Performance Evaluation for IDT xGen 16S Amplicon Panel v2 on DNBSEQ-G50 Platform

IDT product compatible with MGI DNBSEQ sequencing platform enables microbial community identification

This study, based on IDT xGen 16S Amplicon Panel v2 and MGI DNBSEQ-G50 sequencing platform, accurately analyzed bacterial species and corresponding abundances in the ZymoBIOMICS™ Microbial Community DNA Standard (D6305).

Moreover, combination with 16S SNAP APP analysis tool developed by IDT has made it achieve an effective data utilization closed to 100%, showing an improved species typing results.

Recommended application: Microbial Community Identification

Recommended sequencers: DNBSEQ-G50，DNBSEQ-G400 & DNBSEQ-G99ARS

- ### High-precision in microbial identification

  IDT xGen 16S Amplicon Panel v2 covers the hyper variable V1-V9 regions in 16S rRNA gene, exhibiting high precision for microbial identification.

- ### Perfect compatibility with DNBSEQ sequencing platform

  MGIEasy Universal Library Conversion kit enables IDT solutions perfectly work on DNBSEQ sequencing platform.

- ### High-quality sequencing data

  DNBSEQ sequencing technology exhibits many excellent features such as high accuracy, low duplication rate and low index hopping rate. Moreover, DNBSEQ-G50 has become a classic model for microbial sequencing as it's flexible, compact and fast.

- ### Automatic operation compatible

  MGI provides automatic solutions for experiment process, which can significantly save labor cost and improve efficiency.

Background

16S rRNA is the RNA component of prokaryotic ribosome 30S subunits, which is encoded by the 16S rRNA gene. The gene sequence in this region is highly conserved due to slow evolution. Thus, this region is commonly used in phylogeny studies of different types of bacterial and archaebacterial [1,2]. Bacterial 16S rRNA gene has 9 hypervariable regions (V1-V9). Each region is as long as about 30~100 base pairs that are involved in the secondary structure of small ribosomal subunit. These hypervariable regions can be used for bacterial identification of different species as they are highly specific[3].

In this study, the hypervariable regions could be amplified by designing universal primers in the conserved regions and sequenced for bacterial species identification. Currently, 16S rRNA gene sequencing is normally used for microbial identification, typing and quantitation of complex biological mixture, such as environmental samples and intestinal samples[4,5]. Conventionally, species identification is based on sequencing of 1 or 2 of the 9 hypervariable regions, which can only provide limited information. IDT xGen 16S Amplicon Panel v2 is supported by xGen amplicon technology, which can achieve single-tube amplification of the 9 hypervariable regions in 16S rRNA genetic region. In addition, compared to to only cover V3, V4 region design, IDT' s overlapping "tilling" design can collect more valid information in shorter reads (upper part of Figure 1).

This panel uses 23 pairs of primers and amplifies amplicons with an average length of 425 bp that overlaps all 9 regions (V1-V9) of 16S rRNA gene and sequenced with the PE150 recipe. This solution is suitable for microbial community detection and identification in medical, environmental, agricultural forensic settings (lower part of Figure 1).

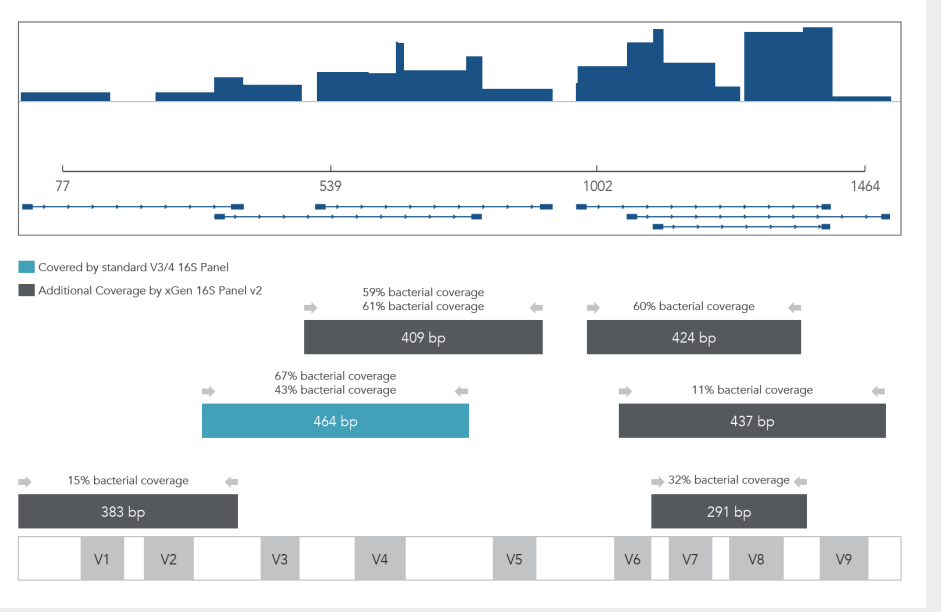

Figure 1. Visualization display of Integrative Genomics Viewer (IGV), IDT xGen 16S Amplicon Panel v2 covers 16S rRNA V1-V9 region (upper); IDT xGen 16S Amplicon Panel v2 uses 6 amplicons , 23 pairs primers compared to the V3 and V4(blue) amplicons design (below).

Study Description

Benefiting from its core DNBSEQ sequencing technology, MGI sequencing platform has the characteristics of high accuracy, low duplication rate and low index hopping rate.The purpose of this study is to evaluate the performance of IDT xGen 16S Amplicon Panel v2 designed for microbial identification on DNBSEQ sequencing platform. Based on the commercially available ZymoBIOMICS™ Microbial Community DNA Standard, the results showed that this DNBSEQ sequencing platform based IDT solution can output stable species identification results even when the input DNA is as low as 10 pg or data size is as low as 100k reads. Moreover, it effectively improves data utilization with the bioinformatic analysis software 16S SNAP APP.

Materials and Methods

DNA Sample Collection

In order to test the performance of IDT xGen 16S Amplicon Panel v2 on DNBSEQ sequencing platform, this study carried out a performance evaluation based on ZymoBIOMICS™ Microbial Community DNA Standard (D6305). This standard was prepared by mixing 8 different microbial DNA with a certain ratio (Table 1), where Avg. GC refers to the average GC value of genomic DNA, Gram strain refers to gram-negative/positive, and 16S gDNA Abun. refers to the abundance of 16S gDNA.

In order to test the performance of this panel, this study serially diluted the standard and selected varied initial input amounts (10 ng, 5 ng, 1 ng and 10 pg) for library prep, sequencing and bioinformatic analysis.

Species	Avg. GC (%)	Gram Stain	16S gDNA Abun. (%)
Pseudomonas aeruginosa	66.2	-	4.2
Escherichia coli	56.8	-	10.1
Salmonella enterica	52.2	-	10.4
Lactobacillus fermentum	52.8	+	18.4
Enterococcus faecalis	37.5	+	9.9
Staphylococcus aureus	32.7	+	15.5
Listeria monocytogenes	38	+	14.1
Bacillus subtilis	43.8	+	17.4
Saccharomyces cerevisiae	38.4	Yeast	NA
Cryptococcus neoformans	48.2	Yeast	NA

Table 1. Microbial components in D6305

Library prep and sequencing

This study was carried out in MGI Customer Experience Center (CEC), Wuhan. Three experimental duplicates (2 at 10 ng) were set for each initial DNA input amount. The variable regions of 16S rRNA gene were enriched through 2 cycles of PCR (cycle NO. is 18 and 7, respectively) with the IDT xGen 16S Amplicon Panel v2. As to detailed operation procedures, please refer to related instruction. Afterwards, the MGIEasy Universal Library Conversion kit (App-A) was used to convert the amplified library to be compatible with DNBSEQ sequencing platform through several cycles of PCR. Eventually, DNA Nanospheres (DNB) were prepared and sequenced on DNBSEQ-G50 platform with paired-end 150bp (PE150) recipe and ensure data size is about 15 M reads for each sample.

Bioinformatic analysis

The generated sequencing reads are analyzed mainly by the IDT 16S SNAP APP, an automated workflow that analyzes microbial community structure based on the variable regions of 16S rRNA in the mode of paired-end sequencing. This workflow begins with primer trimming using

Cutadapt followed by quality filtering, denoising, paired-end merging and chimera removal using DADA2 resulting in amplicon sequence variants (ASVs), ASV read pairs undergo dereplication using VSEARCH to form a smaller, unique sequence set, which is used to query the custom 16S database (RDP11.5, https:/rdp.cme.msu.edu) using BLAST for high identity matches. Unlike other analytics tools, 16S SNAP APP tool can assemble individual reads from different regions of different strains into a longer "template". Using assembled "template" from V1-V9 region to compare the 16S database, can increase the accuracy of test results and the effective utilization of data.

Finally, the sequencing results were classified using RDP Classifier to generate and output a standard format abundance table for subsequent analysis (Figure 2).

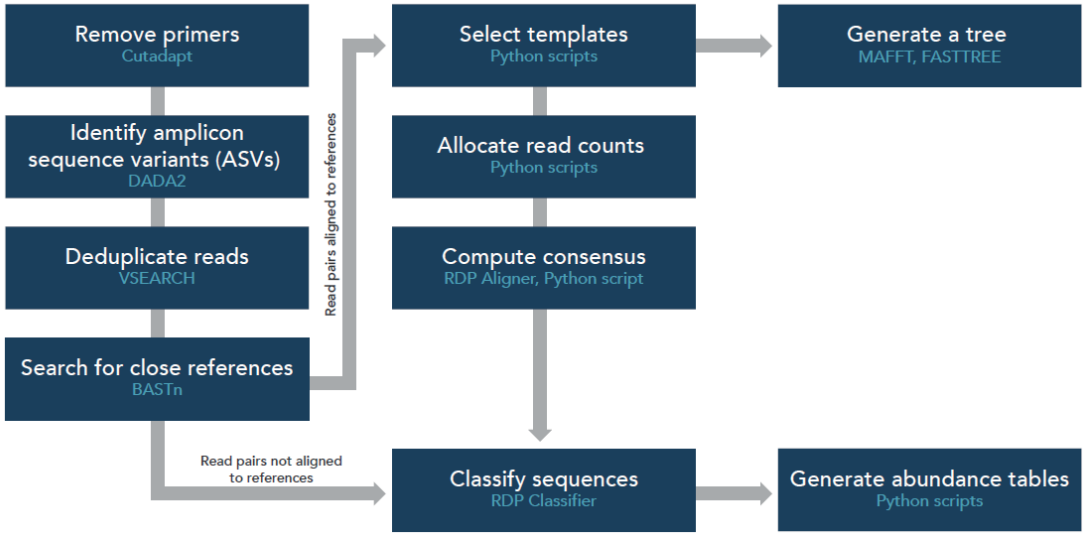

Figure 2. Analytic workflow of 16S SNAP APP

Sample collection	Library prep & Sequencing	Bioinformatics Analysis	Result analysis
ZymoBIOMICS™ Microbial Community DNA Standard（D6305） DNA standard	 IDT xGen 16S Amplicon Panel v2 MGIEasy Universal Library Conversion kit (App-A)  DNBSEQ-G50 genetic sequencer	16S SNAP APP	A comprehensive evaluation of IDT xGen 16S Amplicon Panel v2

Results

High-quality library prep

In this study, IDT xGen 16S Amplicon Panel v2 was used for library preparation with varied initial DNA input amounts (10 pg, 1 ng, 5 ng and 10 ng). The results showed the library concentrations and yields in varied initial DNA input amounts (Figure 3A). According to the gel electrophoresis of Agilent 2100, library obtained at an initial input of 1 ng is highly qualified, with an expected amplicon size and no primer dimer or other non-specific peaks (Figure 3B).

A

Input	UDI#	Library Output Conc. (ng/μL)	Library Output (ng)
10pg	1	0.7	14.2
	2	0.6	11.6
	8	0.7	14.9
1ng	13	52	1040
	21	57	1140
	39	54.4	1088
5ng	44	52	1040
	54	51	1020
	73	39	780
10ng	82	34.2	684
	84	45.2	904
NC	93	0	0

B

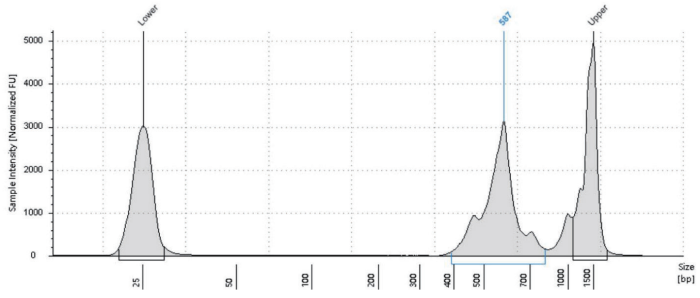

Figure 3. Library prep results and library quality in varied initial sample input amounts. Figure A shows the library concentrations and yields obtained at varied sample input amounts (3 duplicates at 10 pg, 1 ng, 5 ng, and 2 at 10 ng). Figure B shows the Agilent 2100 result of the library when the initial sample input amount was 1 ng.

Stable performance displayed

16S SNAP APP, a IDT self-developed analysis tool, was applied for data analysis after sequencing. The measured data show that, based on IDT xGen 16S Amplicon Panel v2 and DNBSEQ-G50 sequencing platform, the precision and relative species abundance with varied initial sample input amounts were consistent and as expected (Table 2). For example, at an input amount of 10 pg, the relative abundances of *Pseudomonas aeruginosa*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus fermentum*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Listeria monocytogenes* and *Bacillus subtilis* were 6.6%±0.1%, 12.9%±0.8%, 9.5%±0.9%, 8.1%±0.2%, 11.6%±0.2%, 17.4%±0.2%, 15.1%±0.3% and 18.8%±1.0%, respectively, which were consistent with theoretical proportions. Similar results were observed at other input amounts (Table 2).

Data analysis by intercepting different amounts of data (~15M reads and ~100k reads) showed that data size as low as ~100k reads still yielded a satisfactory relative species abundance, which had no significant difference from that in the data size of ~15M reads, and were consistent with theoretical proportions roughly (Figure 4).

Species composition	Theoretical proportion	10pg (AVE±SEM)	1ng (AVE±SEM)	5ng (AVE±SEM)	10ng (AVE±SEM)
Pseudomonas aeruginosa	4.2%	6.6%±0.1%	6.7%±0.4%	7.2%±0.1%	7.2%±0.1%
Escherichia coli	10.1%	12.9%±0.8%	12.5%±0.8%	13.2%±0.9%	12.4%±0.9%
Salmonella enterica	10.4%	9.5%±0.9%	10.3%±0.8%	11.4%±0.6%	11%±0.4%
Lactobacillus fermentum	18.4%	8.1%±0.2%	8.1%±0.6%	8.2%±0.5%	9.8%±1.0%
Enterococcus faecalis	9.9%	11.6%±0.2%	11.4%±0.3%	10.8%±0.1%	10.9%±0.6%
Staphylococcus aureus	15.5%	17.4%±0.2%	16.3%±0.4%	16.8%±0.5%	15.6%±0.1%
Listeria monocytogenes	14.1%	15.1%±0.3%	15%±0.8%	13.3%±0.1%	14.1%±1.0%
Bacillus subtilis	17.4%	18.8%±1.0%	19.6%±0.5%	19.1%±0.5%	18.9%±1.1%

Table 2. Test results with different sample input amount

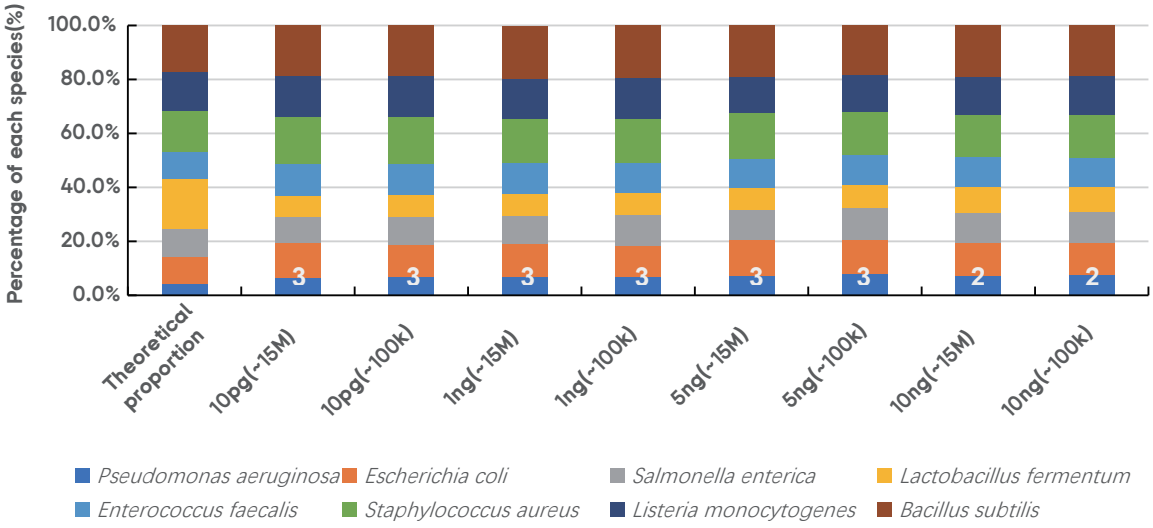

Figure 4. Species abundance calculation with different size of sequencing data. In the data size of both ~ 15M reads and ~100k reads, abundances in varied initial input amounts (10 pg, 1 ng, 5 ng, 10 ng) showed no significant difference and approximated theoretical values. "3" and "2" in the figure stand for 3 and 2 duplicates, respectively.

Effective data utilization rate

Unlike conventional analysis tools for 16S rRNA , IDT 16S SNAP APP software analysis can associate each independent sequencing reads to produce longer sequences for subsequent species classification, thus enhancing both typing precision and effective data utilization rate. Based on IDT xGen 16S Amplicon Panel v2, DNBSEQ-G50 platform and IDT 16S SNAP APP analytical process, this study found that effective data utilization rates are 99.90%, 99.91%, 99.61% and 99.39% respectively in the varied initial input amounts (10 pg, 1 ng, 5 ng and 10 ng), which were all close to 100% (Figure 5).

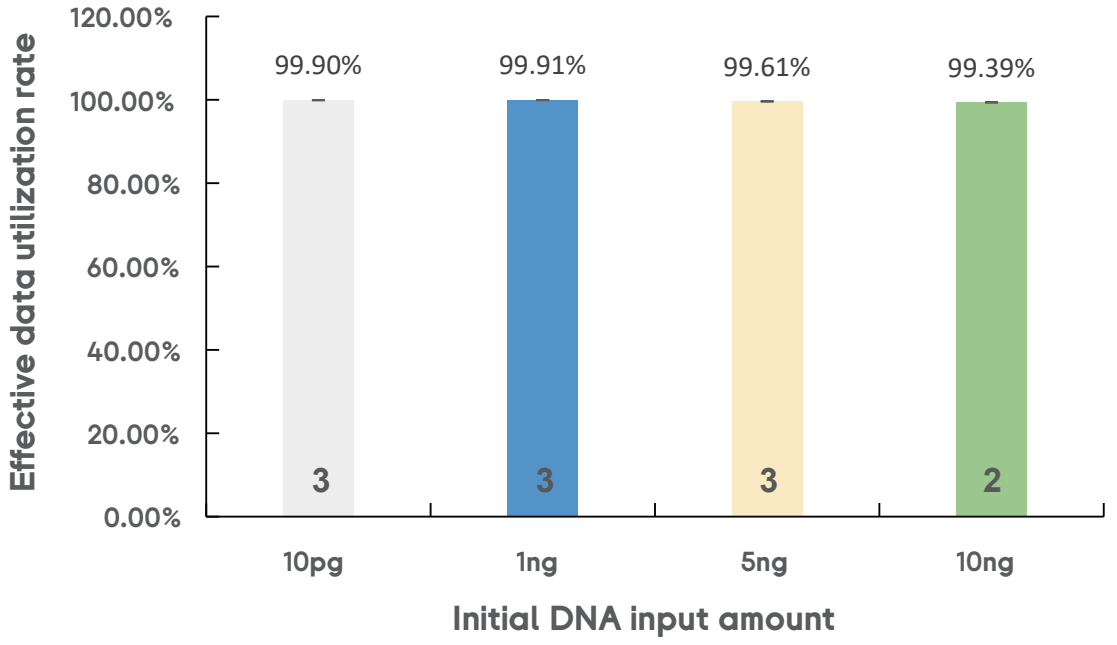

Figure 5. Effective data utilization rates with 16S SNAP APP. Abscissa indicates initial input amount and ordinate indicates effective data utilization rate. 3 duplicates at 10 pg, 1 ng, 5 ng, and 2 at 10 ng.

Summary

The evaluating result of IDT panel on DNBSEQ-G50 sequencer indicated that xGen 16S Amplicon Panel v2 is compatible with MGI's DNBSEQ-G50 platform, which was able to yield stable species testing results in varied initial input amounts and data size. Moreover, using 16S SNAP APP can further efficiently improve data utilization rate.

DNBSEQ-G50 platform is based on DNBSEQ™ proprietary technology, making it capable of yielding high-quality sequencing data. Large and small double slide design meets flexible sequencing demands. The large slides with 500M reads capacity can greatly reduce the sequencing data size and sequencing cost in the application scenario of high depth sequencing. For 100k reads per sample, small slides with 100M reads can theoretically satisfied 16S rDNA sequencing of 1000 samples. DNBSEQ-G50 platform is known for its flexibility, compact size and speed, so has become a classic model of microbial sequencing.

DNBSEQ-G50

Reference

1. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088-5090, doi:10.1073/pnas.74.11.5088 (1977).
2. Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**, 697-703, doi:10.1128/jb.173.2.697-703.1991 (1991).
3. Pereira, F. *et al.* Identification of species by multiplex analysis of variable-length sequences. *Nucleic Acids Res* **38**,e203,doi:10.1093/nar/gkq865 (2010).
4. Schmidt, T. M., DeLong, E. F. & Pace, N. R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* **173**, 4371-4378, doi:10.1128/jb.173.14.4371- 4378.1991 (1991).
5. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* **102**, 11070-11075, doi:10.1073/pnas.0504978102 (2005).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-G50RS	900-000353-00
	MGISP-100RS Automated Sample Preparation System	900-000206-00
	MGISP-960RS Automated Sample Preparation System	900-000146-00
Software	MegaBOLT Bioinformatics analysis accelerator	900-000555-00
	16S SNAP APP	https://github.com/swiftbiosciences/16S-SNAPP-py3
Library Prep	IDT xGen™ Amplicon Core (96RXN)	10009827
	IDT xGen™ 16S Amplicon Panel v2 (96RXN)	10009828
	IDT xGen™ Amplicon UDI Primers Set 1	10009846
	MGIEasy Universal Library Conversion kit (App-A) (16RXN)	1000004155
Sequencing Reagents	DNBSEQ-G50RS High-throughput Sequencing Set (FCL PE150)*	1000019858
	High-Throughput Pair-End Sequencing Primer Kit (App-A)	1000020832
	High-Throughput Barcode Primer 3 Reagent Kit (App-A)	1000014047

MGI Tech Co.,Ltd

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

The copyright of this brochure is solely owned by MGI Tech Co. Ltd.. The information included in this brochure or part of, including but not limited to interior design, cover design and icons, is strictly forbidden to be reproduced or transmitted in any form, by any means (e.g. electronic, photocopying, recording, translating or otherwise) without the prior written permission by MGI Tech Co., Ltd.. All the trademarks or icons in the brochure are the intellectual property of MGI Tech Co., Ltd. and their respective producers.

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Author：Qiwei Wang, Hanfei Zhang

Editor-in-Charge：Qiwei Wang

Reviewer：Yao Jiang

*1. For StandardMPS and CoolMPS: Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, Spain, UK, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland, Portugal, Austria and Romania. Unless otherwise informed, StandardMPS sequencing reagents, and sequencers for use with such reagents are not available in Hong Kong. No purchase orders for StandardMPS products will be accepted in the USA until after January 1, 2023.

2. For HotMPS sequencers: This sequencer is only available in selected countries, and its software has been specially configured to be used in conjunction with MGI's HotMPS sequencing reagents exclusively.

3. For HotMPS reagents: This sequencing reagent is only available in selected countries.