد. در این بررسی ۳ جمعیت ذرت با ساختارهای متفاوت و ۵ روش نقشه‌یابی (کنترل ژنومی، مدل خطی با ماتریس‌های K و Q ، مدل خطی با ماتریس K ، مدل خطی با ماتریس Q و مدل ساده) مورد استفاده قرار گرفت.

اگرچه این روش به طور مؤثری ساختار جمعیت را کنترل می‌کند ولی استفاده از جمعیت‌های دارای ساختار با درجاتی از خویشاوندی بین افراد کارایی این روش را کاهش می‌دهد. به همین دلیل یو و همکاران در سال ۲۰۰۵ مدل خطی مخلوط^۱ را پیشنهاد کردند که اطلاعات مربوط به ساختار جمعیت و ضریب خویشاوندی را در تجزیه و تحلیل لحاظ می‌کند و نسبت به روش‌های فوق از توان آماری بالایی برخوردار است (Yu et al., 2005). مدل خطی مخلوط را می‌توان به صورت زیر نشان داد:

$$P_{i(k)} = \mu + S_k + x_{i(k)} \alpha + \varepsilon_{i(k)}$$

که در آن $P_{i(k)}$ ارزش فنوتیپی فرد i در زیرجمعیت K ، μ میانگین جمعیت، S_k فاکتور تصحیح برای ساختار موجود در جمعیت، $x_{i(k)}$ داده



شکل ۱. مقایسه پنج روش مختلف نقشه‌یابی ارتباطی برای سه صفت در ذرت

نرم‌افزار قادر است به صورت گرافیکی میزان عدم تعادل لینکاژی را نشان دهد (Bradbury *et al.*, 2007). برای انجام نقشه‌یابی ارتباطی باید داده‌های فنوتیپی، داده‌های ژنوتیپی و ماتریس Q به نرم‌افزار معرفی شود. تجزیه ارتباطی در این نرم‌افزار با استفاده از مدل خطی عمومی^۱ و یا مدل خطی مخلوط انجام می‌شود. اگر مدل خطی عمومی برای تعیین ارتباط نشانگر-صفت استفاده شود، فقط ماتریس Q که نشان‌دهنده ساختار جمعیت است، در محاسبات وارد می‌شود و روابط خویشاوندی بین افراد در نظر گرفته نمی‌شود. نقشه‌یابی ارتباطی با مدل خطی مخلوط نیازمند محاسبه ماتریس K (نشان‌دهنده روابط خویشاوندی بین افراد جمعیت) است که با استفاده از داده‌های ژنوتیپی در این نرم‌افزار انجام می‌گیرد و سپس ارتباط بین نشانگر-صفت به همراه احتمال اشتباه نوع اول محاسبه می‌شود.

STRUCTURE نرم‌افزاری است که معمولاً برای تعیین ساختار جمعیت و برآورد ماتریس Q (ماتریسی از مرتبه $n \times p$ که n تعداد افراد موجود در جمعیت و p تعداد زیرجمعیت‌های تعریف‌شده را نشان می‌دهد) مورد استفاده قرار می‌گیرد (Pritchard *et al.*, 2000). این نرم‌افزار از مدل پیشنهاد شده توسط پریچارد و همکاران برای تعیین ساختار جمعیت و اختصاص افراد به زیرجمعیت‌ها استفاده می‌کند. در این روش فرض می‌شود که K زیر جمعیت وجود دارد که هر یک از این زیرجمعیت‌ها با مجموعه‌ای از فراوانی‌های آللی در هر مکان ژنی مشخص شده و در داخل آن‌ها تعادل هاردی-واینبرگ برقرار است. لگاریتم احتمال اینکه فرد X متعلق به هر یک از زیرجمعیت‌های K باشد محاسبه می‌شود و بر این اساس افراد به زیرجمعیت‌های مختلف منتسب می‌شوند به طوری که در تمامی زیرجمعیت‌ها تعادل هاردی-واینبرگ همچنان برقرار است.

در هر ۳ جمعیت مدل ساده (بدون در نظر گرفتن ساختار جمعیت و روابط خویشاوندی افراد) از کارایی کمتری برخوردار بود و مدل خطی مخلوط با مدنظر قرار دادن ساختار و روابط خویشاوندی بیشترین کارایی را داشت. نقشه‌یابی ارتباطی صفت زمان گلدهی با مدل ساده نشان داد ۳۷/۶ درصد از نشانگرها در سطح ۵ درصد معنی‌دار است در حالی که در مدل خطی با ماتریس Q، ۱۴/۱ درصد، در مدل خطی با ماتریس K و Q ۶/۱ درصد و در مدل خطی با ماتریس‌های K و Q فقط ۶ درصد از نشانگرها معنی‌دار بودند. بای‌پلات سطح احتمال مشاهده شده و سطح احتمال تجمعی نشان داد مدل خطی با ماتریس‌های K و Q دارای توزیع یکنواخت بوده و در هر سه سطح از ساختار جمعیت (کم، متوسط و زیاد) نسبت به دیگر روش‌ها اشتباه نوع اول را بهتر کنترل می‌کند و از توان آماری بیشتری برخوردار است (شکل ۱). به علاوه در نظر گرفتن روابط خویشاوندی (مدل خطی با ماتریس K) نسبت به حالتی که فقط ساختار جمعیت در نظر گرفته شده بود (مدل خطی با ماتریس Q)، از کارایی بیشتری برخوردار بود (Yu *et al.*, 2005).

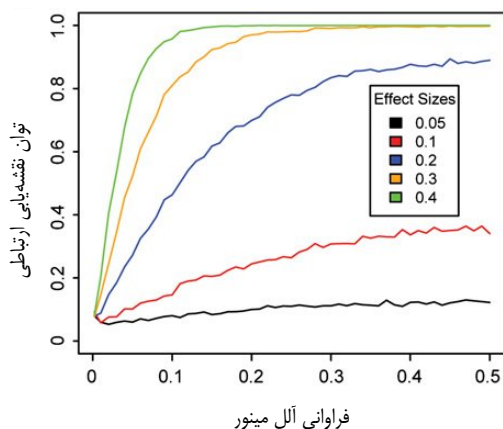
با توجه به فرض صفر (اثر نشانگر برابر صفر است) روش‌هایی که توزیع یکنواختی از مقدار P نشان می‌دهند اشتباه نوع اول را بهتر کنترل می‌کنند. شکل به وضوح نشان می‌دهد که مدل خطی مخلوط به همراه ماتریس‌های K و Q در هر سه سطح مختلف ساختار در جمعیت، از پایداری خوبی برخوردار است (Yu *et al.*, 2005).

بسته‌های نرم‌افزاری برای نقشه‌یابی ارتباطی

امروزه با پیشرفت فناوری اطلاعات نرم‌افزارهای متنوعی برای انجام تجزیه ارتباطی معرفی شده است که به مهمترین آن‌ها اشاره می‌شود.

TASSEL یکی از نرم‌افزارهایی است که معمولاً برای نقشه‌یابی ارتباطی در گیاهان مورد استفاده قرار می‌گیرد و به‌طور مداوم به‌روزرسانی می‌شود. این

1. Generalized Linear Model (GLM)



شکل ۲. رابطه بین توان نقشه‌یابی ارتباطی و فراوانی آللی (Myles *et al.*, 2009).

متأسفانه تئوری‌های ژنتیک کمی نشان می‌دهد که بخش اعظمی از گونه‌ها، اغلب دارای آلل‌ها نادر هستند. علاوه بر این مهم‌ترین عاملی که وراثت‌پذیری گم‌شده را توجیه می‌کند ساختار ژنتیکی صفت مورد مطالعه است. از طرفی توان نقشه‌یابی ارتباطی تابعی از میزان تأثیر QTL بر روی صفت فنوتیپی است. به عبارت دیگر تجزیه ژنتیکی صفتی که توسط تعداد زیادی QTL کوچک‌اثر کنترل می‌شود بسیار دشوارتر از صفتی است که تعداد کمی QTL کوچک‌اثر در کنترل آن دخالت دارند (Myles *et al.*, 2009). در برخی از گونه‌ها به دلیل میزان کم عدم تعادل لینکاژی کاربرد نقشه‌یابی ارتباطی دشوار و پیچیده است. در چنین شرایطی به دلیل میزان کم عدم تعادل لینکاژی تعداد زیادی نشانگر مورد نیاز است با این وجود در برخی از گونه‌ها پلاتفرم‌های تجاری نشانگرها در دسترس نیست و لذا کاربرد نقشه‌یابی ارتباطی در این گونه‌ها محدود به نقشه‌یابی ژن‌های کاندیدااست. با اینکه نقشه‌یابی ارتباطی در جمعیت‌های بزرگتر از دقت و صحت بالایی برخوردار است با این وجود اندازه جمعیت مورد استفاده در مطالعات انجام گرفته نسبتاً کوچک باقی مانده است. در اغلب مطالعاتی که در دهه اخیر صورت گرفته در حدود ۱۰۰ فرد مورد ارزیابی قرار

SPAGeDi نرم‌افزاری برای تعیین روابط خویشاوندی موجود در جمعیت و برآورد ماتریس K (ماتریسی از مرتبه $n \times n$ که عناصر خارج از قطر اصلی F_{ij} (احتمال همسانی اجدادی بر اساس نشانگر) و عناصر قطری برای اینبردها مقدار ۱ و برای غیراینبردها $(1 + F_x) / 5$ می‌باشد و F_x ضریب درون‌زادآوری است (Hardy *et al.*, 2002).

ASREML و MTDFREML نرم‌افزارهایی هستند در مطالعات ژنتیکی حیوانات و در جمعیت‌های بسیار بزرگ از آن‌ها می‌توان استفاده کرد. SAS, QTD, Merlin, EINGENSTRAT و R از دیگر نرم‌افزارهایی است که می‌توان از آن‌ها در تجزیه ارتباطی استفاده کرد.

محدودیت‌های نقشه‌یابی ارتباطی

هدف نقشه‌یابی ارتباطی شناسایی نشانگرهایی است که با یک یا چند صفت کمی ارتباط دارند. این نشانگرها برای ارزیابی ژنوتیپی جمعیتی که داده‌های فنوتیپی آن موجود است، مورد استفاده قرار می‌گیرد. نقشه‌یابی ارتباطی بر این فرض استوار است که تنوع فنوتیپی با تنوع ژنوتیپی ارتباط دارد. با این وجود کاربرد نقشه‌یابی ارتباطی در انسان یکی از مهم‌ترین محدودیت‌های این روش را آشکار کرد. نقشه‌یابی ارتباطی قد انسان ارتباط ۴۰ نشانگر با این صفت را نشان داد که تنها ۵ درصد از وراثت‌پذیری این صفت را توجیه می‌کرد (وراثت‌پذیری گم‌شده) (Maher, 2008). این امر به دلیل آن بود که آلل‌های مؤثر در تنوع این صفت از فراوانی کمی برخوردار بودند. در واقع قدرت نقشه‌یابی ارتباطی تابعی از فراوانی آللی است (شکل ۲). آللهایی با فراوانی کم حتی اگر تأثیر قابل ملاحظه‌ای بر روی تنوع فنوتیپی داشته باشند نمی‌توانند توسط نقشه‌یابی ارتباطی شناسایی شوند.

با همه مزیت‌هایی که نقشه‌یابی ارتباطی نسبت به نقشه‌یابی لینکاژی دارد اما نباید این دو را به عنوان جایگزین یکدیگر در نظر گرفت. کاربرد هر دو روش به صورت مکمل یکدیگر توان تجزیه QTLها به ژن‌ها را تا حد زیادی افزایش خواهد داد. انتخاب ژرم پلاسِم مناسب، ارزیابی فنوتیپی و کیفیت داده‌های فنوتیپی، ارزیابی ژنوتیپی افراد جمعیت و استفاده از نرم‌افزارهای آماری برای تجزیه ارتباطی از مراحل مهم مطالعات نقشه‌یابی ارتباطی برای شناسایی QTLها است. امروزه کاربرد اغلب مطالعات نقشه‌یابی ارتباطی در گیاهان محدود به گونه‌های گیاهی مهم با ژنوم شناخته شده مانند برنج و ذرت بوده است علاوه بر این فقط تعداد محدودی از QTLهای شناسایی شده در سطح ژن همسانه سازی شده‌اند لذا اتمام پروژه‌های توالی‌یابی ژنوم گیاهان به همراه روش‌های ارزیابی فنوتیپی دقیق‌تر در آینده تا حد زیادی می‌تواند وسعت و سرعت مطالعات نقشه‌یابی ارتباطی در گیاهان را افزایش دهد.

REFERENCES

- Agrama H, Eizenga G, Yan W (2007) Association mapping of yield and its components in rice cultivars. *Molecular Breeding* 19: 341-356.
- Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics* 1: 531-539.
- Arbelbide M, Yu J, Bernardo R (2006) Power of mixed-model QTL mapping from phenotypic, pedigree and marker data in self-pollinated crops. *Theoretical and Applied Genetics* 112: 876-884.
- Ataei R, Mohammadi V, Talei AR, Naghavi MR (2016) Association mapping of yield and some important agronomic traits in barley (*Hordeum vulgare* L.). *Modern Genetics Journal* 11: 411-423.
- Bernardo R (2002) Breeding for quantitative traits in plants. Stemma press Woodburypp.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens T M, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635.
- Breseghele F, Sorrells ME (2006) Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Science* 46: 1323-1330.
- Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172: 557-567.

گرفته است. در حالی که میانگین تعداد افراد ارزیابی شده در مطالعات در نقشه‌یابی لینکاژی حدود ۲۵۰ فرد است (Bernardo, 2002). لذا اگر صفت مورد نظر با ژن‌هایی بزرگ‌اثر کنترل شود و نشانگرها عدم تعادل لینکاژی بالایی با این ژن‌ها داشته باشند، در این صورت شناسایی ارتباط نشانگر-صفت در جمعیتی با اندازه کوچک بسیار دشوار خواهد بود (Zhu et al., 2008).

نتیجه‌گیری کلی

نقشه‌یابی ارتباطی و نقشه‌یابی لینکاژی دو روش عمده برای شناسایی QTLها در جمعیت‌های گیاهی است. نقشه‌یابی لینکاژی روشی قدرتمند برای ارزیابی دو آلل با وضوح پایین می‌باشد در حالی که نقشه‌یابی ارتباطی توان ارزیابی چندین آلل با وضوح بسیار بالا را دارد با این حال استفاده موفقیت‌آمیز این روش در جمعیت‌های گیاهی نیاز به تصحیحات خاصی دارد تا ارتباطات کاذب نشانگر-صفت را به حداقل رساند.

- Chandler VL, Brendel V (2002) The maize genome sequencing project. *Plant Physiology* 130: 1594-1597.
- Dunning AM, Durocher F, Healey CS, Teare MD, McBride S E, Carlomagno F, Xu C-F, Dawson E, Rhodes S, Ueda S (2000) The extent of linkage disequilibrium in four populations with distinct demographic histories. *The American Journal of Human Genetics* 67: 1544-1554.
- Falconer DS (1981) Introduction to quantitative genetics. Longman.pp.
- Falconer D S, Mackay T F, Frankham R (1996) Introduction to Quantitative Genetics (4th edn). *Trends in Genetics* 12: 280.
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7: 574-578.
- Ferreira JR, Pereira JF, Turchetto C, Minella E, Consoli L, Delatorre CA (2016) Assessment of genetic diversity in Brazilian barley using SSR markers. *Genetics and molecular biology* 39: 86-96.
- Flint-Garcia SA, Thornsberry JM, IV B (2003) Structure of Linkage Disequilibrium in Plants. *Annual Review of Plant Biology* 54: 357-374.
- Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman M M, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. *The Plant Journal* 44: 1054-1064.
- Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology* 57: 461-485.
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618-620.
- Hill W, Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38: 226-231.
- Jorde L (2000) Linkage disequilibrium and the search for complex disease genes. *Genome Research* 10: 1435-1444.
- Kearsey M, Farquhar A (1998) QTL analysis in plants; where are we now? *Heredity* 80: 137-142.
- Knowler WC, Williams R, Pettitt D, Steinberg AG (1988) Gm3; 5, 13, 14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *American Journal of Human Genetics* 43: 520.
- Laurentin H (2009) Data analysis for molecular characterization of plant genetic resources. *Genetic Resources and Crop Evolution* 56: 277-292.
- Lawrence CJ, Harper LC, Schaeffer ML, Sen TZ, Seigfried TE, Campbell DA (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. *International Journal of Plant Genomics* 2008: 496957.
- Maher B (2008) The case of the missing heritability. *Nature* 456: 18-21.
- Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology* 147: 969-977.
- Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proceedings of the National Academy of Sciences of the United States of America* 102: 2442-2447.
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell*

- Online. 21: 2194-2202.
- Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154: 923-929.
- Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *The American Journal of Human Genetics*. 65: 220-228.
- Pritchard J K, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*. 155: 945-959.
- Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R (2001) Linkage disequilibrium in the human genome. *Nature* 411: 199-204.
- Ritland K (2005) Multilocus estimation of pairwise relatedness with dominant markers. *Molecular Ecology* 14: 3157-3165.
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* 52: 506-516.
- Stich B, Melchinger AE (2010) An introduction to association mapping in plants. *CAB Rev* 5: 1-9.
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences* 98: 9161-9166.
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nature Genetics* 28: 286-289.
- Weir BS (1996) *Genetic data analysis 2: methods for discrete population genetic data*. Sinauer, Massachusetts.
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology* 17: 155-160.
- Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38: 203-208.
- Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *The Plant Genome* 1: 5-20.