Published online 4 October 2019 *Nucleic Acids Research, 2020, Vol. 48, Database issue* ***D659–D667***
doi: 10.1093/nar/gkz854

Animal-ImputeDB: a comprehensive database with multiple animal reference panels for genotype imputation

Wenqian Yang[1,†], Yanbo Yang[1,†], Cecheng Zhao[1], Kun Yang[1], Dongyang Wang[1], Jiajun Yang[1], Xiaohui Niu[1,*] and Jing Gong[1,2,*]

[1]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, P. R. China and [2]College of Biomedicine and Health, Huazhong Agricultural University, Wuhan 430070, P. R. China

Received August 14, 2019; Revised September 19, 2019; Editorial Decision September 20, 2019; Accepted October 01, 2019

ABSTRACT

Animal-ImputeDB (http://gong_lab.hzau.edu.cn/Animal_ImputeDB/) is a public database with genomic reference panels of 13 animal species for online genotype imputation, genetic variant search, and free download. Genotype imputation is a process of estimating missing genotypes in terms of the haplotypes and genotypes in a reference panel. It can effectively increase the density of single nucleotide polymorphisms (SNPs) and thus can be widely used in large-scale genome-wide association studies (GWASs) using relatively inexpensive and low-density SNP arrays. However, most animals except humans lack high-quality reference panels, which greatly limits the application of genotype imputation in animals. To overcome this limitation, we developed Animal-ImputeDB, which is dedicated to collecting genotype data and whole-genome resequencing data of nonhuman animals from various studies and databases. A computational pipeline was developed to process different types of raw data to construct reference panels. Finally, 13 high-quality reference panels including ∼400 million SNPs from 2265 samples were constructed. In Animal-ImputeDB, an easy-to-use online tool consisting of two popular imputation tools was designed for the purpose of genotype imputation. Collectively, Animal-ImputeDB serves as an important resource for animal genotype imputation and will greatly facilitate research on animal genomic selection and genetic improvement.

INTRODUCTION

Genotype imputation is a process to predict and impute missing genotypes in terms of the haplotypes and genotypes in a reference panel (1), which plays essential roles in genome-wide association studies (GWASs) or fine mapping studies of a specific region (2,3). Genotype imputation is based on the assumption that two individuals, even if obviously unrelated, share short panels from a distant common ancestor in their genomes. Thus, it is possible to infer unobserved genotypes in one sample via the reference panel, which includes a large set of markers. Most contemporary imputation tools employ a hidden Markov model (HMM) framework to infer the genotype from the estimated haplotypes in a reference panel (4,5). Imputing genotypes at ungenotyped loci could dramatically boost the density of SNPs, increase the power of association studies, improve the ability to fine-map causal variations, facilitate the combination of different studies, and promote meta-analysis (6). Therefore, genotype imputation has been widely used in all kinds of genetic research, especially in humans (7–10).

In animals, numerous GWASs have been performed for genomic selection and genetic improvement. Although advances in high-throughput sequencing technologies have reduced the cost of whole-genome sequencing, GWASs usually require thousands of genotyped animals or more, resulting in high genotyping costs (11). Considering the high genotyping cost, most genetic studies still use low-density SNP panels. Some studies have adopted genotype imputation to increase SNP density after the fact (12) and have confirmed the accuracy and necessity of genotype imputation in animals (13–15). In Brangus beef cattle, genotype imputation has not only integrated different samples using different 40k SNP chips but also increased the density of SNPs (14). Additionally, genotype imputation substantially increased the genomic prediction accuracies of the es-

*To whom correspondence should be addressed. Tel: +86 27 87285085; Email: gong.jing@mail.hzau.edu.cn
Correspondence may also be addressed to Xiaohui Niu. Tel: +86 27 87285085; Email: niuxiaoh@126.com
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

timated breeding values (EBVs) of ten traits (14). For example, the genomic prediction accuracy of the EBV of calving ease direct (CED) was increased from 0.52 to 0.68 using leave-one-out cross-validation (LOOCV) (14). In a GWAS of lumbar number in Sutai pigs using the original 60K SNP array panel, no significant association between genotypes and lumbar number was observed in 418 Sutai pigs. However, after imputation, a quantitative trait locus (QTL) in *SSC1* was identified with a *P*-value of 9.01×10^{-18}, which was close to the location of the potential causative gene *NR6A1* (15).

A high-quality reference panel is usually a prerequisite for an effective and accurate genome imputation (16). For example, a HapMap 2 CEU reference panel of 60 individuals with 2.1 million markers was applied for genome-wide imputation in humans (17–19). With the rapid development of high-throughput sequencing, the 1000 Genomes Project accumulated low-coverage whole-genome sequencing data of 2504 individuals from 26 populations world-wide (16). Based on this dataset, a reference panel was constructed, which included 5008 haplotypes with over 88 million variants. Recently, the Haplotype Reference Consortium has generated a human reference panel of 64 976 haplotypes with 39 million SNPs by combining 20 studies (20). These high-quality human reference panels make it possible to accurately impute millions of genetic variations for human studies using low-density SNP array panels (21). Recently, some animal reference panels, such as pig and sheep, have been constructed (22,23). However, most animals lack a corresponding high-quality reference panel, which greatly limits the wide application of genotype imputation in animal genetic studies. In addition, the formats of the reference panels vary with the imputation tools, and current genotype imputation tools require researchers to have a certain background knowledge of computer language and bioinformatics, which makes it challenging for general geneticists and biologists to perform genotype imputation. Therefore, it is essential to develop a convenient database to provide these reference panels and imputation tools for animal genetic research.

To address this need, we developed the Animal Imputation Database (Animal-ImputeDB, http://gong_lab.hzau.edu.cn/Animal_ImputeDB/), which is dedicated to collecting publicly available genomic sequencing data of 13 animal species, constructing high-quality reference panels, and providing online genotype imputation tools.

DATA COLLECTION AND PROCESSING

Data collection

To collect as many samples as possible, several steps were taken in the data collection process. We first performed a systematic literature search in the PubMed, ISI Web of Science with the language restricted to English using the following main keywords: 'pig, dog, monkey, duck, chicken, horse, sheep, cattle, buffalo, rabbit, tarpan, panda, or goat' and 'genome sequencing, DNA sequencing, resequencing, genome-wide association study, genomic prediction, or GWAS'. The abstracts of these published studies were downloaded and manually checked by three researchers to select eligible studies. The full texts of all eligible articles were downloaded. Then, all references listed in these articles were also examined to identify more relevant literature. Next, researchers manually checked whether the raw genotypes or sequencing data could be downloaded. In addition, several sequencing deposit databases, such as the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) (24) of the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/sra/) (25) and the BIG Data Center (https://bigd.big.ac.cn/) (26), were also explored to find eligible samples. Finally, genotype or whole-genome sequencing (WGS) data of 13 species were collected from different sources to construct the animal genetic reference panels.

i) Collection of genotype data. Genotype data of pig, horse, cattle, goat, buffalo, chicken, tarpan, and panda in variant call format (VCF) were collected from the Genome Variation Map (GVM, https://bigd.big.ac.cn/gvm/home) (27) of BIG Data Center (BIGD, http://bigd.big.ac.cn/). The genotype data of dog (28), sheep (29) and duck (30) were gathered from the National Human Genome Research Institute (NHGRI, https://www.genome.gov/) (28), European Institute of Bioinformatics (EBI, https://www.ebi.ac.uk/) (31) and DUCKbase (http://duckbase.org/home) (30), respectively.
ii) Collection of WGS data. The WGS data of rabbit (32,33) and monkey (34–36) were obtained from the NCBI SRA.
iii) SNP annotation file. The dbSNP IDs of dog, horse and tarpan were collected from the dbSNP of the NCBI database (ftp://ftp.ncbi.nih.gov/snp/organisms/archive/) (37), and the dbSNP ID of sheep was collected from the Ensembl database (ftp://ftp.ensembl.org/pub/release-75/variation/vcf/ovis_aries/) (38). We downloaded the known variant files and ensured that the reference genome versions were consistent with ours.

Data processing

The raw WGS reads were subjected to quality control using FastQC (version: 0.11.5-Java-1.8.0_92) and cleaned with Trimmomatic (version: 0.36) (39). Subsequently, the clean reads were aligned to the reference genome using Burrows-Wheeler Aligner mem (BWA, version: 0.7.17-r1188) with default parameters (40). The aligned data were merged into a single BAM file, and the processed data were marked for duplicates by using Picard in Genome Analysis Toolkit (GATK, version: 4.0.12.0) (41). The duplicate reads were removed. We further performed variant calling by running HaplotypeCaller and variant refining by variant quality score recalibration (VQSR). Then, running HaplotypeCaller, an intermediate genomic GVCF file for each sample was produced by using GVCF mode, and GenotypeGVCFs in GATK was applied to pool all GVCF files together to create a VCF file of the raw variants. These raw variants identified by GATK were further filtered by using VariantFiltration (35). Default parameters of tools were used in the variant calling methodology. All genotype data were filtered through the following two steps with GATK

and Perl scripts: (i) SNP filtration. SNPs were first selected based on the following criteria: QualByDepth < 5.0, FisherStrand > 15.0, RMSMappingQuality < 50.0, ReadPosRankSumTest < −8.0, MappingQualityRankSumTest < −12.5, StrandOddsRatio > 3.0 (42). Then, the SNPs with a call rate < 0.9 or a minor allele frequency (MAF) < 0.01 were removed. (ii) Sample filtration. The animal samples with a genotype call rate < 0.9 were removed (Figure 1).

The detailed statistics of the genetic variants and sample data of each species in the final dataset are listed in Table 1.

Reference panel construction

Haplotypes of each species were constructed by Beagle (v5.0) (43) using clean SNP data with the default parameters. The reference panels were converted from VCF to M3VCF format by Minimac3 (2). Beagle (v5.0), Impute2, and Minimac3 are the most frequently used tools for genotype imputation. All these tools are similar in accuracy, but differ in memory requirements and computation time (44). Beagle is computationally fast and highly efficient in memory. Minimac3 is also superior to Impute2 in these two aspects (44). Furthermore, Beagle and Minimac3 are widely applied in animal genotype imputation. Therefore, we provided the reference panels in VCF and M3VCF formats corresponding to Beagle and Minimac3 in our database.

SNP annotation

For each species with a SNP annotation file, we mapped the dbSNP ID to the SNPs in our reference panels according to chromosome position. Furthermore, the allele frequency of each SNP was calculated. The above steps were performed with in-house scripts.

IMPLEMENTATION

Animal-ImputeDB (http://gong_lab.hzau.edu.cn/Animal_ImputeDB/) was built based on the Flask (version 1.0.3) framework with AngularJS (version 1.6.1) as the JavaScript library, running on the Apache 2 web server (version 2.4.18) with MongoDB (version 3.4.2) as its database engine. Animal-ImputeDB is available online without registration and optimized for Chrome (recommended), Internet Explorer, Opera, Firefox, Windows Edge and macOS Safari.

DATABASE CONTENT AND USAGE

Samples of 13 species in Animal-ImputeDB

In total, ∼400 million SNPs of 2265 samples from 13 species were deposited in Animal-ImputeDB. The detailed information, including the number of samples per species, the number of chromosomes, genome version, and the number of SNPs, is shown in Table 1 and displayed on the 'Home' page (Figure 2A). The species information, including the basic animal introduction, genome size, and the number of chromosomes, is presented in the 'Species information' module, which can be accessed by clicking the animal photos on the 'Home' page (Figure 2B). The detailed sample information of each species is described in the 'Sample information' module. The information including PubMed ID, journal, publication year of article, sample number, material, technology, platform, data type, and sequencing coverage of the project is provided. Users could obtain more information at NCBI PubMed Central (PMC) by clicking the 'PubMed ID' hyperlink on the list.

The imputation accuracy using reference panels in Animal-ImputeDB

To validate the performance of reference panels and imputation process, we calculated imputation accuracy for seven species with sample size larger than 100 using 5-fold cross-validation strategy. For each species, individuals were randomly divided into five folds. Each time, one-fold was selected as the study population, and the remaining individuals were used as the reference panel. Since most commercial SNP arrays of animals contain about 50k probes (14,15), we randomly selected 50 000 SNPs on autosomes of the study populations and masked other SNPs. Then we used Beagle and Minimac3 to impute the genotypes with default parameters. In this way, we have both the true and imputed genotypes. The imputed SNPs with MAF ≥0.01 and estimated squared correlation ≥0.3 were remained as properly imputed variants and used for the following evaluation. Two values were used to evaluate the accuracy of imputation. One is the concordance rate (CR), which is calculated as the number of genotypes imputed correctly divided by total imputed genotypes per species. The other value is the squared correlation (R^2) between true and imputed genotypes. The accuracy of imputation was the mean CR or R^2 across five folds for each species.

The results of imputation are summarized in Table 2. After imputation, the number of SNPs increased by 8.0–95.8 folds when using 50k markers in the study population. The average CRs for all test species were greater than 0.8. The average R^2 of Beagle ranged from 0.679 for duck to 0.812 for sheep, and the average R^2 of Minimac3 ranged from 0.751 for duck to 0.856 for sheep. These results indicate that our reference panels and used imputation tools have good performances, which can greatly increase the number of SNPs with relatively high accuracy.

Web interface

The Animal-ImputeDB database provides a user-friendly interface. It contains three main modules, namely, 'Imputation' for online genotype imputation, 'Reference Panel' for SNP search, and 'Download' for reference panel download. Users can access the 'Imputation/Reference Panel/Download' modules by clicking the corresponding buttons on the 'Home' page (Figure 2A) or by clicking on the hyperlink embedded in the corresponding animal photo (Figure 2B) in the 'Module' section on the 'Home' page. These 'Imputation/Reference Panel/Download' modules provide the functions 'Species Information/Online Imputation/SNP search/Sample information' (Figure 2C). Animal-ImputeDB provides detailed supporting documentation on the 'Help' page, and it is open to any feedback with email address provided on the 'Contact' page.

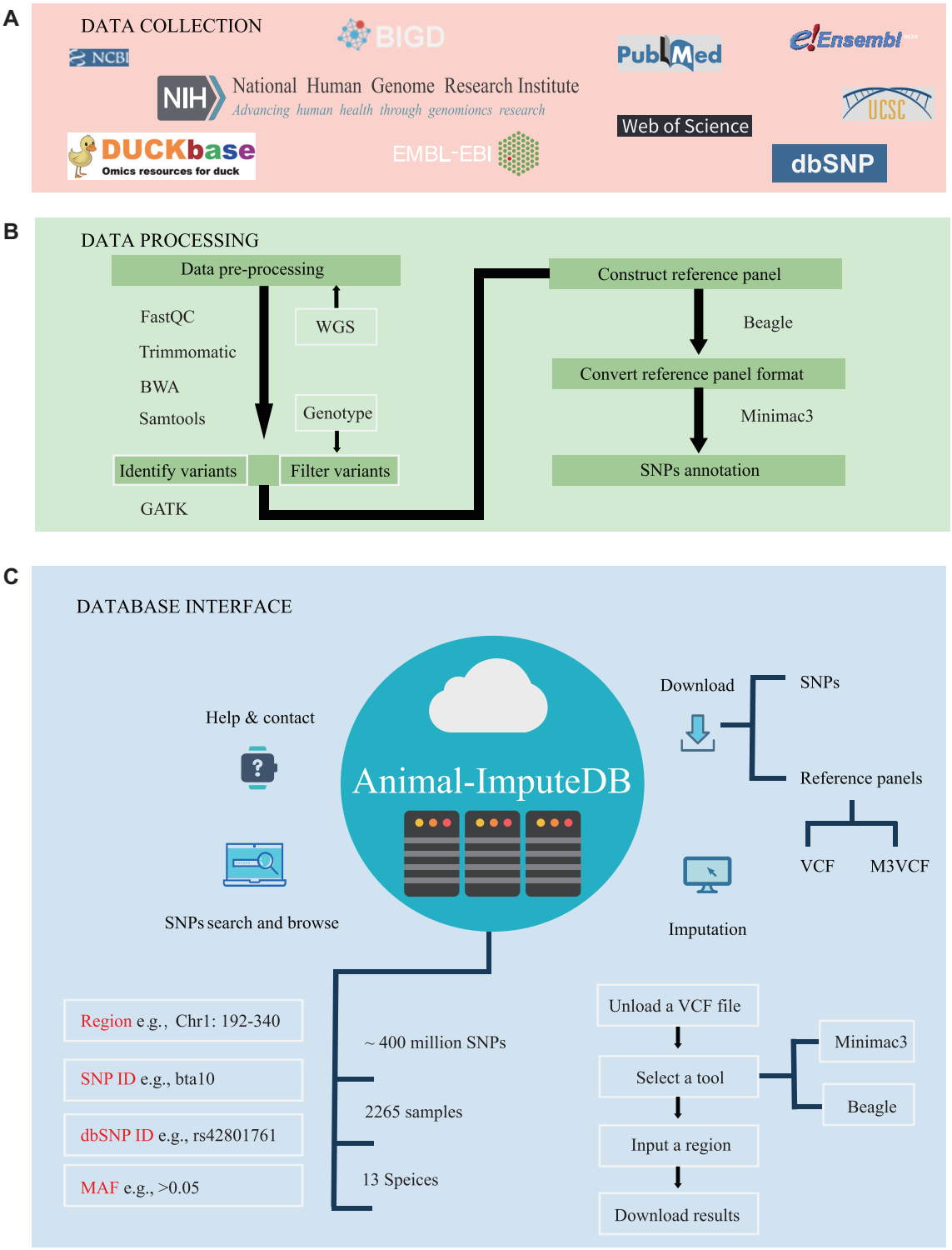
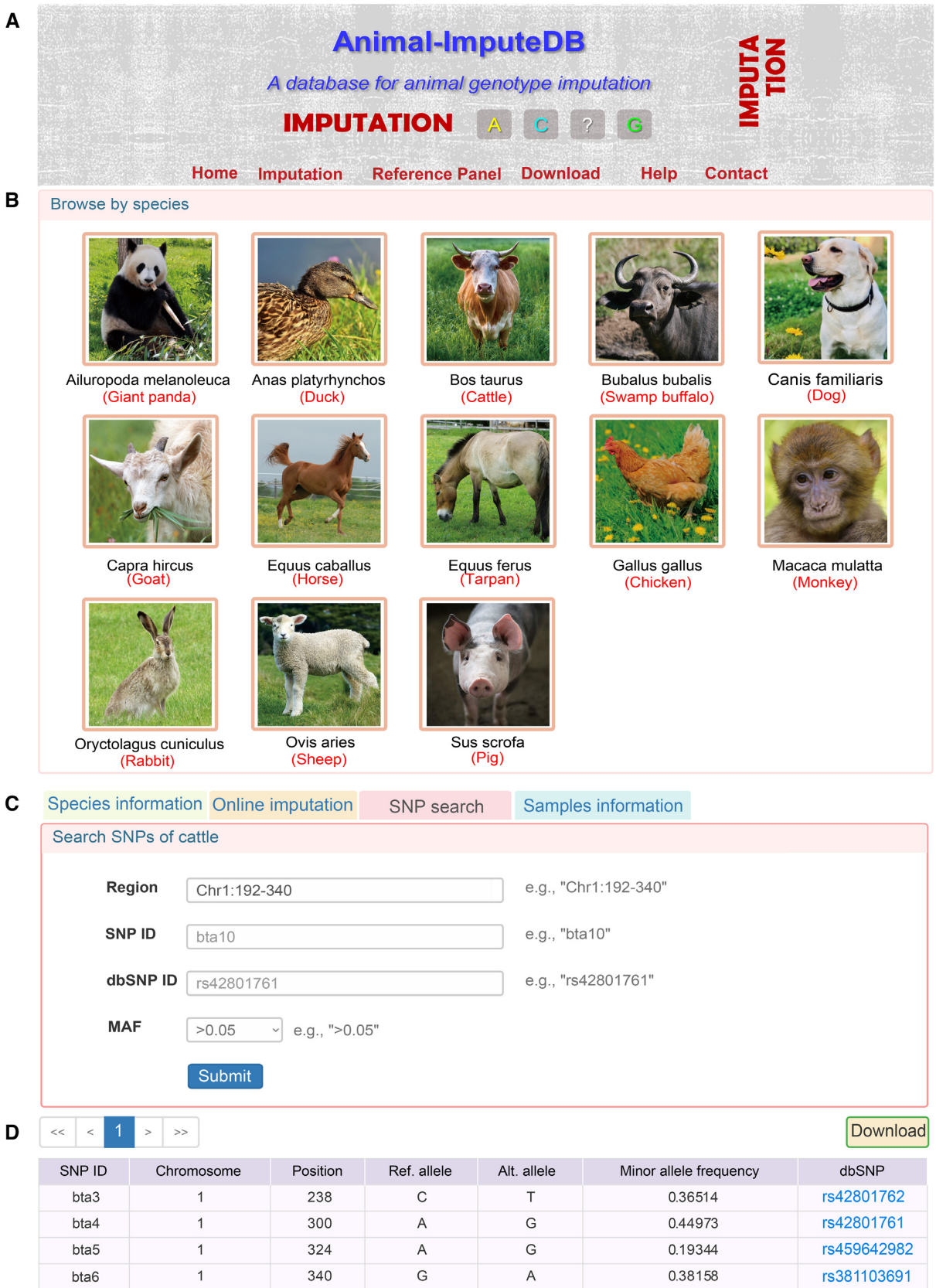


Figure 1. Construction of animal reference panels in Animal-ImputeDB. (**A**) Data collection. (**B**) Data processing. (**C**) Database content and web interface.



SNP ID	Chromosome	Position	Ref. allele	Alt. allele	Minor allele frequency	dbSNP
bta3	1	238	C	T	0.36514	rs42801762
bta4	1	300	A	G	0.44973	rs42801761
bta5	1	324	A	G	0.19344	rs459642982
bta6	1	340	G	A	0.38158	rs381103691

Figure 2. Overview of the Animal-ImputeDB database. (**A**) The main functions in Animal-ImputeDB, including 'Imputation', 'Reference Panel' and 'Download' modules. (**B**) The species included in Animal-ImputeDB. (**C**) The search box of SNP in Animal-ImputeDB. (**D**) An example of search results after inputting 'Chr1:192–420' in the 'SNP search' section of 'cattle'.

Table 1. Data summary in Animal-ImputeDB

Species	No. of chromosome	Reference panel	
		No. of sample	No. of SNPs
Ailuropoda melanoleuca (Giant panda)	28 354 scaffolds	34	4 671 936
Anas platyrhynchos (Duck)	30	106	12 682 400
Bos taurus (Cattle)	30	93	41 808 907
Bubalus bubalis (Swamp buffalo)	24	206	33 245 917
Canis familiaris (Dog)	39	658	61 065 811
Capra hircus (Goat)	30	233	29 889 815
Equus caballus (Horse)	32	53	19 257 635
Equus ferus (Tarpan)	32	19	7 809 754
Gallus gallus (Chicken)	35	103	26 864 273
Ovis aries (Sheep)	27	450	29 889 815
Sus scrofa (Pig)	19	233	40 323 709
Macaca mulatta (Monkey)	21	30	47 332 297
Oryctolagus cuniculus (Rabbit)	22	46	40 420 337

Table 2. The imputation accuracy using reference panels in Animal-ImputeDB

	Beagle imputation results				Minimac3 imputation results			
	No. of imputed SNPs (mean±SD)	Increased fold	CR (mean±SD)	R^2 (mean±SD)	No. of imputed SNPs (mean±SD)	Increased fold	CR (mean±SD)	R^2(mean±SD)
Buffalo	1 618 065±51 924	32.4	0.835±0.010	0.756±0.010	333 402±11 424	6.7	0.900±0.006	0.843±0.006
Chicken	1 637 061±218 238	32.7	0.939±0.031	0.772±0.052	519 892±100 062	10.4	0.946±0.031	0.824±0.036
Dog	449 768±11 343	9	0.871±0.006	0.733±0.012	221 222±8 932	4.4	0.905±0.006	0.799±0.014
Duck	750 920±14 269	15	0.813±0.015	0.679±0.023	293 485±9 285	5.9	0.865±0.012	0.751±0.021
Goat	797 748±20 260	16	0.888±0.009	0.807±0.018	320 904±10 751	6.4	0.920±0.010	0.856±0.018
Pig	4 792 133±390 227	95.8	0.929±0.031	0.751±0.033	2 072 512±327 546	41.5	0.950±0.022	0.818±0.030
Sheep	1 239 606±11 604	24.8	0.859±0.003	0.812±0.002	399 671±14 665	8.0	0.905±0.002	0.856±0.003

CR: concordance rate between true and imputed genotypes.
R^2: squared correlation between true and imputed genotypes.

SNPs of 13 curated species for searching and browsing in Animal-ImputeDB

To support SNP search and browse, the 'Reference Panel' page provides an advanced search box for different species. SNPs can be browsed by inputting the specific chromosomal region (e.g. Chr1: 192–340), SNP ID (e.g. bta10), dbSNP ID (e.g. rs42801761), or MAF (e.g. >0.05). Fuzzy queries are used in the search, and the query results are displayed in a table containing the basic SNP information, including SNP ID, chromosome, position, allele, minor allele frequency, and dbSNP ID. For example, when users select 'Cattle' and enter 'Chr1: 192–340' in the 'Region' box, the query results will be returned as shown in Figure 2D. The returned tables can be sorted by clicking on a specific column header. In addition, the query results can be exported to a tab-separated file and saved by clicking the 'Download' button (Figure 2D). To help users find more detailed SNP information, the dbSNP IDs in the query results are linked to the dbSNP database.

Online imputation for 13 curated species in Animal-ImputeDB

On the 'Imputation' page, Animal-ImputeDB provides an easy-to-use online tool consisting of two free and popular tools, namely, Beagle and Minimac3, for genotype imputation. There are two ways to navigate to the 'imputation' module: (i) by clicking on 'Imputation' in the 'Home' page browser bar and (ii) by clicking on the hyperlink in the corresponding species photo on the 'Home' page. Users can enter or copy pending processed genotype data into the text box (Figure 3A) or upload the genotype data directly via the 'Choose File' button. The genotype data should be input in VCF format with annotation information. An example of genotype data in the VCF format can be obtained by clicking the 'Example' button above the input box. After uploading the candidate genotype data, users should select one of the two tools (Figure 3B), enter the chromosome region, and click the 'Submit' button to submit the inquiry (Figure 3C). Then, the imputation results will be returned as a VCF format file and can be downloaded freely (Figure 3D).

Reference panels of 13 curated species for download in Animal-ImputeDB

Reference panels for 13 species are publicly available on the 'Download' page of Animal-ImputeDB. These 13 reference panels support both VCF and M3VCF file formats (text and binary), so users can download a reference panel in either VCF format or M3VCF format according to their own tool requirements. The M3VCF file is usable by only the Minimac3 tool stores a large reference panel in a compact manner, whereas the VCF file can be widely applied by most popular imputation tools. Our database provides a total of ∼400G data for users to download.

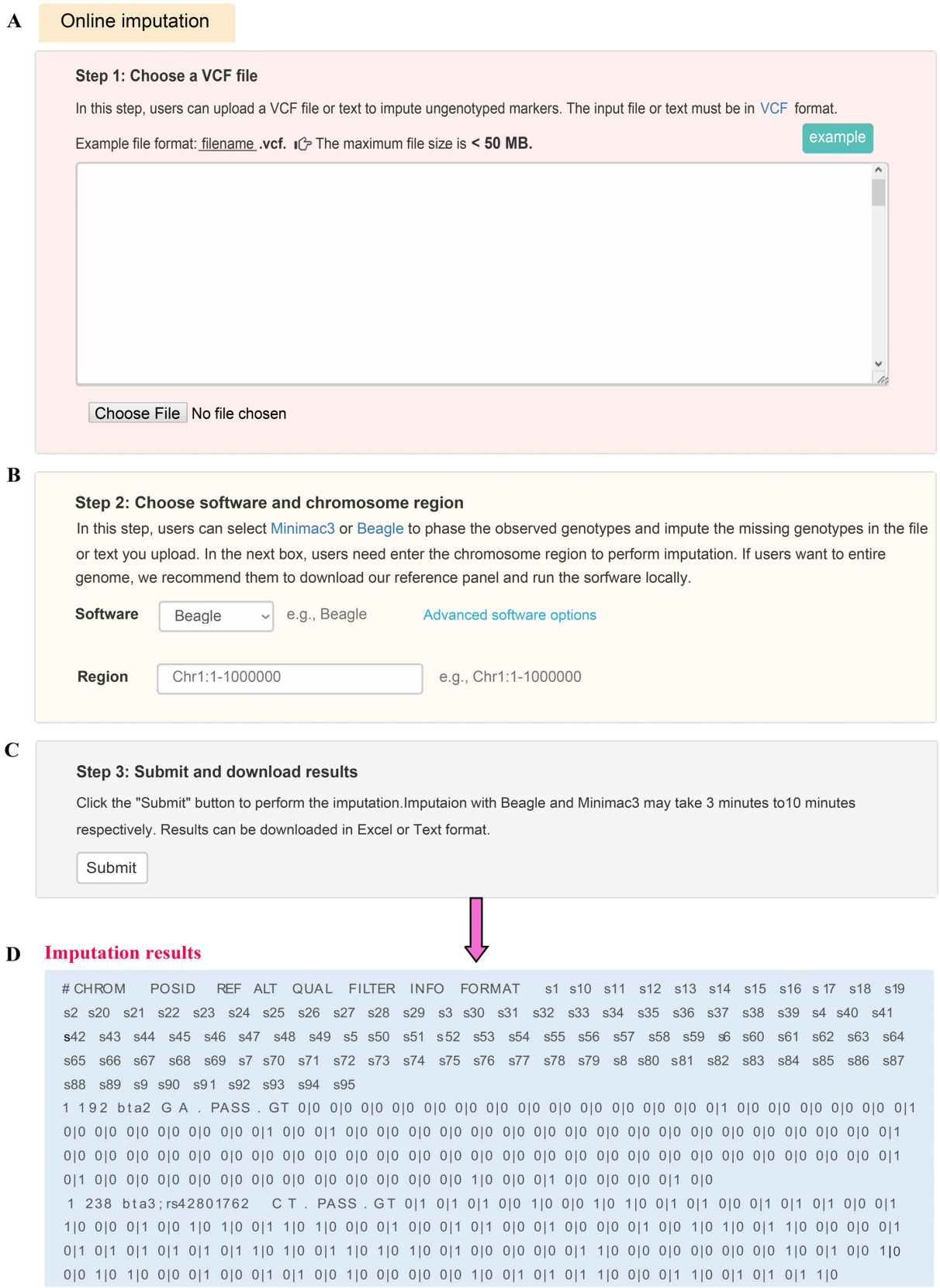


Figure 3. Online imputation tool in the Animal-ImputeDB database. (**A**) Input data through typing genotypes or uploading a VCF file by clicking the 'Choose File' button. (**B**) Select an imputation tool (Beagle or Minimac3) and enter the chromosome region of interest. (**C**) Submit the imputation task to Animal-ImputeDB. (**D**) An example of an imputation result.

SUMMARY AND FUTURE DIRECTIONS

Rapid progress has been seen in animal genome research in recent decades. Several animal-related databases have been widely used by animal researchers, such as AnimalQTLdb (45,46) and AnimalTFDB (47). However, no convenient database is available for animal genotype imputation. In this study, we developed the Animal-ImputeDB database by collecting publicly available data, constructing reference panels of 13 curated species, and designing an easy-to-use online genotype imputation tool. Reference panels of 13 animal species could be downloaded and used for the corresponding animal studies to increase the power of GWAS and to aid fine mapping of causative variants. All the SNPs of reference panels could be browsed and downloaded in our database. For user convenience, we linked the SNPs in Animal-ImputeDB to the NCBI SRA, BIGD, EBI, NCBI dbSNP, and NCBI PMC. With this easy-to-use online tool, researchers without coding experience can also perform genotype imputation easily. We believe that Animal-ImputeDB, with multiple animal reference panels and online imputation tools, will be a valuable resource for the field of animal breeding and genetic improvement.

Recent next-generation sequencing technology and imputation algorithm advances provide us with unprecedented opportunities to construct animal reference panels for genotype imputation. In the future, we will annually perform a systematic literature search to integrate more samples and species into Animal-ImputeDB and continue to update the database. Future database development will focus on response to community needs and functional annotations to improve the efficiency and comprehensiveness of the database. Collectively, we will maintain Animal-ImputeDB as an informative and valuable resource for animal genetic research.

FUNDING

National Natural Science Foundation of China (NSFC) [31970644 to J.G.]; Huazhong Agricultural University Scientific & Technological Self-innovation Foundation [11041810350 to J.G.]; Fundamental Research Funds for the Central University HZAU [2662017JC048 to X.H.N.]. Funding for open access charge: Huazhong Agricultural University Scientific & Technological Self-innovation Foundation [11041810350].

Conflict of interest statement. None declared.

REFERENCES

1. Marchini,J. and Howie,B. (2010) Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.*, **11**, 499–511.
2. Das,S., Forer,L., Schonherr,S., Sidore,C., Locke,A.E., Kwong,A., Vrieze,S.I., Chew,E.Y., Levy,S., McGue,M. *et al.* (2016) Next-generation genotype imputation service and methods. *Nat. Genet.*, **48**, 1284–1287.
3. Das,S., Abecasis,G.R. and Browning,B.L. (2018) Genotype imputation from large reference panels. *Annu. Rev. Genomics Hum. Genet.*, **19**, 73–96.
4. Browning,B.L. and Browning,S.R. (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.*, **84**, 210–223.
5. Pe'er,I., de Bakker,P.I., Maller,J., Yelensky,R., Altshuler,D. and Daly,M.J. (2006) Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat. Genet.*, **38**, 663–667.
6. Li,Y., Willer,C., Sanna,S. and Abecasis,G. (2009) Genotype imputation. *Annu. Rev. Genomics Hum. Genet.*, **10**, 387–406.
7. Genomes Project, C., Abecasis,G.R., Auton,A., Brooks,L.D., DePristo,M.A., Durbin,R.M., Handsaker,R.E., Kang,H.M., Marth,G.T. and McVean,G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
8. Genomes Project, C., Abecasis,G.R., Altshuler,D., Auton,A., Brooks,L.D., Durbin,R.M., Gibbs,R.A., Hurles,M.E. and McVean,G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
9. International HapMap, C., Altshuler,D.M., Gibbs,R.A., Peltonen,L., Altshuler,D.M., Gibbs,R.A., Peltonen,L., Dermitzakis,E., Schaffner,S.F., Yu,F. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52–58.
10. Visscher,P.M., Wray,N.R., Zhang,Q., Sklar,P., McCarthy,M.I., Brown,M.A. and Yang,J. (2017) 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.*, **101**, 5–22.
11. Calus,M.P., Bouwman,A.C., Hickey,J.M., Veerkamp,R.F. and Mulder,H.A. (2014) Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications. *Animal*, **8**, 1743–1753.
12. Habier,D., Fernando,R.L. and Dekkers,J.C. (2009) Genomic selection using low-density marker panels. *Genetics*, **182**, 343–353.
13. Daetwyler,H.D., Capitan,A., Pausch,H., Stothard,P., van Binsbergen,R., Brondum,R.F., Liao,X., Djari,A., Rodriguez,S.C., Grohs,C. *et al.* (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.*, **46**, 858–865.
14. Lopes,F.B., Wu,X.L., Li,H., Xu,J., Perkins,T., Genho,J., Ferretti,R., Tait,R.G. Jr, Bauck,S. and Rosa,G.J.M. (2018) Improving accuracy of genomic prediction in Brangus cattle by adding animals with imputed low-density SNP genotypes. *J. Anim. Breed. Genet.*, **135**, 14–27.
15. Yan,G., Qiao,R., Zhang,F., Xin,W., Xiao,S., Huang,T., Zhang,Z. and Huang,L. (2017) Imputation-based whole-genome sequence association study rediscovered the missing QTL for lumbar number in sutai pigs. *Sci. Rep.*, **7**, 615.
16. Genomes Project, C., Auton,A., Brooks,L.D., Durbin,R.M., Garrison,E.P., Kang,H.M., Korbel,J.O., Marchini,J.L., McCarthy,S., McVean,G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
17. Marchini,J., Howie,B., Myers,S., McVean,G. and Donnelly,P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
18. Wellcome Trust Case Control, C. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
19. International HapMap, C., Frazer,K.A., Ballinger,D.G., Cox,D.R., Hinds,D.A., Gibbs,R.A., Belmont,J.W., Boudreau,A., Hardenbol,P. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
20. McCarthy,S., Das,S., Kretzschmar,W., Delaneau,O., Wood,A.R., Teumer,A., Kang,H.M., Fuchsberger,C., Danecek,P., Sharp,K. *et al.* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.
21. Wood,A.R., Esko,T., Yang,J., Vedantam,S., Pers,T.H., Gustafsson,S., Chu,A.Y., Estrada,K., Luan,J., Kutalik,Z. *et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.*, **46**, 1173–1186.
22. O'Brien,A.C., Judge,M.M., Fair,S. and Berry,D.P. (2019) High imputation accuracy from informative low-to-medium density single nucleotide polymorphism genotypes is achievable in sheep. *J. Anim. Sci.*, **97**, 1550–1567.
23. van den Berg,S., Vandenplas,J., van Eeuwijk,F.A., Bouwman,A.C., Lopes,M.S. and Veerkamp,R.F. (2019) Imputation to whole-genome sequence using multiple pig populations and its use in genome-wide association studies. *Genet. Sel. Evol.*, **51**, 2.
24. Kodama,Y., Shumway,M., Leinonen,R. and International Nucleotide Sequence Database, C. (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–D56.

25. Sayers,E.W., Agarwala,R., Bolton,E.E., Brister,J.R., Canese,K., Clark,K., Connor,R., Fiorini,N., Funk,K., Hefferon,T. *et al.* (2019) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **47**, D23–D28.
26. Data Center Members, BIG. (2019) Database Resources of the BIG Data Center in 2019. *Nucleic Acids Res.*, **47**, D8–D14.
27. Song,S., Tian,D., Li,C., Tang,B., Dong,L., Xiao,J., Bao,Y., Zhao,W., He,H. and Zhang,Z. (2018) Genome Variation Map: a data repository of genome variations in BIG Data Center. *Nucleic Acids Res.*, **46**, D944–D949.
28. Plassais,J., Kim,J., Davis,B.W., Karyadi,D.M., Hogan,A.N., Harris,A.C., Decker,B., Parker,H.G. and Ostrander,E.A. (2019) Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.*, **10**, 1489.
29. Naval-Sanchez,M., Nguyen,Q., McWilliam,S., Porto-Neto,L.R., Tellam,R., Vuocolo,T., Reverter,A., Perez-Enciso,M., Brauning,R., Clarke,S. *et al.* (2018) Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nat. Commun.*, **9**, 859.
30. Zhou,Z., Li,M., Cheng,H., Fan,W., Yuan,Z., Gao,Q., Xu,Y., Guo,Z., Zhang,Y., Hu,J. *et al.* (2018) An intercross population study reveals genes associated with body size and plumage color in ducks. *Nat. Commun.*, **9**, 2648.
31. Cook,C.E., Lopez,R., Stroe,O., Cochrane,G., Brooksbank,C., Birney,E. and Apweiler,R. (2019) The European Bioinformatics Institute in 2018: tools, infrastructure and training. *Nucleic Acids Res.*, **47**, D15–D22.
32. Wang,Z., Zhang,J., Li,H., Li,J., Niimi,M., Ding,G., Chen,H., Xu,J., Zhang,H., Xu,Z. *et al.* (2016) Hyperlipidemia-associated gene variations and expression patterns revealed by whole-genome and transcriptome sequencing of rabbit models. *Sci. Rep.*, **6**, 26942.
33. Carneiro,M., Rubin,C.J., Di Palma,F., Albert,F.W., Alfoldi,J., Martinez Barrio,A., Pielberg,G., Rafati,N., Sayyab,S., Turner-Maier,J. *et al.* (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*, **345**, 1074–1079.
34. Yan,G., Zhang,G., Fang,X., Zhang,Y., Li,C., Ling,F., Cooper,D.N., Li,Q., Li,Y., van Gool,A.J. *et al.* (2011) Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat. Biotechnol.*, **29**, 1019–1023.
35. Zhong,X., Peng,J., Shen,Q.S., Chen,J.Y., Gao,H., Luan,X., Yan,S., Huang,X., Zhang,S.J., Xu,L. *et al.* (2016) RhesusBase popgateway: genome-wide population genetics atlas in rhesus macaque. *Mol. Biol. Evol.*, **33**, 1370–1375.
36. Zhang,S.J., Liu,C.J., Yu,P., Zhong,X., Chen,J.Y., Yang,X., Peng,J., Yan,S., Wang,C., Zhu,X. *et al.* (2014) Evolutionary interrogation of human biology in well-annotated genomic framework of rhesus macaque. *Mol. Biol. Evol.*, **31**, 1309–1324.
37. Sherry,S.T., Ward,M.-H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K.J.N.A.R. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, D308–D311.
38. Cunningham,F., Achuthan,P., Akanni,W., Allen,J., Amode,M.R., Armean,I.M., Bennett,R., Bhai,J., Billis,K., Boddu,S. *et al.* (2019) Ensembl 2019. *Nucleic Acids Res.*, **47**, D745–D751.
39. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
40. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
41. McKenna,A., Hanna,M., Banks,E., Sivachenko,A., Cibulskis,K., Kernytsky,A., Garimella,K., Altshuler,D., Gabriel,S., Daly,M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
42. Bimber,B.N., Raboin,M.J., Letaw,J., Nevonen,K.A., Spindel,J.E., McCouch,S.R., Cervera-Juanes,R., Spindel,E., Carbone,L., Ferguson,B. *et al.* (2016) Whole-genome characterization in pedigreed non-human primates using genotyping-by-sequencing (GBS) and imputation. *BMC Genomics*, **17**, 676.
43. Browning,B.L., Zhou,Y. and Browning,S.R. (2018) A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.*, **103**, 338–348.
44. Browning,B.L. and Browning,S.R. (2016) Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.*, **98**, 116–126.
45. Hu,Z.L., Fritz,E.R. and Reecy,J.M. (2007) AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Res.*, **35**, D604–D609.
46. Hu,Z.L., Park,C.A. and Reecy,J.M. (2016) Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res.*, **44**, D827–D833.
47. Zhang,H.M., Chen,H., Liu,W., Liu,H., Gong,J., Wang,H. and Guo,A.Y. (2012) AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.*, **40**, D144–D149.