

تعیین برخی عوامل مؤثر بر نرخ خطای ایمپوتیشن ژنوتیپی برای استفاده در مطالعات رابطه‌یابی سراسری ژنومی در گاوهای شیری

محمد صاحب‌هنر^۱، محمد مرادی شهر بابک^{۲*}، سیدرضا میرایی آشتیانی^۳، عباس پاکدل^۴، دورین جی گریگ^۵
۱. ۴. ۳. ۲. ۱. دانشجوی دکتری، استادان و دانشیار، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران
۵. استاد، دانشگاه ایالتی آیوا، ایمز، ایالات متحده آمریکا
(تاریخ دریافت: ۹۲/۶/۳ - تاریخ تصویب: ۹۲/۱۲/۱۰)

چکیده

در این مطالعه اثر دو راهبرد متفاوت، تأثیر رابطه‌ی خویشاوندی بین جمعیت مرجع و جمعیت‌های آزمون، و اثر فراوانی آلل نادر بر نرخ خطای ایمپوتیشن با استفاده از جمعیتی شبیه‌سازی شده، بررسی شد. جمعیت مرجع حاوی ۱۰۰۰ فرد و جمعیت‌های آزمون حاوی ۵۰۰ فرد بود. ژنوتیپ افراد در جمعیت مرجع با دو تراشه تعیین شد. تراشه اول تراشه‌ی متراکم حاوی ۷۵۰۰۰ SNP به اضافه QTL‌های مؤثر بر صفات بود. تراشه دوم، تراشه‌ی با تراکم متوسط حاوی ۷۵۰۰ SNP بود. ژنوتیپ افراد جمعیت‌های آزمون بر مبنای تراشه‌ی با تراکم پایین حاوی ۵۰۰ SNP تعیین گردید. ایمپوتیشن ژنوتیپی طی دو راهبرد انجام گرفت. در راهبرد دوم مرحله‌ی، ژنوتیپ تراشه تراکم پایین به‌طور مستقیم از تراشه متراکم ایمپوت شد. در راهبرد سه مرحله‌ی، ابتدا ژنوتیپ تراشه تراکم پایین از تراشه تراکم متوسط ایمپوت و سپس ژنوتیپ تراشه متراکم ایمپوت شد. ایمپوتیشن ژنوتیپی با روش BEAGLE صورت گرفت. بررسی همبستگی بین فراوانی آلل نادر با نرخ خطای ایمپوتیشن نشان داد که نرخ خطای ایمپوتیشن تحت تأثیر فراوانی آلل نادر قرار دارد. نرخ خطای ایمپوتیشن زمانی که از تراشه SNP تراکم متوسط استفاده شد، کاهش یافت. با افزایش فاصله بین جمعیت مرجع و جمعیت‌های آزمون، نرخ خطای ایمپوتیشن افزایش یافت.

کلیدواژگان: ایمپوتیشن ژنوتیپی، تراشه متراکم، شبیه‌سازی، فراوانی آلل نادر.

مقدمه

(Hayes et al., 2009). یکی از فناوری‌های استفاده شده در ارزیابی ژنومی، ایمپوتیشن ژنوتیپی^۱ است که به فرایند پیش‌بینی ژنوتیپ نشانگرهای تعیین ژنوتیپ نشده در جمعیت آزمون^۲ با استفاده از الگوی تنوع هاپلوتیپی جمعیت مرجع^۳، اطلاق می‌گردد (Marchini & Howie, 2010). جمعیت مرجع شامل گروهی از افراد است که برای همه نشانگرهای بررسی شده، تعیین ژنوتیپ شده‌اند (Browning, 2008). در برخی از سناریوهای گوناگون نظیر پیش‌بینی ژنوتیپ‌های گمشده، ترکیب نتایج مطالعات متفاوت ژنومی، ایمپوتیشن ژنوتیپی

پیشرفت در فناوری‌های ژنومی ابزارهای جدیدی را برای افزایش تولید و رفاه حیوانات مزرعه‌ای فراهم کرده است. ارزیابی ژنومی به فرایند پیش‌بینی شایستگی ژنتیکی براساس SNP‌های متراکم در سطح ژنوم اطلاق می‌شود که اکنون به شکل گسترده در صنعت گاو شیری استفاده می‌گردد (Meuwissen et al., 2001). این فناوری به دلیل ظهور تراشه‌های متراکم SNP و افزایش پیشرفت ژنتیکی ناشی از کاهش فاصله نسل و افزایش صحت انتخاب به سرعت فراگیر شده است (Schaeffer, 2006).

افزون بر آن، از آنجاکه تعیین ژنوتیپ با تراشه‌های تراکم پایین هزینه کمی دارد، می‌توان تعداد زیادی از حیوانات را به‌طور مرتب تعیین ژنوتیپ کرد. گروهی از تراشه‌های SNP تجاری برای گاو شیری موجود است. این تراشه‌ها شامل Bovine3K با تراکم ۲۹۰۰ SNP، BovineSNP50 با تراکم ۵۴۰۰۱ SNP و BovineHD با تراکم ۷۷۷۹۶۲ SNP توسط Illumina تولید شده است (Wiggans *et al.*, 2011). افزون بر آن، اخیراً دو تراشه با تراکم ۶۵۰ هزار و سه میلیون SNP توسط Affymetrix تولید شده است (Khatkar *et al.*, 2012). یکی از عوامل مؤثر بر صحت انتخاب ژنومی تراکم نشانگر است. نتایج تعدادی از مطالعات نشان می‌دهد که افزایش تراکم نشانگر به افزایش صحت ارزیابی ژنومی می‌انجامد که دلیل آن توضیح بهتر ساختار عدم تعادل پیوستگی بین نشانگر و QTL است (Solberg *et al.*, 2008, Goddard & Hayes, 2009). با این وجود با افزایش تراکم نشانگر، هزینه تعیین ژنوتیپ نیز افزایش می‌یابد. درضمن، تاکنون تعداد زیادی از حیوانات با تراشه تجاری BovineSNP50 تعیین ژنوتیپ شده‌اند که با ظهور تراشه‌های جدید با تراکم بالاتر، تعیین ژنوتیپ حیوانات ژنوتیپ‌شده یا تعیین ژنوتیپ حیوانات جدید با تراشه جدید گران است (Erbe *et al.*, 2012). استفاده از ایمپوتیشن ژنوتیپی روشی مناسب برای حل مشکل نابرابری تعداد لوکوس‌های ژنوتیپ‌شده است. می‌توان ژنوتیپ گروهی از حیوانات را با تراشه SNP مترکم تعیین کرد و سپس با ایمپوتیشن، ژنوتیپ سایر افراد جامعه که با تراشه‌ای با تراکم کمتر ژنوتیپ شده‌اند، را پیش‌بینی کرد. اگرچه تراشه Bovine50K فقط برای ارزیابی گاوهای نر برتر در برنامه تلقیح مصنوعی و گاوهای ماده مولد گاوهای نر برتر استفاده می‌شود، هزینه پایین Bovine3K تعیین ژنوتیپ گاوهای ماده را امکان‌پذیر ساخته است (Wiggans *et al.*, 2012). بنابراین، می‌توان با استفاده از ایمپوتیشن، ژنوتیپ SNP‌های تراشه Bovine50K را به تراشه Bovine3K اضافه کرد. از آوریل ۲۰۱۰ ژنوتیپ ایمپوتیشن‌شده برای گاوهای ماده به‌منظور ارزیابی ژنتیکی استفاده می‌گردد (Wiggans *et al.*, 2011). حال سؤال این است که با توجه به اهمیت نقش ایمپوتیشن ژنوتیپی در مطالعات ژنومی در گاوهای شیری چه عواملی

مطلوب است اما اکنون این اصطلاح بیشتر در زمینه استفاده از جمعیت مرجع به‌منظور پیش‌بینی ژنوتیپ افرادی که برای زیرمجموعه‌ای از SNP‌ها تعیین ژنوتیپ شده‌اند، کاربرد دارد. سرانجام، استفاده از ژنوتیپ‌های ایمپوتیشن‌شده یا سیلیکونی^۱ سبب بهبود توان مطالعات رابطه‌یابی سراسری ژنومی^۲ می‌شود (Marchini & Howie, 2010). به‌طور کلی، فرایند ایمپوتیشن شامل شکستن ژنوتیپ به هاپلوتیپ پدری و مادری و ردیابی توارث هاپلوتیپ برای پیداکردن ژنوتیپ نامعلوم است. در ایمپوتیشن ژنوتیپی گروهی از نشانگرها وجود دارند که به کمک آن‌ها هاپلوتیپ افراد براساس جمعیت مرجع مشخص می‌گردد. مدل پنهان مارکوف^۳، روشی متداول برای ایمپوتیشن ژنوتیپی است. در این روش اطلاعات قابل مشاهده از وضعیت‌های پنهان ایجاد می‌گردد (Rabiner, 1989). در مبحث ایمپوتیشن، وضعیت پنهان شامل الگوهای هاپلوتیپی و اطلاعات قابل مشاهده در برگیرنده ژنوتیپ‌های ایمپوتیشن‌شده است. منظور از مدل مارکوف در وضعیت‌های پنهان این است که داده مشاهده‌شده در هر نشانگر فقط به وضعیت پنهان آن نشانگر بستگی دارد. چالش عمده در انتخاب ژنومی هزینه تعیین ژنوتیپ در بیشتر گونه‌هاست به‌طوری که تعیین ژنوتیپ همه افراد با استفاده از تراشه مترکم مقرون به صرفه نیست (Habier *et al.*, 2009). ارزش اطلاعات حاصل از تعیین ژنوتیپ باید بیشتر از هزینه تعیین ژنوتیپ باشد. در برنامه‌های انتخاب ژنومی در صنعت گاو شیری، هزینه تعیین ژنوتیپ با استفاده از تراشه مترکم برای گاوهای کاندیدشده تولید اسپرم تجاری مقرون به صرفه است، زیرا استفاده از تراشه‌های مترکم به افزایش صحت گاوهای نر و افزایش فروش می‌انجامد (Boichard *et al.*, 2012). با وجود این، به‌منظور بهره‌وری کامل از انتخاب ژنومی باید تعداد زیادی از حیوانات ماده نیز تعیین ژنوتیپ شوند که این امر نیازمند کاهش هزینه تعیین ژنوتیپ است. بدین منظور می‌توان از تراشه‌های تراکم پایین برای حیوانات ماده استفاده کرد (Dassonneville *et al.*, 2012).

1. In Silico
2. Genome Wide Association Studies
3. Hidden Markov Model

مثال، تبدیل A/T به G/C، T/A، C/G و G/C). از وقوع جهش تکراری جلوگیری به عمل آمد. سپس هاپلو تیپ پدری و مادری براساس الگوریتم نوترکیبی بیان شده برای جمعیت بنیان گذاران ایجاد شد. به منظور ایجاد افراد نسل بعد، یکی از جمعیت‌های نسل ۴۵۰۰۰ به عنوان لاین پدری و جمعیت دیگر به عنوان لاین مادری انتخاب گردید و بین آن‌ها آمیزش تصادفی رخ داد. اندازه مؤثر جمعیت در این نسل برابر ۲۰۰ بود. سپس، آمیزش تصادفی تا هشت نسل ادامه پیدا کرد و طی آن اندازه جمعیت نیز به ۱۰۰۰ (G0) افزایش یافت. این جمعیت به عنوان جمعیت مرجع تعیین شد. افزون بر آن، تراشه‌های تعیین ژنوتیپ در این نسل طراحی گردید که در ادامه نحوه طراحی آنها بیان شده است. برای این مطالعه پنج تکرار انجام گرفت و میانگین و خطای استاندارد نتایج گزارش شد.

طراحی تراشه‌های تعیین ژنوتیپ

افراد جمعیت مرجع با استفاده از تراشه متراکم (HD-REF) و تراکم متوسط (MD-REF) تعیین ژنوتیپ شدند. تراشه HD-REF، شامل ۱۵۰ SNP بود که از میان تنوع‌های غیر عملکردی با فراوانی آلی بیشتر از ۰/۰۱ در هر سانتی‌مورگان انتخاب گردید. در برنامه CORONA به طور پیش فرض دو صفت همبسته شبیه‌سازی می‌شود و در مجموع ۱۸۰۰ QTL هر دو صفت را کنترل می‌کنند. کلیه QTLها به تراشه HD-REF اضافه شد، بنابراین تعداد نهایی لوکوس‌های تراشه HD-REF برابر ۷۶۸۰۰ بود. تراشه MD-REF شامل ۱۵ SNP غیر عملکردی در هر سانتی‌مورگان و در مجموع حاوی ۷۵۰۰ SNP بود که به عنوان تراشه میانی در جمعیت مرجع استفاده شد. افراد جمعیت آزمون با تراشه تراکم پایین (LD-SNP) تعیین ژنوتیپ شدند. تراشه LD-SNP از یک SNP در هر سانتی‌مورگان و در مجموع ۵۰۰ SNP تشکیل شده بود.

ایمپوتیشن ژنوتیپی

به منظور مطالعه تأثیر رابطه بین افراد جمعیت مرجع و جمعیت‌های آزمون بر صحت ایمپوتیشن، آمیزش پنج نسل ادامه پیدا کرد. در هر نسل ۵۰ نر برتر براساس فنوتیپ صفت اول انتخاب و به طور تصادفی با ۵۰ ماده

بر صحت ژنوتیپ‌های ایمپوت شده تأثیرگذار است. در این مطالعه ابتدا به کمک شبیه‌سازی، داده‌هایی با قابلیت استفاده در زمینه مطالعه ایمپوتیشن ژنوتیپی ایجاد شد. سپس، به کمک اطلاعات شبیه‌سازی شده برخی از عوامل تأثیرگذار مشتمل بر: ۱. اثر دو راهبرد متفاوت ایمپوتیشن شامل راهبرد دو مرحله‌ای به صورت پیش‌بینی مستقیم ژنوتیپ‌های تراشه تراکم پایین از تراشه متراکم و راهبرد سه مرحله‌ای به صورت پیش‌بینی ژنوتیپ‌های تراشه تراکم پایین از تراشه متراکم با استفاده از یک تراشه تراکم متوسط، ۲. تأثیر فاصله بین جمعیت مرجع و جمعیت‌های آزمون؛ و ۳. اثر فراوانی آلل نادر^۱ بر نرخ خطای ایمپوتیشن بررسی شده است.

مواد و روش‌ها

شبیه‌سازی جمعیت

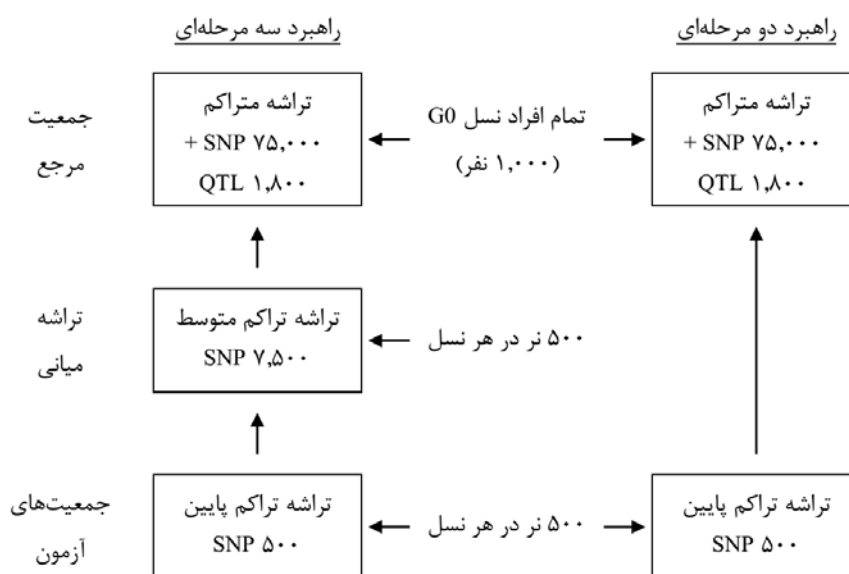
برای شبیه‌سازی از نرم‌افزار CORONA (Sahebbonar et al. منتشر نشده) استفاده گردید. جمعیت بنیان گذاران با اندازه مؤثر ۱۰۰ (۵۰ نر و ۵۰ ماده) ایجاد شد. برای هر فرد، ژنومی شامل پنج کروموزوم هر کدام به طول یک مورگان^۲ شبیه‌سازی گردید. هر رشته کروموزوم از ۱۰^۸ جفت باز تشکیل شده بود. نوع باز آلی برای هر نوکلئوتید مشخص شد (A, T, C, G). افراد این جمعیت برای ۳۰۰۰ نسل به صورت تصادفی آمیزش داده شدند. به منظور تشکیل تصادفی هاپلو تیپ پدری و مادری، نرخ وقوع نوترکیبی به ازای هر کروموزوم دورشته‌ای از توزیع پواسون با میانگین یک به ازای هر مورگان پیروی می‌کرد. محل وقوع نوترکیبی به صورت تصادفی در طول کروموزوم انتخاب شد. در این مرحله از وقوع جهش صرف نظر شده است. سپس، دو جمعیت با اندازه مؤثر ۱۰۰ و نسبت برابر از هر دو جنس از جمعیت بنیان گذاران ایجاد شد. در هر جمعیت، آمیزش تصادفی برای ۴۵۰۰۰ نسل ادامه یافت. جهش با نرخ ۲/۵×۱۰^۸ به ازای هر نوکلئوتید در هر میوز رخ داد. جهش به طور تصادفی در طول کروموزوم به وقوع پیوست. جهش سبب تبدیل یک جفت باز به سه نوع دیگر شد (برای

1. Minor Allele Frequency

2. Morgan

افزایش صحت ایمپوتیشن ابتدا ژنوتیپ‌های تراشه LD-REF از تراشه MD-REF پیش‌بینی شد. سپس با ژنوتیپ‌های ایمپوت شده، ژنوتیپ‌های تراشه LD-SNP از تراشه HD-REF پیش‌بینی گردید که به‌طور خلاصه در شکل ۱ نشان داده شده است.

آمیزش داده شد که به تولید ۱۰۰۰ نفر در هر نسل گردید. ژنوتیپ حیوانات نر در هر نسل طی دو راهبرد دو و سه‌مرحله‌ای، با استفاده از ژنوتیپ افراد جمعیت مرجع ایمپوت گردید. در راهبرد دو مرحله‌ای، با تراشه LD-SNP ژنوتیپ حیوانات نر برای تراشه HD-REF به‌صورت مستقیم ایمپوت شد. در راهبرد سه‌مرحله‌ای، برای



شکل ۱. راهبرد دو و سه‌مرحله‌ای در ایمپوتیشن ژنوتیپی

ایمپوت شده $100 \times$ محاسبه گردید. بازدهی دو راهبرد دو و سه‌مرحله‌ای براساس نرخ خطای ایمپوتیشن و صحت ایمپوتیشن (R^2) مقایسه گردید. صحت ایمپوتیشن از همبستگی بین تعداد آلل نادر^۳ ژنوتیپ واقعی (ژنوتیپ محاسبه شده براساس احتمالات ژنوتیپ پسین^۴) محاسبه می‌شود. اگر $P(AA)$ ، $P(AB)$ و $P(BB)$ احتمالات ژنوتیپی باشد، تعداد آلل نادر (B) از رابطه $0 \times P(AA) + 1 \times P(AB) + 2 \times P(BB)$ محاسبه می‌شود.

نتایج و بحث

میانگین نرخ خطای ایمپوتیشن و صحت ایمپوتیشن ژنوتیپی با راهبرد دو و سه‌مرحله‌ای برای ایمپوت ژنوتیپ‌های تراشه LD-SNP از ژنوتیپ‌های تراشه HD-

ایمپوتیشن ژنوتیپی با نرم‌افزار BEAGLE v3.3.2 (Browning & Browning, 2009) و مدل افراد غیر خویشاوند انجام شد. ژنوتیپ هر کروموزوم به‌صورت جداگانه ایمپوت گردید. نرم‌افزار BEAGLE نیازی به پارامتر اولیه ندارد و براساس مدل گروه‌بندی محلی هاپلوتیپ^۱ طراحی شده است که با مدل پنهان مارکوف، محتمل‌ترین ساختار هاپلوتیپ را براساس نشانگر، آشکار می‌سازد. فراوانی آلل نادر ابتدا برای هر SNP ایمپوت شده تعیین و سپس همبستگی بین فراوانی آلل نادر و نرخ خطای ایمپوتیشن محاسبه شد. فراوانی آلل نادر برای هر SNP به‌صورت فراوانی آللی که فراوانی کمتر دارد، تعریف می‌شود. نرخ خطای ایمپوتیشن برای هر SNP، از درصد ژنوتیپ‌هایی که به‌طور اشتباه ایمپوت شده بود (براساس انطباق^۲ ژنوتیپ ایمپوت شده بر ژنوتیپ واقعی برای هر SNP) به کل SNP‌های

3. Allele Dosage

4. Posterior Genotype Probabilities

1. Localized Haplotype-Clustering Model

2. Concordance

می‌شود، به طوری که نرخ خطای ایمپوتیشن مستقیم تراشه 3K از تراشه 800K را $4/78$ درصد و نرخ خطای ایمپوتیشن زمانی که از تراشه 50K استفاده شد $4/62$ درصد گزارش کردند. در هر دو راهبرد، نرخ خطای ایمپوتیشن در جمعیت‌های آزمون طی نسل‌های متوالی و با افزایش فاصله جمعیت آزمون از جمعیت مرجع افزایش یافت (جدول ۱). Druet *et al.* (۲۰۱۰) اثر رابطه خویشاوندی بر نرخ خطای ایمپوتیشن را با استفاده از ژنوتیپ گاوهای هلشتاین هلندی بررسی کردند. در این مطالعه حیوانات براساس درجه خویشاوندی با افراد جمعیت مرجع گروه‌بندی شدند. نتایج نشان داد که نرخ خطای ایمپوتیشن با افزایش خویشاوندی کاهش می‌یابد. Khatkar *et al.* (۲۰۱۲) نیز در بررسی تأثیر رابطه خویشاوندی بر نرخ خطای ایمپوتیشن، افراد جمعیت آزمون را بر حسب حضور پدر در جمعیت مرجع گروه‌بندی کردند. نتایج نشان داد که نرخ خطای ایمپوتیشن برای جمعیت آزمون که پدران آن‌ها در جمعیت مرجع حضور دارند، کمتر است.

REF براساس میانگین پنج تکرار و هر تکرار شامل پنج کروموزوم، طی پنج نسل در جدول ۱ نشان داده شده است. نرخ خطای ایمپوتیشن در راهبرد سه‌مرحله‌ای کمتر از راهبرد دومرحله‌ای بود. از آنجاکه در استفاده از تراشه‌های تجاری، ژنوتیپ واقعی افراد مشخص نیست، از معیار صحت ایمپوتیشن برای ارزیابی ژنوتیپ‌های پیش‌بینی شده استفاده می‌شود. صحت ایمپوتیشن بدون دانستن ژنوتیپ واقعی و با استفاده از احتمالات ژنوتیپ پسین برای ژنوتیپ‌های ایمپوت شده برآورد می‌گردد. همان‌طور که در جدول ۱ نشان داده شده است صحت ایمپوتیشن در راهبرد سه‌مرحله‌ای بیشتر از راهبرد دومرحله‌ای بود. به نظر می‌رسد که استفاده از تراشه تراکم متوسط به بهبود شناسایی ساختار هاپلوتیپ می‌انجامد. Khatkar *et al.* (۲۰۱۲) در ایمپوت ژنوتیپ‌های تراشه تجاری 3K از تراشه تجاری 800K در گاوهای هلشتاین فریزین استرالیایی، گزارش کردند که استفاده از تراشه تجاری 50K به‌عنوان تراشه میانی سبب کاهش نرخ خطای ایمپوتیشن در مقایسه با ایمپوت مستقیم ژنوتیپ تراشه 3K از تراشه 800K

جدول ۱. مقایسه میانگین نرخ خطای ایمپوتیشن و صحت ایمپوتیشن (R^2) بین راهبرد دو و سه‌مرحله‌ای برای ایمپوت ژنوتیپ‌های تراشه تراکم پایین (LD-SNP) از ژنوتیپ‌های تراشه متراکم (HD-REF)

راهبرد سه‌مرحله‌ای		راهبرد دومرحله‌ای		
R^2	نرخ خطا	R^2	نرخ خطا	
$0/98$ ($0/0002$)	$2/43$ ($0/03$)	$0/94$ ($0/0007$)	$3/35$ ($0/04$)	نسل ۱
$0/98$ ($0/0002$)	$2/96$ ($0/04$)	$0/93$ ($0/0006$)	$3/70$ ($0/04$)	نسل ۲
$0/98$ ($0/0002$)	$3/46$ ($0/04$)	$0/93$ ($0/0007$)	$4/03$ ($0/05$)	نسل ۳
$0/98$ ($0/0002$)	$3/88$ ($0/04$)	$0/93$ ($0/0006$)	$4/37$ ($0/05$)	نسل ۴
$0/98$ ($0/0002$)	$4/28$ ($0/05$)	$0/92$ ($0/0007$)	$4/67$ ($0/06$)	نسل ۵

اعداد داخل پرانتز نشان‌دهنده خطای استاندارد برآورد است.

(Gabriel *et al.*, 2002). در این نواحی شواهد اندکی از وقوع نوترکیبی وجود دارد و در نتیجه عدم تعادل پیوستگی^۱ بالایی بین نشانگرهای ژنتیکی مشاهده می‌شود (Paigen & Petkov, 2010) (عدم تعادل پیوستگی عبارت از وابستگی غیر تصادفی بین آلل‌های لوکوس‌های مجاور (Lewontin, 1964) است). دو عامل

روش‌های ایمپوتیشن ژنوتیپی با شناسایی بخش‌های مشترک هاپلوتیپی بین افراد جمعیت مرجع و افراد جمعیت‌های آزمون، ژنوتیپ‌های گمشده را پیش‌بینی می‌کنند. در حقیقت، ایده ضمنی فرض شده در روش‌های ایمپوتیشن ژنوتیپی این است که ژنوتیپ هر فرد در جمعیت آزمون، ترکیبی از ساختار هاپلوتیپی سایر افراد است (Marchini & Howie, 2010). هاپلوتیپ شامل نواحی کروموزومی با اندازه متغیر می‌باشد که حاوی توالی خاصی از آلل‌های مشاهده شده در جمعیت است

پنج نسل، در جدول ۲ نشان داده شده است. این نتایج براساس میانگین پنج تکرار که هر تکرار شامل پنج کروموزوم است، محاسبه شده است. بررسی نتایج نشان داد که با کاهش فراوانی آلل نادر، نرخ خطای ایمپوتیشن افزایش می‌یابد. براساس نتایج جدول ۲ ایمپوتیشن SNP های نادر در مقایسه با SNP های رایج دشوارتر است. در مطالعه Marchini & Howie (۲۰۱۰) گزارش شد که نرخ خطای ایمپوتیشن با کاهش فراوانی آلل نادر برای SNP ، افزایش می‌یابد. همبستگی بین فراوانی آلل نادر با نرخ خطای ایمپوتیشن با افزایش فاصله بین جمعیت‌های آزمون از جمعیت مرجع کاهش داشت که دلیل آن می‌تواند افزایش نرخ خطای ایمپوتیشن طی نسل‌های متوالی (جدول ۱) باشد.

جهش و نوترکیبی سبب شکستن ساختار هاپلوטיפ می‌شود (Gabriel *et al.*, 2002). بر این اساس می‌توان نتیجه گرفت که روش‌های ایمپوتیشن ژنوتیپی به ساختار عدم تعادل پیوستگی وابسته است. بنابراین، رابطه خویشاوندی بین افراد جمعیت مرجع با جمعیت‌های آزمون بر نرخ خطای ایمپوتیشن تأثیرگذار است به طوری که افزایش نرخ خطای ایمپوتیشن با افزایش فاصله بین جمعیت مرجع از جمعیت‌های آزمون (جدول ۱) ممکن است در اثر وقوع نوترکیبی و از بین رفتن ساختارهای هاپلوטיפی با افزایش فاصله بین جمعیت مرجع و جمعیت‌های آزمون باشد. همبستگی بین فراوانی آلل نادر با نرخ خطای ایمپوتیشن با استفاده از راهبرد دو و سه مرحله‌ای به منظور ایمپوت ژنوتیپ‌های تراشه LD-SNP از ژنوتیپ‌های تراشه HD-REF برای

جدول ۲. میانگین همبستگی بین فراوانی آلل نادر با نرخ خطای ایمپوتیشن با استفاده از راهبرد دو و سه مرحله‌ای به منظور ایمپوت ژنوتیپ‌های تراشه تراکم پایین (LD-SNP) از ژنوتیپ‌های تراشه تراکم (HD-REF)

راهبرد دو مرحله‌ای	راهبرد سه مرحله‌ای	نسل
-۰/۴۲ (۰/۰۱)	-۰/۴۶ (۰/۰۱)	۱
-۰/۳۵ (۰/۰۱)	-۰/۳۷ (۰/۰۱)	۲
-۰/۲۹ (۰/۰۱)	-۰/۳۰ (۰/۰۱)	۳
-۰/۲۲ (۰/۰۱)	-۰/۲۲ (۰/۰۲)	۴
-۰/۱۷ (۰/۰۲)	-۰/۱۶ (۰/۰۲)	۵

اعداد داخل پرانتز نشان‌دهنده خطای استاندارد برآورد است.

فرکانس آللی می‌تواند نقش مهمی در بازدهی مطالعات ژنومی داشته باشد. با توجه به شرح و بسط فراوان در تولید تراشه SNP متراکم در صنعت گاو شیری، وجود گروه‌هایی در جمعیتی که با تراشه‌های گوناگون تعیین ژنوتیپ شده باشند، رایج است.

در این تحقیق برخی از عوامل مؤثر بر نرخ خطای ایمپوتیشن بررسی شد. با استفاده از ایمپوتیشن ژنوتیپی امکان پیش‌بینی و تلفیق اطلاعات ژنوتیپی تراشه‌های گوناگون با تراکم‌های متفاوت امکان‌پذیر است. استفاده از ژنوتیپ‌های ایمپوت شده افزون بر کاهش فراوان در هزینه‌های تعیین ژنوتیپ سبب افزایش صحت انتخاب ژنومی می‌گردد.

اینک مطالعات ژنومی بر مبنای تنوع‌های رایج^۱ پایه‌گذاری شده است. تنوع‌های رایج تنوع‌هایی با فراوانی آلل نادر بیشتر از ۰/۰۵ تعریف می‌شود. اصطلاح وراثت‌پذیری گمشده^۲ بیان می‌کند که تنوع‌های رایج به تنهایی و یا در مجموع سهم اندکی از وراثت‌پذیری را توضیح می‌دهند (Manolio *et al.*, 2009). یکی از عوامل مؤثر بر وراثت‌پذیری گمشده احتمالاً تنوع‌هایی با فراوانی آلل نادر پایین است ($\text{MAF} < 0.05$).

این تنوع‌ها ممکن است با تأثیر فراوان بر واریانس صفت نقش تعیین‌کننده‌ای در وراثت‌پذیری گمشده داشته باشند (McCarthy *et al.*, 2008). از طرف دیگر، با توجه به رابطه بین فرکانس آلل نادر و نرخ خطای ایمپوتیشن (جدول ۲) زمانی که از ژنوتیپ‌های ایمپوت شده در مطالعات ژنومی استفاده می‌شود

1. Common Variants
2. Missing Heritability

REFERENCES

1. Boichard, D., H. Chung, R. Dasonneville, X. David, A. Eggen, S. Fritz, K. J. Gietzen, B. J. Hayes, C. T. Lawley, T. S. Sonstegard, C. P. Van Tassell, P. M. VanRaden, K. A. Viaud-Martinez, G. R. Wiggans, & L. D. C. for the Bovine. (2012). Design of a Bovine low-density SNP array optimized for imputation. *PLoS ONE* 7, 34130.
2. Browning, B. L. & Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics* 84(2):210-223.
3. Browning, S. R. (2008). Missing data imputation and haplotype phase inference for genome-wide association studies. *Human Genetics* 124(5):439-450.
4. Dasonneville, R., S. Fritz, Ducrocq, V. & Boichard, D. (2012). Short communication: Imputation performances of 3 low-density marker panels in beef and dairy cattle. *Journal of Dairy Science*, 95(7), 4136-4140.
5. Druet, T., Schrooten, C. & Roos, De, A. P. W. (2010). Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. *Journal of Dairy Science*, 93(11), 5443-5454.
6. Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, & M. E. Goddard. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, 95(7), 4114-4129.
7. Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly, and D. Altshuler. (2002). The structure of haplotype blocks in the human genome. *Science* 296(5576):2225-2229.
8. Goddard, M. E. and B. J. Hayes. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* 10(6):381-391.
9. Habier, D., R. L. Fernando, and J. C. M. Dekkers. (2009). Genomic selection using low-density marker panels. *Genetics* 182(1):343-353.
10. Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92(2):433-443.
11. Khatkar, M. S., G. Moser, B. J. Hayes, and H. W. Raadsma. (2012). Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. *BMC Genomics* 13:538.
12. Lewontin, R. C. (1964). The interaction of selection and linkage. I. general considerations; heterotic models. *Genetics* 49(1):49-67.
13. Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttacher, A. E., Kong, A., Kruglyak, L., Mardis, C. N., Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, E., Haines, J. L., Mackay, T. F. C., McCarroll, S. A. & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747-753.
14. Marchini, J. & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, 11(7), 499-511.
15. McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, 9(5), 356-369.
16. Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819-1829.
17. Paigen, K. & Petkov, P. (2010). Mammalian recombination hot spots: properties, control and evolution. *Nature Reviews Genetics*, 11(3), 221-233.
18. Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.
19. Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123(4), 218-223.
20. Solberg, T. R., Sonesson, A. K., Woolliams, J. A. & Meuwissen, T. H. E. (2008). Genomic selection using different marker types and densities. *Journal of Animal Science*, 86(10), 2447-2454.
21. Wiggans, G. R., Cooper, T. A., Vanraden, P. M., Olson, K. M. & Tooker, M. E. (2012). Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. *Journal of Dairy Science*, 95(3), 1552-1558.
22. Wiggans, G. R., Vanraden, P. M. & Cooper, T. A. (2011). The genomic evaluation system in the United States: Past, present, future. *Journal of Dairy Science*, 94(6), 3202-3211.