

بهبود ارزیابی ژنومی گاوهای هلستاین با استفاده از ماتریس روابط ژنومی بر پایه اطلاعات هاپلوتایپ‌ها

مهرنوش فروتن^۱ - سعید انصاری مهباری^{۲*} - فلاویو شنکل^۳ - مهدی سرگلزایی^۴

تاریخ دریافت: ۱۳۹۶/۰۴/۲۴

تاریخ پذیرش: ۱۳۹۶/۰۹/۲۵

چکیده

امروزه استفاده از اطلاعات هاپلوتایپ به منظور تخمین ارزش اصلاحی ژنومی (GEBV) با صحت و دقت بالاتر مورد توجه قرار گرفته است. استفاده از شباهت بین هاپلوتایپ‌ها نسبت به مستقل در نظر گرفتن اثرات هاپلوتایپ‌ها می‌تواند منجر به قابلیت پیش‌گویی بالاتر برآوردهای ژنومی گردد. هدف از این مطالعه بررسی صحت و اریبی GEBV به روش بهترین برآورد ناریب خطی ژنومی (GBLUP) با استفاده از ماتریس روابط خویشاوندی ژنومی بر پایه اطلاعات نشانگرهای چندشکلی تک نوکلئوتیدی (G_{SNP}) و هاپلوتایپ (G_{HAP}) و ماتریس هیبرید شامل اطلاعات هاپلوتایپ‌ها و SNPها (G_{SNP,HAP}) بود. ماتریس‌های روابط خویشاوندی ژنومی با استفاده از تراشه ۵۰K برای جمعیت گاو شیری هلستاین آمریکای شمالی برای صفات تولید چربی شیر، تولید شیر، نمره سلول بدنی، تیپ و روزهای بازتشدیل گردید. مقادیر صحت GEBV با استفاده از ماتریس G_{SNP} و G_{HAP} به ترتیب از ۰/۴۹ (روزهای باز) تا ۰/۷۳ (تولید چربی شیر) و ۰/۵۰ (روزهای باز) تا ۰/۷۳ (تولید چربی شیر) متغیر بود. استفاده از ماتریس G_{HAP} در مقایسه با دیگر ماتریس‌های مورد بررسی در این مطالعه منجر به برآوردهای ناریب‌تری در تمام صفات شد. استفاده از ماتریس G_{HAP} نسبت به G_{SNP} موجب ۶ تا ۱۵ درصد کاهش اریبی برآوردها در صفات مختلف گردید. یکی از دلایل بهبود اریبی برآوردها می‌تواند برآورد بهتر واریانس ژنتیکی در سطح جمعیت توسط هاپلوتایپ نشانگرها نسبت به نشانگر انفرادی باشد. بر اساس نتایج حاضر ماتریس G_{HAP} می‌تواند جایگزین مناسبی برای G_{SNP} در ارزیابی‌های ژنومی به منظور دستیابی به برآوردهای ارزش اصلاحی ناریب‌تر و دقیق‌تر باشد.

واژه‌های کلیدی: اریبی برآوردها، صحت، گاو هلستاین، ماتریس روابط خویشاوندی ژنومی، هاپلوتایپ

مقدمه

بتوان به حداکثر پیشرفت ژنتیکی ممکن در زمان کوتاهی دست یافت. اگرچه آزمون عملکرد نتایج از جمله روش‌های موثر در این مسیر است لیکن پیشرفت‌های فناوری زیستی نشان داده است با بهره‌وری از نشانگرهای مولکولی مرتبط با ژن‌های کمی، علاوه بر افزایش صحت انتخاب، در سنین اولیه و حتی بلافاصله پس از تولد، این امکان وجود دارد که حیوانات با ارزش ژنتیکی بالا شناسایی، نگهداری و مورد استفاده قرار داده شوند. به این ترتیب با کاهش فاصله نسل و افزایش شدت انتخاب، پیشرفت ژنتیکی به حداکثر خواهد رسید (۱۷).

در سال ۱۹۹۷ نجاتی جوارمی و همکاران استفاده از ماتریس روابط خویشاوندی مبتنی بر اطلاعات نشانگری را جهت افزایش صحت برآورد ارزش اصلاحی بر اساس روش بهترین برآورد ناریب خطی (BLUP) پیشنهاد نمودند. در سال ۲۰۰۱ روش انتخاب ژنومی توسط میوسن و همکاران ارائه شد و طی آن ارزش اصلاحی کل با استفاده از تعداد زیادی نشانگر ژنتیکی در دو جایگاه مجاور چند آلی در کل ژنوم برآورد گردید (۱۰). در این روش فرض بر این است که

پیشرفت‌های قابل توجه در توسعه و استفاده از نشانگرهای مولکولی طی سالیان اخیر موجب شده است که تحقیقات متعددی در زمینه امکان استفاده از نشانگرهای مرتبط با جایگاه‌های ژنی صفات کمی به اجرا درآید (۱۹). به طور کلی درگله‌های مولد و هسته‌های اصلاح نژاد، برنامه‌های به‌نژادی به صورتی طراحی و به اجرا در می‌آیند تا از طریق افزایش دقت و صحت در برآوردهای ارزش اصلاحی

۱ و ۲- به ترتیب دانش‌آموخته دکتری و دانشیار دانشکده کشاورزی، دانشگاه صنعتی اصفهان، اصفهان، ایران
۳- استاد مرکز بهبود ژنتیکی حیوانات، دانشکده علوم زیستی، دانشگاه گوتلف، کانادا
۴- دانشیار مرکز بهبود ژنتیکی حیوانات، دانشکده علوم زیستی، دانشگاه گوتلف، محقق شرکت Semex و شرکت HiggsGene Solutions، کانادا
(Email: s.ansari@cc.iut.ac.ir)
* نویسنده مسئول:

صحت بالاتری را برای ارزش اصلاحی ژنومی برآورد شده نسبت به ماتریس G_{SNP} گزارش نمودند. فردوسی و همکاران نیز در سال ۲۰۱۶ در مطالعه‌ی خود شباهت بین هاپلوتایپ‌ها را با در نظر گرفتن همپوشانی بین قطعات هاپلوتایپ تخمین زده و بیان نمودند که استفاده از اطلاعات هاپلوتایپ‌های حاوی چندین SNP می‌تواند منجر به افزایش صحت برآوردها نسبت به روش G_{SNP} شود.

نشانه‌های SNP موجود در تراشه‌ها معمولاً بر اساس فراوانی آللی متوسط تا بالا انتخاب می‌شوند و بنابراین حاصل جهش‌های قدیمی هستند. جهش‌ها در ابتدا از فراوانی پایینی برخوردار هستند و سپس به مرور زمان به‌عنوان مثال در اثر انتخاب و رانش تصادفی، فراوانی تعدادی از آنها افزایش می‌یابد و تعدادی از آنها ممکن است قبل از رسیدن به یک فراوانی مطلوب قابل توجه از جمعیت حذف شوند (۱۱). ماتریس روابط خویشاوندی ژنومی G_{SNP} با استفاده از نشانه‌ها با فراوانی بالای آلل نادر ساخته می‌شوند. این موضوع حاکی از آن است که G_{SNP} قادر است روابط ژنومی افراد در نسل‌های قدیمی را نشان دهد و بنابراین ممکن است تغییرات حاصل به دلیل انتخاب در نسل‌های جدید را به طور دقیق منعکس نکند (۱۱). لذا این فرضیه مطرح می‌شود که احتمالاً تشکیل ماتریس روابط ژنومی از طریق محاسبه روابط بین افراد در جمعیت رفرنس و همچنین بین جمعیت رفرنس و جمعیت تایید به‌ترتیب با استفاده از اطلاعات SNPها و هاپلوتایپ‌ها می‌تواند منجر به افزایش صحت و کاهش اریبی برآوردهای ارزش اصلاحی ژنومی گردد.

مطالعه حاضر با هدف بررسی صحت و اریبی در برآورد ژنومی ارزش اصلاحی با استفاده از روش GBLUP با کاربرد ماتریس روابط خویشاوندی ژنومی بر پایه اطلاعات نشانه‌های چند شکلی تک نوکلئوتیدی (G_{SNP}) و هاپلوتایپ (G_{HAP}) و ماتریس هیبرید شامل اطلاعات هاپلوتایپ‌ها و SNPها ($G_{SNP,HAP}$) برای اولین بار با استفاده از داده‌های واقعی گاوهای نر آمریکای شمالی به اجرا درآمد.

مواد و روش‌ها

در این مطالعه از جمعیت گاوهای نر هلشتاین آمریکای شمالی متولد شده در سال‌های ۱۹۶۰ الی ۲۰۱۱ استفاده شد. اطلاعات ژنوتیپ و فنوتیپ حیوانات از شرکت CDN کانادا دریافت گردید. حیوانات ژنوتایپ شده در دو دسته رفرنس و تایید دسته‌بندی شدند. جمعیت رفرنس شامل گاوهای نر متولد شده بین سال‌های ۱۹۶۰ تا ۲۰۰۷ بود. حیوانات که در سال‌های ۲۰۰۷ تا ۲۰۱۱ متولد شده بودند و دارای پروف رسمی در سال ۲۰۱۵ بودند به‌عنوان جمعیت تایید در نظر گرفته شدند. اندازه جمعیت رفرنس و تایید برای صفات تولید شیر، چربی شیر، نمره سلول بدنی و تیپ به‌ترتیب ۶۵۱۵ و ۱۲۹۴ گاو نر بود. همچنین جمعیت رفرنس و تایید برای صفت روزهای باز به‌ترتیب

نشانه‌ها در تعدادی از جایگاه‌ها در عدم تعادل فاز گامتی^۱ (LD) با آلل‌های QTL مؤثر بر صفات مورد انتخاب هستند. کاربرد کنونی برآوردهای ارزش اصلاحی ژنومی به طور عمده بر اساس SNP بوده و ژنوتیپ SNPها به طور مستقیم در مدل‌های پیش‌بینی ژنومی استفاده می‌شوند (۱۹).

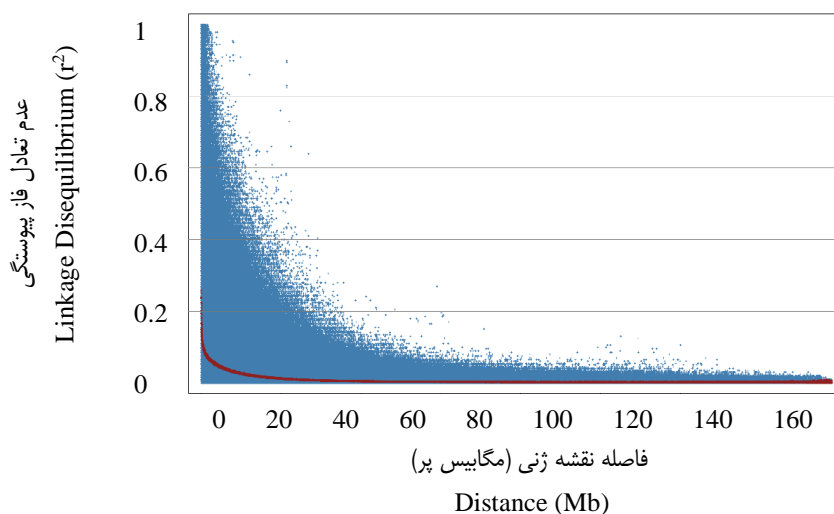
مطالعات متعدد با استفاده از روش‌های متفاوت برای تعیین بلوک هاپوتایپ و روابط بین هاپلوتایپ‌ها در سطح ژنوم به بررسی استفاده از اطلاعات هاپلوتایپ در ارزیابی ژنومی پرداخته‌اند. کالوس و همکاران (۲۰۰۸) با مقایسه استفاده از هاپلوتایپ نشانه‌ها به‌صورت کواریت در مدل و یا استفاده از تک SNP بیان نمودند که استفاده از اطلاعات هاپلوتایپ‌ها می‌تواند منجر به برآوردهای ژنومی دقیق‌تری گردد هرچند این برتری با افزایش تراکم نشانه‌ها کاهش می‌یابد. بهبود ارزیابی ژنومی با استفاده از اطلاعات هاپلوتایپ نسبت به تک SNP می‌تواند به این علت باشد که در هنگام استفاده از اطلاعات هاپلوتایپ نشانه‌ها، عدم تعادل فاز گامتی خصوصاً بین فواصل کوچک ژنوم^۲ بین SNPهای موجود در هر هاپلوتایپ در نظر گرفته می‌شود در صورتی که در حالت تک SNP این اثرات نادیده گرفته می‌شود (۸) و به این ترتیب در حالت استفاده از تک SNP، تغییر ترتیب SNPها تأثیری بر نتایج نهایی ماتریس G ندارد (۶). این در حالی است که امکان دارد یک آلل مطلوب QTL در LD با یک آلل SNP در یک هاپلوتایپ در یک جمعیت باشد و همین آلل در LD با آلل دیگر SNP در هاپلوتایپی دیگر جمعیت دیگر باشد. استفاده از این فاز متفاوت بین SNP و QTL نیاز به در اختیار داشتن اطلاعات هاپلوتایپی می‌باشد.

با این حال مستقل در نظر گرفتن هاپلوتایپ‌ها در ارزیابی ژنومی موجب افزایش تعداد اثرات قابل برآوردی شده و در نتیجه کاهش برتری این روش نسبت به استفاده از SNPها را می‌تواند به‌همراه داشته باشد. روشی که برای حل این مشکل پیشنهاد می‌شود استفاده از شباهت بین هاپلوتایپ‌ها و مدل نویسی کواریانس اثرات ژنومی می‌باشد. هیکی و همکاران (۲۰۱۳) با استفاده از داده‌های شبیه‌سازی شده ماتریس روابط ژنومی بین افراد را از طریق محاسبه شباهت بین هاپلوتایپ‌ها تشکیل دادند. هرچند در مطالعه‌ی آنها صحت برآورد ارزش اصلاحی افزایش چندانی نیافت اما آنها همبستگی بالاتری بین اجزا قطری و غیر قطری ماتریس G_{HAP} و ماتریس ژنومی با استفاده از اطلاعات QTLها (G_{QTL}) نسبت به G_{SNP} و G_{QTL} گزارش نمودند. میوسن و همکاران (۲۰۱۴) با تعریف ژنوتیپ SNP کاذب^۳ دو آللی برای هاپلوتایپ‌های مشاهده شده در جمعیت و سپس ساخت ماتریس روابط خویشاوندی ژنومی با استفاده از این SNPهای کاذب

- 1- Linkage disequilibrium
- 2- Short range linkage disequilibrium
- 3- Pseudo SNP

شدند. در نهایت تعداد ۴۲۸۹۲ نشانگر SNP مورد بررسی قرار گرفت. متوسط عدم تعادل فاز پیوستگی (r^2) میان نشانگرها در داده‌های ژنوتیپ مورد بررسی در شکل ۱ آورده شده است. بیشترین کاهش LD در فواصل کمتر از ۱۰۰ kb مشاهده شد. در جمعیت مورد بررسی میانگین r^2 در فاصله ۵۰ kb بین دو نشانگر حدود ۰/۲۱ بود. روند کاهش در میزان LD با افزایش فاصله بین نشانگرها در جمعیت مورد بررسی مطابق با الگوی کاهش مورد انتظار برای جمعیت گاو هلشتاین و در توافق با مطالعات پیشین بود (۱).

شامل ۵۳۱۵ و ۱۳۲۲ گاو نر بود. حیوانات با استفاده از تراشه ۵۰k ژنوتیپ شده بودند. تنها نشانگرهای موجود بر کروموزوم‌های اتوزوم مورد بررسی قرار گرفت. از آنجایی که نشانگرها با فراوانی پایین ممکن است تنها در یک یا چند خانواده از یک جمعیت دیده شوند که موجب ارتباط تصادفی بین آن نشانگرها و صفات مورد بررسی شده و در نتیجه موجب کاهش صحت برآورد اثر نشانگر می‌شود (۵)، لذا اطلاعات نشانگرهایی که دارای فراوانی آلی نادر کمتر از ۰/۰۲ بودند و در تعادل هاردی واینبرگ قرار نداشتند از داده‌های ژنوتیپ کل حذف



شکل ۱- توزیع عدم تعادل پیوستگی در جمعیت گاو شیری هلشتاین آمریکای شمالی
Figure 1- Distribution of linkage disequilibrium in North American Holstein cattle

طریق محاسبه ضریب رگرسیون پروف ارزش اصلاحی ژنومی بر ارزش‌های اصلاحی برآوردی و همبستگی بین پروف ارزش اصلاحی ژنومی و ارزش‌های اصلاحی برآوردی در جمعیت تایید محاسبه گردید.

تشکیل ماتریس‌های روابط ژنومی (G)

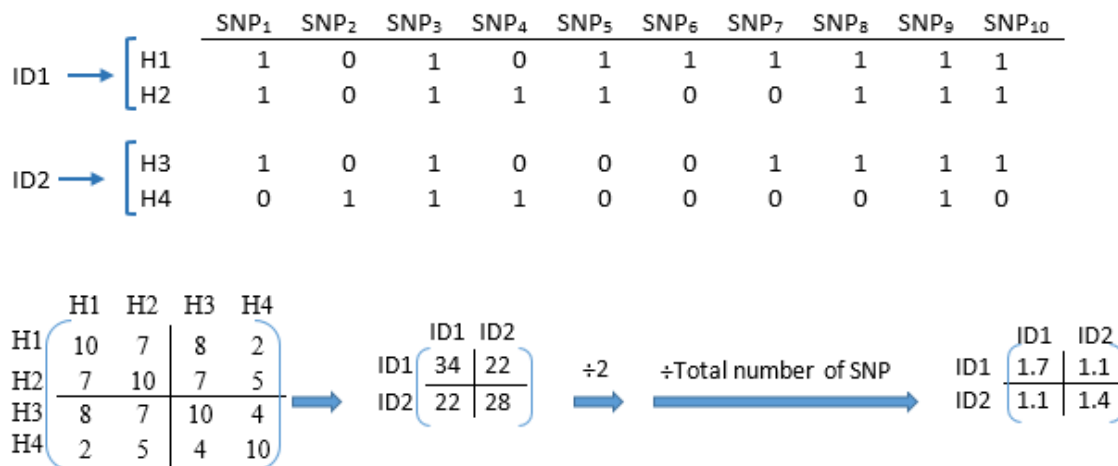
۱- ماتریس G_{HAP} : به منظور ایجاد ماتریس شباهت ژنومی بر پایه اطلاعات هاپلوتایپ‌ها، ابتدا برای هر کروموزوم ماتریس گامتی (M_{chr}) به طور ساده از طریق یافتن نواحی مشترک بین هاپلوتایپ‌های افراد ایجاد شد. سپس ماتریس وزنی M به صورت میانگین ماتریس‌های M_{chr} برای همه کروموزوم‌ها محاسبه شد. در نهایت ماتریس روابط ژنومی بر پایه اطلاعات هاپلوتایپ (G_{HAP}) از طریق حاصل جمع ۴ جزء متناظر ماتریس M محاسبه شده بر اساس شباهت هاپلوتایپ‌ها بین دونفر (پدری فرد ۱ و مادری فرد ۲، پدری فرد ۱ و مادری فرد ۲، مادری فرد ۱ و پدری فرد ۲، مادری فرد ۱ و مادری فرد ۲) تقسیم بر ۲ به دست آمد (شکل ۲).

برآورد ارزش اصلاحی ژنومی

ارزش اصلاحی حیوانات با استفاده از روش بهترین برآورد ناریب خطی ژنومی طبق رابطه زیر برآورد شد:

$$y = 1\mu + Zu + e \quad \text{رابطه ۱}$$

که در آن y بردار پروف رسمی گاوهای نر در سال ۲۰۱۱، Z ماتریس ضرایبی مربوط به اثرات تصادفی، u بردار اثرات تصادفی با واریانس $G\sigma_u^2$ و μ میانگین، 1 وکتور ضرایب حاوی عدد ۱ و e باقیمانده تصادفی برای حیوان می‌باشد. تمام ارزیابی‌ها با استفاده از نرم‌افزار SNP1101 (۱۶) صورت گرفت. پروف ارزش‌های اصلاحی بر اساس ارزیابی ژنتیکی در سال ۲۰۱۵ به منظور بررسی صحت و اریبی ارزش اصلاحی ژنومی برآورد شده مورد استفاده قرار گرفت. قابلیت اطمینان پروف ارزش اصلاحی برای صفات تولید چربی شیر، تولید شیر، نمره سلول بدنی، تیپ و روزهای باز به ترتیب ۹۳/۳۶، ۹۳/۴۲، ۸۳/۷۸ و ۷۴/۲۱ بود. اریبی و صحت برآوردها به ترتیب از



شکل ۲- توصیف روش مورد استفاده برای ساخت ماتریس روابط ژنومی بر پایه اطلاعات هاپلوتایپها (G_{HAP}). به عنوان مثال ژنوم با ۱۰ نشانگر تک نوکلئوتیدی و دو فرد مورد نظر است.

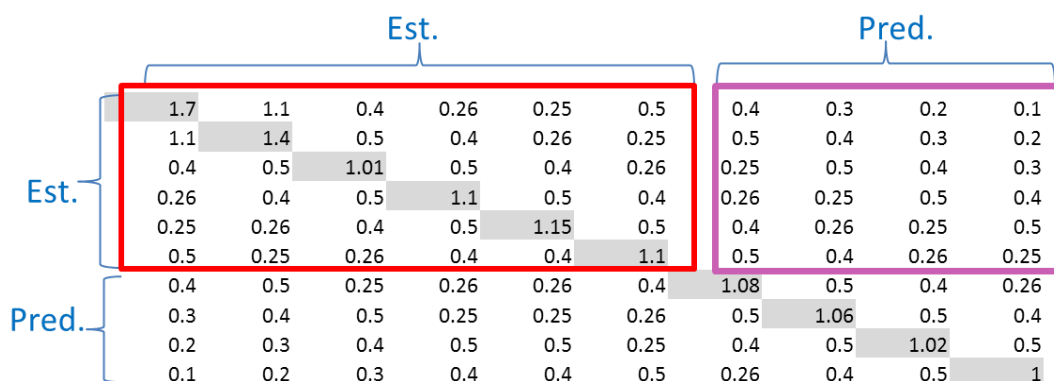
Figure 2- Description of the method used to construct haplotype based relationship matrix (G_{HAP}). For example, the genome with 10 SNP for 2 individuals is considered

۳- ماتریس هیبرید G_{SNP,HAP}: این ماتریس با استفاده از مجموعه اطلاعات موجود در ماتریس‌های روابط خویشاوندی ژنومی بر پایه SNPها و هاپلوتایپها محاسبه شد. به این ترتیب ماتریس G_{SNP,HAP} به ترتیب بر اساس محاسبه شباهت بین افراد در جمعیت رفرنس با استفاده از اطلاعات SNPها و شباهت بین افراد در جمعیت تایید با استفاده از اطلاعات هاپلوتایپها به دست آمد (شکل ۳).

۲- ماتریس G_{SNP}: در مطالعه‌ی حاضر G_{SNP} به روش اول وان رادن (۲۰۰۸) جهت محاسبه ماتریس روابط خویشاوندی ژنومی مورد نظر بود. برای تشکیل ماتریس از فرمول زیر استفاده شد:

$$G = ZZ' / 2 \sum p(1-p)$$

که در آن ماتریس Z حاوی 0 - 2p برای ژنوتیپ هموزیگوت، 1 - 2p برای ژنوتیپ هتروزیگوت و 2 - 2p برای ژنوتیپ هموزیگوت دیگر می‌باشد. منظور از p فراوانی آلی هر SNP می‌باشد (۱۹).



شکل ۳- طرح کلی روش مورد استفاده در تشکیل ماتریس G_{SNP,HAP}، از طریق محاسبه روابط بین افراد در جمعیت رفرنس و همچنین بین جمعیت رفرنس و جمعیت تایید به ترتیب با استفاده از اطلاعات SNPها و هاپلوتایپها

Figure 3- Schematic pattern of the method used to construct hybrid matrix, G_{SNP,HAP} (estimation of similarity between individuals in the estimation group using SNP-based (G_{SNP}) and similarity between estimation and prediction group using haplotype-based relationship matrix (G_{HAP}))

نتایج و بحث

یکی از روش‌های رایج در ارزیابی ژنومی روش GBLUP است. روش GBLUP بر اساس مدل‌های خطی مختلط بیان شده و در آن از ماتریس روابط خویشاوندی ژنومی (G) بر اساس اطلاعات نشانگرها استفاده می‌شود (۱۹). در مقایسه با روش‌های رایج که از ماتریس روابط خویشاوندی بر پایه شجره (A) استفاده می‌کنند، روش GBLUP قادر به در نظر گرفتن عدم تعادل بین نشانگر و ژن‌های مسبب، تفرق مندلی و برقراری ارتباط ژنتیکی بین اجداد مشترک نامشخص در شجره می‌باشد. بنابراین ماتریس G دارای برتری نسبت به ماتریس روابط خویشاوندی شجره در ارزیابی ژنتیکی است (۱۸). در این مطالعه ۳ روش تشکیل ماتریس G مورد بررسی قرار گرفت. میزان صحت ارزش‌های اصلاحی برآوردی با استفاده از ماتریس‌های روابط ژنومی مورد بررسی برای صفات مختلف در جدول ۱ آورده شده است.

صحت برآورد ارزش اصلاحی ژنومی با استفاده از ماتریس G_{SNP} برای صفات مورد بررسی در توافق با نتایج مطالعه‌ی وان رادن و همکاران (۲۰۰۹) با استفاده از جمعیت گاو هلشتاین آمریکای شمالی بود. بالاترین صحت برآورد برای صفت تولید چربی شیر و کمترین صحت برآورد برای صفت روزهای باز به‌دست آمد.

صحت برآورد ارزش اصلاحی با استفاده از اطلاعات هاپلوتایپ‌ها در مقایسه با ماتریس G_{SNP} در تمام صفات مورد بررسی به جز صفت تیپ در حدود ۱ درصد افزایش یافت. صحت برآورد ارزش اصلاحی برای صفت تولید چربی شیر با استفاده از ماتریس G_{HAP} و G_{SNP} به ترتیب ۰/۷۳ و ۰/۷۲ به‌دست آمد. دروس و همکاران (۲۰۱۰) صحت برآورد ارزش اصلاحی ژنومی برای این صفت را با استفاده از ماتریس روابط ژنومی و با استفاده از اطلاعات هاپلوتایپ اجدادی با طول ۱۵ هاپلوتایپ به ازای هر جایگاه و مدل زنجیره مارکوف به ترتیب ۰/۶۷ و ۰/۷۰ برآورد نمودند. تفاوت در صحت برآورد ارزش اصلاحی در این دو مطالعه می‌تواند به دلیل تفاوت در اندازه جمعیت رفرنس و قابلیت اطمینان رکوردهای به‌دست آمده در هر یک از این دو مطالعه باشد. هیکی و همکاران (۲۰۱۳) با استفاده از داده‌های شبیه‌سازی شده صحت برآورد ارزش اصلاحی ژنومی با استفاده از ماتریس G_{HAP} را برای صفتی با وراثت‌پذیری ۰/۲۵ و توزیع گاما برای اثرات QTL مشابه با ماتریس G_{SNP} و کمتر از ماتریس G_{QTL} گزارش نمودند. این محققین مدعی شدند که عدم بهبود صحت برآورد ارزش اصلاحی ژنومی در مطالعه‌ی آنها می‌تواند به دلیل مدل شبیه‌سازی به‌کار گرفته شده باشد چرا که در آن مطالعه تمام QTL‌های شبیه‌سازی شده بدون در نظر گرفتن زمینه ژنتیکی و جمعیتی حیوانات، دارای اثر ژنتیکی افزایشی یکسانی بودند (۹). در مطالعه‌ی حاضر از داده‌های

واقعی و صفات با ساختار ژنتیکی متفاوت برای بررسی صحت برآورد ارزش اصلاحی ژنومی با استفاده از اطلاعات هاپلوتایپ در مقایسه با ماتریس G_{SNP} استفاده شد. به طور کلی صحت برآورد ارزش اصلاحی هنگام استفاده از ماتریس G_{HAP} نسبت به ماتریس G_{SNP} در تمام صفات مورد بررسی افزایش ناچیزی یافت. این افزایش اندک در میزان صحت احتمالاً به دلیل برتری هاپلوتایپ‌ها در توضیح LD در جمعیت رفرنس نسبت به نشانگر انفرادی می‌باشد. به طور کلی هاپلوتایپ‌ها قادرند اطلاعات دقیق‌تری در مورد نواحی ژنومی با منشا جدی یکسان (IBD) نسبت به نشانگرهای انفرادی فراهم کنند و هاپلوتایپ‌های طویل‌توانایی بالاتری از این لحاظ در مقایسه با هاپلوتایپ‌های کوتاه‌تر دارند (۹). عدم افزایش قابل توجه در صحت برآوردها با استفاده از ماتریس G_{HAP} در این مطالعه نشان از بهینه نبودن همبستگی بین روابط تشابه آلل‌های با منشا غیر جدی (IBS) و جدی (IBD) می‌باشد، لذا به منظور افزایش هرچه بیشتر صحت برآورد ارزش‌های اصلاحی ژنومی به‌کارگیری عوامل تصحیح‌کننده جهت بهینه نمودن همبستگی بین ارتباطات IBS و IBD در تشکیل ماتریس G_{HAP} توصیه می‌شود. در این راستا در نظرگرفتن فراوانی آللی نشانگرهای مشابه در هر هاپلوتایپ و همچنین طول هر هاپلوتایپ به صورت بهینه می‌تواند از جمله عوامل مؤثر در بهبود همبستگی بین ارتباطات IBS و IBD در محاسبه روابط خویشاوندی بین افراد باشد.

معیار مهم دیگر در ارزیابی ارزش‌های ژنتیکی تعیین شده، اریبی برآوردها می‌باشد. ضریب رگرسیون نزدیک به یک نشان از ناریب بودن برآوردها و مقادیر کوچکتر و یا بزرگتر از یک به ترتیب اریب روبه بالا و پایین را نشان می‌دهد. اگرچه روش GBLUP اساساً روشی ناریب جهت برآورد ارزش‌های اصلاحی می‌باشد اما با ثابت بودن تمام عوامل، تنها عامل ایجادکننده‌ی تفاوت در میزان اریبی برآوردها در مطالعه‌ی حاضر، مربوط به نوع ماتریس ارتباط ژنومی استفاده شده می‌باشد. همانطور که در شکل ۴ مشاهده می‌شود استفاده از ماتریس G_{HAP} نسبت به ماتریس G_{SNP} منجر به برآوردهایی با کمترین اریبی شد. ماتریس‌های ارتباط ژنومی مختلف، واریانس ژنتیکی و باقیمانده را به صورت‌های مختلفی نشان می‌دهند و یا به عبارت دیگر واریانس ژنتیکی را برای جمعیت پایه در نسل‌های متفاوت برآورد می‌کنند (۹). با توجه به ارتباط بین صحت (r) و اریبی

$$(b) \text{ برآورد ارزش اصلاحی } \left(r = b \frac{\sigma_g}{\sigma_{g\bar{g}}} \right) \text{ که در آن } \sigma_g \text{ و } \sigma_{g\bar{g}} \text{ به}$$

ترتیب جذر واریانس ژنتیکی واقعی و برآوردی است (۳) و تشابه

1- Identical by descent

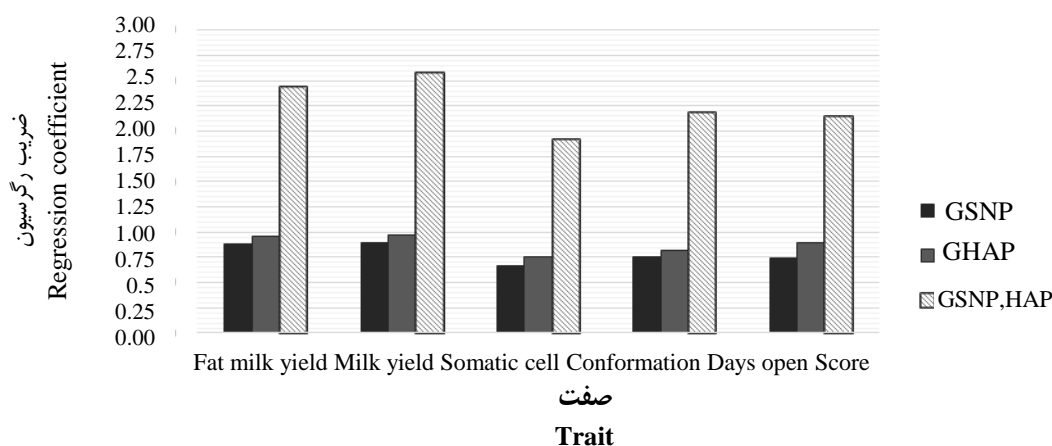
2- Identical by state

صحت برآورد ارزش اصلاحی از طریق ماتریس G_{SNP} و G_{HAP} در مطالعه‌ی حاضر، ضریب رگرسیون نزدیک تر به یک در زمان استفاده از اطلاعات هاپلوتایپ‌ها می‌باشد.

جدول ۱- برآورد صحت ارزش اصلاحی ژنومی با استفاده از ماتریس روابط خویشاوندی ژنومی بر پایه اطلاعات نشانگرهای چندشکلی تک نوکلئوتیدی (G_{SNP}) و هاپلوتایپ (G_{HAP}) و ماتریس هیبرید شامل اطلاعات هاپلوتایپ‌ها و SNPها ($G_{SNP,HAP}$) در جمعیت گاو شیری آمریکای شمالی

Table 1- Accuracies of estimates of genomic breeding values based on SNP-based (G_{SNP}), haplotype-based (G_{HAP}) and hybrid of SNP and haplotype based ($G_{SNP,HAP}$) relationship matrix in North Holstein dairy cattle

صفت Trait	G_{SNP}	G_{HAP}	$G_{SNP,HAP}$
چربی شیر Milk fat yield	0.72	0.73	0.72
تولید شیر Milk yield	0.70	0.71	0.70
نمره سلول بدنی Somatic cell score	0.61	0.62	0.61
تیپ conformation	0.56	0.54	0.56
روزهای باز Days open	0.49	0.50	0.49



شکل ۴- ضریب رگرسیون پروف رسمی ارزش اصلاحی سال ۲۰۱۵ بر برآورد ارزش اصلاحی ژنومی در صفات مختلف با استفاده از ماتریس روابط خویشاوندی ژنومی بر پایه اطلاعات نشانگرهای چندشکلی تک نوکلئوتیدی (G_{SNP}) و هاپلوتایپ (G_{HAP}) و ماتریس هیبرید شامل اطلاعات هاپلوتایپ‌ها و SNPها ($G_{SNP,HAP}$) در جمعیت گاو شیری آمریکای شمالی

Figure 4- Regression coefficient of de-regressed EBV (dEBV) based on the 2015 genetic evaluation (dEBV2015) on estimated genomic breeding value in different traits using SNP-based (G_{SNP}), haplotype-based (G_{HAP}) and hybrid of SNP and haplotype based ($G_{SNP,HAP}$) relationship matrix in North Holstein dairy cattle

است را بهتر نشان می‌دهند و این امر می‌تواند اثرات اپیستاتیک موضعی میان ژن‌ها درون یک هاپلوتایپ را نیز شامل شود. با توجه به ماهیت چند آلی هاپلوتایپ‌ها، استفاده از هاپلوتایپ بهتر می‌تواند LD بین QTL‌های چند آلی را توضیح دهد و این امر خصوصاً در صفاتی با وراثت‌پذیری پایین که تحت کنترل تعداد زیادی ژن‌های چندآلی هستند اهمیت بیشتری می‌یابد. فردوسی و همکاران (۲۰۱۶) نشان دادند که استفاده از هاپلوتایپ‌های متشکل از چندین نشانگر SNP

استفاده از ماتریس G_{HAP} نسبت به G_{SNP} موجب ۶ تا ۱۵ درصد کاهش ارایی برآوردها در صفات مختلف شد. بیشترین بهبود در ارایی برآوردها در صفت روزهای باز با وراثت‌پذیری ۰/۱ مشاهده شد. پارکرگادیس و همکاران (۲۰۱۴) در توافق با مطالعه‌ی حاضر بیان داشتند که استفاده از اطلاعات هاپلوتایپ خصوصاً در صفات با وراثت‌پذیری پایین می‌تواند موجب افزایش قابلیت پیش‌گویی گردد. هاپلوتایپ‌ها احتمالاً اثرات چندین QTL که جدا نمودن آنها دشوار

جمعیت رفرنس و تایید و به عبارت دیگر بهینه نبودن همبستگی بین روابط IBS و IBD می باشد. همچنین با توجه به اینکه دو ماتریس G_{SNP} و G_{HAP} به گونه‌ای متفاوت واریانس ژنتیکی را تعیین می‌کنند، احتمالاً آریب رو به بالای برآوردهای حاصل از ماتریس هیبرید به دلیل مقیاس متفاوت عناصر این ماتریس‌ها می‌باشد. با توجه به نتایج حاضر، مطالعات بیشتری در رابطه با نحوه ترکیب اطلاعات هاپلوتایپ‌ها و SNP به منظور بهبود صحت و دقت برآوردهای ارزش اصلاحی ژنومی لازم می‌باشد.

نتیجه‌گیری کلی

بر اساس نتایج این مطالعه، ماتریس G_{HAP} می‌تواند به‌عنوان یک روش جایگزین به منظور کاهش آریبی برآوردهای ارزش اصلاحی ژنومی در ارزیابی‌های رایج مورد استفاده قرار گیرد. در زمانی که از اطلاعات هاپلوتایپ‌ها در تشکیل ماتریس روابط ژنومی استفاده شد، آریبی برآوردها خصوصاً در صفاتی با وراثت‌پذیری پایین که اخیراً تحت انتخاب قرار داشتند به طور قابل توجهی کاهش یافت. بهبود اندک در صحت برآوردها با استفاده از ماتریس G_{HAP} حاکی از بهینه نبودن همبستگی بین روابط IBS و IBD در این ماتریس می‌باشد. بر اساس نتایج حاضر و به منظور بهبود قابل توجه در صحت و دقت ارزیابی‌ها، مطالعات بیشتری در رابطه با چگونگی تعریف هاپلوتایپ به منظور بهینه نمودن همبستگی بین ارتباطات IBS و IBD توصیه می‌شود. همچنین لازم است که مطالعات بیشتری جهت ارزیابی اثر انتخاب، وراثت‌پذیری و ساختار ژنتیکی صفات بر عملکرد ماتریس G_{HAP} صورت گیرد.

سپاسگزاری

بدین‌وسیله از شرکت Canadian Dairy Network جهت فراهم نمودن داده‌های ژنوتیپ جمعیت گاو هلشتاین آمریکای شمالی تشکر و قدردانی می‌گردد.

می‌تواند موجب افزایش احتمال حداکثر درست‌نمایی ارزیابی ژنومی نسبت به SNP انفرادی شود. ماتریس روابط ژنومی G_{SNP} با استفاده از نشانگرها با فراوانی بالای آلل نادر ساخته می‌شوند. بنابراین G_{SNP} قادر است روابط ژنومی افراد در نسل‌های قدیمی را نشان دهد و بنابراین ممکن است تغییرات حاصل به دلیل انتخاب در نسل‌های جدید را به طور دقیق منعکس نکند (۱۱). از طرف دیگر هاپلوتایپ‌ها خصوصاً هاپلوتایپ‌های طولانی‌تر بهتر می‌توانند جهش‌ها و روابط ژنومی در نسل‌های اخیر را نشان دهند. جمعیت گاو هلشتاین آمریکای شمالی تحت انتخاب شدید برای صفات تولیدی و تیپ برای سال‌های متمادی بوده است. تغییر راهبردهای انتخاب از سال ۲۰۰۵ موجب افزایش وزن بر صفات تولیدمثلی و سلامت در شاخص انتخاب شده است (۱۲). بنابراین می‌توان انتظار داشت که استفاده از اطلاعات هاپلوتایپ‌ها به جای نشانگرهای SNP منجر به برآورد دقیق‌تر و نارایب‌تر ارزش اصلاحی ژنومی برای صفات روزهای باز و نمره سلول بدنی در بین صفات مورد بررسی شود. استفاده از ماتریس G_{HAP} به جای G_{SNP} موجب کاهش آریبی برآوردها در حدود ۱۵ و ۹ درصد به ترتیب در صفت روزهای باز و نمره سلول بدنی شد (شکل ۴).

با توجه به قابلیت بالاتر ماتریس G_{HAP} در نشان دادن جهش‌ها و روابط ژنومی در نسل‌های اخیر، در این مطالعه برای اولین بار فرضیه تشکیل ماتریس هیبرید برای امکان‌سنجی استفاده توأم از مزایای هاپلوتایپ و SNP مطرح شد. صحت برآورد ارزش‌های اصلاحی در حالت استفاده از ماتریس هیبرید مشابه ماتریس G_{SNP} و بین ۰/۴۹ (صفت روزهای باز) تا ۰/۷۰ (صفت تولید چربی شیر) متغیر بود. هایبر و همکاران (۲۰۰۷) صحت برآورد ارزش اصلاحی ژنومی مشاهده شده را تحت تاثیر میزان LD در جمعیت رفرنس و روابط خویشاوندی بین افراد در جمعیت رفرنس و تایید دانستند. میزان LD در جمعیت رفرنس در ماتریس هیبرید بر اساس اطلاعات نشانگرهای انفرادی و روابط خویشاوندی بین افراد در جمعیت رفرنس و تایید بر اساس اطلاعات هاپلوتایپ‌ها به‌دست می‌آید. با توجه به یکسان بودن میزان LD به‌دست آمده در دو ماتریس هیبرید و G_{SNP} ، تشابه صحت در این دو ماتریس نشان از عدم برتری اطلاعات هاپلوتایپ در مقایسه با نشانگر انفرادی در نشان دادن روابط خویشاوندی بین افراد در

منابع

1. Bohmanova, J., M. Sargolzaei, and F. S. Schenkel. 2010. Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics*, 11:421.
2. Calus, M. P. L., T. H. E. Meuwissen, A. P. W. de Roos, and R. F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, 178:553-561.
3. Dekkers, J. C. M. 2005. *Animal Breeding Strategies*. Available at <http://www.anslab.iastate.edu/Class/AnS652X/Chapter4.pdf>.
4. de Roos, A. P. W., C. Schrooten, and T. Druet. 2011. Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. *Journal of Dairy Science*, 94:4708-4714.
5. Edriss, V., B. Gulbrandtsen, M. S. Lund, and G. Su. 2013. Effect of marker-data editing on the accuracy of

- genomic prediction. *Journal of Animal Breeding and Genetics*, 130:128-135.
6. Ferdosi, M. H., J. Henshall, and B. Tier. 2016. Study of the optimum haplotype length to build genomic relationship matrices. *Genetic Selection Evolution*, 48:75.
 7. Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177: 2389-2397.
 8. Habier, D., R. L. Fernando, and D. J. Garrick. 2013. Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics*, 194:597-607.
 9. Hickey, J. M., B. P. Kinghorn, B. Tier, S. A. Clark, J. H. J. van der Werf, and G. Gorjanc. 2013. Genomic evaluations using similarity between haplotypes. *Journal of Animal Breeding and Genetic*, 130: 259-269.
 10. Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157:1819-1829.
 11. Meuwissen, T. H., J. Odegard, I. Andersen-Ranberg, and E. Grindflek. 2014. On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. *Genetic Selection Evolution*, 46:49.
 12. Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *Journal of Dairy Science*, 88:1255-1263.
 13. Nejati Javaremi, A., C. Smith, and J. P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science*, 7: 1738-1745.
 14. Parker Gaddis, K. L., J. B. Cole., J. S. Clay, and C. Maltecca. 2014. Genomic selection for producer-recorded health event data in US dairy cattle. *Journal of Dairy Science*, 97:3197-3199.
 15. Sargolzaei, M., and F. S. Schenkel. 2009. QMSim: A large-scale genome simulator for livestock. *Bioinformatics*, 25(5):680-681.
 16. Sargolzaei, M., 2014. SNP1101 User's Guide. Version 1.0. HiggsGene Solutions Inc., Guelph, Ontario, Canada.
 17. Schaeffer, L. R. 2006. Strategy for applying genome wide selection in dairy cattle. *Journal of Animal Breeding and Genetic*, 123(4):218-223.
 18. Su, G., O. F. Christensen, L. Janss, and M. S. Lund. 2014. Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *Journal of Dairy Science*, 97:6547-6559.
 19. VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91:4414-4423.
 20. VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92:16-24.



Improving Genomic Evaluation of Holstein Cattle Using a Haplotype-Based Relationship Matrix

M. Forutan¹- S. Ansari Mahyari^{2*}- F. Schenkel³- M. Sargolzaei⁴

Received: 15-07-2017

Accepted: 16-12-2017

Introduction With the advent of high throughput genotyping technologies, interest has grown in using genomic information to estimate breeding values. Genomic selection, in which genetic markers across the whole genome are used to estimate breeding values of individuals, is routinely applied in dairy cattle breeding programs. In dairy cattle, genomic selection has resulted in a substantial increase in the rate of genetic gain compared to traditional selection. This has been achieved mainly by reducing the generation interval, which became possible because of the higher accuracies of genomic breeding values (GEBV) estimated early in life compared to parent averages. Since single nucleotide polymorphism (SNP) genotypes are bi-allelic and, therefore, their information content is not high, so SNP-based methods may not effectively capture the linkage disequilibrium (LD) between SNPs and multi-allelic quantitative trait loci (QTLs). Haplotypes are in general “multi-allelic” and compared to individual SNPs may better capture LD with multi-allelic QTL. Furthermore, most of the SNPs in the chips are old mutations. This may imply that SNP-based relationship matrix traces very old relationships from distant relatives and, therefore, may not trace changes due to recent selection accurately. It has also been hypothesized that using similarity between haplotypes to model the covariance between genomic effects can result in better predictive ability than modeling covariance based on SNP genotypes. So the objective of this study was to investigate the accuracy and bias of GEBV using genomic best linear unbiased prediction (GBLUP) with alternate genomic relationship matrices (G) based on SNPs and haplotypes information.

Materials and Methods The North American Holstein genotype data was provided by the Canadian Dairy Network (CDN). The Holstein bulls with official domestic proofs were genotyped using the Illumina Bovine SNP 50 TM Chip. The genotyped Holstein bulls were classified as estimation group and prediction group. All bulls in prediction group were those bulls who were born from 2007 to 2011 and had official proof in 2015. De-regressed EBV (dEBV) based on the 2015 genetic evaluation (dEBV2015) were used for validation purposes. The estimation group included bulls born mainly between 1960 to 2007. The analyzed traits were fat yield, milk yield, somatic cell score, conformation, and days open. The GEBVs for mentioned traits were estimated based on GBLUP models using SNP1101 software. Three G matrices were built: 1) SNP genotype based relationship matrix (G SNP), 2) haplotype based relationship matrix (GHAP), and 3) hybrid matrix composing of SNP genotype and haplotype based relationship matrices (G SNP, HAP). Accuracy of prediction was calculated as Pearson's correlation between estimated GEBV and dEBV2015 for prediction group. Bias of prediction was also calculated as regression coefficient of dEBV2015 on GEBV for prediction group.

Results and Discussion Observed differences between alternate G matrices were larger in prediction bias than in prediction accuracy. Accuracy of genomic predictions based on GHAP were 0.73, 0.71, 0.62, 0.54, and 0.50 for fat yield, milk yield, somatic cell score, conformation, and days open, respectively. Accuracy of genomic predictions based on G SNP also were 0.72, 0.70, 0.61, 0.56, and 0.49, respectively. Although using GHAP instead of G SNP did not significantly improve the accuracy of prediction, it resulted in estimates with 6 (conformation) to 15% (days open) less bias and closer to one over the different traits. Genomic selection based on GHAP instead of G SNP seems to improve bias of prediction for traits under recent selection, such as days open. This might also be due to the fact that this trait is lowly heritable. Although accuracy of genomic predictions based on G SNP, HAP were similar to G SNP, using G SNP, HAP resulted in higher biases than GHAP or G SNP.

Conclusion Based on these findings, GHAP might be an alternative approach to reduce bias of genomic

1 and 2- PhD student and Associate professor, respectively, Department of Animal Science, College of Agriculture, Isfahan University of Technology, Iran

3- Professor, Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Canada

4- Associate professor, Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Semex Alliance and HiggsGene Solutions Inc., Canada

(*- Corresponding author email: s.ansari@cc.iut.ac.ir)

predictions in routine genomic evaluations. Genomic selection based on GHAP instead of GSNP improved bias of prediction especially for low heritable traits under recent selection, such as days open. Small gain achieved with GHAP compared to GSNP may imply that the correlation between IBS and IBD in GHAP is not optimal for maximization of prediction accuracy. Based on current results further research is needed to investigate the use of haplotype length and allele frequency of markers in haplotype segments in the definition of haplotype similarity. In addition, this study suggests further research to assess the effect of recent selection, heritability and genetic architecture of the traits on the performance of haplotype based relationship matrix.

Keywords: Accuracy, Bias of prediction, Genomic relationship matrix, Haplotype, Holstein cattle