دانشگاه علوم کشاورزی و منابع طبیعی ساری
پژوهشهای تولیدات دامی

"مقاله پژوهشی"

اثر اندازه جمعیت مرجع و روش‌های استنباط ژنوتیپ بر صحت استنباط در جمعیت‌های خالص و آمیخته

یحیی محمدی[1] و جواد احمدپناه[2]

۱- استادیار و عضو هیات علمی گروه علوم دامی دانشگاه ایلام (نویسنده مسوول: mohamadi_yahya@yahoo.com)
۲- استادیار بخش تحقیقات علوم دامی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی کرمانشاه، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرمانشاه، ایران
تاریخ دریافت: ۱۳۹۸/۱۲/۱۱ تاریخ پذیرش: ۱۳۹۹/۰۳/۱۸
صفحه: ۱۰۹ تا ۱۱۴

چکیده

استنباط ژنوتیپ به‌عنوان یک روش تبدیل تراشه‌های با تراکم پایین به تراشه‌های با تراکم بالا به‌منظور افزایش صحت انتخاب ژنومی در حیوانات است. در پژوهش حاضر به‌منظور بررسی صحت استنباط ژنوتیپ سه جمعیت آمیخته (سناریو ۱)، خالص (سناریو ۲) و آمیخته + خالص (سناریو ۳) به کمک نرم‌افزار QMSim شبیه سازی شد. از دو روش استنباط ژنوتیپ Beagle و Flmpute برای دو نوع تراشه با تراکم کم و زیاد استفاده شد. اندازه جمعیت مرجع برای هر سناریو ۲۵۰، ۵۰۰ و ۱۰۰۰ حیوان انتخاب شد. نتایج پژوهش حاضر نشان داد که برای تمام سناریوها با افزایش اندازه جمعیت مرجع از ۲۵۰ به ۵۰۰ حیوان، افزایش صحت استنباط ژنوتیپ بیشتر و از ۵۰۰ به ۱۰۰۰ حیوان این افزایش، کندتر شد. صحت استنباط ژنوتیپ به‌کمک روش Flmpute نسبت به Beagle برای اندازه جمعیت مرجع کوچک (۲۵۰ حیوان) بیشتر نمایان شد. در اندازه‌های جمعیت مرجع ۵۰۰ و ۱۰۰۰ حیوان در تمام سناریوها افزایش صحت Flmpute نسبت به‌روش Beagle معنی‌دار نبود. برای سناریو ۱ میزان صحت استنباط ژنوتیپ نسبت به سناریو ۲ بیشتر برآورد شد. افزایش صحت استنباط ژنوتیپ در سناریو ۳ در مقایسه با سناریو ۱ بی‌معنی شد. به‌طور کلی نتایج پژوهش حاضر نشان داد که در کشورهای در حال توسعه که جمعیت حیوانی تعیین ژنوتیپ شده کوچک در دسترس هست، برای افزایش صحت انتخاب ژنومی استفاده از روش استنباط ژنوتیپ Flmpute به کمک جمعیت آمیخته با افزایش ارتباط خویشاوندی بین جمعیت مرجع و هدف می‌توان به‌عنوان یک راهکار مناسب استفاده شود.

واژه‌های کلیدی: انتخاب ژنومی، استنباط ژنوتیپ، جمعیت آمیخته، شبیه‌سازی، صحت پیش‌بینی

مقدمه

انتخاب حیوانات به‌کمک برآورد ارزش اصلاحی ژنومی[1] (GEBV) به‌عنوان یک روش استاندارد برای بهبود ژنتیکی بسیاری از گونه‌های دامی مورد توافق می‌باشد. در انتخاب ژنومی از عدم تعادل پیوستگی (LD) بین نشانگرهای SNP شناخته شده و جهش‌های علی[2] ناشناخته برای تخمین ارزش اصلاحی ژنومی کمک گرفته می‌شود (۱). در انتخاب ژنومی از معادلات پیش بینی ژنومی در جمعیت مرجع بزرگ برای تخمین ارزش‌های اصلاحی ژنومی برای کاندیداهای انتخاب که فقط دارای اطلاعات ژنومی هستند استفاده می‌شود (۱۴). انتخاب ژنومی به‌ویژه در هنگام استفاده ار روش‌های سنتی برآورد ارزش اصلاحی که اطلاعات شجره در دسترس نیست بسیار مهم می‌باشد (۱). در بسیاری از پژوهش‌ها گزارش شده است که با افزایش اندازه جمعیت مرجع و استفاده از تراشه‌های نشانگری با تراکم بالا می‌توان صحت GEBV را در جمعیت‌های مرجع افزایش داد (۱۳،۹). تراشه‌های نشانگری برای توجیه کردن LD بین نشانگرها و QTLها مؤثر و مفید می‌باشند. گزارش شده که برای حصول LD مناسب در داخل یک نژاد به تراشه با تراکم ۵۰ هزار کیلو بازی برای حداکثر صحت پیش بینی نیاز است و در داخل گله‌های مختلط تراکم تراشه باید بیشتر باشد (۷). این امر به‌ویژه زمانی که اندازه جمعیت مرجع کوچک و دیگر نژادها هم در حیوانات جمعیت مرجع برای برآورد اثرات نشانگر وجود دارند، بیشتر احساس می‌شود. اگرچه هزینه تعیین ژنوتیپ به‌صورت محسوس کاهش پیدا نموده است اما با این وجود باید گفته شود که تهیه نشانگرهای با تراکم بالا به‌ویژه برای کشورهای در حال توسعه و گله‌های کوچک یک امر پرهزینه است. در این کشورها یک روش مقرون به‌صرفه برای تبدیل تراشه‌های با تراکم پایین نشانگر به تراشه‌های با تراکم بالا استفاده از روش استنباط ژنوتیپ (ایمپیوتیشن) است (۱). تعداد مناسب نشانگرهای SNP و یک الگوریتم مناسب برای انتخاب آن‌ها در یک تراشه با چگالی کم و تبدیل این تراشه به‌کمک استنباط ژنوتیپ برای ایجاد یک تراشه با چگالی زیاد هنوز ناشناخته است (۱). فاکتورهای زیادی صحت استنباط ژنوتیپ را تحت تأثیر قرار می‌دهند که از آن‌ها می‌توان به اندازه جمعیت مرجع و رابطه خویشاوندی بین جمعیت مرجع و هدف اشاره نمود (۲۷،۲۲). هابیر و همکاران (۸) گزارش کردند که جایگذاری نشانگرهایSNP با فواصل یکسان در تمام سطح ژنوم برای تراشه‌های کم تراکم[3] و کمک از روش استنباط ژنوتیپ می‌توان به یک تراشه با تراکم و دقت بالا دست یافت. تلاش‌های دیگری در راستای استفاده از تراشه‌های با تراکم پایین براساس جایگذاری یکسان نشانگرها در تمام سطح ژنوم و حداکثر نمودن فرآوانی آلل‌های کمیاب مطرح شده است (۳،۲). برای استنباط ژنوتیپ نشانگرهای SNP روش‌های متعددی وجود دارند. این گونه روش‌ها براساس شجره، تفرق آلل‌ها، عدم تعادل پیوستگی (LD)، اطلاعات جمعیت و ترکیب متقابل آنها می‌باشند. از بین روش‌های موجود، می‌توان روش بگل (Beagle) (۵)، فیندهپ (Findhap) (۲۶) و پدایمپیوت (PedImpute) (۲۱) را نام برد. در بعضی موارد که بین افراد ارتباط خویشاوندی قوی

1- Genomic estimated breeding values (GEBV) 2- Causal mutations 3- Sparse

[DOR: 20.1001.1.22518622.1399.11.30.8.0]
[DOI: 10.52547/rap.11.30.109]

وجود ندارد به اطلاعات شجره برای ایجاد استنباط ژنوتیپ نیاز نیست. برای ایجاد استنباط ژنوتیپ در این گونه جوامع، نرم‌افزار Flmpute پیشنهاد می‌شود (۲۲). از طرف دیگر Beagle، یک ابزار نرم‌افزاری مؤثر و حافظه‌ای الکترونیکی است که از درون‌یابی‌خطی[1] برای ایمپیوت کردن واریانت‌های ژنوتیپ نشده با تنظیمات پیش‌فرض استفاده می‌کند (۱۱). تراشه نشانگرهای SNP و همچنین روش استنباط ژنوتیپ بیشتر برای جمعیت‌های خالص طراحی گردیده است. برای کشورهای در حال توسعه که اکثر گله‌های گاو شیری آمیخته و از طرفی حیوانات تعیین ژنوتیپ شده کمی در دسترس می‌باشد، روش استنباط ژنوتیپ برای ایجاد تراشه با تراکم بالا که حداکثر صحت پیش‌بینی ژنومی را ایجاد نماید، یک اصل پذیرفته شده است. هدف از این مطالعه بررسی صحت استنباط ژنوتیپ در سه جمعیت خالص، آمیخته و خالص+ آمیخته به‌کمک روش‌های Flmpute و Beagle، بررسی صحت استنباط ژنوتیپ برای دو تراشه با تراکم پایین و تراشه با تراکم بالا در راهبردهای متفاوت سه جمعیت ۳ و در نهایت بررسی صحت استنباط ژنوتیپ برای سناریوهای مختلف با اندازه‌های جمعیت مرجع متفاوت تحت روش‌های Flmpute و Beagle می‌باشد.

مواد و روش‌ها

به‌کمک نرم‌افزار QMsim (۲۳) سه جمعیت گاو شیری آمیخته (سناریو ۱)، خالص (سناریو ۲) و خالص+ آمیخته (سناریو ۳) شبیه‌سازی شد. بعد از شبیه‌سازی، از طریق دو روش متفاوت، Flmpute (۲۲) و Beagle (۶) به‌عنوان نرم‌افزارهای استنباط ژنوتیپ، به پیش‌بینی نقاط گم‌شده از طریق روابط فامیلی و الگوریتم‌های بر پایه جمعیت برای تأثیر روی صحت انتخاب ژنومی، اقدام شد. افراد جمعیت مرجع با استفاده از دو تراشه با تراکم ۷۹۰۰ نشانگر SNP (تراشه تراکم پایین) و تراشه با تراکم ۳۵۰۰۰ هزار نشانگر SNP (تراکم بالا) تعیین ژنوتایپ شدند. به‌منظور مطالعه تأثیر روابط بین افراد جمعیت مرجع و جمعیت هدف بر صحت استنباط ژنوتیپ، آمیزش پنج نسل ادامه پیدا نمود. در هر نسل ۲۰ حیوان نر برتر به‌طور تصادفی با ۲۰۰ حیوان ماده آمیزش داده شد که به تولید ۴۰۰ حیوان در هر نسل گردید. ژنوتیپ حیوانات نر طی دو راهبرد دو و سه‌مرحله‌ای، با استفاده از ژنوتیپ افراد جمعیت مرجع استنباط ژنوتیپی گردیدند. در راهبر دومرحله‌ای با تراشه تراکم پایین ژنوتیپ حیوانات نر برای تراشه با تراکم بالا ایمپیوت شد. در روش سه‌مرحله‌ای ابتدا ژنوتیپ‌های تراشه بالا از تراشه متوسط ایمپیوت شدند و در مرحله بعدی تراشه تراکم بالا از تراکم متوسط ایجاد شدند. راهبرد دو و سه‌مرحله‌ای در جدول ۱ نشان داده شد. در راهبرد دو مرحله‌ای بعد از ایمپیوت شدن تعداد ۳۴۱۷۸ نشانگر باقی ماند. در روش سه مرحله‌ای تعداد نشانگرهای باقی مانده بعد از ایمپیوتیشن ۳۴۰۲۰ عد بود.

جدول ۱- مراحل راهبرد دو و سه مرحله‌ای در استنباط ژنوتیپی

Table 1. Two and three-step strategy in genotypic imputation

راهبرد دومرحله‌ای		راهبرد سه‌مرحله‌ای		
تراشه تراکم پایین با ۷۹۰۰ نشانگر	→	تمام افراد نسل پایه (۴۰۰ حیوان)	← تراشه تراکم پایین با ۷۹۰۰ نشانگر	جمعیت مرجع
⇩		۲۰۰ نر در هر نسل	← تراشه تراکم متوسط	تراشه میانی
تراشه تراکم بالا با ۳۵۰۰۰ نشانگر	→	۲۰۰ نر در هر نسل	← تراشه تراکم بالا با ۳۵۰۰۰ نشانگر	جمعیت‌های آزمون

روابط خویشاوندی بین جمعیت مرجع و جمعیت هدف با توجه به اندازه جمعیت مرجع برای سناریوهای مختلف در جدول ۲ نشان داده شده‌اند. صحت استنباط ژنوتیپ به‌کمک ضریب همبستگی پیرسن[2] بین ژنوتیپ‌های مشاهده و ایمپیوت شده برآورد شد. برای معنی‌داری دو به دوی این روش‌ها از آزمون Hotelling Williams- t-test استفاده شد.

جدول ۲- خویشاوندی بین سناریوهای متفاوت تحت اندازه‌های متفاوت جمعیت مرجع

Table 2. Relationship between different scenarios under different sizes of reference population

	سناریوهای استنباط ژنوتیپ		
اندازه جمعیت مرجع	سناریو ۱	سناریو ۲	سناریو ۳
۲۵۰	۰/۲۵	۰/۱۲	۰/۲۶
۵۰۰	۰/۲۳	۰/۱۱	۰/۲۵
۱۰۰۰	۰/۲۲	۰/۱۰	۰/۲۳

سناریو ۱، ۲ و ۳ به‌ترتیب جمعیت آمیخته، جمعیت خالص و جمعیت آمیخته+ خالص

نتایج و بحث

صحت استنباط ژنوتیپ برای تمام سناریوهای با افزایش اندازه جمعیت مرجع افزایش پیدا نمود (جدول ۳ و ۴). در مطالعات مختلفی گزارش شده است که با افزایش اندازه جمعیت مرجع صحت استنباط ژنوتیپ افزایش خواهد یافت که نتایج پژوهش حاضر در مطابقت با این پژوهش‌ها بود (۲۰،۱۰،۱۶). افزایش صحت استنباط ژنوتیپ برای تمام سناریوها برای افزایش اندازه جمعیت مرجع از ۲۵۰ به ۵۰۰ حیوان شیب بیشتری دیده شد در مقابل افزایش اندازه جمعیت مرجع از ۵۰۰ به ۱۰۰۰ حیوان شیب افزایش، کاهشی شد (جدول ۳ و ۴). در مطالعه قریشی‌فر و همکاران (۱۰) افزایش اندازه مرجع از ۴۲ به ۲۰۲ حیوان باعث افزایش صحت

1- Interpolation

2- Pearsons correlation

[DOR: 20.1001.1.22518622.1399.11.30.8.0]
[DOI: 10.52547/rap.11.30.109]

استنباط ژنوتیپ به‌میزان ۸ واحد در حالیکه با افزایش جمعیت مرجع از ۲۰۲ به ۷۳۶ حیوان این افزایش فقط ۲ واحد گزارش شد. افزایش میزان صحت در اندازه جمعیت مرجع کوچک به اندازه جمعیت مرجع متوسط (۲۵۰ به ۵۰۰ حیوان) احتمالاً به‌دلیل افزایش تعداد هاپلوتایپ‌های مشترک در جمعیت مرجع باشد (۹)، در حالی‌که افزایش جمعیت مرجع از حالت متوسط به بزرگ (۵۰۰ به ۱۰۰۰ حیوان) احتمالاً باعث افزایش مقدار ناچیز هاپلوتیپ مشترک در جمعیت مرجع می شود (۱).

برای اندازه جمعیت مرجع کوچک (۲۵۰ حیوان) صحت استنباط ژنوتیپ به‌کمک روش Flmpute نسبت به Beagle بیشتر بود (جدول ۳). در اندازه‌های جمعیت مرجع ۵۰۰ و ۱۰۰۰ حیوان در تمام سناریوها افزایش صحت Flmpute نسبت به‌روش Beagle معنی‌دار نبود ($p>0/05$). در یک مطالعه که بر روی گاوهای آمیخته هلشتاین آفریقای جنوبی انجام شد، نتایج آنها نشان داد که صحت استنباط ژنوتیپ در روش Flmpute در مقایسه با روش Beagle بیشتر برآورد شد که با نتایج پژوهش حاضر مطابقت دارد (۱). همچنین در مطالعه قریشی‌فر و همکاران (۱۰) که صحت استنباط ژنوتیپ را بر روی گاومیش‌های آبی به‌کمک دو روش Flmpute و Beagle مقایسه نمودند، نتایج آن‌ها نشان داد که اندازه جمعیت مرجع نسبت به نوع نرم‌افزار تأثیر بیشتری بر روی صحت استنباط ژنوتیپ دارد. دلیل افزایش صحت در نرم‌افزار Flmpute در مقایسه با روش Beagle احتمالاً به دخالت دادن روابط خویشاوندی در جمعیت مرجع و هدف، اطلاعات پیوستگی میان خانواده‌ها و همچنین اطلاعات LD موجود در جمعیت است. همچنین گزارش شده است که با توجه به استفاده روش Flmpute از اطلاعات شجره، احتمالاً بهبود کیفیت فاز و توالی‌یابی ژنوتایپ باعث افزایش صحت استنباط ژنوتیپ در این روش نسبت به روش‌های دیگر گردد (۱۷). در مطالعه‌ای گزارش شده است که افزایش صحت استنباط ژنوتیپ برای مقایسه روش Flmpute نسبت به Beagle برای اندازه جمعیت مرجع کوچک معنی‌دار گزارش شد، و برای اندازه جمعیت مرجع بزرگتر از ۴۰۰ حیوان این افزایش صحت معنی‌دار نبود (۱۰). متوسط صحت استنباط ژنوتیپ برای روش Beagle برای سناریوی ۱، ۲ و ۳ به‌ترتیب ۰/۷۲، ۰/۵۹ و ۰/۷۴ و برای روش Flmpute به‌ترتیب ۰/۸۱، ۰/۶۹ و ۰/۸۲ بدست آمد. افزایش صحت سناریو ۱ در مقایسه با سناریو ۲ معنی‌دار ($p\leq 0/05$). اما در مقایسه با سناریو ۳ معنی‌دار نبود ($p> 0/05$). که این عمل را می‌توان به عدم خویشاوندی افراد خالص و آمیخته در جمعیت مرجع نسبت داد (سناریو ۳). در مطالعه حاضر صحت استنباط ژنوتیپ برای اندازه جمعیت مرجع ۱۰۰۰ حیوان در سناریو ۱ برای روش Flmpute و Beagle به‌ترتیب ۰/۸۷ و ۰/۸۶ برآورد شد که این تفاوت صحت بسیار ناچیز و بی‌معنی شد ($p>0/05$). در یک مطالعه با روش Beagle برای اندازه جمعیت مرجع ۱۰۰۰ حیوان از گاوهای شیری سوئدی برای تراشه با تراکم ۵۰ هزار کیلوباز به تراشه با تراکم بالاتر صحت استنباط ژنوتیپ تغییر ناچیز برآورد شد (۱۳). در مطالعه ما و همکاران (۱۳) صحت استنباط ژنوتیپ روش Beagle نسبت به Flmpute برای اندازه جمعیت مرجع ۳۰۰۰ حیوان، بیشتر گزارش شد که در تناقض با نتایج پژوهش حاضر بود که دلیل آن احتمالاً به نوع داده استفاده شده در پژوهش مربوط باشد زیرا که در مطالعه آنها از داده‌های واقعی گاوهای سوئدی و Finnish Red استفاده شده بود. در توافق با پژوهش حاضر که افزایش صحت استنباط ژنوتیپ روش Flmpute نسبت به Beagle مطالعه میلانسی و همکاران (۱۵) و قریشی‌فر و همکاران (۱۰) بود.

برای سناریو ۱ (جمعیت آمیخته) میزان صحت استنباط ژنوتیپ نسبت به سناریو ۲ (جمعیت خالص) بیشتر به‌دست آمد. دلیل افزایش صحت در سناریو ۱ نسبت به سناریو ۲ احتمالاً به خویشاوندی بیشتر افراد جمعیت مرجع در سناریو ۱ مرتبط است (جدول ۲).گزارش شد که هر چه بین جمعیت مرجع و هدف رابطه خویشاوندی بیشتر باشد صحت استنباط ژنوتیپ بیشتر برآورد می‌گردد (۱۱). در مطالعه دیگر گزارش شد، که صحت استنباط ژنوتیپ برای تراشه‌های با تراکم پایین بیشتر تحت تأثیر استراتژی استفاده شده برای استنباط ژنوتیپ ژنوتیپ‌ها است. انتخاب جمعیت مرجع و الگوریتم‌های استنباط ژنوتیپ در مرحله بعدی اهمیت قرار دارند (۱). اندازه جمعیت مرجع بزرگ عموماً باعث افزایش صحت استنباط ژنوتیپ می‌شود (۹). در هنگام استفاده از تراشه با تراکم پایین دارا بودن اجداد مشترک در جمعیت مرجع و تایید نسبت به اندازه جمعیت مرجع بر روی صحت استنباط ژنوتیپ تأثیر بیشتری دارد (۲۷). برای رسیدن به حداکثر صحت در جمعیت مرجع برای کشورهای در حال توسعه باید حیوانات آمیخته در این جمعیت حداکثر ارتباط و خویشاوندی را با هم داشته باشند. در مطالعه الیلو و همکاران (۱) صحت استنباط ژنوتیپ در حالت استفاده از تراشه با تراکم پایین برای جمعیت گاوهای آمیخته آفریقای جنوبی در مقایسه با جمعیت خالص که ارتباط خویشاوندی بیشتر با هم داشتند، صحت استنباط ژنوتیپ بیشتر برآورد نمودند. در سناریو ۳ (آمیخته+ خالص) اضافه نمودن جمعیت خالص به جوامع آمیخته صحت استنباط ژنوتیپ را در تمام حالات در مقایسه با سناریو ۱ (جمعیت آمیخته) به مقدار جزئی افزایش داده است. دلیل این امر که اضافه نمودن حیوانات خالص به جمعیت آمیخته (سناریو۳) باعث افزایش معنی‌دار در صحت استنباط ژنوتیپ نگردیده را احتمالاً می‌توان به دلیل خویشاوندی پایین جمعیت خالص با افراد جمعیت آمیخته نسبت داد که صحت را زیاد تحت تأثیر قرار نداد. در توافق با پژوهش حاضر مطالعات زیادی گزارش نمودند که استفاده از چند نژاد به‌جای یک نژاد در جمعیت مرجع باعث افزایش صحت استنباط ژنوتیپ می شود (۱۹،۴). دلیل افزایش صحت را می‌توان به تنوع بیشتر از تفرق هاپلوتیپ‌ها در تراشه جمعیت مرجع چند نژاد نسبت داد که قادر است یک فیت شدن بهتر برای عدم تعادل پیوستگی با جمعیت هدف ایجاد نماید (۱۱). در مطالعه اوگاوا و همکاران (۱۸) که بر روی گاوهای سیاه ژاپنی[1] انجام گردید، برای تراشه با تراکم ۱۰ کیلوبازی به تراشه با تراکم ۵۰ کیلوبازی افزایش صحت استنباط ژنوتیپ ۰/۹۷ گزارش شد که این افزایش صحت برای جمعیت مرجع با تعداد ۴۹۴ حیوان بود.

1- Japanese Black cattle

[DOR: 20.1001.1.22518622.1399.11.30.8.0]
[DOI: 10.52547/rap.11.30.109]

جدول ۳- صحت استنباط ژنوتیپ در سناریوهای متفاوت تحت اندازه جمعیت مرجع متفاوت برای تراشه با تراکم پایین
Table 3. .Accuracy of imputation in different scenarios under different reference population sizes for low-density chip

تراشه	نرم‌افزار	اندازه جمعیت مرجع	سناریوهای استنباط ژنوتیپ: سناریو ۱	سناریو ۲	سناریو ۳
تراشه با تراکم پایین	Beagle	۲۵۰	۰/۵۸	۰/۴۳	۰/۶۱
		۵۰۰	۰/۷۳	۰/۶۲	۰/۷۴
		۱۰۰۰	۰/۸۶	۰/۷۱	۰/۸۷
	Flmpute	۲۵۰	۰/۷۶	۰/۶۵	۰/۷۹
		۵۰۰	۰/۸۱	۰/۶۸	۰/۷۷
		۱۰۰۰	۰/۸۷	۰/۷۴	۰/۸۹

سناریو ۱، ۲ و ۳ به‌ترتیب جمعیت آمیخته، جمعیت خالص و جمعیت آمیخته+ خالص

صحت استنباط ژنوتیپ تحت سناریوهای متفاوت برای تراشه با تراکم بالا در جدول شماره ۳ نشان داده شده است. در تمام سناریوها صحت استنباط ژنوتیپ برای تراشه با تراکم بالا نسبت به تراشه با تراکم پایین بیشتر برآورد شد. در تراشه‌های با تراکم پایین افزایش اندازه جمعیت مرجع در مقایسه با تراشه‌های با تراکم بالا افزایش صحت استنباط ژنوتیپ بیشتر مشهود بود. برای تراشه با تراکم پایین افزایش اندازه جمعیت مرجع از ۲۵۰ حیوان به ۱۰۰۰ حیوان باعث افزایش صحت استنباط ژنوتیپ برای روش‌های Beagle و Flmpute در حالت سناریو ۱ به‌ترتیب به‌میزان ۰/۲۸ و ۰/۱۱ شد. اما این میزان افزایش برای تراشه با تراکم بالا با افزایش اندازه جمعیت مرجع برای دو روش فوق به‌ترتیب ۰/۱۵ و ۰/۱۲ در حالت سناریو ۱ برآورد شد. میزان صحت استنباط ژنوتیپ برای سناریو ۳ نسبت به سناریو ۲ تفاوت بسیار ناچیز نشان‌داد، اما تفاوت صحت نسبت به سناریو اول معنی‌دار بود ($p\leq$۰/۰۵) (جدول ۴).

جدول ۴- صحت استنباط ژنوتیپ در سناریوهای متفاوت تحت اندازه جمعیت مرجع متفاوت برای تراشه با تراکم بالا
Table 4. Accuracy of imputation in different scenarios under different reference population sizes for high-density chip

تراشه	نرم‌افزار	اندازه جمعیت مرجع	سناریوهای استنباط ژنوتیپ: سناریو ۱	سناریو ۲	سناریو ۳
تراشه با تراکم بالا	Beagle	۲۵۰	۰/۷۳	۰/۵۶	۰/۷۴
		۵۰۰	۰/۸۴	۰/۶۷	۰/۸۵
		۱۰۰۰	۰/۸۸	۰/۷۳	۰/۹۱
	Flmpute	۲۵۰	۰/۸۳	۰/۶۹	۰/۸۲
		۵۰۰	۰/۸۹	۰/۷۲	۰/۹۰
		۱۰۰۰	۰/۹۵	۰/۸۴	۰/۹۷

سناریو ۱، ۲ و ۳ به‌ترتیب جمعیت آمیخته، جمعیت خالص و جمعیت آمیخته+ خالص

در جداول ۳ و ۴ برای سناریو ۱ نسبت به سناریو ۲ به‌دلیل افزایش روابط خویشاوندی صحت استنباط ژنوتیپ بیشتر برآورد شد. ولی افزایش صحت سناریو ۳ نسبت به سناریو ۱ اختلاف بسیار ناچیز و بی‌معنی شد ($p>$ ۰/۰۵). به‌طور کلی نتایج این پژوهش نشان داد که صحت استنباط ژنوتیپ برای تمام سناریوهای با افزایش اندازه جمعیت مرجع افزایش پیدا نمود. افزایش صحت استنباط ژنوتیپ برای تمام سناریوها برای افزایش اندازه جمعیت مرجع از ۲۵۰ به ۵۰۰ حیوان شیب بیشتری دیده –شد و در مقابل افزایش از ۵۰۰ به ۱۰۰۰ حیوان شیب افزایشی، کاهش یافت. صحت استنباط ژنوتیپ به‌کمک روش Flmpute نسبت به Beagle برای اندازه جمعیت مرجع کوچک (۲۵۰ حیوان) بیشتر خود را نشان‌داد. در اندازه‌های جمعیت مرجع ۵۰۰ و ۱۰۰۰ حیوان در تمام سناریوها این افزایش خیلی چشمگیر نبود. برای سناریو ۱ میزان صحت استنباط ژنوتیپ نسبت به سناریو ۲ به‌دلیل بالا بودن رابطه خویشاوندی بین جمعیت مرجع و هدف بیشتر به‌دست آمد. در سناریو ۳ اضافه نمودن جمعیت خالص به جوامع آمیخته صحت استنباط ژنوتیپ را در تمام حالات در مقایسه با سناریو ۱ به‌مقدار جزئی افزایش داد. بنابراین با توجه به نتایج پژوهش حاضر می‌توان گفت که در کشورهای در حال توسعه که جمعیت حیوانی تعیین ژنوتیپ شده کوچک در دسترس هست، برای افزایش صحت انتخاب ژنومی استفاده از روش استنباط ژنوتیپ Flmpute به‌کمک جمعیت آمیخته با افزایش ارتباط خویشاوندی بین جمعیت مرجع و هدف می‌توان به‌عنوان یک راهکار مناسب دیده شود.

منابع

1. Aliloo, H., R. Mrode, A.M. Okeyo, G. Ni, M.E. Goddard and J.P. Gibson. 2018. The feasibility of using low-density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa. Journal Dairy Science, 101: 9108-9127.
2. Boichard, D., H. Chung, R. Dassonneville, 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, and Bovine L.D. Consortium. 2012. Design of a bovine low-density SNP array optimized for imputation. PLoS One, 7: e34130.

3. Bolormaa, S., K. Gore, J.H.J. van der Werf, B.J. Hayes and H.D. Daetwyler. 2015. Design of a low density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. Animal Genetics, 46: 544-556.
4. Bouwman, A.C. and R.F. Veerkamp. 2014. Consequences of splitting whole genome sequencing effort over multiple breeds on imputation accuracy. BMC Genetic, 15: 105.
5. Browning, B., Zhou, Y. and S. Browning. 2018. A one-penny imputed genome from next-generation reference panels. The American Journal of Human Genetics, 103: 338-348.
6. Browning, B.L. and S.R. Browning. 2016. Genotype imputation with millions of reference samples. American Journal of Human Genetics, 98: 116-126.
7. de Roos, A.P.W., B.J. Hayes, R.J. Spelman and M.E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein– Friesian, Jersey and Angus cattle. Genetics, 179: 1503-1512.
8. Habier, D., R.L. Fernando and J.C.M. Dekkers. 2009. Genomic selection using low-density marker panels. Genetics, 182: 343-353.
9. Hozé, C., M.N. Fouilloux, E. Venot, F. Guillaume, R. Dassonneville, S. Fritz, V. Ducrocq, F. Phocas, D. Boichard and P. Croiseau. 2013. High-density marker imputation accuracy in sixteen French cattle breeds. Genetic Selection Evolution, 45: 33.
10. Ghoreishifar S.M., H. Moradi-Shahrbabak, M. Moradi-Shahrbabak, E.L. Nicolazzi, J.L. Williams, D. Iamartino and A. Nejati- Javaremi. 2018. Accuracy of imputation of single-nucleotide polymorphism marker genotypes for water buffaloes (Bubalusbubalis) using different reference population sizes and imputation tools. Livestock Science, 216: 174-182.
11. Korkuc, P., D. Arneds and G.A. Brockmann. 2019. Finding the optimal imputation strategy for small cattle populations animal breeding biology and molecular genetics. Frontiers in Genetics, 10: 52-59.
12. Ma, P., R.F. Brøndum, Q. Zhang, M.S. Lund and G. Su. 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. Journal Dairy Science, 96: 4666-4677.
13. Ma, P., R.F. Brondum, Q. Zhang, M.S. Lund and G. Su. 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. Journal Dairy Science, 96: 4666-4677.
14. Meuwissen, T., B. Hayes and M. Goddard. 2016. Genomic selection: A paradigm shift in animal breeding. Animal Frontiers, 6: 6-14.
15. Milanesi, M., D. Vicario, A. Stella, A. Valentini, P. Ajmone-Marsan, S. Biffani, F. Biscarini, G. Jansen and E.L. Nicolazzi. 2015. Imputation accuracy is robust to cattle reference genome updates. Animal Genetics, 46: 69-72.
16. Mohammadi, Y. and M. Mokhtari. 2017. Genomic selection accuracy parametric and nonparametric statistical methods with additive and dominance genetic architectures. Research on Animal Production, 8(18): 161-167 (In Persian).
17. Mulder H., M. Calus, T. Druet and C. Schrooten. 2012. Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. Journal of Dairy Science, 95: 876-889.
18. Ogawa, S., H. Matsuda, Y. Taniguchi, T. Watanabe, A. Takasuga, Y. Sugimoto and H. Iwaisaki. 2016. Accuracy of imputation of single nucleotide polymorphism marker genotypes from low-density panels in Japanese Black cattle. Journal Animal Science, 87: 3-12.
19. Pausch, H., I.M. MacLeod, R. Fries, R. Emmerling, P.J. Bowman and H.D. Daetwyler. 2017. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. Genetic Selection Evolution, 49: 1-10.
20. Pausch, H., B. Aigner, R. Emmerling, C. Edel, K.U. Götz and R. Fries. 2013. Imputation of high-density genotypes in the Fleckvieh cattle population. Genetic Selection Evolution, 45: 3.
21. Nicolazzi, E., S. Biffani and G. Jansen. 2013. Short communication: Imputing genotypes using PedImpute fast algorithm combining pedigree and population information. Journal of Dairy Science, 96: 2649-2653.
22. Sargolzaei, M., J.P. Chesnais and F.S. Schenkel. 2014. A new approach for efcient genotype imputation using information from relatives. BMC Genomics, 15.
23. Sargolzaei M. and F.S. Schenkel. 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics, 25: 680-1.
24. Van Binsbergen, R., M.C. Bink, M.P. Calus, F.A. van Eeuwijk, B.J. Hayes and I. Hulsegge. 2014. Accuracy of imputation to whole-genome sequence frontiers in genetics data in Holstein Friesian cattle. Genetic Selection Evolution, 46: 41.
25. VanRaden P., C. Sun and J. O'Connell. 2015. Fast imputation using medium or low-coverage sequence data. BMC Genetics, 16(82): 2039-2042.
26. Ventura, R.V., D. Lu, F.S. Schenkel, Z. Wang, C. Li and S.P. Miller. 2014. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle1. Journal Animal Science, 92: 1433-1444.
27. Wang, Z. and N. Chatterjee. 2017. Increasing mapping precision of genome wide association studies: to genotype and impute, sequence, or both? Genome Biology, 18: 17-19.

Effect of Reference Population Size and Imputation Methods on the Accuracy of Imputation in Pure and Mixed Populations

Yahya Mohammadi[1] and Javad Ahmadpanah[2]

1- Assistant Professor and Faculty Member, Department of Animal Science, Faculty of Agriculture, Ilam University, Ilam, Iran (Corresponding author: mohamadi_yahya@yahoo.com)
2- Assistant Professor, Animal Science Research Department, Kermanshah Agricultural and Natural Resources Research and Education Center, AREEO, Kermanshah, Iran
Received: March 1, 2020 Accepted: June 7, 2020

Abstract

Imputation as a method of creating low-density chips to high-density chips has been introduced to increase the accuracy of genomic selection in animals. In the current study, to investing imputation accuracy, three populations of mixed (scenario 1), pure (scenario 2) and mixed + pure (scenario 3) were simulated using QMSim. Two methods of imputation including Beagle and Flmpute were used for two types of low and high density chips. Selected reference population sizes for each scenario were 250, 500 and 1000 animals. The results showed that in all considered scenarios, the accuracy of imputation raised by increasing the reference population size from 250 to 500 animals, but decreased by increasing the reference population size from 500 to 1000 animals. The accuracy of imputation using Flmpute method was greater than that of Beagle for the small reference population (250 animals). In all scenarios and reference population sizes of 500 and 1000 animals, increased accuracy in Flmpute method was not significant in compared to the Beagle method. The accuracy of the imputation was higher for scenario 1than for scenario 2. Also the increase in the accuracy of the imputation in Scenario 3 was not significant in compared to Scenario 1. Generally, the results of the current study showed that in developing countries where small genotyped animal populations are available, to increase the accuracy of genomic selection, using Flmpute method and mixed population and increasing the relationship between the reference and the target population could be a suitable approach.

Keywords: Genomic selection, Imputation, Mixed population, Prediction accuracy, Simulation

[DOR: 20.1001.1.22518622.1399.11.30.8.0]
[DOI: 10.52547/rap.11.30.109]