ژنتیک نوین
دوره هشتم، شماره ۳، پاییز ۱۳۹۲
صفحه ۳۰۴- ۲۹۹

مکان‌یابی نواحی ژنومی کنترل کننده صفات کمی با استفاده همزمان از نشانگرهای کل ژنوم در جمعیت‌های حاصل از تلاقی لاین‌های خالص

Mapping quantitative trait loci incorporating the markers of entire genome in populations derived from inbred line crosses

علی اسمعیلی زاده کشکوئیه*

دانشیار، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، ایران

Esmailizadeh Koshkoiyeh A*

Associate Professor, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran

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

(تاریخ دریافت: ۹۱/۳/۳۰- تاریخ پذیرش: ۹۱/۱۰/۲۴)

چکیده

در این تحقیق، یک روش مدل مختلط خطی برای نقشه یابی نواحی ژنومی کنترل کننده صفات کمی (QTL) در تلاقی لاین های خالص پیشنهاد شد که اثرات مرتبط با نشانگر را به صورت تصادفی در نظر می گیرد. این روش دارای دو مرحله است. در مرحله اول، معنی دار بودن واریانس ناشی از اثرات تصادفی نشانگرها (واریانس ژنوم) با استفاده از آزمون نسبت احتمال باقیمانده (REMLRT) مورد آزمون قرار می گیرد. معنی دار بودن واریانس ژنوم بیانگر حضور حداقل یک QTL در سطح ژنوم است. در مرحله دوم، بهترین پیش بینی نااریب خطی (BLUPs) نشانگرها برای مکان یابیQTL استفاده می شود. نشانگر دارای بالاترین مقدار مطلق BLUP در بخش ثابت مدل وارد شده و دو مرحله فوق تکرار می شوند تا زمانیکه مولفه واریانس ژنوم غیرمعنی دار شود. دقت این روش با استفاده از داده های شبیه سازی شده مورد مطالعه قرار گرفت. نتایج نشان داد که این روش دقت بالایی در تشخیص QTL در جوامع با اندازه متوسط و نسبتاً بزرگ حاصل از تلاقی لاین های خالص دارد. اگرچه دقت این روش در تشخیص QTL در نمونه های نسبتاً کوچک کم بود (شبیه سایر روش های موجود)، مزیت اصلی این روش توانایی آن در وارد نمودن همزمان کل نشانگرها در آنالیز در نتیجه غلبه بر مشکل آزمون های چندگانه است.

واژه‌های کلیدی

لاین های خالص، مدل مختلط خطی، نقشه یابی ژنتیکی، QTL

[Downloaded from mg.genetics.ir on 2026-11-03]

[DOR: 20.1001.1.20084439.1392.8.3.9.5]

مقدمه

توسعه انواع نشانگرهای مولکولی به خصوص نشانگرهای تک نوکلئوتیدی (SNPs)، نقشه‌یابی نواحی ژنومی مرتبط با تنوع در صفات کمی (QTL) را میسر نموده است. دستیابی به جمعیت‌های مناسب در حال تفرق، سیستم نشانگرهای ژنتیکی مناسب و تجزیه و تحلیل پیشرفته آماری سه عامل کلیدی در حصول به نتایج مطلوب در مکان‌یابی QTL می‌باشند. با توجه به اینکه انتخاب والدین مناسب و تهیه جمعیت نقشه‌یابی از اصول اولیه و اساسی مکان‌یابی QTL می‌باشد، والدین بایستی از نظر بسیاری از صفات مورد نظر تنوع کافی داشته باشند. در تهیه جمعیت نقشه‌یابی، کاراترین جمعیت‌ها جمعیت‌هایی هستند که از تلاقی بین لاین‌های خالص حاصل شده‌اند.

نقشه‌یابی درون فاصله‌ای[1] مبتنی بر روش حداکثر درستنمایی پیشنهاد شده توسط Lander and Botstein (1989) یا مدل ساده مبتنی بر رگرسیون (Haley and Knott 1992; Martinez and Curnow 1992)، روش های رایج برای شناسایی نواحی ژنومی مرتبط با تنوع در صفات کمی QTL هستند. نتایج یک تحقیق نشان داد که روش نقشه یابی درون فاصله‌ای مبتنی بر رگرسیون ساده فقط زمانی که تعداد اندکیQTL با اثر زیاد در نظر گرفته شوند قادر به تشخیص QTL است و برای شناسایی QTLهای کوچک اثر نیاز به اندازه نمونه‌های بسیار بزرگ است (Esmailizadeh et al. 2008). این روش‌ها در حالتی که تعداد نسبتاً زیادی QTL روی ژنوم قرار دارند، کارایی پایینی در تشخیص و برآورد میزان اثر QTL دارند. از دیگر معایب این روش‌ها، عدم توانایی در تشخیص QTLهایی است که در فاصله نزدیکی نسبت به هم روی یک گروه پیوستگی قرار دارند. برای رفع این مشکلات، روش‌های نقشه‌یابی درون فاصله ای مرکب و نقشه‌یابی QTLهای چندگانه پیشنهاد شده‌اند (Jansen 1993; Zeng 1994). در این روش‌ها، تعدادی از نشانگرها بصورت متغیرهای کمکی خطی در مدل وارد می‌شوند. این متغیرهای کمکی بعنوان QTL های نزدیک به QTL تحت آزمون، با جذب اثرات ژنتیکی این QTLها، این اثرات را از واریانس باقیمانده (خطا) جدا می نمایند. در نتیجه، دقت آنالیز QTL افزایش می یابد. اما مشکل اساسی این روش‌ها، یافتن راهی برای انتخاب متغیرهای کمکی است. از آنجا که این روش‌ها مبتنی بر مدل‌های ثابت خطی هستند، امکان در نظر گرفتن همزمان تمامی نشانگرها را نداشته و لذا در هر مرتبه آنالیز، فقط دسته‌ای از نشانگرها را به عنوان متغیر کمکی در نظر می‌گیرند. اگر تعداد زیادی نشانگر وجود داشته باشد و تمام آن‌ها بطور همزمان در این مدل ها وارد شوند، به دلیل همبستگی زیاد نشانگرها، میزان همپوشانی چندگانه زیاد خواهد بود. از طرفی تعداد نشانگرها ممکن است حتی بیش از تعداد کل افراد آزمایش باشند که در این صورت روش حداقل مربعات معمولی مورد استفاده برای حل معادلات مربوط به مدل ثابت خطی جواب واحدی نخواهد داشت (Xu 2003). در صورت تصادفی منظور نمودن نشانگرها در مدل‌های مختلط، مشکلات مذکور وجود نخواهد داشت. لذا هدف از انجام این تحقیق، توسعه یک روش مبتنی بر مدل های مختلط خطی با در نظر گرفتن همزمان کل نشانگرها به منظور آنالیز QTL بود.

مواد و روش‌ها

با فرض اینکه دو لاین همخون والدی aa و AA که برای جایگاه QTL و نشانگرها هموزیگوت هستند با هم تلاقی یابند و نتاج F_1 با یکی از لاین‌های والدی تلاقی داده شوند (Aa × AA) و داده های ژنوتیپی برای m نشانگر و داده‌های فنوتیپی صفت کمی مورد نظر بر روی n فرد حاصل از تلاقی برگشتی جمع آوری شوند. لذا مدل مختلط خطی برای توصیف این داده‌ها بصورت زیر خواهد بود:

$$y=Xb+Zu+e \qquad [۱]$$

در این مدل، y بردار ستونی حاوی مقادیر فنوتیپی صفت، b بردار اثرات ثابت، u بردار اثرات تصادفی نشانگرها، X و Z به ترتیب ماتریس های طرح برای اثرات ثابت و تصادفی و e بردار ستونی انحرافات باقیمانده هستند.

بر اساس مدل [۱]، تمامی نشانگرها بطور همزمان در مدل بصورت متغیرهای کمکی تصادفی در نظر گرفته شده‌اند. بنابراین، فقط یک مولفه واریانس برای کل نشانگرهای ژنوم برآورد می شود. مولفه‌های واریانس (مولفه واریانس ناشی از نشانگرها (Var_M) و مولفه واریانس باقیمانده (Var_R) با استفاده از الگوریتم

[1] Interval mapping

[Downloaded from mg.genetics.ir on 2026-11-03]

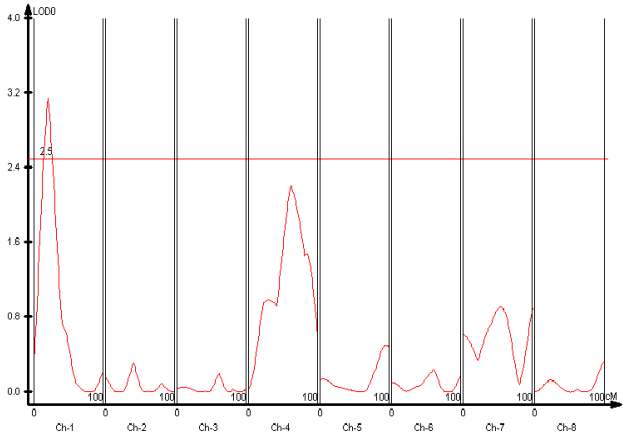
[DOR: 20.1001.1.20084439.1392.8.3.9.5]

AI برآورد شده و پس از حل معادلات مدل مختلط، بهترین پیش بینی نارایب خطی (BLUP) برای تک تک نشانگرها بدست می آید (Gilmour et al. 1995). برای آزمون معنی‌دار بودن مولفه واریانس ناشی از نشانگر، آزمون نسبت حداکثر احتمال باقیمانده (REMLRT) که یک آزمون یک طرفه است مورد استفاده قرار می گیرد. منفی دو برابر تغییر در لگاریتم نسبت احتمال به صورت ترکیبی از توزیع χ_0^2 و χ_1^2 با مقدار آستانه ۲/۷۹ برای α= ۰/۰۵ می باشد (Stram and Lee 1994). در این روش، معنی‌دار بودن مولفه واریانس نشانگر، دلیلی بر حضور QTL است. از تقسیم BLUP به انحراف معیار خطای پیش‌بینی برای هر نشانگر مقادیری بدست می‌آیند که دارای توزیع نرمال هستند. بنابراین می توان احتمال انحراف اثر نشانگر از صفر را بدست آورده و آن‌ها را به لگاریتم نسبت احتمالات (LOD) تبدیل نمود. نشانگر دارای بالاترین رتبه LOD بصورت متغیر کمکی در بخش ثابت مدل وارد می‌شود (در بردارb) و مدل [۱]، مجددا برازش داده می‌شود. معنی‌دار بودن مولفه واریانس ناشی از نشانگر با استفاده از REMLRT مورد آزمون قرار می‌گیرد. فرآیند فوق تکرار می‌شود تا زمانی که مولفه واریانش ناشی از نشانگر معنی‌دار نشود.

به منظور مطالعه قدرت روش فوق در نقشه‌یابی QTL، جوامع حاصل از تلاقی برگشتی لاین‌های خالص با اندازه نمونه‌های متفاوت (۱۲۵، ۲۵۰ و ۵۰۰) با ۸ کروموزوم به طول ۱۰۰ سانتی مورگان (cM) و ۶ نشانگر بر روی هر کروموزوم به فاصله cM ۲۰ به تعداد ۱۰۰۰ تکرار شبیه سازی شدند. یک QTL در فاصله ۲۰cM روی کروموزوم شماره ۱ (QTL1)، یک QTL در فاصله ۶۰cM روی کروموزوم شماره ۴ (QTL2) و یک QTL روی کروموزوم شماره ۷ در فاصله ۱۰۰cM (QTL3) شبیه سازی شدند. یک صفت کمی دارای توزیع نرمال و واریاس خطای برابر واحد در نظر گرفته شد. هر QTL شبیه سازی شده، میزان ۱۰ درصد از کل واریانس فنوتیپی جمعیت حاصل از تلاقی برگشتی را بیان می‌نمود.

نتایج و بحث

داده های مربوط به یکی از نمونه های شبیه سازی شده (اندازه نمونه برابر ۱۲۵) به روش نقشه‌یابی درون فاصله‌ای مبتنی بر حداکثر درستنمایی آنالیز و لگاریتم حداکثر احتمال (LOD) بدست آمده برای ۸ کروموزوم شبیه سازی شده در شکل ۱ نشان داده شده است. همانگونه که در این شکل مشخص است فقط یکی از سه QTL شبیه‌سازی شده به حد آستانه معنی‌دار (LOD=۲/۵) رسیده و دو QTL دیگر به حد آستانه نرسیده‌اند.



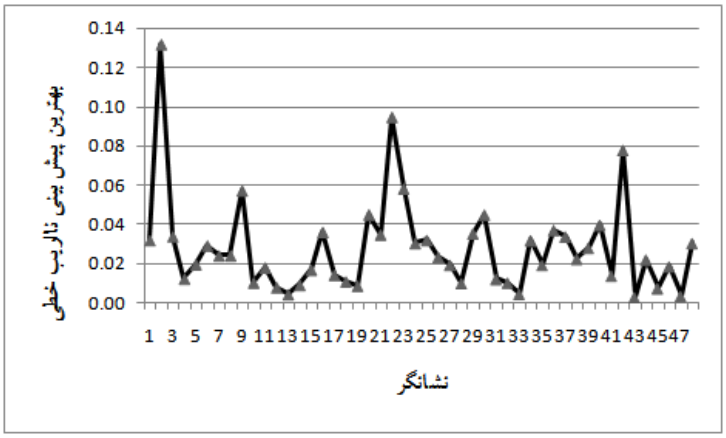
شکل۱- نمودار مربوط به لگاریتم نسبت احتمال (LOD) بدست آمده از روش نقشه‌یابی درون فاصله‌ای (داده‌های بدست آمده برای یکی از ۱۰۰۰ تکرار با اندازه نمونه ۱۲۵) برای کروموزوم‌های مختلف. ch-1 تا ch-8 به ترتیب بیانگر کروموزوم‌های شماره ۱ تا ۸ می‌باشند.

قدر مطلق BLUP بدست آمده برای نشانگرها (نتایج مربوط به یک تکرار) برای اندازه نمونه ۱۲۵، ۲۵۰ و ۵۰۰ به ترتیب در شکل‌های ۲ تا ۴ نشان داده شده است. همانگونه که در این شکل ها مشخص است، نشانگرهای نزدیک به موقعیت QTL مقدار (قدر مطلق) BLUP بسیار بالاتری نسبت به سایر نشانگرها دارند. در روش پیشنهادی در این تحقیق تعداد اندکی مدل به منظور انتخاب مدل ژنتیکی مناسب برای توصیف داده‌ها مورد بررسی قرار می‌گیرند. بنابراین، بر خلاف روش نقشه‌یابی درون فاصله ای مشکل آزمون‌های چندگانه[1] وجود ندارد. پس از رد فرض صفر مبنی بر عدم وجود QTL، به منظور تعیین موقعیت QTL، نمودار مربوط به BLUP نشانگرها مورد استفاده قرار می‌گیرد.

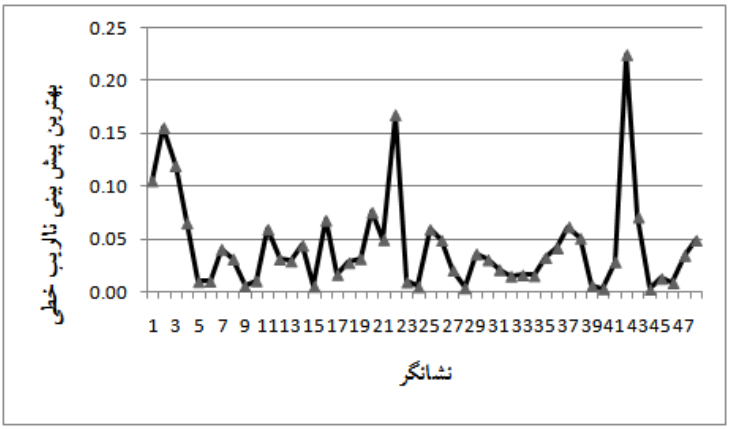
اساس انتخاب نشانگر در این مرحله، رسیدن BLUP به یک حد آستانه معنی‌دار از پیش تعیین شده نیست. بلکه نشانگر دارای بالاترین مقدار BLUP در نظر گرفته شده و بصورت اثر ثابت در مدل آماری وارد می‌شود.

[1] Multiple testing

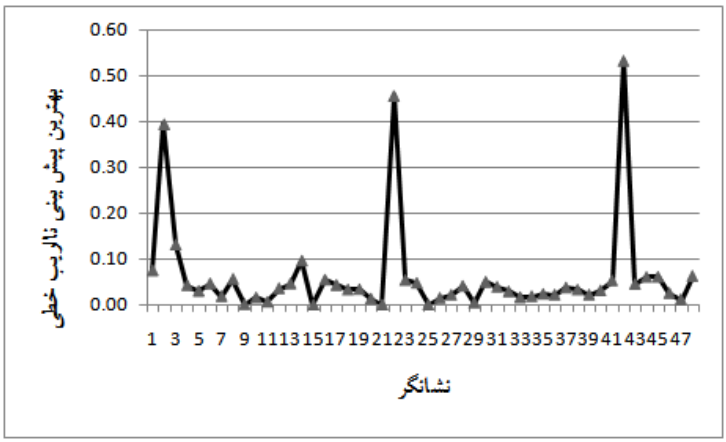
[Downloaded from mg.genetics.ir on 2026-11-03]
[DOR: 20.1001.1.20084439.1392.8.3.9.5]



شکل۲- نمودار مربوط به قدر مطلق BLUPهای بدست آمده برای هر نشانگر (داده‌های بدست آمده برای یکی از ۱۰۰۰ تکرار با اندازه نمونه ۱۲۵).



شکل۳- نمودار مربوط به قدر مطلق BLUPهای بدست آمده برای هر نشانگر (داده‌های بدست آمده برای یکی از ۱۰۰۰ تکرار با اندازه نمونه ۲۵۰).



شکل۴- نمودار مربوط به قدر مطلق BLUPهای بدست آمده برای هر نشانگر (داده‌های بدست آمده برای یکی از ۱۰۰۰ تکرار با اندازه نمونه ۵۰۰).

نتایج مربوط به لگاریتم نسبت احتمال (LogL) برای مدل فاقد اثرات نشانگرها (مدل پایه) و مدل حاوی اثرات نشانگرها، آزمون نسبت حداکثر احتمال باقیمانده (REMLRT) و مولفه‌های واریانس برآورد شده ژنوم (Var_M) و باقیمانده (Var_R) در جدول۱ آمده است. با افزایش اندازه نمونه برای هر دو مدل قدر مطلق لگاریتم نسبت احتمال، برآورد مولفه‌های واریانس و آزمون نسبت حداکثر احتمال باقیمانده افزایش یافت (جدول۱). بالا بودن آزمون نسبت حداکثر احتمال باقیمانده بیانگر قدرت بالای آزمون در رد فرض صفر مبنی بر عدم وجود QTL در کل ژنوم می‌باشد.

[Downloaded from mg.genetics.ir on 2026-11-03]

[DOR: 20.1001.1.20084439.1392.8.3.9.5]

جدول۱- لگاریتم نسبت احتمال (LogL) برای مدل‌های مختلف، آزمون نسبت حداکثر احتمال باقیمانده (REMLRT) و مولفه‌های واریانس برآورد شده ژنوم (VarM) و باقیمانده (VarR)

اندازه نمونه	مدل پایه		مدل با اثرات نشانگر			REMLRT
	LogL	Var_T	LogL	Var_M	Var_R	
۱۲۵	−۸۴/۰	۱/۳۸۰۱	−۷۵/۵	۰/۰۰۸۲	۰/۹۶۴۷	۱۷/۰
۲۵۰	−۱۶۸/۱	۱/۳۹۴۳	−۱۴۴/۲	۰/۰۰۹۲	۰/۹۴۷۰	۴۷/۸
۵۰۰	−۳۴۳/۶	۱/۴۴۱۹	−۲۸۶/۰	۰/۰۰۹۴	۰/۹۹۴۷	۱۱۵/۲

اندازه اثر و خطای معیار برآورد شده برای QTLهای شبیه‌سازی شده در جدول ۲ ارائه شده است. اگرچه اندازه اثر شبیه‌سازی شده برای هر سه QTL یکسان بود (۰/۷۵۵۹ واحد انحراف معیار) اما در تمامی حالات اندازه اثر برآورد شده برای QTL3 بیشتر از مقدار شبیه‌سازی شده بود. از طرفی با افزایش اندازه نمونه این اریبی افزایش پیدا کرده است (جدول۲). این ممکن است به دلیل موقعیت خاص در نظر گرفته شده برای این QTL یعنی در انتهای کروموزوم باشد. یکی از حالات دشوار در تشخیص QTL حالتی است که QTL در انتهای کروموزوم قرار دارد. در حالیکه ساده ترین حالات زمانی است که QTL در وسط کروموزوم قرار داشته باشد که این حالت برای QTLهای شماره ۲ و ۳ در نظر گرفته شده بود. جالب اینکه روش توسعه داده شده در این تحقیق کارایی بالاتری در تشخیص QTL واقع شده در انتهای کروموزوم داشت (جدول۳). در این تحقیق دو نوع تعریف از قدرت آزمون، مورد استفاده قرار گرفت. تعریف اول مربوط به توانایی آزمون در تشخیص هر کدام از سه QTL بود. طبق این تعریف، درصد تکرارهایی که در آن‌ها هر کدام از سه QTL شبیه‌سازی شده تشخیص داده شده بودند، مورد محاسبه قرار گرفت. تعریف دوم مربوط متوسط توانایی روش پیشنهادی در تشخیص کل QTLهای شبیه‌سازی شده بود لذا تعداد QTL تشخیص داده شده تقسیم بر تعداد کل QTLهای شبیه‌سازی شده (عدد۳) شد. نتایج حاصل از آنالیز داده‌های شبیه‌سازی شده نشان داد که توانایی روش پیشنهادی در این تحقیق در نقشه‌یابی QTL تابعی از اندازه (در صورت ثابت بودن میزان اثر QTL) نمونه می‌باشد. برای اندازه نمونه حدود ۵۰۰، روش پیشنهادی دارای توانایی بسیار بالایی در تشخیص QTL بود (جدول ۳).

مشکل اصلی در آنالیز QTL انتخاب بهترین مدل برای توصیف داده ها است و در این فرآیند تعداد بسیار زیادی مدل مورد ارزیابی و در نهایت انتخاب می‌شوند Jansen 1993; Hoeschele et al. 1997; Piepho and Gauch 2001; Broman and Speed (2002; Xu 2003) بنابراین، این یک آزمون استاندارد نبوده و مشکل آزمون های چندگانه بوجود خواهد آمد. روش های بیزی نیاز به تعیین آستانه‌های معنی دار ندارند و قادرند مدل های مختلف را مقایسه کنند (Esmailizadeh 2004). اما عیب اصلی روش‌های بیزی، زمان طولانی مورد نیاز برای انجام محاسبات مربوطه و نیاز به نرم افزار های اختصاصی است. در روش پیشنهادی در این تحقیق تعداد بسیار اندکی مدل در نظر گرفته می‌شوند. از طرفی روش مورد استفاده برای آزمون مولفه واریانس ناشی از نشانگر (آزمون REMLRT) یکی از روش‌های شناخته شده با کارآیی بالا برای انتخاب مدل در آمار می‌باشد. علاوه بر این، این روش بر خلاف روش‌های نقشه‌یابی درون فاصله‌ای مرکب یا نقشه یابی QTL های چندگانه، کل نشانگرهای ژنوم را بطور همزمان در مدل وارد می‌نماید که این می‌تواند منجر به افزایش صحت و دقت در برآورد پارامترهای QTL شود (Jansen 1993; Zeng 1994). در این روش فرض بر این بود که QTL دقیقا در جایگاه نشانگر قرار دارد که البته در مورد بسیاری از نشانگرها از جمله ریزماهواره‌ها درست نمی باشد و ممکن است منجر به اریبی در برآورد موقعیت QTL شود، اما با توجه به ایجاد نقشه‌های پیوستگی اشباع از نشانگرهای تک نوکلوتیدی (SNPs) و در دسترس بودن روش

[Downloaded from mg.genetics.ir on 2026-11-03]
[DOR: 20.1001.1.20084439.1392.8.3.9.5]

جدول۲- اندازه اثر و خطای معیار برآورد شده برای QTL

اندازه نمونه	QTL1	QTL2	QTL3
۱۲۵	۰/۷۱۴۹ ± ۰/۰۹۰۰	۰/۷۳۲۰ ± ۰/۰۹۰۱	۰/۸۳۲۸ ± ۰/۰۹۰۱
۲۵۰	۰/۷۱۴۸ ± ۰/۱۲۵۷	۰/۶۷۸۳ ± ۰/۱۲۵۷	۰/۸۱۴۱ ± ۰/۱۲۵۸
۵۰۰	۰/۶۷۸۱ ± ۰/۱۷۸۹	۰/۷۰۶۳ ± ۰/۱۷۸۸	۰/۸۰۸۰ ± ۰/۱۷۹۶

جدول۳- قدرت روش مدل مختلط در تشخیص QTL

اندازه نمونه	قدرت شناسایی هر کدام از سه QTL شبیه سازی شده بر حسب درصد[1]			قدرت روش مدل مختلط[2]
	QTL1	QTL2	QTL3	
۱۲۵	۲۲/۸	۱۸/۹	۳۶/۴	۰/۲۶
۲۵۰	۷۱/۱	۷۲/۳	۹۰/۱	۰/۷۸
۵۰۰	۹۷/۱	۹۶/۴	۹۹/۷	۰/۹۸

[1] قدرت عبارت بود از درصد تکرارهایی که در آن‌ها هر QTL تشخیص داده شده است.

[2] تعدد QTL تشخیص داده شده تقسیم بر تعداد کل QTLهای شبیه سازی شده.

های سریع و نسبتا ارزان تعیین ژنوتیپ افراد برای این نشانگرها، روش پیشنهادی در این تحقیق برای آنالیز QTL با استفاده از این نشانگرها کاربرد خواهد داشت. بهرحال قدرت روش REMLRT برای آزمون مولفه واریانس ناشی از نشانگر در حالیکه تعداد بسیار زیادی از این نوع نشانگرها (برای مثال، ده ها هزار نشانگر) استفاده می‌شوند، بایستی مورد تحقیق قرار گیرد.

منابع

Broman KW, Speed TP (2002) A model selection approach for the identification of quantitative trait loci in experimental crosses. Journal of the Royal Statistical Society 64: 641-656.

Esmailizadeh K A (2004) Bayesian model averaging as a robust method for gene mapping. The first congress on animal science. Karaj, Iran (In Farsi).

Esmailizadeh KA, Mohammad Abadi MR, Asadi Foozi M (2008) Mapping quantitative trait loci in livestock using simple linear regression. Iranian Journal of Animal Sciences 39:83-93 (In Farsi).

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51:1440-1450.

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in lines crosses using flanking markers. Heredity 69:315-324.

Hoeschele I, Uimari P, Grignola FE, Zhang Q, Gage KM (1997) Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics 147:1445-1457.

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205-211.

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Martinez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theoretical and Applied Genetics 85:480-488.

Piepho HP, Gauch HG (2001) Marker pair selection for mapping quantitative trait loci. Genetics 157:433-444.

Stram DO, Lee JW (1994) Variance components testing in the longitudinal mixed effects model. Biometrics 50:1171-1177.

Xu S (2003) Estimating polygenic effects using markers of the entire genome. Genetics 163:789-801.

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

[Downloaded from mg.genetics.ir on 2026-11-03]

[DOR: 20.1001.1.20084439.1392.8.3.9.5]