Assessment of the performance of different imputation methods for low-coverage sequencing in Holstein cattle

Jun Teng,1 Changheng Zhao,1 Dan Wang,1 Zhi Chen,2 Hui Tang,1 Jianbin Li,3 Cheng Mei,4 Zhangping Yang,2 Chao Ning,1* and Qin Zhang1*

1Shandong Provincial Key Laboratory of Animal Biotechnology and Disease Control and Prevention, College of Animal Science and Technology, Shandong Agricultural University, Tai’an 271018, China
2College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China
3Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Sciences, Jinan 250000, China
4Dongying Shenzhou AustAsia Modern Dairy Farm Co. Ltd., Dongying 257000, China

Abstract

Low-coverage sequencing (LCS) followed by imputation has been proposed as a cost-effective genotyping approach for obtaining genotypes of whole-genome variants. Imputation performance is essential for the effectiveness of this approach. Several imputation methods have been proposed and successfully applied in genomic studies in human and other species. However, there are few reports on the performance of these methods in livestock. Here, we evaluated a variety of imputation methods, including Beagle v4.1, GeneImp v1.3, GLIMPSE v1.1.0, QUILT v1.0.0, Reveel, and STITCH v1.6.5, with varying sequencing depth, sample size, and reference panel size using LCS data of Holstein cattle. We found that all of these methods, except Reveel, performed well in most cases with an imputation accuracy over 0.9; on the whole, GLIMPSE, QUILT, and STITCH performed better than the other methods. For species with no reference panel available, STITCH followed by Beagle would be an optimal strategy, whereas for species with reference panel available, QUILT would be the method of choice. Overall, this study illustrated the promising potential of LCS for genomic analysis in livestock.

Key words: low-coverage sequencing, genotype imputation method, Holstein cattle

Introduction

Single nucleotide polymorphisms are the most widely used molecular genetic markers for dissecting complex traits and predicting unobserved phenotypes. Often, SNP genotyping is based on commercial SNP arrays in livestock. However, SNP arrays can only genotype limited known variants, and causal mutations are often not included by design. In contrast, whole-genome sequencing (WGS) can provide complete variant information of the whole genome (Meuwissen and Goddard, 2010; Brøndum et al., 2015; MacLeod et al., 2016). In the past decades, the rapid development and cost reduction of next-generation sequencing has allowed the establishment of large-scale high-coverage WGS projects (Wetterstrand, 2020). However, high-coverage sequencing (HCS) of large cohorts is still too expensive for large-scale genomic analysis, especially for livestock (Fraser et al., 2018). An alternative strategy is to genotype the target population with SNP array and then to impute SNP array data to the WGS level based on a reference population with HCS (Druet et al., 2014; van Binsbergen et al., 2014; Veerkamp et al., 2016; Butty et al., 2019; Liu et al., 2020; Fernandes Júnior et al., 2021). However, this kind of strategy is inhibited in livestock due to the following limitations: (1) insufficient imputation accuracy due to the wide marker density span from array level to sequence level (VanRaden et al., 2013); (2) requirement of large reference population with close genetic relationship to the target population, which is not available for most livestock species; and (3) no commercial SNP array for some livestock species (e.g., donkey).

Another alternative strategy to obtain large-scale HCS data is to perform low-coverage sequencing (LCS) at 1× or less and then impute the LCS data to high-coverage WGS data (Li et al., 2011; Zan et al., 2019). This strategy is more pronounced for species without commercially available SNP array (Buerkle and Gompert, 2013). The LCS is a cost-effective strategy for getting whole-genome variants (Pasaniuc et al., 2012; VanRaden et al., 2015; Gilly et al., 2019; Davies et al., 2021) and has been successfully applied in GWAS, genomic prediction, and population genetics analysis in human (Cai et al., 2015; Gilly et al., 2016; Liu et al.,
2018), mice (Nicod et al., 2016), pig (Yang et al., 2021), coral (Fuller et al., 2020), large yellow croaker (Zhang et al., 2021), and maize (Zheng et al., 2018).

Imputation of missing genotypes is necessary for LCS data due to the high missing rates. A series of attractive imputation methods have been proposed and assessed by different research groups. Beagle, proposed by Browning and Browning (2009) and Browning et al. (2018), is a popular genotype imputation software, which is based on the theory of haplotype frequency model with parsimonious state-space (Li and Stephens, 2003). Although it was not designed for LCS data, it has achieved satisfactory results for LCS data of some species (Gilly et al., 2019; Lou et al., 2021, Rubinacci et al., 2021) in the case that a haplotype reference panel was available. Methods designed specifically for LCS data include GeneImp (Spiliopoulou et al., 2017), GLIMPSE (Rubinacci et al., 2021), QUILT (Davies et al., 2021), Reveel (Huang et al., 2016) and STITCH (Davies et al., 2016). GeneImp was proposed to impute genotype likelihoods computed from ultralow coverage sequencing data to a dense reference panel. GLIMPSE performs genotype imputation with a Gibbs sampling procedure at all variable positions of the reference panel. QUILT uses Gibbs sampling to partition reads into maternal and paternal sets, facilitating rapid haploid imputation using large reference panels. Reveel uses a simplified model to leverage the underlying complex linkage disequilibrium structure. STITCH was designed for imputation without requiring a haplotype reference panel. It is based on a Hidden Markov model and adapts to the fact that SNPs in sequences are not independent of each other. Thus far, these imputation methods have been successfully applied to LCS data in human and other species (Davies et al., 2016; Nicod et al., 2016; Liu et al., 2018; Jiang et al., 2021; Yang et al., 2021; Zhang et al., 2021), but whether they are suitable for livestock, especially Holstein cattle, has received less attention.

In this study, we investigated the performance of different imputation methods for LCS data of Holstein cattle with respect to different sequencing depths, sample sizes, and reference panel sizes. We demonstrated that, except Reveel, all these methods performed well in most cases; overall, QUILT and STITCH performed better and would be the choice methods for imputation for LCS data of Holstein cattle.

**MATERIALS AND METHODS**

**Animals and Low-Coverage WGS**

Blood samples of 314 Holstein cows were collected from Dongying Shenzhou AustAsia Modern Dairy Farm Co., Ltd. in Shandong Province, China. Total DNA was isolated using the QIAamp DNA Investigator kit (QIAGEN) following the manufacturer’s instruction. The DNA quality was evaluated by spectrophotometry and agarose gel electrophoresis. The guidelines of experimental animal management of Shandong Agricultural University (SDAU) were followed throughout the study, and the experimental protocols were approved by the Experimental Animal Care and Use Committee of SDAU.

The DNA templates were ultrasonically sheared using a Covaris E220 (Covaris) to yield ~150 bp fragments, and then prepared for sequencing libraries following the workflow of the NEBNext Ultra DNA Library Preparation Protocol (New England Biolabs). Multiple Ampure Bead XP cleanups (Beckman Coulter) were conducted to remove any adapter dimer that might have developed. The quality and concentration of libraries were determined on an Agilent Bioanalyzer 2100 (Agilent Technologies). The quality-controlled genomic library for each sample was PE150 sequenced using the Illumina NovaSeq 6000 sequencing system. The average sequencing depth was 1.49× (1.03× to 2.02×).

In addition, WGS data for 486 Holstein cattle with an average sequencing depth 16.86× (5.05× to 54.10×) were downloaded from the NCBI Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra). The downloaded WGS data sets for Holstein cattle are available in the NCBI SRA under accession numbers PRJNA176557, PRJNA210521, PRJNA277147, PRJNA343262, PRJNA431934 and PRJNA478565. Thus, in total we had 800 animals with sequencing data.

**Sequence Data Preprocessing**

To avoid artificial bias of reads, quality control was performed by Trimmomatic v0.39 (Bolger et al., 2014). The high-quality paired-end reads were mapped to the bovine reference genome ARS-UCD1.2 using BWA v0.7.17 (Li and Durbin, 2009). The file format conversions, sort, and index in the analysis were all carried out using Samtools v1.11 (Li et al., 2009). All PCR duplicates were removed using Picard tools (Broad, 2019). The Genome Analysis Toolkit (GATK) version 3.8 (McKenna et al., 2010) was used for realignment and recalibration, following GATK’s best practices.

**SNP Array Genotyping**

The 314 Holstein cows were also genotyped with GeneSeek Genomic Profiler Bovine 100K SNP array (Neogen GeneSeek Operations), which included 95,256 SNPs over the whole genome. Genotype quality control
was carried out with PLINK1.9 (Chang et al., 2015), and SNPs with call rate less than 95% or without chromosomal position were excluded.

**Genotype Imputation**

We used the following imputation methods for LCS data: Beagle v4.1 (Browning and Browning, 2016), GenelImp v1.3 (Spiliopoulou et al., 2017), GLIMPSE v1.1.0 (Rubinacci et al., 2021), QUILT v1.0.0 (Davies et al., 2021), Reveel (Huang et al., 2016), and STITCH v1.6.5 (Davies et al., 2016). All parameters of different imputation methods were shown in Supplemental Table S1 (https://doi.org/10.6084/m9.figshare.17169335.v1). The WGS data of 1,059 Holstein cattle obtained from the 1000 Bull Genomes Project Run 8 (Hayes and Daetwyler, 2019) was used as a reference panel and converted to different formats according to the different software requirements. Because STITCH can work with or without a reference panel, we considered both cases (i.e., STITCH and STITCH_REF) for without and with a reference panel, respectively.

**Comparison of Different Imputation Methods in Different Scenarios**

We compared the performance of different imputation methods in different scenarios with different sequencing depths, sample sizes, and reference panel sizes (Table 1). We first considered a standard scenario with sequencing depth of 1×, sample size (number of animals with LCS data) of 800, and reference panel size of 1,059, and then varied these parameters as described below.

**Sequencing Depth.** We randomly down-sampled paired-reads from the bam files of the 800 sequenced animals to produce different sequence data sets with depth of 1.0×, 0.5×, 0.4×, 0.3×, 0.2×, and 0.1×, respectively, using DownsmaplSam in Picard tools (Broad, 2019). To test the possible sampling error due to random sampling, we performed 3 repeated samplings for the depth of 0.1×.

**Sample Size.** We randomly sampled 600, 400, 200, and 100 individuals, from the 800 sequenced animals to form samples with smaller sizes. Because the sampling errors for sample size 200 were negligible (Supplemental Table S2, https://doi.org/10.6084/m9.figshare.17169341.v2), it was reasonable to conclude that the sampling errors for sample size 400 or 600 would also be negligible, because the larger the sample, the smaller the negligible error would be. Therefore, we thought it was not necessary to check sampling errors for size 400 or 600. As for sample size 100, because we needed to keep 80 animals that were sequenced with high-coverage (for accuracy evaluation) in the sample for each sampling, it was not meaningful to repeat the random sampling. Therefore, only for the sample size of 200, 3 repeated samplings were performed.

**Reference Panel Size.** Different reference panels with sizes 800, 600, 400, and 200 were generated by sampling from the original reference panel with 1,059 individuals. For the panel size of 200, 3 repeated samplings were performed. Considering both computing time and the robustness of the imputation performances across chromosomes (Jattawa et al., 2016; Korkuč et al., 2019), we only conducted the comparisons for 3 chromosomes (i.e., chromosomes 1, 11, and 21).

**Evaluation of Imputation Accuracy**

To evaluate the imputation accuracy, we selected 80 animals from the 486 animals with sequence data downloaded from the NCBI SRA database, which had sequencing depth greater than 25.08×, with an average of 29.83×, as validation individuals. The SNP calling and genotyping were performed using HaplotypeCaller in GATK (McKenna et al., 2010). We used 2 statistics to measure the imputation accuracy (i.e., $r^2$ and imputation quality score; IQS). The $r^2$ is the squared correlation between the expected dosages (posterior

---

**Table 1. Scenarios used to evaluate imputation performance**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
<th>Sequencing depth</th>
<th>Sample size</th>
<th>Reference panel size</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Standard scenario with fixed sequencing depth, sample size, and reference panel size</td>
<td>1×</td>
<td>800</td>
<td>1,059</td>
</tr>
<tr>
<td>S2</td>
<td>Variable sequencing depth and sample size, fixed reference panel size</td>
<td>1×, 0.5×, 0.4×, 0.3×, 0.2×, 0.1×</td>
<td>800, 600, 400, 200, 100</td>
<td>1,059</td>
</tr>
<tr>
<td>S3</td>
<td>Variable sequencing depth and reference panel size, fixed sample size</td>
<td>1×, 0.5×, 0.4×, 0.3×, 0.2×, 0.1×</td>
<td>800</td>
<td>1,059, 800, 600, 400, 200</td>
</tr>
<tr>
<td>S4</td>
<td>Imputation accuracy against minor allele frequency</td>
<td>1×</td>
<td>800</td>
<td>1,059</td>
</tr>
</tbody>
</table>
expectation of the imputed allele dosages) and the known true genotypes (Browning and Browning, 2009). The IQS is the concordance (proportion of agreement between the imputed genotypes and known true genotypes) adjusted for chance agreement (Lin et al., 2010). The 2 statistics were calculated for each SNP and then averaged across all SNPs. In addition, for the 314 animals that were both sequenced and SNP-array genotyped, imputation accuracy was also evaluated for the imputed array genotypes.

RESULTS

Performance of Different Imputation Methods

We first compared the 6 methods in the standard scenario as follows: sequencing depth = 1×, sample size = 800, and reference panel size = 1,059. Table 2 shows the performance of different methods in terms of number of SNPs and imputation accuracy for chromosome 1 (the results for the other 2 chromosomes are in Supplemental Table S3, https://doi.org/10.6084/m9.figshare.17169344.v2; and Supplemental Table S4, https://doi.org/10.6084/m9.figshare.17169347.v2). The number of SNPs commonly presented in all individuals revealed by different methods differed considerably, ranging from over one million (GeneImp, GLIMPSE, and QUILT) to less than half of one million (STITCH or STITCH_REF). In general, the more SNPs identified, the higher were the percentage of SNPs that were overlapped with the SNPs in the 100K SNP array (5,201 SNPs on chromosome 1), ranging from about 95% (GeneImp, GLIMPSE and QUILT) to about 30% (STITCH and STITCH_REF).

The r² and IQS were very close to each other for all methods. Reveel gave very low imputation accuracy (r² = 0.33 and 0.53 for SNP array and HCS, respectively). The other methods all yielded 0.95 or higher accuracy, except for GeneImp, which yielded an r² of 0.89 for imputed SNP array genotype. STITCH, STITCH_REF, and QUILT gave very similar results and the highest accuracy (r² = 0.99 and 0.98 for SNP array and HCS, respectively); the addition of a reference panel did not actually improve STITCH in this scenario. In view of the poor performance of Reveel, we discarded it in our follow-up analysis.

Effects of Sequencing Depth and Sample Size on the Imputation Performance for HCS Genotypes

The r² of the 5 methods for imputed HCS genotypes under different sequencing depths and sample sizes, keeping the reference panel size equal to 1,059, for chromosome 1 are illustrated in Figure 1 (the results for the other 2 chromosomes are in Supplemental Figure S1, https://doi.org/10.6084/m9.figshare.17169242.v1; and Supplemental Figure S2, https://doi.org/10.6084/m9.figshare.17169260.v1). As expected, the imputation accuracy generally increased along with the increase of the sequencing depth and sample size. GeneImp and QUILT were relatively robust to different sample sizes and sequencing depths. For all sample sizes (100–800) and sequencing depths (0.1× to 1×), GeneImp gave imputation accuracies of over 0.935 and below 0.955 (Figure 1B), and QUILT gave imputation accuracies of over 0.970 and below 0.980 (Figure 1D). Sample size had small effect on Beagle and GLIMPSE, with very small improvement when increasing sample size for all sequencing depths; however, sequencing depth had considerable effect, where the accuracy increased by 0.08 to 0.10 when the sequencing depth increased from 0.1× to 1× for different sample sizes (Figure 1A, 1C). To achieve an accuracy over 0.90, a sequencing depth higher than 0.3× or 0.2× was required for Beagle and

Table 2. Imputation performance of different methods for chromosome 1 in the following scenario: sequencing depth = 1×, sample size = 800, and reference panel size = 1,059

<table>
<thead>
<tr>
<th>Method</th>
<th>No. of SNPs¹</th>
<th>Overlap with 100K SNP array (%)²</th>
<th>(r²)³</th>
<th>IQS⁴</th>
<th>Array</th>
<th>HCS</th>
<th>Array</th>
<th>HCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beagle</td>
<td>815,803</td>
<td>4,698 (90.33)</td>
<td>0.9584</td>
<td>0.9384</td>
<td>0.9531</td>
<td>0.9356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GeneImp</td>
<td>1,153,990</td>
<td>5,003 (96.20)</td>
<td>0.8946</td>
<td>0.9485</td>
<td>0.9141</td>
<td>0.9547</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLIMPSE</td>
<td>1,115,838</td>
<td>4,981 (95.77)</td>
<td>0.9772</td>
<td>0.9646</td>
<td>0.9728</td>
<td>0.9652</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUILT</td>
<td>1,148,727</td>
<td>5,098 (96.29)</td>
<td>0.9871</td>
<td>0.9749</td>
<td>0.9800</td>
<td>0.9737</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reveel</td>
<td>606,989</td>
<td>1,701 (32.71)</td>
<td>0.3329</td>
<td>0.5317</td>
<td>0.3795</td>
<td>0.5806</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STITCH_REF</td>
<td>463,951</td>
<td>1,668 (32.07)</td>
<td>0.9894</td>
<td>0.9818</td>
<td>0.9863</td>
<td>0.9850</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STITCH</td>
<td>473,230</td>
<td>1,563 (30.05)</td>
<td>0.9900</td>
<td>0.9815</td>
<td>0.9858</td>
<td>0.9849</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹Number of SNPs revealed by different methods that were commonly presented in all individuals.
²There are 5,201 SNPs on chromosome 1 in the GGP 100K SNP array.
³Squared correlation between the known typed genotypes from SNP array or high-coverage sequencing (HCS).
⁴IQS = imputation quality score.
GLIMPSE, respectively. STITCH (without a reference panel) was greatly affected by both sample size and sequencing depth (Figure 1F). The accuracies varied from less than 0.1 (sample size = 100 and sequencing depth = 0.1×) to 0.99 (sample size = 800 and sequencing depth = 1×) with varying sample size and sequencing depth. It was more sensitive to sample size with lower sequencing depth and also more sensitive to sequencing depth with smaller sample size. However, it should be noted that STITCH gave accuracies over 0.95 when the sequencing depth was 1× and sample size was ≥200, or when the sequencing depth was ≥0.4× and the sample size was ≥400. Adding a reference panel to STITCH (STITCH_REF) greatly improved STITCH when the sequencing depth was lower than 0.4× or when the sample size was smaller than 600. With STITCH_REF, an accuracy over 0.95 could be achieved when the sequencing depth was ≥0.3× and the sample size was ≥400, the sequencing depth was ≥0.2× and the sample size was ≥600, or the sequencing depth was ≥0.1× and the sample size was ≥800.

The IQS of the 5 methods were almost the same as the r² in all of the scenarios investigated (Supplemental Figure S3, https://doi.org/10.6084/m9.figshare.17169263.v1; Supplemental Figure S4, https://doi.org/10.6084/m9.figshare.17169272.v1; and Supplemental Figure S5, https://doi.org/10.6084/m9.figshare.17169275.v1; for chromosomes 1, 11, and 21). For the sequencing depth of 0.1×, the imputation accuracies from the 3 repeated samplings were very similar (Supplemental Table S5 for chromosome 1 and sample size = 200, https://doi.org/10.6084/m9.figshare.17169350.v2), as did the accuracies from the 3 repeated samplings for the sample size of 200 (Supplemental Table S2, https://doi.org/10.6084/m9.figshare.17169341.v2), indicating that the sampling errors due to the random sampling of the reads and of the individuals were negligible.

In addition, we compared the numbers of SNPs obtained by the 5 imputation methods (Figure 2 for chromosome 1, Supplemental Figure S6 for chromosome 11, https://doi.org/10.6084/m9.figshare.17169278.v2; and Supplemental Figure S7 for chromosome 21, https://doi.org/10.6084/m9.figshare.17169296.v2). In general, the sample size had limited effect on the numbers of SNPs for all of the methods investigated. GeneImp, GLIMPSE, and QUILT were also robust to the sequencing depth, whereas Beagle, STITCH, and STITCH_REF were very much sensitive to the se-
quencing depth. Adding a reference panel to STITCH could increase the number of SNPs, especially when the sample size was small.

**Effect of Reference Panel Size on the Accuracy of Imputation for HCS Genotypes**

A reference panel is necessary for Beagle, GeneImp, GLIMPSE, and QUILT, and can be beneficial for STITCH. We tested the effect of reference panel size on the imputation accuracy in the scenarios with sample size equal to 800 but differed in sequencing depth [Figure 3 (r²) and Supplemental Figure S10 (IQS) for chromosome 1, https://doi.org/10.6084/m9.figshare.17169305.v1; Supplemental Figures S8 (r²) and S11 (IQS) for chromosome 11, https://doi.org/10.6084/m9.figshare.17169284.v1 and https://doi.org/10.6084/m9.figshare.17169308.v1; and Supplemental Figures S9 (r²) and S12 (IQS) for chromosome 21, https://doi.org/10.6084/m9.figshare.17169299.v1 and https://doi.org/10.6084/m9.figshare.17169311.v1]. Again, the r² and IQS were almost the same. We observed that STITCH_REF yielded the highest and stable accuracy under different size of reference panel. The accuracies of the other 4 methods increased with the increase of the reference panel size. Beagle and GeneImp were more significantly affected by the reference panel size, and the lower the sequencing depth, the greater was the effect of the reference panel size. For GLIMPSE and QUILT, the effect of the reference panel size was relatively small, especially when the sequencing depth was 0.2× or higher. For the panel size of 200, the results from the 3 repeated samplings were also quite consistent (Supplemental Table S6, https://doi.org/10.6084/m9.figshare.17169353.v2).

**Effect of Minor Allele Frequency on the Accuracy of Imputation for HCS Genotypes**

We used a data set with sample size of 800, sequencing depth of 1×, and reference panel size of 1,059 to evaluate the effect of minor allele frequency (MAF) on accuracy of different methods. The SNPs were divided into 13 bins according to their MAF as follows: [0–0.005], [0.005–0.01], [0.01–0.02], [0.02–0.05], [0.05–0.10], [0.10–0.15], [0.15–0.20], [0.20–0.25], [0.25–0.30], [0.30–0.35], [0.35–0.40], [0.40–0.45], and [0.45–0.50]. The average accuracies for each bin were calculated and plotted in Figure 4. As expected, imputation accuracy improved as MAF increased. The effect of MAF was obvious for rare variants with MAF <0.05, especially for IQS, which decreased more rapidly than r² in

**Figure 2.** Number of SNPs revealed by different imputation methods for different sequencing depths and sample sizes for chromosome 1 (reference panel size = 1,059).
this case. Beagle and GeneImp were more sensitive to low MAF, whereas GLIMPSE and QUILT were very little affected by MAF. However, for common variants (MAF > 0.05), STITCH, STITCH_REF, and QUILT performed better than the other methods.

Running Time

The running time of the 5 methods in the scenario of sample size = 800, sequencing depth = 1×, and reference panel size = 1,059 is shown in Figure 5. All of the analyses were conducted on a Linux cluster (Intel Xeon Gold 6254 CPU @ 3.10 GHz) with 10 computation nodes. Beagle was the slower than the other methods, followed by QUILT, which was about 20 to 30% faster than Beagle. GeneImp was the fastest and took only about one-fifth of the time of Beagle. GLIMPSE took approximately double the time of GeneImp. The running time of STITCH and STITCH_REF was between that of GLIMPSE and GeneImp.

STITCH and Beagle Strategy

As mentioned above, STITCH discovered fewer SNPs than the other methods, which indicated that there was still a high missing rate after STITCH imputation, leading to a small number of SNPs shared by all individuals. STITCH_REF did not improve this much. We tried to perform a second imputation to the results of STITCH using Beagle. The performance of STITCH and Beagle was evaluated in the scenarios of sample size = 800 with different sequencing depths. We kept those SNPs in the data set that were missing in ≤10% or ≤20% of the individuals after STITCH imputation, and then imputed these missing genotypes by Beagle. Table 3 shows the number of SNPs shared by all individuals for chromosome 1 after Beagle imputation and the corresponding imputation accuracies (the results for the other 2 chromosomes are in Supplemental Table S7, https://doi.org/10.6084/m9.figshare.17169356.v1; and Supplemental Table S8, https://doi.org/10.6084/m9.figshare.17169359.v1). Compared with the STITCH results, for missing rate ≤10%, the numbers of SNPs increased from 75.64% (for sequencing depth of 1×) to 12.245% (for sequencing depth of 0.1×), whereas the imputation accuracies were only slightly reduced, except for the case of sequencing depth of 0.1×. For missing rate ≤20%, the numbers of SNPs further increased from 1.82% (for sequencing depth of 1×) to 35.67% (for sequencing depth of 0.1×), and the accuracies further slightly increased. These results indicated that STITCH and Beagle was an effective strategy to
make up for STITCH’s shortage of yielding less SNPs than other methods, while maintaining its advantage of high accuracy.

**DISCUSSION**

With the rapid reduction of sequencing cost and development of methods and algorithms for handling whole-sequence data, people pay more and more attention to the variants of whole genome rather than the limited variants of SNP arrays (Meuwissen and Goddard, 2010). Low(ultralow)-coverage sequencing followed by missing genotype imputation has been proposed as a cost-effective approach for obtaining genotypes of whole-genome variants (Pasaniuc et al., 2012; Gilly et al., 2019; Martin et al., 2021). The imputation

![Diagrams](image1)

Figure 4. Squared correlations ($r^2$) and imputation quality score (IQS) of different imputation methods as a function of minor allele frequencies (MAF; sample size = 800, sequencing depth = 1×). (A), (B), and (C) show the $r^2$ of chromosomes 1, 11, and 21. (D), (E), and (F) show the IQS of chromosomes 1, 11, and 21. The SNPs were divided into 13 bins according to their MAF as follows: [0–0.005], [0.005–0.01], [0.01–0.02], [0.02–0.05], [0.05–0.10], [0.10–0.15], [0.15–0.20], [0.20–0.25], [0.25–0.30], [0.30–0.35], [0.35–0.40], [0.40–0.45], and [0.45–0.50].

![Diagrams](image2)

Figure 5. Running time per individual of different imputation methods for chromosome 1 (A), 11 (B), and 21 (C; sample size = 800; sequencing depth = 1×). All analyses were conducted on a Linux cluster (Intel Xeon Gold 6254 CPU @ 3.10 GHz) with 10 computation nodes.
performance (e.g., accuracy, number of SNPs discovered, running time) is essential for the effectiveness of the LCS data, which is related to several factors (i.e., imputation method, sequencing depth, number of sequenced individuals or sample size), and whether there is a reference panel available and the size of the panel. In this study, we evaluated the imputation performance of LCS data of Holstein cattle with respect to these factors.

Among the imputation methods investigated in this study [i.e., Beagle, GeneImp, GLIMPSE, QUILT, Reveel, and STITCH (with or without a reference panel)], Reveel resulted in much lower imputation accuracy than the other methods. This is probably because it was not designed for extremely low-coverage sequence data (such as 1× or lower). Huang et al. (2016) reported very high imputation accuracy (over 0.99) of Reveel for human LCS data; however, the sequencing depth of the sequencing data they used was ≥2×.

Beagle, GeneImp, and GLIMPSE mainly leverage the information from a reference haplotype panel to impute missing genotypes; therefore, their accuracies were more affected by sequencing depth than by sample size (Figure 1A, 1B, and 1C). Among the 3 methods, Beagle performed the worst in terms of imputation accuracy (in almost all scenarios), number of SNPs discovered (Table 2), and running time (Figure 5). GeneImp and GLIMPSE had their advantages in different scenarios. When the reference panel was large (e.g., 1,059 in this study), GeneImp was more robust to sequencing depth and could achieve an accuracy of over 0.935 for sequencing depth as low as 0.1×, whereas GLIMPSE gave higher accuracy in cases of sequencing depth ≥0.5× (Figure 1). However, when the reference panel was small (e.g., ≤800), GLIMPSE outperformed GeneImp for all sequencing depths (Figure 3). Another advantage of GLIMPSE is that it is very robust to MAF. It performed quite well (accuracy >0.9) even for SNPs with MAF lower than 0.001. However, GLIMPSE required almost double the running time of GeneImp (Figure 5).

STITCH performed quite well when the sequencing depth was ≥0.4× and the sample size was ≥400 (Figure 1F), even without a reference panel. In particular, for sequencing depth of 1×, it resulted in accuracies close to 1 for all sample sizes, which were the highest accuracies among all methods. Adding a reference panel to STITCH (STITCH_REF) improved its performance, particularly in cases of lower sequencing depth or smaller sample size (Figure 1E). Such improvement was consistent with size of the reference panel (i.e., even a small reference panel could benefit STITCH). The running time of STITCH was between that of Genelmp and GLIMPSE. A disadvantage of STITCH was that there was still a high proportion of unimputed SNPs after imputing with STITCH, and thus it provided the least number of SNPs commonly shared by all individuals among the methods of interest in this study. After performing a second imputation with Beagle, the number of SNPs increased greatly (Table 3) without losing the accuracy.

QUILT was designed specifically for imputation using large reference panels. Our results showed that, with a reference panel of size as large as 1,059, its overall performance was the best among all methods; the imputation accuracies remained at a high level (over 0.97) across all sequencing depths and sample sizes (Figure 1D) and the numbers of SNPs were among the highest across all scenarios (Figure 2). When the reference panel size decreased, its accuracy also decreased, but still higher than 0.95 in all cases (Figure 3). A disadvantage of QUILT was that it needed more running time than the other methods except Beagle (Figure 5). However, its running time was still in the acceptable range.

<table>
<thead>
<tr>
<th>Sequencing depth</th>
<th>STITCH</th>
<th>STITCH + Beagle1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of SNPs</td>
<td>r²</td>
</tr>
<tr>
<td>0.1×</td>
<td>3,219</td>
<td>0.8634</td>
</tr>
<tr>
<td>0.2×</td>
<td>118,774</td>
<td>0.9646</td>
</tr>
<tr>
<td>0.3×</td>
<td>227,211</td>
<td>0.9749</td>
</tr>
<tr>
<td>0.4×</td>
<td>301,633</td>
<td>0.9785</td>
</tr>
<tr>
<td>0.5×</td>
<td>336,882</td>
<td>0.9799</td>
</tr>
<tr>
<td>1×</td>
<td>473,230</td>
<td>0.9815</td>
</tr>
</tbody>
</table>

1SNP missing rate after imputation with STITCH.
2Squared correlation between the known typed genotypes from SNP array or high-coverage sequencing (HCS).
3IQS = imputation quality score.
It should be noted that we focused imputation for LCS data in dairy cattle population in this study. In comparison with populations of other species, the major properties in a typical dairy population are (1) the animals in the population are basically descendants of a few ancestor bulls, and (2) the effective population size is relatively small (Brotherstone and Goddard, 2005). These properties may result in higher imputation accuracy for a dairy cattle population than for populations of other species under the same conditions due to the closer relationship among individuals in the population and the stronger linkage disequilibrium among SNPs in the genome.

CONCLUSIONS

We compared the performance of 6 methods, Beagle v4.1, GeneImp v1.3, GLIMPSE v1.1.0, QUILT v1.0.0, Revel, and STITCH v1.6.5, for imputation of LCS data with sequencing depth of 1× or lower of Holstein cattle. Our results indicated that Revel was not suitable for our data due to its very low imputation accuracy. On the whole, Beagle was not competitive with GeneImp, GLIMPSE, QUILT, and STITCH, although its imputation accuracy was acceptable (over 0.90) in most cases. GeneImp, GLIMPSE, QUILT, and STITCH each had their advantages in relevant situations. When a large reference panel was available, GeneImp and QUILT were very robust to sequencing depth and sample size and produced imputation accuracies near (GeneImp) or higher than (QUILT) 0.95 even for very low-coverage (0.1×) sequencing data and very small size (100). GLIMPSE performed very well in general when the sequencing depth was higher than 0.1×. STITCH, with or without a reference panel, produced the highest accuracy when the sequencing depth was higher than 0.4× and the sample size was larger than 400. Taken overall, considering imputation accuracy, number of SNPs produced, and computing time, STITCH followed by Beagle would be an optimal strategy in the absence of a reference panel, whereas QUILT would be the method of choice in a case of a reference panel.

ACKNOWLEDGMENTS

This work was supported by the Yangzhou University Interdisciplinary Research Foundation for Animal Science Discipline of Targeted Support (yznxk202016; Yangzhou, China), the Project of Genetic Improvement for Agricultural Species (Dairy Cattle) of Shandong Province (2019LZGC011; Jinan, China), Shandong Provincial Natural Science Foundation (ZR2020QC175 and ZR2020QC176; Jinan, China), and the National Natural Science Foundation of China (32002172; Beijing). The authors thank the Institute of Animal Science and Veterinary Medicine of Shandong Academy of Agricultural Sciences and Dongying Shenzhou AustrAsia Modern Dairy Farm Co., Ltd. for providing animal samples and the Supercomputing Center of Shandong Agricultural University for computing technical support. The authors have not stated any conflicts of interest.

REFERENCES


ORCIDS
Jun Teng https://orcid.org/0000-0003-2573-0758
Zhi Chen https://orcid.org/0000-0003-3924-1891
Jianbin Li https://orcid.org/0000-0002-9587-2666
Zhangping Yang https://orcid.org/0000-0002-5479-8447
Chao Ning https://orcid.org/0000-0001-8247-1700