MGI's AgriHigh Low-pass WGS Package Facilitates Molecular Breeding of Modern Agriculture

MGI has developed the AgriHigh Low-pass WGS based on many years of technology accumulation. In this application note, this package was utilized to perform low-pass WGS on 384 pig ear samples, and the results were highly consistent with the precision of solid-phase array or the concordance of high-depth WGS. This solution provides a new option for large-scale breeding of animals and plants, and can achieve the whole-process of genotyping in an ultra-low cost, fast, automated, and high-throughput manner.

Recommended application: Agrigenomics
Recommended models: DNBSEQ-G400RS, DNBSEQ-T7RS (Sequencing platform)
MGISP-NE384RS, AlphaTool, MGISP-960RS, MGISP-100RS
(Automated system)

- A large-scale, high-throughput and low-cost low-pass whole genome solution

  The AgriHigh Low-pass WGS Package provides a whole-process solution that is extremely simple (3-step library preparation), throughput flexible (96-1,536 samples/day), automation friendly, and analysis efficient. This solution is capable of detecting million-level SNP and InDel loci, providing sufficient genetic information for livestock and poultry genetic improvement and molecular breeding.

- Efficient and high-quality sequencing data output

  DNBSEQ sequencing technology exhibits many excellent features such as high accuracy, low duplication rate and low index hopping rate.

Background

Germplasm resources are the material basis for human survival and agricultural development, as well as for crop, livestock and poultry genetic improvement and related basic researches. Throughout the history of agricultural breeding worldwide, biological breeding has gone through traditional breeding (hybrid breeding), mutation breeding, ploidy breeding (haploid breeding, polyploid breeding), cell engineering and chromosome engineering breeding, molecular marker assisted breeding, whole genome breeding, and gene editing breeding, etc[1].

Genome editing technology facilitates the exploration of functional genes and the cultivation of new varieties, while whole genome selection (GS) molecular breeding provides a broad platform for the application of functional genomes. Whole genomic selection technology is based on genotype and large amount of phenotypic data from high-density molecular markers covering the whole genome.

It estimates the breeding value of each phenotype related marker, and then calculates the whole genome estimated breeding values (GEBVs), to detect all loci related to phenotype variation, realize accurate analysis of the contributions of multiple minor quantitative trait loci (QTLs), and achieve high-throughput selection of all loci. By aggregating wanted genotypes at the whole genome level, it improves important agronomic traits[2].

Whole GS can not only shorten the breeding time, but also significantly improve the breeding efficiency. GS has been widely applied in dairy cow breeding, and its application in pig, sheep, and beef cattle breeding is being carried out steadily. With the continuous changes in breeding objectives and the increasing emphasis on green, healthy, and environment-friendly animal production, whole GS breeding can undoubtedly play a more important role in breeding industry innovation.

In the past decade, the rapid development of high-throughput genotyping technologies (including single nucleotide polymorphism (SNP) array[3], reduced representation sequencing (RRS) [4, 5], and whole genome sequencing (WGS)[6]), has significantly enhanced the application of genome prediction in plants[7] and animals[8]. Although low-density SNP data from array or RRS can satisfy most application scenarios, achieving higher SNP density through WGS can theoretically improve the precision of genome prediction[9, 10]. However, the high cost of using WGS to sequence thousands of individuals limits its widespread use in genome prediction of various species. The method of combining low-pass WGS with data imputation came into application[11]. Rather than rely on high-pass genotyping of a limited number of gene loci, this method collects information from millions of randomly sampled low-confidence-coefficient variants, and then using a reference group library to fill this information in possible genotypes. This reference group library consists of a large number of high-depth WGS datasets, which represent potential haplotypes in the species group.

MGI's DNBSEQ sequencing platform is increasingly favored by global customers due to its high precision and sensitivity, low duplication rate, and low index hopping rate, etc. In the field of agricultural research and application, this sequencing platform has been widely used for species genome analysis, trait-related signal pathway analysis, and so on. The low-pass WGS technology based on the DNBSEQ sequencing platform has been successfully applied to genetic analysis and cultivation of selected traits in species such as pig, cow, and fish[12, 13]. MGI has launched a total solution of AgriHigh low-pass WGS. This solution is designed specially for large-scale breeding scenarios of animals and plants, achieving a fast, automated, and high-throughput genotyping process at ultra-low cost. Germplasm resources are the basis of livestock and poultry crop breeding, and the preservation and processing of samples are particularly important during the breeding process. Based on MGI's capabilities in biological sample library, automated equipment, sequencing platform, and bioinformatic processing, MGI has developed a complete solution, namely AgriHigh Low-pass WGS Package, to meet the demand of large-scale biological breeding process.

This solution can achieve an annual throughput of 12,500 to 400,000 by flexibly combining different types of automated equipment and sequencers, meeting the diverse needs of customers.

An Overview of the AgriHigh Low-pass WGS Solution

Sample storage

It often takes some time to collect enough samples and appropriate storage conditions are needed to ensure the integrity of samples. MGI provides a variety of sample storage products, such as the MGICLab-RG series (suitable for temporary storage of blood samples from 4 ℃ ~ -20 ℃), MGICLab-LT series (providing -80 ℃ environment, suitable for long-term storage of blood, tissues, nucleic acid and other samples), and MGICLab-LN Pro series (used for long-term sample storage, providing stable liquid nitrogen low-temperature condition). These products can be used for the storage of hundreds of thousands or even millions of samples, providing reliable sample management solutions.

Nucleic acid extraction

In the medium throughput AgriHigh low-pass WGS package, it is recommended to use the MGIEasy Magnetic Beads Genomic DNA Extraction Kit (96RXN) for automated sample extraction on the MGISP-960 automated system. The range of animal tissue sample input is 5 mg~ 20 mg. In the high throughput AgriHigh low-pass WGS package, it is recommended to use the MGIEasy Genomic DNA Extraction Prepacked Kit (MGISP-NE384) (384RXN) for automated sample extraction on the MGISP-NE384 automated system, with a sample input range of 5 mg ~20 mg. The related automation scripts have been optimized and installed into the system and customer verification is not needed. Please refer to the relevant instructions for detailed operations.

Library preparation

200 ng gDNA is recommended as input and the MGIEasy Large-scale PCR-Free FS Library Prep Set for Low-pass WGS is used for library preparation. Library preparation includes fragmentation, end repair, adaptor ligation, and sample pooling of equal volume. Please refer to the relevant instructions for detailed operations. The MGISP-960 automatic system is used in both medium and high throughput solutions to complete library preparation. The relevant automatic scripts have been optimized and customer verification is not needed.

Sequencing

The prepared libraries were made into DNA nanoballs (DNBs) using a one-step method, mixed with equal mass, and then loaded onto DNBSEQ-G400RS (medium package) or DNBSEQ-T7RS (high package) to complete paired-ended 150 bp (PE150) sequencing.
Taken pig as an example, the recommended data size for each sample is 3GB (1~1.5×). As to PE150, the maximum sample size per chip for DNBSEQ-G400RS and DNBSEQ-T7RS is 96 and 384, respectively.

Bioinformatics analysis

The BOLT Low-pass WGS Software (BOLT_Low-pass) was self-developed for bioinformatics analysis. BOLT_Low-pass is an efficient software based on low-pass sequencing (0.1×~10×) for genotype imputation. During analysis, the MegaBOLT Bioinformatics analysis accelerator is highly recommended and can greatly improve the analyzing speed by accelerating various steps like raw sequencing data handling, quality control, alignment, sorting and other steps.

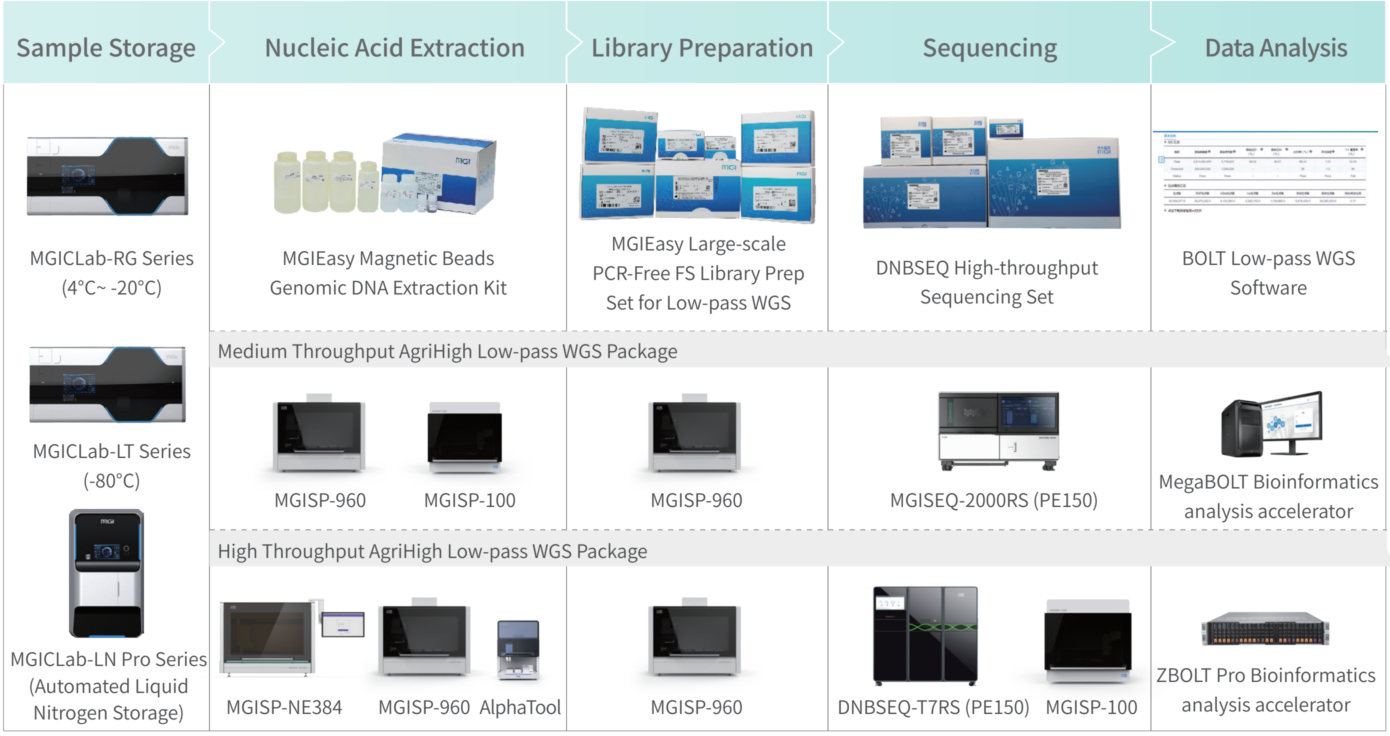

Figure 1. Overview of the whole process of AgriHigh Low-pass WGS solution.

Research Overview

The aim of this study is to verify the feasibility of applying Low-pass WGS genotyping in livestock, poultry and crop breeding. In this research, ear samples from two pig breeds, Yorkshare pigs and Meishan pigs, were selected to investigate SNP distribution with 1~1.5× WGS. Our comprehensive analysis showed that the results from Low-pass WGS were extremely similar to those of solid-state array and 50× WGS, with typing concordance of 97.8±1.1% and 98.0±0.5%, respectively. This study implied that Low-pass WGS is an alternative and efficient method for genome research and molecular breeding in large-scale livestock and poultry.

Results

The AgriHigh Low-pass WGS Package can achieve high-quality nucleic acid extraction and library preparation

In this research, the high throughput AgriHigh Low-pass WGS solution was selected to carry out Low-pass WGS analysis of 384 pig individuals with ear samples that had chip results (190 Yorkshare pigs and 194 Meishan pigs). This study utilized MGISP-NE384, AlphaTool and MGISP-960 automated system to complete DNA extraction and library preparation of 384 samples within one day, DNBSEQ-T7 to finish sequencing within 24 hours, ZBLOT Pro to finish analysis within 6 hours. The overall turnaround time (TAT) was less than 3 days.

The extraction was carried out using MGISP-NE384, with yields exceeding 2μg and A260/A280 ranging from 1.7 to 2.0, which met the experimental requirements (Table 1, Figure 2, the extraction and electrophoresis results of 8 randomly selected samples are shown).

Sample NO.	DNA concentration (ng/μL)	260/280	260/230
S1	258.3	1.87	2.15
S2	235.2	1.89	2.35
S3	241.4	1.88	2.40
S4	263	1.90	2.30
S5	250.1	1.89	2.38
S6	218.5	1.87	2.24
S7	292.4	1.88	2.22
S8	218.9	1.87	2.24

Table 1. The DNA concentration and purity of 8 randomly selected pig ear samples.

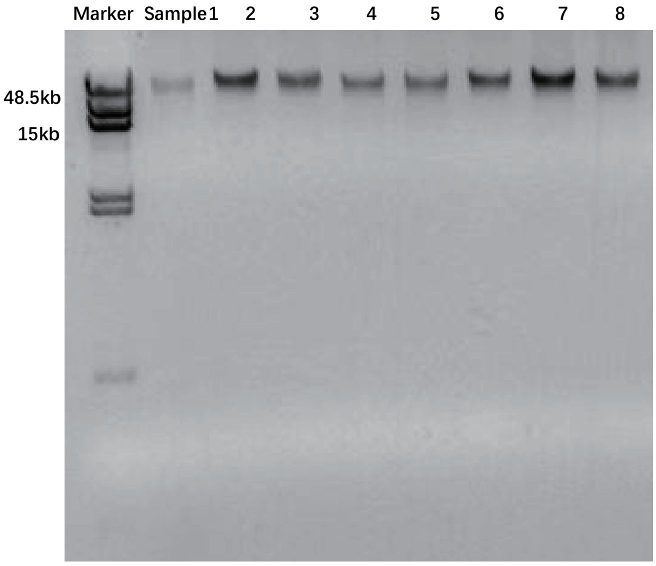

Figure 2. The electrophoresis result of DNA from 8 randomly selected pig ear samples.

During library preparation, 200 ng of genomic DNA was taken as input for fragmentation, end repair, adaptor ligation and sample mixing. A total of 48 samples were pooled into one group, resulting in 8 mixed products for subsequent product purification and DNB preparation (Table 2). The 8 DNBs were mixed of equal mass for subsequent genetic sequencing.

DNA input	NO.	Yield of purified product (ng)	OS V4 DNB input amount (ng)	DNB concentration OSV4(ng/μL)
200ng	1	115	24	18
	2	105	24	16
	3	109	24	20
	4	92	24	18
	5	114	24	20
	6	103	24	18
	7	89	24	21
	8	98	24	18

Table 2. Parameters of key steps during library preparation.

The AgriHigh Low-pass WGS Package offers high-quality sequencing data

Libraries were sequenced using the DNBSEQ-T7RS genetic sequencer, generating 5,634 M high-quality sequencing reads with an overall Q30 of 96.49%. The quality for each cycle is shown in Figure 3A. For each sample, the average data size is 14.7 M reads, mapping ratio is 99.6±0.3%, sequence depth is 1.5±1×, genome coverage is 60.8±16.9%, and the genome coverage under 1× sequence depth data is 53.7±2.5% (Figure 3B). By simulating the genome coverage at different sequence depth, it can be seen that the coverage gradually increased along with the increasing of sequence depth (Figure 3C).

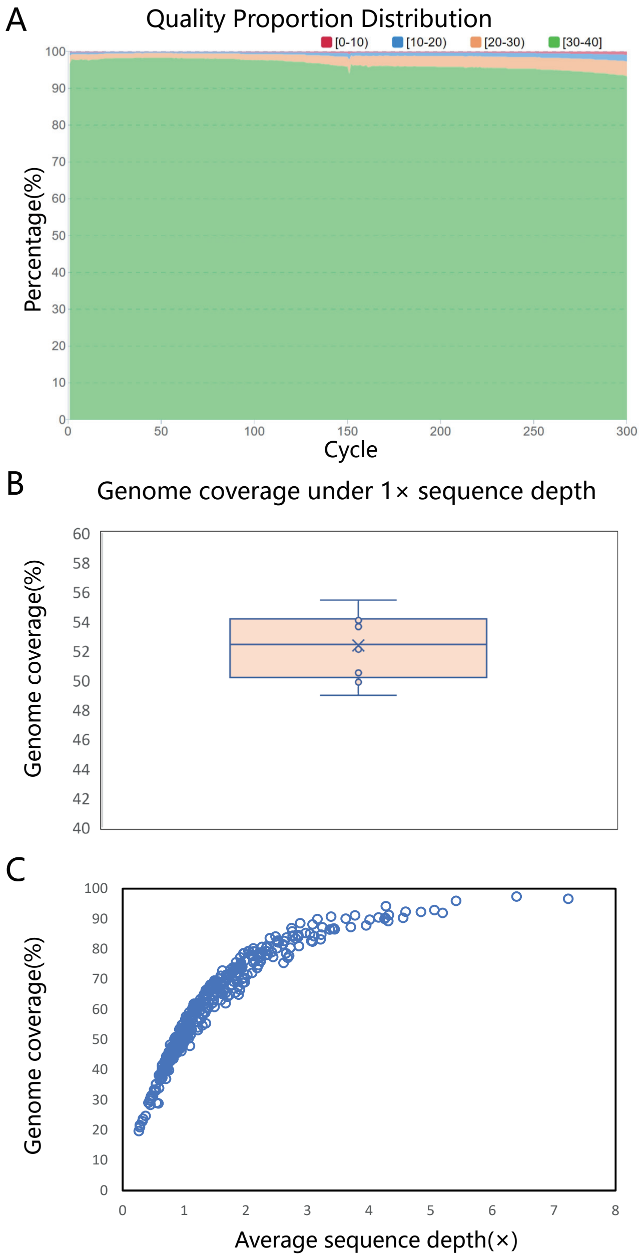

Figure 3. A) The quality of sequencing data; B) the genome coverage at 1× average sequence depth; C) The genome coverage at different sequence depth.

The AgriHigh Low-pass WGS data was comparable to solid-state array and high-pass WGS

A self-built Reference Panel covering 30,475,252 SNP loci and 4,123,565 InDel loci has be used for imputation of the sequencing results. Among all individuals, the actual average number of detected SNPs and InDels was 8,524,950±1,705,032, and 1,293,998±175,194, respectively. This low-pass whole genome sequencing solution far exceeded the number of SNPs detected by the chip array (21,382±4,485), and covered over ~90% of SNP loci on the chip array (Figure 4A). The RTG Tools were used to validate the precision of 20,133±4,055 SNPs shared by the low-pass sequencing (1×) imputation results and the chip array (Figure 4A). The results showed that the precision was 98.7±1.1% in all samples, 98.4±0.8% in the Yorkshare pig group, and 97.2±1.0% in the Meishan pig group (Figure 4B).

Meanwhile, the high-depth sequencing (50×) was conducted on 6 Yorkshare pigs and 2 Meishan pigs, and the common detected loci of the low-pass sequencing (1×) imputation results and the whole genome high-depth were compared using the GATK Genotype Concordance tool. The results showed that, the SNP and InDel Concordance of low-pass and high-pass data in the Yorkshare pig group were 98.1±0.5% and 97.8±0.6%, respectively, and the SNP and InDel Concordance in the Meishan pig group were 98.7% and 97.5%, respectively (Figure 4C).

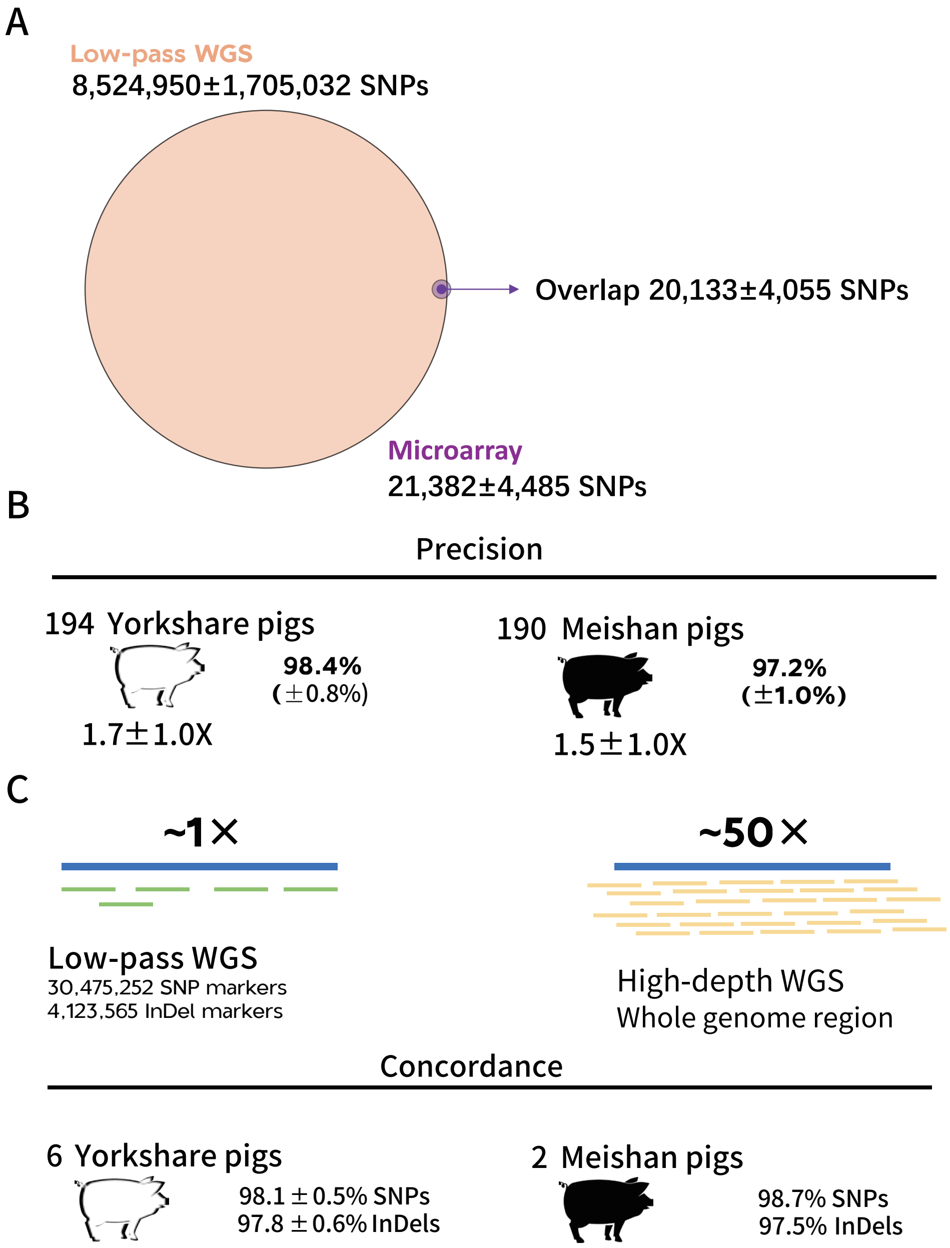

Figure 4. The overlapped SNPs number shared by the solid-chip array and the average number of SNPs for each individual detected by the AgriHigh Low-pass WGS solution(A); the comparison of precision between those two solutions (B); the comparison of concordance of SNP and InDel between AgriHigh Low-pass WGS and high-pass sequencing (50×) (C).

Summary

The AgriHigh Low-pass WGS Package aims to address agricultural challenges under global population growth, climate change and environmental pressures. MGI provides a total solution from nucleic acid extraction to data analysis for the agricultural field through reducing sequencing costs and making technological innovation. The goal is to shorten breeding cycle, improve breeding efficiency, promote the breeding of varieties with superior performance, reduce resource consumption, and promote the sustainable development of agriculture.

Genetic Sequencer DNBSEQ-G400 RS

Genetic Sequencer DNBSEQ-T7RS

References

1. Bawa, G., Liu, Z., Yu, X., Tran, L. P. & Sun, X., Introducing single cell stereo-sequencing technology to transform the plant transcriptome landscape. *TRENDS PLANT SCI* 29 249 (2024).
2. Desta, Z. A. & Ortiz, R., Genomic selection: genome-wide prediction in plant improvement. *TRENDS PLANT SCI* 19 592 (2014).
3. LaFramboise, T., Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *NUCLEIC ACIDS RES* 37 4181 (2009).
4. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J. L., Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLOS ONE* 7 e32253 (2012).
5. Chen, Q. et al., Genotyping by genome reducing and sequencing for outbred animals. *PLOS ONE* 8 e67500 (2013).
6. Giani, A. M., Gallo, G. R., Gianfranceschi, L. & Formenti, G., Long walk to genomics: History and current approaches to genome sequencing and assembly. *COMPUT STRUCT BIOTEC* 18 9 (2020).
7. Parveen, R. et al., Understanding the genomic selection for crop improvement: current progress and future prospects. *MOL GENET GENOMICS* 298 813 (2023).
8. Georges, M., Charlier, C. & Hayes, B., Harnessing genomic information for livestock improvement. *NAT REV GENET* 20 135 (2019).
9. Meuwissen, T. & Goddard, M., Accurate prediction of genetic values for complex traits by whole-genome resequencing. *GENETICS* 185 623 (2010).
10. Iheshiulor, O. O., Woolliams, J. A., Yu, X., Wellmann, R. & Meuwissen, T. H., Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *GENET SEL EVOL* 48 15 (2016).
11. Martin, A. R. et al., Low-coverage sequencing cost-effectively detects known and novel variation in underrepresented populations. *AM J HUM GENET* 108 656 (2021).
12. Yang, R. et al., Accelerated deciphering of the genetic architecture of agricultural economic traits in pigs using a low-coverage whole-genome sequencing strategy. *GIGASCIENCE* 10 (2021).
13. Zhuangbiao Zhang, A. W. H. H., The efficient phasing and imputation pipeline of lowcoverage whole genome sequencing data using a highquality and publicly available reference panel in cattle. *Animal Research and One Health* 1 4 (2023).

Recommended Ordering Information

Medium Throughput AgriHigh Low-pass WGS Package

Category	Product	Cat. NO.
Sample Storage	Low-temperature Automated Biobanking System MGICLab- RG Series	Customized product
	Ultra-low Temperature Automated Biobanking System MGICLab-LT Series	Customized product
	Automated Liquid Nitrogen Storage System MGICLab-LN55K Pro	900-000863-00
Automatic Instruments	DNA Sequencing Library Preparation System MGISP-100RS	900-000206-00
	MGISP-960RS High-throughput Automated Sample Preparation System-Configuration 7	900-000152-00
Genetic Sequencer	Genetic Sequencer DNBSEQ-G400 RS*	900-000170-00
Extraction Reagents	MGIEasy Magnetic Beads Genomic DNA Extraction Kit (96RXN)	940-000972-00
Library Prep	MGIEasy Large-scale PCR-Free FS Library Prep Set for Low-pass WGS (96RXN)	940-002162-00
Sequencing Reagents	DNBSEQ OneStep DNB Make Reagent Kit V4.0 (4 RXN/Kit) *	940-001654-00
	DNBSEQ-G400RS High-throughput Rapid Sequencing Set (FCL PE150) *	940-001356-00
	CPAS Barcode Primer 3 Reagent Kit*	1000020834
Bioinformatic Analysis	MegaBOLT Bioinformatics analysis accelerator	970-000555-00
	BOLT Low-pass WGS Software Analysis Package (96 reports)	970-000519-00

*Unless otherwise informed, this StandardMPS sequencing reagent is not available in Germany, UK, Sweden, and Switzerland.

Recommended Ordering Information

High Throughput AgriHigh Low-pass WGS Package

Category	Product	Cat. NO.
Sample Storage	Low-temperature Automated Biobanking System MGICLab- RG Series	Customized product
	Ultra-low Temperature Automated Biobanking System MGICLab-LT Series	Customized product
	Automated Liquid Nitrogen Storage System MGICLab-LN55K Pro	900-000863-00
Automatic Instruments	Automated Nucleic Acid Extractor MGISP-NE384RS	900-000358-00
	DNA Sequencing Library Preparation System MGISP-100RS	900-000206-00
	MGISP-960RS High-throughput Automated Sample Preparation System-Custom Configuration 5-V3	940-000150-00
	AlphaTool liquid handler	960-001238-00
Genetic Sequencer	MGIDL-T7RS DNB Loader*	900-000134-00
	Genetic Sequencer DNBSEQ-T7RS*	900-000242-00
	Data Analysis Appliance	900-000406-00
Extraction Reagents	24 Reagent Adaptor	962-000068-00
	MGIEasy Genomic DNA Extraction Prepacked Kit (MGISP-NE384) (384RXN)	940-000974-00
Library Prep	MGIEasy Large-scale PCR-Free FS Library Prep Set for Low-pass WGS (Plate) (384RXN)	940-002049-00
Sequencing Reagents	DNBSEQ OneStep DNB Make Reagent Kit V4.0 (4 RXN/Kit)*	940-001654-00
	DNBSEQ-T7 High-throughput Sequencing Set (FCL PE150) V3.0*	900-000268-00
	CPAS Barcode Primer 3 Reagent Kit*	1000020834
Bioinformatic Analysis	ZBOLT Pro Bioinformatics analysis accelerator	900-000460-00
	BOLT Low-pass WGS Software Analysis Package (96 reports)	970-000519-00

*Unless otherwise informed, this StandardMPS sequencing reagent is not available in Germany, UK, Sweden, and Switzerland.

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083

Version: August 2024

☎ +86-4000-688-114
en.mgi-tech.com
✉ MGI-service@mgi-tech.com

Authors: Dou Hu, Lin Yang,
Shaohong Zhang, Xinming Liang
Editor-in-Charge: Qiwei Wang
Reviewer: Yao Jiang