

مکان‌یابی نواحی ژنومی کنترل کننده صفات کمی با استفاده همزمان از نشانه‌های 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 نقشه‌یابی نواحی ژنومی کنترل کننده صفات کمی (QTL) در تلاقی لاین‌های خالص پیشنهاد شد که اثرات مرتبط با نشانگر را به صورت تصادفی در نظر می‌گیرد. 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 می‌گیرد. معنی دار بودن واریانس ژنوم بیانگر حضور حداقل یک QTL در سطح ژنوم است. در مرحله دوم، بهترین پیش‌بینی نااریب خطی (BLUPs) نشانگرها برای مکان‌یابی QTL استفاده می‌شود. نشانگر دارای بالاترین مقدار مطلق BLUP در بخش ثابت مدل وارد شده و دو مرحله فوق تکرار می‌شوند تا زمانیکه مولفه واریانس ژنوم غیرمعنی دار شود. دقت این روش با استفاده از داده‌های شبیه‌سازی 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 لاین‌های خالص دارد. اگرچه دقت این روش در تشخیص QTL در نمونه‌های نسبتاً کوچک کم بود (شبیه سایر روش‌های موجود)، مزیت اصلی این روش توانایی آن در وارد نمودن همزمان کل نشانگرها در آنالیز در نتیجه غلبه بر مشکل آزمون‌های چندگانه است.

واژه‌های کلیدی

لاین‌های خالص،
مدل مختلط خطی،
نقشه‌یابی ژنتیکی،
QTL

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^۱ 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)، روش‌های رایج برای شناسایی نواحی ژنومی مرتبط با تنوع در صفات کمی QTL هستند. نتایج یک تحقیق نشان داد که روش نقشه‌یابی درون فاصله‌ای مبتنی بر رگرسیون ساده فقط زمانی که تعداد اندکی QTL 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 فاصله‌ای 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 QTL‌های نزدیک به QTL تحت آزمون، با جذب اثرات ژنتیکی این QTL‌ها، این اثرات را از واریانس باقیمانده

(خطا) جدا می‌نمایند. در نتیجه، دقت آنالیز QTL افزایش می‌یابد. 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 مدل‌ها 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 مدل‌های 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 ($Aa \times AA$) و داده‌های ژنوتیپی برای m نشانگر و داده‌های فنوتیپی صفت کمی مورد نظر بر روی n فرد حاصل از تلاقی برگشتی جمع‌آوری شوند. لذا مدل مختلط خطی برای توصیف این داده‌ها بصورت زیر خواهد بود:

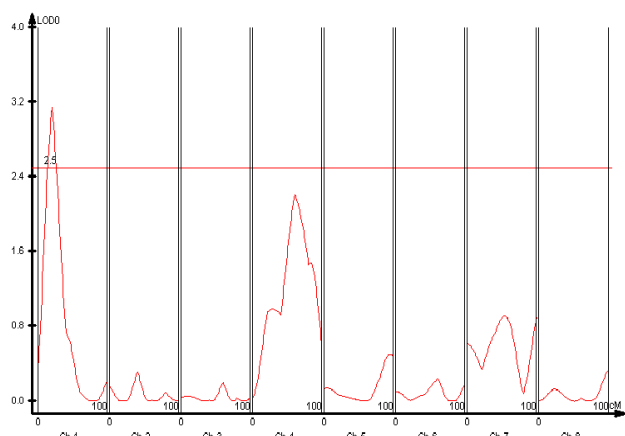
$$y = Xb + Zu + e \quad [1]$$

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

بر اساس مدل [1]، تمامی نشانگرها بطور همزمان در مدل بصورت متغیرهای کمکی تصادفی در نظر گرفته شده‌اند. بنابراین، فقط یک مولفه واریانس برای کل نشانگرهای ژنوم برآورد می‌شود. مولفه‌های واریانس (مولفه واریانس ناشی از نشانگرها (Var_M) و مولفه واریانس باقیمانده (Var_R)) با استفاده از الگوریتم

¹ Interval mapping

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 شبیه سازی شده به حد آستانه معنی دار (LOD=۲/۵) رسیده و دو QTL دیگر به حد آستانه نرسیده اند.



شکل ۱- نمودار مربوط به لگاریتم نسبت احتمال (LOD) بدست آمده از روش نقشه یابی درون فاصله ای (داده های بدست آمده برای یکی از ۱۰۰۰ تکرار با اندازه نمونه ۱۲۵) برای کروموزوم های مختلف. ch-1 تا ch-8 به ترتیب بیانگر کروموزوم های شماره ۱ تا ۸ می باشند.

قدر مطلق BLUP بدست آمده برای نشانگرها (نتایج مربوط به یک تکرار) برای اندازه نمونه ۱۲۵، ۲۵۰ و ۵۰۰ به ترتیب در شکل های 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 داده ها مورد بررسی قرار می گیرند. بنابراین، بر خلاف روش نقشه یابی درون فاصله ای مشکل آزمون های چندگانه^۱ 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 می گیرد. اساس انتخاب نشانگر در این مرحله، رسیدن BLUP به یک حد آستانه معنی دار از پیش تعیین شده نیست. بلکه نشانگر دارای بالاترین مقدار BLUP در نظر گرفته شده و بصورت اثر ثابت در مدل آماری وارد می شود.

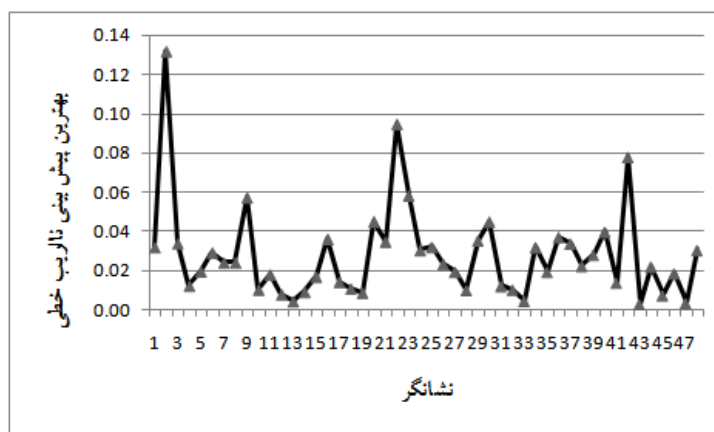
AI برآورد شده و پس از حل معادلات مدل مختلط، بهترین پیش بینی نارایب خطی (BLUP) برای تک تک نشانگرها بدست می آید (Gilmour et al. 1995). برای آزمون معنی دار بودن مولفه واریانس ناشی از نشانگر، آزمون نسبت حداکثر احتمال باقیمانده (REMLRT) که یک آزمون یک طرفه است مورد استفاده قرار می گیرد. منفی دو برابر تغییر در لگاریتم نسبت احتمال به صورت ترکیبی از توزیع χ^2_0 و χ^2_1 با مقدار آستانه ۲/۷۹ برای $\alpha=0.05$ می باشد (Stram and Lee 1994). در این روش، معنی دار بودن مولفه واریانس نشانگر، دلیلی بر حضور QTL است. از تقسیم BLUP به انحراف معیار خطای پیش بینی برای هر نشانگر مقادیری بدست می آیند که دارای توزیع نرمال هستند. بنابراین می توان احتمال انحراف اثر نشانگر از صفر را بدست آورده و آن ها را به لگاریتم نسبت احتمالات (LOD) تبدیل نمود. نشانگر دارای بالاترین رتبه LOD بصورت متغیر کمکی در بخش ثابت مدل وارد می شود (در بردار b) و مدل [۱]، مجدداً برازش داده می شود. معنی دار بودن مولفه واریانس ناشی از نشانگر با استفاده از REMLRT مورد آزمون قرار می گیرد. فرآیند فوق تکرار می شود تا زمانی که مولفه واریانس ناشی از نشانگر معنی دار نشود.

به منظور مطالعه قدرت روش فوق در نقشه یابی QTL، جوامع حاصل از تلاقی برگشتی لاین های خالص با اندازه نمونه های 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 cM ۲۰ روی کروموزوم شماره ۱ (QTL1)، یک QTL در فاصله cM ۶۰ روی کروموزوم شماره ۴ (QTL2) و یک QTL روی کروموزوم شماره ۷ در فاصله cM ۱۰۰ (QTL3) شبیه سازی شدند. یک صفت کمی دارای توزیع نرمال و واریانس 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 می نمود.

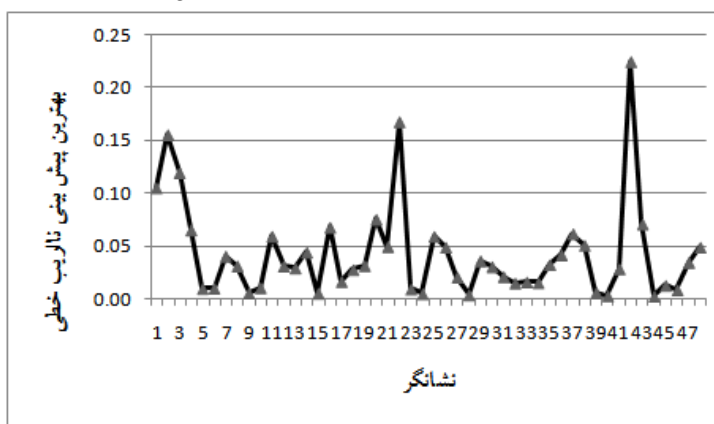
نتایج و بحث

داده های مربوط به یکی از نمونه های شبیه سازی شده (اندازه نمونه برابر ۱۲۵) به روش نقشه یابی درون فاصله ای مبتنی بر

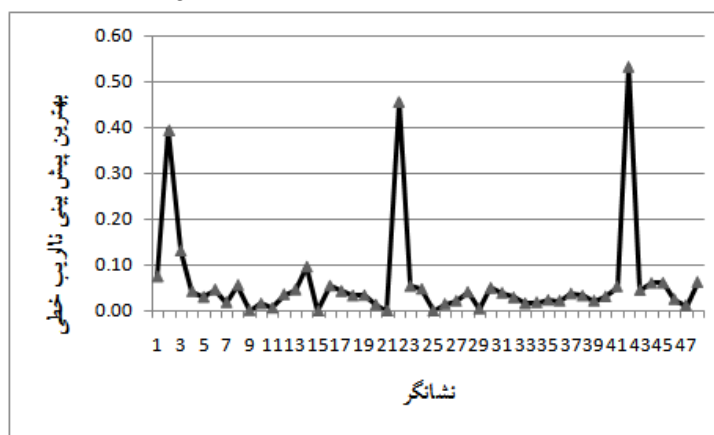
^۱ Multiple testing



شکل ۲- نمودار مربوط به قدر مطلق BLUPهای بدست آمده برای هر نشانگر (داده‌های بدست آمده برای یکی از ۱۰۰۰ تکرار با اندازه نمونه ۱۲۵).



شکل ۳- نمودار مربوط به قدر مطلق BLUPهای بدست آمده برای هر نشانگر (داده‌های بدست آمده برای یکی از ۱۰۰۰ تکرار با اندازه نمونه ۲۵۰).



شکل ۴- نمودار مربوط به قدر مطلق BLUPهای بدست آمده برای هر نشانگر (داده‌های بدست آمده برای یکی از ۱۰۰۰ تکرار با اندازه نمونه ۵۰۰).

لگاریتم نسبت احتمال، برآورد مولفه‌های واریانس و آزمون نسبت حداکثر احتمال باقیمانده افزایش یافت (جدول ۱). بالا بودن آزمون نسبت حداکثر احتمال باقیمانده بیانگر قدرت بالای آزمون در رد فرض صفر مبنی بر عدم وجود QTL در کل ژنوم می‌باشد.

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 جدول ۱ آمده است. با افزایش اندازه نمونه برای هر دو مدل قدر مطلق

جدول ۱- لگاریتم نسبت احتمال (LogL) برای مدل های مختلف، آزمون نسبت حداکثر احتمال باقیمانده (REMLRT) و مولفه های واریانس برآورد شده ژنوم (VarM) و باقیمانده (VarR)

REMLRT	مدل با اثرات نشانگر			مدل پایه		اندازه نمونه
	Var _R	Var _M	LogL	Var _T	LogL	
۱۷/۰	۰/۹۶۴۷	۰/۰۰۸۲	-۷۵/۵	۱/۳۸۰۱	-۸۴/۰	۱۲۵
۴۷/۸	۰/۹۴۷۰	۰/۰۰۹۲	-۱۴۴/۲	۱/۳۹۴۳	-۱۶۸/۱	۲۵۰
۱۱۵/۲	۰/۹۹۴۷	۰/۰۰۹۴	-۲۸۶/۰	۱/۴۴۱۹	-۳۴۳/۶	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 می شوند (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 آستانه های معنی دار ندارند و قادرند مدل های مختلف را مقایسه کنند (Esmailizadeh 2004). اما عیب اصلی روش های 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 می شوند. از طرفی روش مورد استفاده برای آزمون مولفه واریانس ناشی از نشانگر (آزمون REMLRT) یکی از روش های شناخته شده با کارایی بالا برای انتخاب مدل در آمار می باشد. علاوه بر این، این روش بر خلاف روش های نقشه یابی درون فاصله ای مرکب یا نقشه یابی QTL های چندگانه، کل نشانگرهای ژنوم را بطور همزمان در مدل وارد می نماید که این می تواند منجر به افزایش صحت و دقت در برآورد پارامترهای QTL شود (Jansen 1993; Zeng 1994). در این روش فرض بر این بود که QTL دقیقاً در جایگاه نشانگر قرار دارد که البته در مورد بسیاری از نشانگرها از جمله ریزماهواره ها درست نمی باشد و ممکن است منجر به اریبی در برآورد موقعیت QTL شود، اما با توجه به ایجاد نقشه های پیوستگی اشباع از نشانگرهای تک نوکلئوتیدی (SNPs) و در دسترس بودن روش

اندازه اثر و خطای معیار برآورد شده برای QTL های شبیه سازی شده در جدول ۲ ارائه شده است. اگرچه اندازه اثر شبیه سازی 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 شبیه سازی شده بود. از طرفی با افزایش اندازه نمونه این اریبی افزایش پیدا کرده است (جدول ۲). 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 QTL های 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 (جدول ۳). 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 آن ها هر کدام از سه QTL شبیه سازی شده تشخیص داده شدند، مورد محاسبه قرار گرفت. تعریف دوم مربوط متوسط توانایی روش پیشنهادی در تشخیص کل QTL های شبیه سازی شده بود لذا تعداد QTL تشخیص داده شده تقسیم بر تعداد کل QTL های شبیه سازی شده (عدد ۳) شد. نتایج حاصل از آنالیز داده های شبیه سازی شده نشان داد که توانایی روش پیشنهادی در این تحقیق در نقشه یابی QTL تابعی از اندازه (در صورت ثابت بودن میزان اثر QTL) نمونه می باشد. برای اندازه نمونه حدود ۵۰۰،

جدول ۲- اندازه اثر و خطای معیار برآورد شده برای QTL

اندازه نمونه	QTL1	QTL2	QTL3
۱۲۵	0.7149 ± 0.0900	0.7320 ± 0.0901	0.8328 ± 0.0901
۲۵۰	0.7148 ± 0.1257	0.6783 ± 0.1257	0.8141 ± 0.1258
۵۰۰	0.6781 ± 0.1789	0.7063 ± 0.1788	0.8080 ± 0.1796

جدول ۳- قدرت روش مدل مختلط در تشخیص QTL

اندازه نمونه	قدرت شناسایی هر کدام از سه QTL شبیه سازی شده بر حسب درصد ^۱			قدرت روش مدل مختلط ^۲
	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

^۱ قدرت عبارت بود از درصد تکرارهایی که در آن‌ها هر QTL تشخیص داده شده است.

^۲ تعداد QTL تشخیص داده شده تقسیم بر تعداد کل QTL‌های شبیه سازی شده.

حالی‌که تعداد بسیار زیادی از این نوع نشانگرها (برای مثال، ده ها هزار نشانگر) استفاده می‌شوند، بایستی مورد تحقیق قرار گیرد.

های سریع و نسبتاً ارزان تعیین ژنوتیپ افراد برای این نشانگرها، روش پیشنهادی در این تحقیق برای آنالیز QTL با استفاده از این نشانگرها کاربرد خواهد داشت. بهر حال قدرت روش REMLRT برای آزمون مولفه واریانس ناشی از نشانگر در

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.