

Evaluation of MGI's MTB-Explorer Software using Whole-Genome Sequencing Data of Isolated Pure Culture of *Mycobacterium tuberculosis* on DNBSEQ-G99RS Platform

■ Highlights

Ultra-Fast Workflow

The workflow is combined with MGI's ultra-fast sequencer DNBSEQ-G99ARS, and thus report can be released within 24hrs including extraction of gDNA, library preparation using MGIEasy Fast FS DNA Library Prep Set followed by sequencing, and analysis by MTB-Explorer Software(hereinafter called MTBDR).

Excellent data quality

MGI's proprietary DNBSEQ™ sequencing technology generates high-quality sequencing data for downstream analysis. MGI's proprietary DNBSEQ™ technology is based on DNA nanoballs and works by generating multiple copies of the same circular template using rolling circle amplification. Further, DNA nanoballs are generated in a tube out of the flow cells, before loading into patterned flow cells, which prevents issues with optical duplicates or ExAmp duplicates observed in other sequencing platform utilizing cluster generation approach inside the flow cells.

High degree of automation of library construction and data analysis

Combined with MGI's liquid handling systems such as MGISP-100RS automated library preparation system, library construction can be fully automated library preparation. Following sequencing on DNBSEQ-G99ARS the data analysis and reporting is also fully automated by MTB-Explorer Software that is integrated into bioinformatics module of DNBSEQ-G99ARS platform.

Complete analysis function

The MGI's proprietary developed software, **MTB-Explorer Software**, cater fully automated data analysis and reporting functions including genome assembly, typing identification, drug resistance prediction, and evolutionary traceability analysis of individual strains of *Mycobacterium tuberculosis*.

■ Introduction

Tuberculosis (TB) is an infectious disease that causes morbidity and mortality globally, especially in poor resource settings. Inappropriate use of antibiotics in treatment of drug susceptible TB patients, sub-optimal treatment regimens and failure to complete treatment in drug susceptible TB patients leads to drug resistance. Hence, drug-resistant tuberculosis (TB) is an emerging health problem and becoming a very serious obstacle for global TB control programmes. TB patients with drug resistance may be induced by exposure to multidrug- and extensively drug-resistant tuberculosis (MDR/XDR-TB) strains. Multidrug-resistant TB (MDR-TB), caused by *Mycobacterium tuberculosis* complex strains (MTBC) that are resistant to at least isoniazid and rifampicin, has become a great threat in many parts of the world [1-3]. Controlling the high prevalence of drug-resistant TB largely depends on a timely laboratory diagnosis with novel methods such as NGS, because traditional TB drug susceptibility testing (DST) relies on solid or liquid culture, which may take weeks or months to yield results. Few molecular testings for predicting single common anti-TB drugs are being used globally, but sensitivity and specificity reported for predicting multi-drug resistance is very low. Hence, Whole-genome sequencing (WGS), as a molecular diagnostic tool, has been greatly developed in TB research for the prediction of MDR and XDR with high sensitivity and specificity. Based on the independently developed reagents, automated sample preparation system.

To cater the ultra-fast WGS and reporting of Multi-Drug Resistance Prediction of TB specimens, MGI has developed a software, MTB-Explorer Software, that is integrated into bioinformatics module of DNBSEQ-G99ARS platform. Thus, the total workflow starting from extraction of gDNA, automated library construction by MGISP-100RS, sequencing by ultra-fast DNBSEQ-G99ARS and fully automated data analysis and reporting functions including genome assembly, typing identification, drug resistance prediction, and evolutionary traceability analysis of individual strains of *Mycobacterium tuberculosis* could be done within 24hrs.

Hence, this experiment has been designed to evaluate the performance of DNBSEQ-G99ARS sequencing platform for rapid and accurate sequencing, and fully automated data analysis and reporting by integrated MTB-Explorer Software for the precise identification, monitoring and early warning, traceability analysis, and precise prevention and control of tuberculosis infection.

■ Methods

The detailed depiction of the library preparation with MGIEasy Fast FS DNA Library Prep Set to prepare ready-to-sequence libraries using the Isolated Pure Culture of *Mycobacterium tuberculosis*, starting from sample to sequencing and data analysis with reporting can be found in Figure 1.

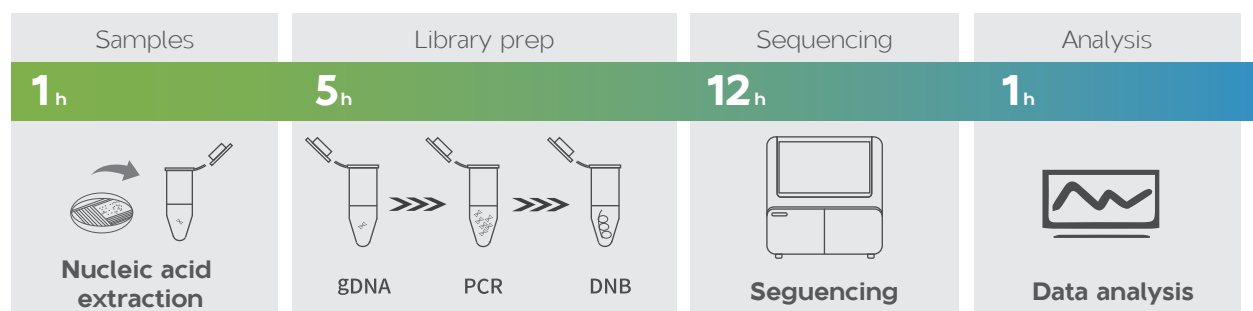


Figure 1. Workflow of the Library Preparation, sequencing and automated data analysis and reporting by DNBSEQ-G99ARS Platform.

■ Workflow

Sample information

Sixteen of pure *Mycobacterium tuberculosis* culture samples with rifampicin resistance from National Drug Reference Standards have been used in this evaluation. (The information of the reference strains can be found at <http://aoc.nifdc.org.cn/sell/home/search.html>).

Whole-Genome Library Preparation

The ready to sequence DNA libraries of the sixteen samples were prepared using MGIEasy Fast FS DNA Library Prep Set according to the manufacturer's instructions. The quality of gDNA was determined spectrophotometrically using a Nanodrop One (Thermo Fisher Scientific, Waltham MA, USA). All libraries were quantified with the Qubit dsDNA BR Assay Kit using Qubit 3.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). library size was analyzed by Agilent 2100 Bioanalyzer Instrument.

Sequencing and data QC

Final libraries were sequenced in Paired-end 2x150 bp mode on the DNBSEQ-G99ARS (Figure 1) system using DNBSEQ-G99RS High-throughput Sequencing Set (G99 SM FCL PE150). DNBSEQ-G99ARS is one of the fastest models among small and medium throughput sequencers worldwide, capable of completing PE150 sequencing within 12 hours. It is equipped with a built-in computing module to achieve data analysis integration. The data output targeted per sample is 5M reads and recommended sequencing parameters of MTB WGS on DNBSEQ-G99RS is mentioned in Table 1.

Table 1. Recommended sequencing parameters of MTb WGS on DNBSEQ-G99RS

Intended use	Recommended read length	Single sample reads	Throughput/FC
Whole-genome sequencing of <i>Mycobacterium tuberculosis</i>	PE150	5 M	16 samples

The total data output was 102.15M reads and $\geq 92.55\%$ of bases were having $\geq Q30$ score indicating that the data output and sequencing data quality were excellent.

Overview of analysis results

Following the sequencing, demultiplexing of the cal files was done by instrument control software of the DNBSEQ-G99ARS. Further data analysis including data QC/genome assembly and prediction of drug resistance etc. was performed by MTB-Explorer Software. The MTB-Explorer Software is an analysis software specifically designed for the analysis of data of *Mycobacterium tuberculosis*. It is integrated into bioinformatics module of DNBSEQ-G99ARS platform and can perform automated analysis functions such as genome assembly, type and spoligotype identification, drug resistance prediction, systematic evolutionary analysis of *M. tuberculosis*, and automatically generate analysis results/reports.

MTB-Explorer Software automatically initiated the analysis of the raw sequencing data using the fastq files generated by instrument control software of the DNBSEQ-G99ARS instrument control software of the DNBSEQ-G99ARS and generated a report after the completion of all analysis functions mentioned above. Rifampicin (RIF) resistance were identified in all sixteen samples included in this evaluation.

Here the results of Sample 1 were demonstrated as an example. The Rifampicin (RIF) and isoniazid (INH) resistances were identified in Sample 1 and the associated mutation information were shown in Table 2.

Table 2. Drug resistance and mutation information associated with resistance in Sample 1 (Strain number: CMCC 94002):

Variants associated with resistance in Sample 1 (Strain number: CMCC 94002)													
Variant (common_name)	Gene	Genome Position	Ref nt	Alt nt	Codon change	Amino Acid change	Number of reads for the position (coverage)	Variant frequency (%)	Variant Depth	Drug	WHO Confidence	Reference	
rpoB_S450L	rpoB	761155	C	T	c.1349C>T	p.Ser 450Leu	147	100	147	RIF	Assoc w R	WHO 2021	
rpsL_K43R	rpsL	781687	A	G	c.128A>G	p.Lys 43Arg	148	100	148	STM	Assoc w R	WHO 2021	
inhA_c-777t (fabG1_c-15t)	fabG1	1673425	C	T	c.-15C>T	NA	153	100	153	ETH	Assoc w R	WHO 2021	
inhA_c-777t (fabG1_c-15t)	fabG1	1673425	C	T	c.-15C>T	NA	153	100	153	INH	Assoc w R	WHO 2021	
pncA_V130G	pncA	2288853	A	C	c.389T>G	p.Val 130Gly	147	100	147	PZA	Assoc w R	WHO 2021	
embB_M306V	embB	4247429	A	G	c.916A>G	p.Met 306Val	123	100	123	EMB	Assoc w R	WHO 2021	

The report overview was presented in a graphical format, as shown in Figure 2. The title at the top represents the name and lineage of the sample. The colored outer ring represents the number of variations related to anti-tuberculosis drug resistance of 13 WHO endorsed anti-tuberculosis drugs that are included in the WHO catalogue database: The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance [4]. The gray inner ring represents the number of corresponding drug resistance gene variations detected in sample 1. Sample 1 was predicted to be resistant to rifampicin and isoniazid, and a total of 5 variations related to drug resistance were detected.



Spoligotyping is currently one of the most frequently used approaches for studying the phylogeography of *Mycobacterium tuberculosis* Complex. Given the binary format of the data, the spoligotyping results can easily be interpreted by MTB-Explorer Software. Spoligotyping was performed to identify the lineages of the sixteen isolates used in this study. Taking sample 1 as an example, the serotype and spoligotype of the sample were identified by typing and these results were shown in Figure 3. The clade identified is from East-Asian which is accurately predicted.

Figure 2. Typing identification results demonstrating that the sample 1 lineage is from East-Asian (Beijing) clade.

Mutation and drug resistance prediction

Based on the WHO catalogue database (The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance [4]) released by WHO in 2021, the drug-resistant mutations of the sixteen samples were annotated and classified into two categories: "Variants associated with resistance" and "Variants not associated with resistance or uncertain significance", realizing the drug resistance prediction. Taking sample 1 as an example, the drug-resistant associated mutations were shown in Table 2. Here the results of are shown in Figure 4 and Table 3.

5. Summary of variants												
Total number of variants	Number of mutations in WHO catalogue		Number of drug resistance associated variants		Number of drug resistance unassociated and uncertain significance variants		Number of other mutations					
1719	26		5		21		1693					

5.1 Variants associated with resistance												
Variant (common_name)	Gene	Genome Position	Ref nt	Alt nt	Codon change	Amino Acid change	Number of reads for the position (coverage)	Variant frequency (%)	Variant Depth	Drug	WHO Confidence	Reference
rpoB_S450L	rpoB	761155	C	T	c.1349C>T	p.Ser450...	147	100.00	147	RIF	Assoc w R	WHO 2021
rpsL_K43R	rpsL	781687	A	G	c.128A>G	p.Lys43Arg	148	100.00	148	STM	Assoc w R	WHO 2021
inhA_c-77...	fabG1	1673425	C	T	c.-15C>T	NA	153	100.00	153	ETH	Assoc w R	WHO 2021
inhA_c-77...	fabG1	1673425	C	T	c.-15C>T	NA	153	100.00	153	INH	Assoc w R	WHO 2021
pncA_V130G	pncA	2288853	A	C	c.389T>G	p.Val130...	147	100.00	147	PZA	Assoc w R	WHO 2021
embB_M3...	embB	4247429	A	G	c.916A>G	p.Met130...	123	100.00	123	EMB	Assoc w R	WHO 2021

5.2 Variants not associated with resistance or uncertain significance												
Variant (common_name)	Gene	Genome Position	Ref nt	Alt nt	Codon change	Amino Acid change	Number of reads for the position (coverage)	Variant frequency (%)	Variant Depth	Drug	WHO Confidence	Reference
gyrA_E21Q	gyrA	7362	G	C	c.61G>C	p.Glu21Gln	154	100.00	154	LEV	Not ass...	WHO 2021
gyrA_E21Q	gyrA	7362	G	C	c.61G>C	p.Glu21Gln	154	100.00	154	MXF	Not ass...	WHO 2021
gyrA_S95T	gyrA	7585	G	C	c.284G>C	p.Ser95Thr	131	100.00	131	LEV	Not ass...	WHO 2021
gyrA_S95T	gyrA	7585	G	C	c.284G>C	p.Ser95Thr	131	100.00	131	MXF	Not ass...	WHO 2021
gyrA_G668D	gyrA	9304	G	A	c.2003G>A	p.Gly668...	137	100.00	137	LEV	Not ass...	WHO 2021
gyrA_G668D	gyrA	9304	G	A	c.2003G>A	p.Gly668...	137	100.00	137	MXF	Not ass...	WHO 2021
mshA_A1...	mshA	575907	C	T	c.560C>T	p.Ala187Val	130	100.00	130	ETH	Uncertal...	WHO 2021

Figure 4. Screenshot of Drug resistance prediction results of Sample 1: Overall mutations identified, Variants associated with resistance and Variants not associated with resistance or uncertain significance.

Table 3. Drug resistance and mutation information not associated with resistance or uncertain significance in Sample 1 (Strain number: CMCC 94002):

Variants not associated with resistance or uncertain significance in Sample 1 (Strain number: CMCC 94002)												
Variant (common_name)	Gene	Genome Position	Ref nt	Alt nt	Codon change	Amino Acid change	Number of reads for the position (coverage)	Variant frequency (%)	Variant Depth	Drug	WHO Confidence	Reference
gy-rA_E21Q	gyrA	7362	G	C	c.61G>C	p.-Glu21Gln	154	100	154	LEV	Not associated with R	WHO 2021
gy-rA_E21Q	gyrA	7362	G	C	c.61G>C	p.-Glu21Gln	154	100	154	MXF	Not associated with R	WHO 2021
gy-rA_S95T	gyrA	7585	G	C	c.284G>C	p.Ser95Thr	131	100	131	LEV	Not associated with R	WHO 2021
gy-rA_S95T	gyrA	7585	G	C	c.284G>C	p.Ser95Thr	131	100	131	MXF	Not associated with R	WHO 2021
gy-rA_G668D	gyrA	9304	G	A	c.2003G>A	p.Gly668Asp	137	100	137	LEV	Not associated with R	WHO 2021
gy-rA_G668D	gyrA	9304	G	A	c.2003G>A	p.Gly668Asp	137	100	137	MXF	Not associated with R	WHO 2021
mshA_A187V	mshA	575907	C	T	c.560C>T	p.Ala187Val	130	100	130	ETH	Uncertain significance	WHO 2021
mshA_A187V	mshA	575907	C	T	c.560C>T	p.Ala187Val	130	100	130	INH	Not associated with R	WHO 2021
ccsA_I245M	ccsA	620625	A	G	c.735A>G	p.Ile245Met	154	100	154	AMI	Not associated with R	WHO 2021
ccsA_I245M	ccsA	620625	A	G	c.735A>G	p.Ile245Met	154	100	154	CAP	Not associated with R	WHO 2021
mmpL5_I948V	mmpL5	775639	T	C	c.2842A>G	p.Ile948Val	151	100	151	BDQ	Not associated with R	WHO 2021

Variant (common_name)	Gene	Genome Position	Ref nt	Alt nt	Codon change	Amino Acid change	Number of reads for the position (coverage)	Variant frequency (%)	Variant Depth	Drug	WHO Confidence	Reference
mmpL5_I948V	mmpL5	775639	T	C	c.2842 A>G	p.Ile948Val	151	100	151	CFZ	Not associated	WHO 2021
mmpL5_T794I	mmpL5	776100	G	A	c.2381C>T	p.Thr794Ile	130	100	130	BDQ	Not associated	WHO 2021
mmpL5_T794I	mmpL5	776100	G	A	c.2381C>T	p.Thr794Ile	130	100	130	CFZ	Not associated	WHO 2021
mmpL5_D767N	mmpL5	776182	C	T	c.2299 G>A	p.Asp767Asn	132	100	132	BDQ	Not associated	WHO 2021
mmpL5_D767N	mmpL5	776182	C	T	c.2299 G>A	p.Asp767Asn	132	100	132	CFZ	Not associated	WHO 2021
rpsL_t-165c	rpsL	781395	T	C	c.-165T>C	NA	143	100	143	STM	Not associated	WHO 2021
fbiC_P42OL	fbiC	1304189	C	T	c.1259C>T	p.Pro420Leu	155	100	155	DLM	Uncertain significance	WHO 2021
Rv1258c_581_ins_1_t_tg	Rv1258c	1406760	T	TG	c.580_581insC	p.Glu194fs	100	99	99	INH	Not associated	WHO 2021
Rv1258c_581_ins_1_t_tg	Rv1258c	1406760	T	TG	c.580_581insC	p.Glu194fs	100	99	99	PZA	Not associated	WHO 2021
Rv1258c_581_ins_1_t_tg	Rv1258c	1406760	T	TG	c.580_581insC	p.Glu194fs	100	99	99	STM	Not associated	WHO 2021
rrs_c-187t	mcr3	1471659	C	T	n.41C>T	NA	124	100	124	AMI	Not associated	WHO 2021
rrs_c-187t	mcr3	1471659	C	T	n.41C>T	NA	124	100	124	CAP	Not associated	WHO 2021
rrs_c-187t	mcr3	1471659	C	T	n.41C>T	NA	124	100	124	KAN	Not associated	WHO 2021
rrs_c-187t	mcr3	1471659	C	T	n.41C>T	NA	124	100	124	LZD	Not associated	WHO 2021
rrs_c-187t	mcr3	1471659	C	T	n.41C>T	NA	124	100	124	STM	Not associated	WHO 2021
katG_Y597H	katG	2154323	A	G	c.1789T>C	p.Tyr597His	135	100	135	INH	Uncertain significance	WHO 2021
katG_N596S	katG	2154325	T	C	c.1787A>G	p.Asn596Ser	130	100	130	INH	Uncertain significance	WHO 2021

Variant (common_name)	Gene	Genome Position	Ref nt	Alt nt	Codon change	Amino Acid change	Number of reads for the position (coverage)	Variant frequency (%)	Variant Depth	Drug	WHO Confidence	Reference
kat_G_R463L	katG	2154724	C	A	c.1388G>T	p.Arg463Leu	146	100	146	INH	Not associated	WHO 2021
PPE35_L896S	PPE35	2167926	A	G	c.2687T>C	p.Leu896Ser	151	99.34	150	PZA	Not associated	WHO 2021
Rv1979c_a-129g	Rv1979c	2223293	T	C	c.-129A>G	NA	175	100	175	BDQ	Not associated	WHO 2021
Rv1979c_a-129g	Rv1979c	2223293	T	C	c.-129A>G	NA	175	100	175	CFZ	Not associated	WHO 2021
Rv3236c_T102A	Rv3236c	3612813	T	C	c.304A>G	p.Thr102Ala	149	100	149	PZA	Not associated	WHO 2021
aft_B_D397G	aftB	4267647	T	C	c.1190A>G	p.Asp397Gly	125	100	125	AMI	Not associated	WHO 2021
aft_B_D397G	aftB	4267647	T	C	c.1190A>G	p.Asp397Gly	125	100	125	CAP	Not associated	WHO 2021
whiB6_-74_del_1_g_c_g	whiB6	4338595	GC	G	c.-75delG	NA	127	100	127	AMI	Not associated	WHO 2021
whiB6_-74_del_1_g_c_g	whiB6	4338595	GC	G	c.-75delG	NA	127	100	127	CAP	Not associated	WHO 2021
whiB6_-74_del_1_g_c_g	whiB6	4338595	GC	G	c.-75delG	NA	127	100	127	STM	Not associated	WHO 2021
gid_E92D	gid	4407927	T	G	c.276A>C	p.Glu92Asp	122	100	122	STM	Not associated	WHO 2021

Information on mutations identified in sample 1, but not included in the WHO catalogue was displayed in the "other mutations" results, as shown in Figure 5.

5.3 Other mutations								
Gene	Genome Position	Ref nt	Alt nt	Codon change	Amino Acid change	Number of reads for the position (coverage)	Variant frequency (%)	Variant Depth
dhxA	1456	G	A	c.1456G>A	p.Glu486Lys	169	100.00	169
NA	1849	C	A	c.-203C>A	NA	142	100.00	142
NA	1977	A	G	c.-75A>G	NA	137	100.00	137
recF	4013	T	C	c.734T>C	p.Ile245Thr	132	100.00	132
NA	11820	C	G	c.-648C>G	NA	133	100.00	133
Rv0008c	11879	A	G	c.433T>C	p.Ser145Pro	153	99.35	152

Figure 5. Screenshot of Other mutations results of Sample 1.

Summary

MGI's MTB-Explorer Software portfolio based on the sequencing of sixteen reference standard sample using WGS method on DNBSEQ-G99ARS platform was demonstrated as an accurate predictive software with the characteristics of fast whole process together with excellent data quality of DNBSEQ-G99ARS. complete analysis functions for the precise identification, monitoring and early warning, traceability analysis, and precise prevention and control of tuberculosis infection. Thus, MTB-Explorer Software is demonstrated to provide a powerful tool for the identification of individual strains of Mycobacterium tuberculosis to diagnostic companies and genomic researchers as well.

Ordering informatio

Products	Specifications	Item No.
Device		
MGISP-100RS Automated Sample Preparation System	Standard configuration	900-000070-00
DNBSEQ-G99ARS Genetic Sequencers	High configuration	900-000560-00
Kits		
MGEasy Fast FS DNA Library Prep Set	96 RXN	940-000027-00
MGEasy Fast FS DNA Library Prep Set	16 RXN	940-000029-00
DNBSEQ One-step DNB preparation kit (OS_DB)	4 RXN	1000026466
DNBSEQ-G99RS High-throughput Sequencing Set (G99 SM FCL PE150)	FCL PE150	940-000410-00
Software		
MTB-Explorer Software	/	970-000385-00

References

1. Kerubo G et al. (2016) Drug susceptibility profiles of pulmonary Mycobacterium tuberculosis isolates from patients in informal urban settlements in Nairobi, Kenya. BMC Infectious Diseases 16: 583.
2. Urassa W, Mugusi F, Villamor E, Msamanga G, Moshiri C, Bosch R, Saathoff E, Fawzi W. Primary antimicrobial resistance among Mycobacterium tuberculosis isolates from HIV seropositive and HIV seronegative patients in Dar es Salaam Tanzania. BMC Res Notes. 2008;1:58. doi: 10.1186/1756-0500-1-58.
3. Wells CD, Cegielski JP, Nelson LJ, Laserson KF, Holtz TH, Finlay A, Castro KG, Weyer K. HIV Infection and Multidrug-Resistant Tuberculosis—The Perfect Storm. J Infect Dis. 2007;196:S86–107. doi: 10.1086/518665.
4. Walker TM, Miotto P, Kser CU, Fowler PW, Knaggs J, Iqbal Z, Hunt M, Chindelevitch L, Farhat M, Cirillo DM, Comas I, Posey J, Omar SV, Peto TE, Suresh A, Uplekar S, Laurent S, Colman RE, Nathanson CM, Zignol M, Walker AS; CRYPTIC Consortium; Seq&Treat Consortium; Crook DW; Ismail N, Rodwell TC. The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance: A genotypic analysis. Lancet Microbe. 2022 Apr;3(4):e265–e273. doi: 10.1016/S2666-5247(21)00301-3. PMID: 35373160; PMCID: PMC7612554.

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone,Yantian District,Shenzhen,China

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

