

مدل سازی ساختار کوواریانس آثار نشانگرهای DNA در بین صفات

همبسته با استفاده از یک مدل تحلیل عاملی تصادفی

علی اسماعیلی زاده کشکوئی*^۱

۱- استادیار و عضو هیئت علمی بخش علوم دامی، دانشکده کشاورزی،

دانشگاه شهید باهنر کرمان

* نویسنده مسئول مکاتبات، پست الکترونیکی: aliesmaili@mail.uk.ac.ir

(تاریخ دریافت: ۸۸/۱/۲۹ - تاریخ پذیرش: ۸۹/۸/۹)

چکیده

هدف از انجام این تحقیق، ارائه یک روش برای تجزیه چند صفتی داده‌های حاصل از آزمایش‌های نقشه یابی جایگاه‌های ژنی مرتبط با صفات کمی (QTL) و مدل سازی آثار پلیوتروپی QTL بود. اولین مرحله از تحقیق، توسعه یک مدل ریاضی و روش آماری به منظور انتخاب یک مدل مناسب برای اثر ژن بود. روش پیشنهادی در این تحقیق، مبتنی بر مدل‌های مختلط خطی بود که در آن آثار نشانگرهای DNA در بین صفات همبسته بصورت تصادفی در نظر گرفته شده و مدل مورد استفاده برای ساختار کوواریانس این آثار، یک مدل تحلیل عاملی بود. برای کوواریانس آثار باقیمانده مربوط به صفات مختلف مدل غیرساختاری استفاده شد. برای بررسی کارایی روش پیشنهادی از داده‌های شبیه‌سازی شده مولکولی و فنوتیپی استفاده شد. اثر تراکم نشانگر و اندازه نمونه بر کارایی مدل تحلیل عاملی در تشخیص QTL های پلیوتروپی مورد بررسی قرار گرفت. کارایی مدل با افزایش اندازه نمونه افزایش یافت. کاهش فاصله نشانگرها از ۲۰ سانتی مورگان به ۵ سانتی مورگان کارایی مدل را افزایش داد اما کاهش فاصله نشانگرها به ۱ سانتی مورگان تاثیر چندانی در کارایی مدل نداشت. نتایج حاصل از شبیه‌سازی نشان داد که مدل پیشنهادی توانایی بالایی در مدل سازی آثار نشانگرهای DNA در بین صفات همبسته و در نتیجه تشخیص QTL های پلیوتروپی دارد.

واژه‌های کلیدی

آثار پلیوتروپی،

صفات کمی،

مدل تحلیل عاملی،

نشانگرهای DNA،

QTL

مقدمه

اطلاعات مربوط به صفات مختلف، نتایج حاصل از تجزیه چندصفتی قابلیت اعتماد بیشتری دارند.

چندین روش برای تجزیه چند صفتی QTL پیشنهاد شده است (۳، ۴، ۸، ۱۱، ۱۲). برای مثال، نات و هالی (۱۱) استفاده از تجزیه چند صفتی QTL در قالب حداقل مربعات را پیشنهاد کردند. اما این روش از مدل تک ژنی استفاده می‌کند، بعبارت دیگر وقتی اثر پلیوتروپی یک QTL مورد مطالعه قرار می‌گیرد، اثر سایر ژن‌های ژنوم نادیده گرفته می‌شود. علاوه بر این همانگونه که نات (۱۰) اشاره کرده است نتایجی از این روش بدست می‌آید که مسلماً^۲ درست نمی‌باشند. دلیل عدم درستی نتایج روش فوق این است که تجزیه تک تک صفات بیانگر وجود QTL برای همه صفات در یک نقطه از ژنوم است اما تجزیه صفات بطور همزمان بیانگر وجود QTL در نقطه دیگری از ژنوم است که بر اساس تجزیه تک صفتی مدرکی مبنی بر وجود QTL در آن نقطه نبوده است (۱۰). در اکثر روش‌های موجود برای تجزیه چند صفتی QTL، یک یا بیش از یک ترکیب خطی از صفات اولیه از طریق تجزیه به مولفه‌های اصلی یا تحلیل عاملی بدست آمده و متعاقب آن از روش مرسوم تجزیه تک متغیره صفات جدید به روش نقشه یابی درون فاصله‌ای (۱۳) صورت می‌گیرد. برای مثال، ولر و همکاران (۲۱) روش تجزیه مولفه‌های اصلی و تبدیل متعارف^۲ را پیشنهاد کردند. در این روش صفات به ترکیبات خطی جدیدی تبدیل می‌شوند و تعدادی از این ترکیبات خطی که بیشترین واریانس صفات را بیان می‌کنند، به عنوان صفات جدید برای تجزیه QTL مورد استفاده قرار می‌گیرند. یکی از معایب این روش‌ها مشکل بودن تفسیر نتایج بر حسب صفات اولیه و عدم استفاده از آزمون آماری مناسب برای آثار پلیوتروپی QTL است. علاوه بر این، روش‌های موجود مبتنی بر مدل‌های ثابت بوده و امکان مدل‌سازی آثار کل نشانگرها بطور همزمان وجود ندارد. لذا هدف از انجام این تحقیق توسعه یک روش مبتنی بر مدل‌های مختلط برای آزمون اثر پلیوتروپی QTL و مدل‌سازی آثار همزمان کل نشانگرهای ژنوم بوده است.

بسیاری از خصوصیات مهم در علوم کشاورزی (خصوصیات مرتبط با تولید کمی و کیفی محصولات دامی و گیاهی) و پزشکی (خصوصیات مرتبط با بیماری‌هایی مثل انواع سرطان، دیابت، چاقی و...) دارای مکانیسم وراثتی پیچیده بوده و اصطلاحاً^۱ صفات کمی نام دارند. فن آوری‌های مولکولی به شکل نشانگرهایی که تفاوت افراد را در سطح DNA آنها نشان می‌دهند، می‌توانند نقش مهمی در بهبود ژنتیکی صفات پیچیده از طریق انتخاب به کمک نشانگر یا ژن برای اصلاح دام و گیاه، یا توسعه آزمون‌های ژنتیکی برای تشخیص بیماری‌های ژنتیکی یا تولید دارو برای این بیماری‌ها، داشته باشند. هر ناحیه ژنومی مرتبط با یک صفت کمی اصطلاحاً^۱ یک QTL^۱ نامیده می‌شود.

یک ناحیه ژنومی خاص ممکن است بیش از یک صفت را تحت تاثیر قرار دهد که گفته می‌شود QTL اثر پلیوتروپی دارد. تجزیه‌های چندمتغیره امکان تهیه نقشه ژنتیکی و مدل‌سازی آثار پلیوتروپی QTL را فراهم می‌سازند. بطور معمول در همه مطالعات ژنومی از جمله آزمایشات مربوط به مکان‌یابی QTL، تعداد زیادی از صفات رکوردبرداری می‌شوند که برخی از آنها همبستگی ژنتیکی بالایی دارند. تجزیه‌های تک صفتی بطور معمول روی داده‌ها انجام می‌شوند اما تجزیه‌های چند صفتی به دلیل عدم وجود روش مناسب و پیچیده بودن، کمتر مورد استفاده قرار می‌گیرند. این در حالیست که در تجزیه چند صفتی، به دلیل استفاده از اطلاعات مربوط به همبستگی ژنتیکی و فنوتیپی بین صفات، قدرت آزمون و دقت برآورد پارامترها افزایش می‌یابد (۳، ۸، ۱۱، ۱۲). جیانگ و زنگ (۸) نشان دادند که اگر یک QTL دارای اثر پلیوتروپی روی دو صفت باشد، آنگاه تجزیه دو صفت بطور همزمان، منجر به افزایش توانایی آزمون و بهبود دقت تهیه نقشه ژن می‌شود. اگر مدل پلیوتروپی، مدل درستی باشد، انتظار می‌رود که برازش این مدل توانایی بالایی در شناسایی QTL داشته باشد و انحراف معیار برآورد موقعیت QTL نیز به دلیل استفاده از چند صفت در برآورد یک پارامتر (موقعیت مشابه QTL موثر بر صفات مختلف) کاهش یابد (۸، ۱۱). بنابراین، به دلیل تجمع

² Canonical transformation

¹ Quantitative trait loci

مواد و روش‌ها

مدل آماری برای تجزیه چند صفتی QTL

با فرض اینکه تعداد p صفت روی n فرد اندازه گیری شوند و افراد برای m نشانگر تعیین ژنوتیپ شوند، بردار ستونی y_j با ابعاد $1 \times (np)$ حاوی مشاهدات بوده و از اتصال بردارهای مربوط به هر صفت بدست می‌آید. i امین عنصر بردار y_j به مشاهدات صفت j ام روی فرد i ام مربوط می‌شود. بنابراین، مدل مختلط خطی برای توصیف داده‌های بردار ترکیبی صفات مختلف بصورت مدل زیر خواهد بود:

$$y_j = X_j \tau_j + Z_j u_j + e_j \quad (1)$$

در اینجا X_j ، ماتریس طرح برای صفت j است که مشاهدات را به اثر عوامل ثابت مرتبط می‌سازد؛ τ ، بردار اثر عوامل ثابت برای صفت j ؛ تعداد q_j اثر عامل ثابت با صفت j مرتبط هستند، بطوریکه X_j و τ_j به ترتیب دارای ابعاد $n \times q_j$ و $1 \times q_j$ برای هر صفت است؛ u_j ، بردار آثار تصادفی مرتبط با صفت j با ابعاد $1 \times b_j$ است؛ Z_j ، ماتریس طرح برای صفت j است که مشاهدات را به آثار تصادفی مرتبط می‌سازد. با فرض اینکه فقط یک مشاهده برای هر صفت روی هر فرد وجود دارد، $Z_j = I$ بوده و مدل مختلط را می‌توان بصورت زیر نوشت:

مدل (۲)

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_p \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_p \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix}$$

فرض می‌شود که توزیع مشترک آثار تصادفی و خطاهای تصادفی (u_j, e_j) توزیع نرمال با میانگین صفر و ماتریس واریانس-کوواریانس زیر است:

مدل (۳)

$$\theta_j \begin{pmatrix} G_j(\gamma) & 0 \\ 0 & R_j(\phi) \end{pmatrix}$$

در اینجا θ_j پارامتر مقیاس بوده که در حالت چند صفتی بصورت عدد ثابت یک در نظر گرفته می‌شود. اما در مدلهایی که

فقط یک واریانس خطا وجود دارد، θ_j برابر واریانس باقیمانده (σ_e^2) است. γ و ϕ بردارهای حاوی پارامترهای واریانس هستند.

با مجزا نمودن نمودن آثار تصادفی نشانگرها (آثار QTL) و سایر آثار تصادفی از جمله آثار پلی ژنیک، مدل ۱ بصورت مدل زیر در می‌آید:

$$y_j = X_j b_j + Z_{0j} u_{0j} + Z_{gj} u_{gj} + e_j \quad (4)$$

در این مدل، y_j ؛ داده‌های فنوتیپی مربوط به j امین صفت، X_j ؛ ماتریس طرح مربوط به اثر عوامل ثابت، b_j ؛ بردار اثر عوامل ثابت، Z_{0j} ؛ ماتریس طرح برای آثار تصادفی غیر از نشانگرها (u_{0j}) که ماتریس واریانس آن برابر G_0 است، u_{gj} ؛ بردار آثار تصادفی نشانگرها برای صفت j ؛ e_j ؛ بردار خطاهای آزمایشی برای صفت j ؛ ماتریس اثر متقابل نشانگر و صفت برابر است با حاصل ضرب کرومکس ماتریس نشانگر (G_p) در یک ماتریس واحد (I_m). در اینجا عناصر قطری G_p واریانس نشانگرها برای هر صفت و عناصر خارج قطری آن کوواریانس نشانگرها بین هر جفت صفت هستند. ساختارهای کوواریانس متفاوتی را برای اثر متقابل نشانگر و صفت (u_{gj}) می‌توان در نظر گرفت. ساختار کوواریانس مورد استفاده در تحقیق حاضر برای اثر متقابل نشانگر و صفت (u_{gj})، مدل تحلیل عاملی بود. تجزیه عاملی در واقع یکی از روش‌های تجزیه چند متغیره است که اولین بار توسط روانشناسان به منظور مطالعات روانشناسی مورد استفاده قرار گرفت (۱۴). از این تکنیک به منظور توصیف ساختار کوواریانس تعداد زیادی متغیر بر حسب تعداد اندکی از کمیت‌های تصادفی غیرقابل مشاهده بنام "عامل" استفاده می‌شود. تحلیل عاملی مبتنی بر یک مدل آماری بوده و مدل تحلیل عاملی از لحاظ عامل‌های مشترک بصورت خطی است (۹).

مدل تحلیل عاملی برای اثر متقابل نشانگر و صفت بصورت مدل زیر بود:

$$u_g = (\lambda_1 \otimes I_m) f_1 + \dots + (\lambda_k \otimes I_m) f_k + \delta \quad (5)$$

در مدل ۵، f_1, \dots, f_k کمیت‌های تصادفی هستند که عامل نامیده می‌شوند. ضرایب λ عبارتند از بار عامل‌ها، δ عبارت است از

آورده و آنها را به لگاریتم نسبت احتمالات (LOD^۵) تبدیل نمود (۲۰). نقطه‌ای از ژنوم که بالاترین رتبه LOD را برای عامل دارد، محتمل‌ترین موقعیت QTL با اثر پلیوتروپی می‌باشد. نشانگر متصل به این QTL بصورت اثر ثابت در مدل وارد می‌شود و مدل تحلیل عاملی مجدداً در نظر گرفته می‌شود. فرآیند فوق تکرار می‌شود تا زمانی که مدل تحلیل عاملی معنی‌دار نشود. واضح است که اگر مدل تحلیل عاملی در اولین مرحله از برازش مدل معنی‌دار نشود بیانگر عدم وجود QTL با اثر مشترک روی صفات مختلف می‌باشد.

مطالعه مدل با استفاده از داده‌های شبیه‌سازی شده

به منظور بررسی کارایی مدل، در شناسای QTL های پلیوتروپی، حجم زیادی از داده‌ها شبیه‌سازی و مورد مطالعه قرار گرفتند. پارامترهای شبیه‌سازی شامل دو بخش، پارامترهای مربوط به صفت کمی و پارامترهای مربوط به اطلاعات نقشه‌یابی بود. در این تحقیق ۵ صفت کمی دارای توزیع نرمال و واریانس خطای برابر واحد در نظر گرفته شدند. به هر فرد از جمعیت، یک مقدار تصادفی از این توزیع اختصاص داده شد. یک ژنوم متشکل از ۵ کروموزوم هر کدام به طول ۱۲۰ سانتی مورگان در نظر گرفته شد. چهار سطح تراکم نشانگر مورد مطالعه قرار گرفت که عبارت بودند از تراکم کم (هر ۲۰ سانتی مورگان یک نشانگر بعبارت دیگر ۷ نشانگر روی هر کروموزوم و ۳۵ نشانگر در طول ژنوم)، تراکم متوسط (هر ۱۰ سانتی مورگان یک نشانگر بعبارت دیگر ۱۳ نشانگر روی هر کروموزوم و ۶۵ نشانگر در طول ژنوم)، تراکم زیاد (هر ۵ سانتی مورگان یک نشانگر بعبارت دیگر ۲۵ نشانگر روی هر کروموزوم و ۱۲۵ نشانگر در طول ژنوم) و تراکم خیلی زیاد (هر یک سانتی مورگان یک نشانگر بعبارت دیگر ۱۲۱ نشانگر روی هر کروموزوم و ۶۰۵ نشانگر در طول ژنوم). از لحاظ موقعیت QTL موثر بر صفات مختلف سه حالت در نظر گرفته شد. وضعیت اول، اثر پلیوتروپی QTL روی ۲ صفت بود که یک QTL روی کروموزوم ۱ (۲۰ سانتی مورگان) روی صفات ۱ و ۲ تاثیر داشت. برای هر کدام از صفات ۲ و ۳ یک QTL به ترتیب روی کروموزوم‌های ۲ (۲۰ سانتی مورگان) و ۳ (۴۰ سانتی

بردار آثار باقیمانده. مدل ۵ به شکل ماتریسی بصورت زیر خواهد بود:

$$\mathbf{u}_g = (\mathbf{\Lambda} \otimes \mathbf{I}_m) \mathbf{f} + \boldsymbol{\delta} \quad (\text{مدل ۶})$$

در این مدل، \mathbf{f} بردار مربوط به عامل ها، $\mathbf{\Lambda}$ ماتریس مربوط به بار عامل‌ها، $\boldsymbol{\delta}$ بردار آثار باقیمانده یا عدم برازش برای مدل تحلیل عاملی است. بنابراین، پارامترهای مربوط به مدل ۴ (e، $\boldsymbol{\delta}$ و \mathbf{f}) دارای توزیع چندمتغیره نرمال با میانگین صفر و ماتریس واریانس-کوواریانس زیر بود:

$$\begin{pmatrix} \mathbf{I}_K \otimes \mathbf{I}_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \otimes \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{pmatrix} \quad (\text{مدل ۷})$$

در اینجا $\boldsymbol{\Psi}$ ماتریس قطری است که عناصر قطری آن واریانس اختصاصی برای صفت i هستند. برای برآورد پارامترهای مربوط به مدل تحلیل عاملی (بار عامل‌ها و مولفه‌های واریانس آثار باقیمانده) از الگوریتم پیشنهادی توسط تامپسون و همکاران (۱۹) استفاده شد. به‌منظور توسعه یک مدل برای آثار پلیوتروپی، اولین مرحله برازش مدل ۴ بود که در آن تمامی نشانگرها بطور همزمان بصورت آثار تصادفی وارد شده بودند. دو نوع ساختار کوواریانس برای G_p در نظر گرفته شد (مدل قطری یا DIAG^۳ و مدل تحلیل عاملی با یک عامل یا FA1^۴). مدل DIAG بیانگر این مطلب است که صفات از لحاظ اثر نشانگرها مستقل از هم بوده و واریانس ناشی از آثار نشانگرها برای آنها متفاوت است. برای مقایسه مدل FA1 و مدل DIAG، از آماره آزمون نسبت درست‌نمایی (دو برابر اختلاف لگاریتم درست‌نمایی) استفاده شد. این آزمون دارای توزیع مربع کای بوده و درجه آزادی آن برابر اختلاف تعداد پارامترهای دو مدل است. با استفاده از پارامترهای برآورد شده، بهترین پیش-بینی نارایب خطی (BLUP) از اثر نشانگرها برای هر عامل بدست آمد. با تقسیم BLUP به انحراف معیار خطای پیش‌بینی برای هر نشانگر مقادیری بدست می‌آیند که دارای توزیع نرمال هستند. بنابراین، می‌توان احتمال انحراف اثر نشانگر از صفر را بدست

³ Diagonal model

⁴ Factor analysis model

⁵ Logarithm of odds

نتایج

نقشه یابی درون فاصله ای-تجزیه چند صفتی در این تحقیق به منظور حصول اطمینان از شبیه سازی صحیح داده های چندصفتی QTL، در ابتدا تجزیه توام صفات بر اساس نقشه یابی درون فاصله ای ارائه شده توسط جیانگ و زنگ (۸) که مبتنی بر روش حداکثر درستنمایی است، انجام شد. مقدار عددی نسبت لگاریتم درستنمایی (LOD) بدست آمده برای موقعیت QTL در حالت تجزیه توام صفات نسبت به تجزیه تک صفتی بیشتر بود و بنابراین منجر به شناسایی بهتر موقعیت QTL گردید. این اثر خصوصاً زمانی که QTL اثر پلیوتروپی روی چندین صفت داشت، کاملاً مشهود بود. بهرحال اشکال اساسی این روش تجزیه این است که مدل مورد استفاده، مدل تک ژنی است. یعنی در هر مرتبه تجزیه، فرض بر آن است که فقط یک QTL بر هر صفت موثر است و تمامی QTL ها را بطور همزمان مدل سازی نمی کند. این می تواند منجر به اریبی در برآورد میزان اثر و موقعیت QTL شود.

مورگان) شبیه سازی شد. برای صفت ۴ یک QTL روی کروموزوم ۴ (۲۰ سانتی مورگان) در نظر گرفته شد. برای صفت ۵ تعداد یک QTL روی کروموزوم ۵ (۴۰ سانتی مورگان) شبیه سازی شد (جدول ۱). وضعیت دوم، اثر پلیوتروپی QTL روی ۳ صفت بود. این حالت شبیه وضعیت اول بود ولی QTL شبیه سازی شده روی کروموزوم ۱ (۲۰ سانتی مورگان) روی صفات ۱، ۲ و ۳ تأثیر داشت. در وضعیت سوم، اثر پلیوتروپی QTL روی ۴ صفت در نظر گرفته شد. این حالت شبیه وضعیت اول بود اما QTL شبیه سازی شده روی کروموزوم ۱ (۲۰ سانتی مورگان) اثر پلیوتروپی روی صفات ۱، ۲، ۳ و ۴ داشت. جمعیت شبیه سازی شده یک جمعیت حاصل از تلاقی برگشتی بود. تمام وضعیت های سه گانه فوق برای سطوح مختلف تراکم نشانگر و چهار اندازه نمونه مختلف ۱۰۰، ۲۰۰، ۴۰۰ و ۶۰۰ و به تعداد ۱۰۰ تکرار شبیه سازی شدند. توارث تمامی جایگاه های نشانگرها با فرض جور شدن تصادفی آلل ها و استقلال وقایع نوترکیبی از یکدیگر تعیین شد. بنابراین، تابع نقشه مورد استفاده تابع هالدن (۷) بود.

جدول ۱- پارامترهای استفاده شده برای شبیه سازی QTL در وضعیت اثر پلیوتروپی روی دو صفت

صفت ۱	صفت ۲	صفت ۳	صفت ۴	صفت ۵	
۱ (۲۰)	۱ (۲۰)	-	۴ (۲۰)	-	پارامترهای QTL
۰/۵	۰/۷	-	۰/۷	-	موقعیت *
-	-	-	-	-	اندازه اثر **
-	۲ (۲۰)	۳ (۴۰)	-	۵ (۴۰)	پارامترهای QTL
-	۰/۷	۰/۷	-	۰/۷	موقعیت
-	-	-	-	-	اندازه اثر
%۵	%۲۰	%۱۰	%۱۰	%۱۰	وراثت پذیری ***

* عدد داخل پرانتز موقعیت QTL بر حسب سانتی مورگان و عدد خارج پرانتز شماره کروموزوم است.

** اندازه اثر QTL به واحد انحراف معیار فنوتیپی.

*** بیانگر بخشی از واریانس فنوتیپی در جمعیت حاصل از تلاقی برگشتی که ناشی از واریانس اثر کل QTL های موثر بر هر صفت است.

جدول ۲- توانایی آماری* مشاهده شده برای آزمون مدل پلیوتروپی

وضعیت پلیوتروپی				فاصله نشانگرها**	اندازه نمونه
چهار صفت	سه صفت	دو صفت	سه صفت		
۷۳	۶۲	۵۹	۲۰	۱۰۰	
۸۰	۷۱	۶۷	۱۰		
۸۳	۷۷	۷۲	۵		
۸۶	۷۹	۷۴	۱		
۸۸	۸۳	۷۸	۲۰	۲۰۰	
۹۲	۸۷	۸۱	۱۰		
۹۵	۸۹	۸۵	۵		
۹۶	۹۱	۸۹	۱		
۹۴	۹۱	۸۵	۲۰	۴۰۰	
۹۵	۹۳	۸۷	۱۰		
۹۷	۹۵	۹۰	۵		
۹۹	۹۴	۹۲	۱		
۹۷	۹۵	۹۱	۲۰	۶۰۰	
۹۸	۹۷	۹۳	۱۰		
۱۰۰	۹۷	۹۴	۵		
۱۰۰	۹۹	۹۴	۱		

* درصد تکرارهایی که در آنها مدل تحلیل عاملی معنی دار بود ($P < 0.05$).

** فاصله نشانگرها بر حسب سانتی مورگان.

اختصاصی ژنوم (یک مولفه واریانس اختصاصی ژنوم برای هر صفت) و ۵ بار عامل. از آنجا که میزان اثر شبیه سازی شده QTL برای صفت ۱ کم بود، واریانس ناشی از آثار ژنوم برای این صفت کمتر از سایر صفات بود. از طرفی برای صفت ۵ فقط یک QTL اختصاصی در نظر گرفته شده بود لذا بار عامل روی این صفت در همه حالات شبیه سازی شده نزدیک به صفر بود.

توانایی آزمون پلیوتروپی

در این تحقیق آزمون پلیوتروپی عبارت بود از مقایسه مستقیم دو مدل (مدل قطری و مدل کوواریانس تحلیل عاملی) برای آثار نشانگرها در بین صفات مختلف با استفاده از آزمون نسبت حداکثر درستنمایی باقیمانده. در واقع دو برابر تفاوت لگاریتم نسبت درستنمایی دو مدل فوق دارای توزیع مربع کای با درجه آزادی برابر با تفاوت تعداد پارامترهای دو مدل می باشد. معنی دار شدن مدل تحلیل عاملی در مقایسه با مدل قطری، بیانگر وجود QTL با اثر پلیوتروپی بود. بنابراین، در این تحقیق توانایی مدل تحلیل عاملی در شناسایی QTL پلیوتروپی (کارایی آزمون پلیوتروپی) عبارت بود از درصد تکرارهایی که در آنها مدل

مدل تحلیل عاملی تصادفی

تخمین پارامترهای مدل

زمانیکه کوواریانس آثار نشانگر در بین صفات در نظر گرفته نشده بود (مدل DIAG برای اثر متقابل نشانگر و صفت) در هر وضعیت برای داده های شبیه سازی شده لازم بود تعداد ۲۰ پارامتر برآورد شود. این پارامترها عبارت بودند از ۱۵ مولفه واریانس- کوواریانس برای آثار باقیمانده (۵ واریانس و ۱۰ کوواریانس باقیمانده بین صفات) و ۵ مولفه واریانس ژنوم (یک مولفه واریانس ژنوم برای هر صفت). لازم به ذکر است که مدل کوواریانس مورد استفاده در این تحقیق برای آثار باقیمانده، مدل غیر ساختاری^۶ بود. زمانیکه کوواریانس آثار نشانگر بین صفات در نظر گرفته شده بود (مدل FAI با یک عامل برای اثر متقابل نشانگر و صفت) در هر وضعیت برای داده های شبیه سازی شده لازم بود تعداد ۲۵ پارامتر برآورد شود. این پارامترها عبارت بودند از ۱۵ مولفه واریانس- کوواریانس برای آثار باقیمانده (۵ واریانس و ۱۰ کوواریانس باقیمانده بین صفات) و ۵ مولفه واریانس

⁶ Unstructured

برای هر صفت. بر اساس این تعریف، در تمام سه وضعیت شبیه-سازی شده، با افزایش اندازه نمونه و تراکم نشانگر توانایی مدل در شناسایی موقعیت صحیح QTL، افزایش یافت (جدول ۳ تا ۵). افزایش اثر پلیوتروپی QTL از دو صفت به چهار صفت توانایی مدل در شناسایی موقعیت QTL را برای همه صفات به غیر از صفت ۵ افزایش داد. این مورد به خصوص برای صفت ۱ قابل توجه بود. QTL های شبیه سازی شده برای صفت ۵ هیچ ارتباطی با QTL های سایر صفات نداشت. توانایی مدل در شناسایی موقعیت صحیح QTL برای این صفت در هر سه وضعیت تقریباً "یکسان بود.

کوواریانس تحلیل عاملی برای اثر متقابل صفت و نشانگر معنی دار بوده است. بر اساس تعریف فوق، توانایی آزمون پلیوتروپی به میزان زیادی به اندازه نمونه و تا حدودی به تراکم نشانگر بستگی داشت بطوریکه با افزایش تراکم نشانگر در طول ژنوم، توانایی مدل افزایش می یافت. از طرفی با افزایش اثر پلیوتروپی QTL از ۲ صفت تا ۴ صفت، کارایی مدل افزایش یافت (جدول ۲).

توانایی مدل در شناسایی موقعیت QTL

توانایی مدل مورد استفاده در این تحقیق در شناسایی موقعیت QTL روی ژنوم، برابر بود با نسبت QTL هایی که موقعیت آنها بطور صحیح شناسایی شده، به کل QTL های شبیه سازی شده

جدول ۳- توانایی * مدل تحلیل عاملی در شناسایی موقعیت QTL در حالت اثر پلیوتروپی بر دو صفت

اندازه نمونه	فاصله نشانگرها**	صفت ۱	صفت ۲	صفت ۳	صفت ۴	صفت ۵
۱۰۰	۲۰	۰/۳۲	۰/۵۰	۰/۳۹	۰/۳۸	۰/۳۵
	۱۰	۰/۴۱	۰/۶۱	۰/۴۸	۰/۴۶	۰/۴۹
	۵	۰/۴۲	۰/۶۲	۰/۴۹	۰/۴۵	۰/۵۰
	۱	۰/۴۳	۰/۶۳	۰/۵۱	۰/۴۵	۰/۵۱
۲۰۰	۲۰	۰/۵۷	۰/۷۴	۰/۶۵	۰/۶۴	۰/۶۶
	۱۰	۰/۶۳	۰/۸۵	۰/۷۱	۰/۷۳	۰/۷۱
	۵	۰/۶۵	۰/۸۸	۰/۷۶	۰/۷۵	۰/۷۷
	۱	۰/۶۵	۰/۹۰	۰/۷۸	۰/۷۹	۰/۷۸
۴۰۰	۲۰	۰/۷۵	۰/۸۴	۰/۷۲	۰/۷۱	۰/۶۹
	۱۰	۰/۸۱	۰/۸۹	۰/۷۸	۰/۸۰	۰/۷۸
	۵	۰/۸۳	۰/۹۰	۰/۸۰	۰/۷۹	۰/۸۱
	۱	۰/۸۵	۰/۹۲	۰/۸۰	۰/۸۲	۰/۸۳
۶۰۰	۲۰	۰/۸۱	۰/۹۱	۰/۷۷	۰/۷۸	۰/۷۷
	۱۰	۰/۸۸	۰/۹۵	۰/۸۵	۰/۸۳	۰/۸۴
	۵	۰/۹۰	۰/۹۶	۰/۸۶	۰/۸۵	۰/۸۴
	۱	۰/۹۰	۰/۹۷	۰/۸۸	۰/۸۷	۰/۸۸

* نسبت QTL هایی که موقعیت آنها بطور صحیح شناسایی شده به کل QTL های شبیه سازی شده برای هر صفت.

** فاصله نشانگرها بر حسب سانتی مورگان.

جدول ۴- توانایی* مدل تحلیل عاملی در شناسایی موقعیت QTL در حالت اثر پلیوتروپی بر سه صفت

اندازه نمونه	فاصله نشانگرها**	صفت ۱	صفت ۲	صفت ۳	صفت ۴	صفت ۵
۱۰۰	۲۰	۰/۴۲	۰/۶۲	۰/۵۹	۰/۴۰	۰/۳۶
	۱۰	۰/۵۰	۰/۷۳	۰/۷۱	۰/۴۵	۰/۴۸
	۵	۰/۵۲	۰/۷۸	۰/۷۴	۰/۴۵	۰/۵۲
	۱	۰/۵۳	۰/۷۹	۰/۷۷	۰/۴۶	۰/۵۴
۲۰۰	۲۰	۰/۶۵	۰/۷۹	۰/۷۸	۰/۶۳	۰/۶۵
	۱۰	۰/۷۰	۰/۸۷	۰/۸۳	۰/۷۴	۰/۷۳
	۵	۰/۷۲	۰/۸۹	۰/۸۵	۰/۷۴	۰/۷۵
	۱	۰/۷۱	۰/۹۲	۰/۸۵	۰/۷۷	۰/۷۷
۴۰۰	۲۰	۰/۸۰	۰/۸۹	۰/۸۱	۰/۷۳	۰/۷۰
	۱۰	۰/۸۵	۰/۹۳	۰/۸۷	۰/۷۹	۰/۷۸
	۵	۰/۸۷	۰/۹۳	۰/۹۰	۰/۸۱	۰/۷۹
	۱	۰/۸۷	۰/۹۵	۰/۹۰	۰/۸۲	۰/۸۲
۶۰۰	۲۰	۰/۸۳	۰/۹۳	۰/۹۱	۰/۸۰	۰/۷۹
	۱۰	۰/۸۷	۰/۹۶	۰/۹۵	۰/۸۴	۰/۸۵
	۵	۰/۹۰	۰/۹۹	۰/۹۸	۰/۸۷	۰/۸۶
	۱	۰/۹۱	۱/۰۰	۰/۹۹	۰/۸۸	۰/۸۶

* نسبت QTL هایی که موقعیت آنها بطور صحیح شناسایی شده به کل QTL های شبیه سازی شده برای هر صفت.

** فاصله نشانگرها بر حسب سانتی مورگان.

جدول ۵- توانایی* مدل تحلیل عاملی در شناسایی موقعیت QTL در حالت اثر پلیوتروپی بر چهار صفت

اندازه نمونه	فاصله نشانگرها**	صفت ۱	صفت ۲	صفت ۳	صفت ۴	صفت ۵
۱۰۰	۲۰	۰/۵۱	۰/۷۰	۰/۷۱	۰/۷۴	۰/۳۹
	۱۰	۰/۶۰	۰/۸۱	۰/۸۰	۰/۸۲	۰/۴۷
	۵	۰/۶۲	۰/۸۵	۰/۸۳	۰/۸۱	۰/۴۹
	۱	۰/۶۲	۰/۸۶	۰/۸۵	۰/۸۴	۰/۴۹
۲۰۰	۲۰	۰/۶۹	۰/۸۵	۰/۸۴	۰/۸۳	۰/۶۷
	۱۰	۰/۷۵	۰/۹۱	۰/۸۷	۰/۸۸	۰/۷۱
	۵	۰/۷۷	۰/۹۴	۰/۹۰	۰/۸۹	۰/۷۴
	۱	۰/۷۷	۰/۹۴	۰/۹۲	۰/۹۱	۰/۷۶
۴۰۰	۲۰	۰/۸۰	۰/۸۹	۰/۹۱	۰/۹۲	۰/۷۲
	۱۰	۰/۸۵	۰/۹۳	۰/۹۵	۰/۹۵	۰/۷۹
	۵	۰/۸۷	۰/۹۳	۰/۹۷	۰/۹۸	۰/۸۰
	۱	۰/۸۷	۰/۹۵	۰/۹۶	۰/۹۸	۰/۸۲
۶۰۰	۲۰	۰/۸۶	۰/۹۶	۰/۹۵	۰/۹۵	۰/۷۸
	۱۰	۰/۸۹	۰/۹۸	۰/۹۸	۰/۹۹	۰/۸۲
	۵	۰/۹۲	۱/۰۰	۰/۹۹	۱/۰۰	۰/۸۱
	۱	۰/۹۴	۱/۰۰	۱/۰۰	۱/۰۰	۰/۸۳

* نسبت QTL هایی که موقعیت آنها بطور صحیح شناسایی شده به کل QTL های شبیه سازی شده برای هر صفت.

** فاصله نشانگرها بر حسب سانتی مورگان.

بحث

در این تحقیق، یک روش مبتنی بر مدل‌های مختلط خطی برای شناسایی QTL دارای اثر پلیوتروپی ارائه شد. در روش پیشنهادی، مدل مورد استفاده برای کوواریانس آثار نشانگرها در بین صفات، مدل تحلیل عاملی و مدل مورد استفاده برای مدل سازی آثار غیر ژنتیکی بین صفات (باقیمانده‌های مدل)، مدل غیرساختاری است. تجزیه و تحلیل داده‌های شبیه سازی شده با استفاده از این روش نشان داد که این روش دارای کارایی زیادی در شناسایی QTL پلیوتروپی است. تجزیه عاملی، یکی از روش‌های آماری چند متغیره است که فرض آن این است که داده‌های مشاهده شده تحت تاثیر تعداد اندکی عامل مشترک هستند (۹). فرض تحقیق حاضر بر این بود که در تجزیه چند صفتی QTL، این عوامل مشترک همان QTL های دارای اثر پلیوتروپی هستند.

برای مدل سازی چندصفتی خصوصاً در جوامع آزمایشی با اندازه کوچک که معمولاً در بررسی QTL استفاده می‌شوند، استفاده از روش‌های آماری با هدف کاهش تعداد پارامترهای مورد نیاز برای برازش مدل، اهمیت زیادی دارد (۱۵، ۱۶). در روش مورد استفاده در این تحقیق، ساختار خاص مورد استفاده برای کوواریانس آثار نشانگر در بین صفات مختلف، مدل تحلیل عاملی بود. هدف از استفاده از این ساختار، مدل سازی کوواریانس آثار نشانگرها در بین صفات بر حسب یک عامل و در نتیجه کاهش تعداد پارامترها در هر مرتبه از برازش مدل بود. از آنجا که روش پیشنهادی مبتنی بر مدل‌های مختلط خطی است، با مقایسه مدل DIAG مبنی بر عدم همبستگی آثار نشانگرها در بین صفات و مدل FA1 مبنی بر وجود همبستگی آثار نشانگرها، اهمیت کوواریانس آثار نشانگرها در بین صفات قابل آزمون بود. نتایج حاصل از داده‌های شبیه سازی شده نشان داد که مدل فوق توانایی بالایی خصوصاً زمانیکه اندازه نمونه نسبتاً زیاد است، در شناسایی QTL پلیوتروپی دارد. از آنجا که همبستگی بین آثار یک ژن دارای اثر پلیوتروپی برابر +۱ یا -۱ است (۶)، استفاده از مدل تحلیل عاملی برای مدل سازی QTL پلیوتروپی، مدل مناسبی بنظر می‌رسد.

مدل تحلیل عاملی برای مدل سازی آثار متقابل ژنوتیپ و محیط در تجزیه داده‌های آزمایش‌های مربوط به بررسی عملکرد واریته-

های گیاهان زراعی در محیط‌های مختلف مورد استفاده قرار گرفته است (۱، ۱۷، ۱۸). هدف اصلی از این تجزیه‌ها، برآورد صحیح و دقیق میانگین عملکرد واریته‌ها و کمک به تفسیر و فهم اثر متقابل واریته \times محیط است (۱۷). بنابراین، در این آزمایش‌ها فقط عوامل مشترک دارای اهمیت هستند اما در مدل سازی اثر متقابل نشانگر \times صفت با استفاده از مدل تحلیل عاملی در تحقیق حاضر، هم عوامل مشترک و هم عوامل اختصاصی دارای اهمیت می‌باشند. از نقطه نظر اصلاح ژنتیکی، عوامل مشترک (QTL های پلیوتروپی) برای استفاده در برنامه‌های انتخاب به کمک نشانگر اهمیت زیادی دارند. چون در صورت وجود همبستگی نشانگر با تعدادی از صفات، انتخاب بر اساس نشانگر، منجر به تغییر ژنتیکی در همه صفات خواهد شد. علاوه بر این، شناسایی QTL های پلیوتروپی برای بررسی ماهیت همبستگی ژنتیکی بین صفات نیز اهمیت دارد. از طرفی، شناسایی ژن‌های دارای اثر اختصاصی بر صفات در طراحی برنامه‌های بهبود ژنتیکی زمانیکه هدف انتخاب تغییر ژنتیکی یک صفت بدون تغییر صفت دیگر است، مهم می‌باشد.

اگر بعد از برازش مدل تحلیل عاملی، واریانس ژنتیکی اختصاصی ناشی از QTL اختصاصی (QTL هایی که فقط یک صفت خاص را تحت تاثیر قرار می‌دهند) برای برخی صفات معنی‌دار باشد، می‌توان نتیجه گرفت که علاوه بر QTL پلیوتروپی، جایگاه‌های ژنی اختصاصی نیز بر این صفات تاثیرگذار هستند. زمانی که یک یا چند QTL پلیوتروپی وجود داشته باشند و QTL با اثر اختصاصی وجود نداشته باشد، آنگاه برازش مدل تحلیل عاملی برای آثار QTL منجر به برآورد واریانس صفر برای واریانس ژنتیکی اختصاصی برای این صفات خواهد شد. در این حالت خاص، ساختار واریانس و کوواریانس مدل تحلیل عاملی دارای رتبه کامل نخواهد بود و استفاده از روش‌های استاندارد برآورد مولفه‌های واریانس مثل روش مبتنی بر اطلاعات متوسط^۷ (۵) امکان پذیر نخواهد بود. خصوصیت ویژه الگوریتم تامپسون و همکاران (۱۹) که در این تحقیق برای برآورد پارامترهای مدل تحلیل عاملی مورد استفاده قرار گرفت این است که امکان منظور نمودن مقدار عددی صفر را برای واریانس اختصاصی صفات

⁷ Average Information

نتیجه‌گیری

شناسایی جایگاه‌های ژنی دارای آثار پلیوتروپی اهمیت زیادی در طراحی برنامه‌های اصلاحی و بهبود ژنتیکی صفات همبسته دارد. از طرفی شناسایی جایگاه‌های ژنی اختصاصی برای صفات مختلف زمانی که هدف اصلاح نژاد تغییر یک صفت بدون تاثیر بر صفات دیگر است، دارای اهمیت است. نتایج این تحقیق نشان داد که برای تمایز ژنهای اختصاصی از ژنهای دارای اثر پلیوتروپی، مدل تحلیل عاملی مدل مناسبی است.

سپاسگزاری

هزینه انجام این تحقیق از محل اعتبار ویژه پژوهشی دانشگاه شهید باهنر کرمان به نگارنده مقاله تامین گردیده است.

منابع

1. Cullis BR, Gogel BJ, Verbyla AP, and Thompson R (1998) Spatial analysis of multi-environment early generation trials. *Biometrics*, 54: 1-18.
2. Falconer DS, and Mackay TFC (1996) *Introduction to quantitative genetics*. 4 ed. Longmans Green, Harlow, Essex, UK, 467 pp.
3. Gilbert H, and Le Roy P (2003) Comparison of three multitrait methods for QTL detection. *Genetics Selection Evolution*, 35: 281-304.
4. Gilbert H, and Le Roy P (2004) Power of three multitrait methods for QTL detection in crossbred populations. *Genetics Selection Evolution*, 36: 347-361.
5. Gilmour AR, Thompson R, and Cullis BR (1995) Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 51:1440-1450.
6. Goddard ME (2001) The validity of genetic models underlying quantitative traits. *Livestock Production Science*, 72:117-127.
7. Haldane JBS (1919) The combination of linkage values, and the calculation of distance between linked factors. *Journal of Genetics*, 8: 299-309.
8. Jiang C, and Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics*, 140: 1111-1124.
9. Johnson RA, and Wichern DW (1998) *Applied multivariate analysis*. 4 Ed. Prentice Hall, Englewood Cliffs, New Jersey, 594 pp.
10. Knott SA (2005) Regression-based quantitative trait loci mapping: Robust, efficient and effective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360: 1435-1442.

فراهم می‌کند. بنابراین، مزیت استفاده از این روش علاوه بر همگرایی سریع مدل تحلیل عاملی، امکان برآورد پارامترهای مدل‌هایی است که رتبه آنها کاهش یافته^۸ است.

یکی دیگر از مزایای روش مورد استفاده در تحقیق حاضر، امکان تجزیه همزمان تعداد زیادی صفت است. همچنین به دلیل استفاده از مدل‌های مختلط به جای مدل‌های ثابت، آثار کل نشانگرهای ژنوم بطور همزمان در مدل وارد می‌شوند. این موضوع می‌تواند در برآورد صحیح‌تر موقعیت و میزان اثر QTL موثر باشد. حداکثر درست‌نمایی باقیمانده (REML^۹) در برآورد مولفه‌های واریانس و بهترین پیش‌بینی ناریب خطی (BLUP^{۱۰}) سال‌های متمادی است که برای برآورد پارامترهای ژنتیکی و پیش‌بینی ارزش‌های اصلاحی و انتخاب ژنتیکی استفاده می‌شوند. یکی دیگر از ویژگی‌های این تحقیق نسبت به سایر تحقیقات در این زمینه، استفاده از این تکنیک‌ها به منظور تجزیه داده‌های مولکولی بود.

بر اساس فالکونر و مک کی (۲)، منابع همبستگی ژنتیکی شامل پلیوتروپی و پیوستگی ژنتیکی (یعنی دو ژن موثر بر صفات مجزا در فاصله نزدیکی روی یک کروموزوم واقع شده‌اند و لذا این ژن‌ها بطور مستقل در فرآیند میوز از هم تفرق حاصل نمی‌کنند) هستند. در این تحقیق، کارایی مدل در تمایز QTL‌های پیوسته و پلیوتروپی مورد بررسی قرار نگرفت و لذا این جنبه از تحقیق نیاز به بررسی و مطالعه دارد. علاوه بر این، همبستگی محیطی بین صفات می‌تواند بر نتایج مدل‌سازی و شناسایی QTL‌های پلیوتروپی تاثیر داشته باشد (۱۱). اگر چه در روش ارائه شده در این تحقیق، مدل‌سازی همبستگی محیطی (آثار باقیمانده) بین صفات با استفاده از مدل کوواریانس غیر ساختاری بطور همزمان با مدل‌سازی آثار نشانگرها در بین صفات قابل انجام است و از نظر آماری تمام اطلاعات موجود بطور همزمان استفاده می‌شوند اما اثر همبستگی محیطی متفاوت بین صفات بر کارایی مدل در شناسایی QTL پلیوتروپی نیاز به تحقیق دارد.

^۸ Reduced Ranked models

^۹ Residual maximum likelihood

^{۱۰} Best linear unbiased prediction

11. Knott SA, and Haley CS (2000) Multitrait least squares for quantitative trait loci detection. *Genetics*, 156: 899-911.
12. Korol AB, Ronin YI, Itskovich AM, Peng J, and Nevo E (2001) Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. *Genetics*, 157: 1789-1803.
13. Lander ES, and Botstein D (1989) Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics*, 121: 185-199.
14. Lawley DN, and Maxwell AE (1971) Factor analysis as a statistical method. 2nd ed. Butterworth, London.
15. Piepho H-P (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. *Genetics*, 156: 2043-2050.
16. Sillanpaa MJ, and Corander J (2002) Model choice in gene mapping: What and why. *Trends in Genetics*, 18: 301-307.
17. Smith A, Cullis B, and Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, 57: 1138-1147.
18. Smith AB, Cullis BR, and Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *Journal of Agricultural Science*, 143: 1-14.
19. Thompson R, Cullis BR, Smith A, and Gilmour A (2003) A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Australian & New Zealand Journal of Statistics*, 45: 445-459.
20. Verbyla AP, Eckermann PJ, Thompson R, and Cullis BR (2003) The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. *Australian Journal of Agricultural Research*, 54: 1395-1408.
21. Weller JI, Wiggans GR, VanRaden PM, and Ron M (1996) Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *Theoretical Applied Genetics*, 92: 998-1002.

