Journal of Applied Microbiology

Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

A proof-of-concept study of an automated solution for clinical metagenomic next-generation sequencing

Y. Luan[1,2,3], H. Hu[1,2,3], C. Liu[4], B. Chen[4], X. Liu[1,2,3], Y. Xu[1,2,3], X. Luo[1,2,3], J. Chen[4], B. Ye[4], F. Huang[4], J. Wang[4] and C. Duan[1,2,3]

1 Department of Clinical Laboratory, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China
2 Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China
3 RNA Biomedical Institute, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China
4 Matridx Biotechnology Co., Ltd, Hangzhou, China

Keywords
automation, diagnosis of infectious diseases, metagenomic next-generation sequencing, PCR-free library preparation.

Correspondence
Chaohui Duan, Department of Clinical Laboratory, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, 107 YanJiangXi Road, Guangzhou, Guangdong Province, China 510120.
E-mail: 1725012289@qq.com
Jun Wang, Matridx Biotechnology Co., Ltd. Building 5, 208 Zhenzhong Road, Hangzhou, Zhejiang Province, China 310030.
E-mail: wangjun@matridx.com

Yi Luan, Huiling Hu and Chao Liu, contributed equally to this work.

2021/2377: received 15 November 2020, revised 6 January 2021 and accepted 11 January 2021

doi:10.1111/jam.15003

Abstract

Aims: Metagenomic next-generation sequencing (mNGS) has been utilized for diagnosing infectious diseases. It is a culture-free and hypothesis-free nucleic acid test for diagnosing all pathogens with known genomic sequences, including bacteria, fungi, viruses and parasites. While this technique greatly expands the clinical capacity of pathogen detection, it is a second-line choice due to lengthy procedures and microbial contaminations introduced from wet-lab processes. As a result, we aimed to reduce the hands-on time and exogenous contaminations in mNGS.

Methods and Results: We developed a device (NGSmaster) that automates the wet-lab workflow, including nucleic acid extraction, PCR-free library preparation and purification. It shortens the sample-to-results time to 16 and 18·5 h for DNA and RNA sequencing respectively. We used it to test cultured bacteria for validation of the workflow and bioinformatic pipeline. We also compared PCR-free with PCR-based library prep and discovered no differences in microbial reads. Moreover we analysed results by automation and manual testing and found that automation can significantly reduce microbial contaminations. Finally, we tested artificial and clinical samples and showed mNGS results were concordant with traditional culture.

Conclusion: NGSmaster can fulfil the microbiological diagnostic needs in a variety of sample types.

Significance and Impact of the Study: This study opens up an opportunity of performing in-house mNGS to reduce turnaround time and workload, instead of transferring potentially contagious specimen to a third-party laboratory.

Introduction

Shotgun metagenomic next-generation sequencing (mNGS) has been used in both laboratory research and clinical diagnosis of infectious diseases, such as lower respiratory infection (Young *et al.* 2015; Qi *et al.* 2019), bloodstream infection (Blauwkamp *et al.* 2019), meningitis (Miller *et al.* 2019), and parasitic infections (Pallen 2014). mNGS is a 'culture-free' and 'hypothesis-free' diagnostic technique for a wide array of potential pathogens, including bacteria, fungi, viruses and parasites. It can theoretically detect all micro-organisms with known genomic sequences directly from patient samples (Chiu and Miller 2019). When applied clinically, mNGS has been shown to improve the diagnostic rate of suspected infections and result in actionable antibiotic treatment and patient care (Wilson *et al.* 2014). Based on next-generation sequencing (NGS) and bioinformatic analysis,

mNGS has the potential to identify a large repertoire of micro-organisms for which genomic sequences are available. Therefore, it is especially suitable for the immunocompromised cohort in which polymicrobial and atypical infections are common (Forbes *et al.* 2018). However, despite its vast potential, mNGS has several limitations that may hinder its use as a routine microbiological test: (i) the turnaround time (TAT) is typically longer than 36 h since it is currently performed as a 'send-out' test. Even for in-house testing, the TAT is usually longer than 24 h, depending on sequencing platform and library preparation methods (Miao *et al.* 2018; Gu *et al.* 2019). Specialized PCR laboratory, equipment and trained personnel are required to handle the laboratory procedures; (ii) there are no established standards for the experimental workflow and quality control measures. Clinical laboratories that offer mNGS service generally adopt different criteria; (iii) interpretation of results is often challenging since mNGS not only detects nucleic acids in the sample, but also contaminating nucleic acids from the container, environment, and reagents; (iv) samples collected from nonsterile sites usually harbour a normal flora, and it is difficult to differentiate between colonizing and causal micro-organisms (Chiu and Miller 2019). As a result, expert opinions towards this technique are diversified. While many studies pointed out its value in diagnosing infections, several studies suggested limited utility of mNGS as currently performed in the United States (Kufner *et al.* 2019; Hogan *et al.* 2020). In our opinion, mNGS is not yet ready for implementation as a first-line diagnostic test in a clinical setting. Two major obstacles exist: (i) slow TAT that prevents timely diagnosis and treatment for critically ill patients; (ii) microbial contaminations that obscure the interpretation of results. Although pipetting workstations for NGS experiments are common, however, they generally lack flexibility, which is crucial for mNGS since different types of sample require distinctive handling. Patient specimen at a tertiary hospital is normally collected at different time points during a day. Once a workstation starts processing, new samples cannot be added and therefore must wait for the current run to finish. Moreover two workstations are often needed (one for experiments pre-PCR, the other for procedures post-PCR) to minimize aerosol contamination.

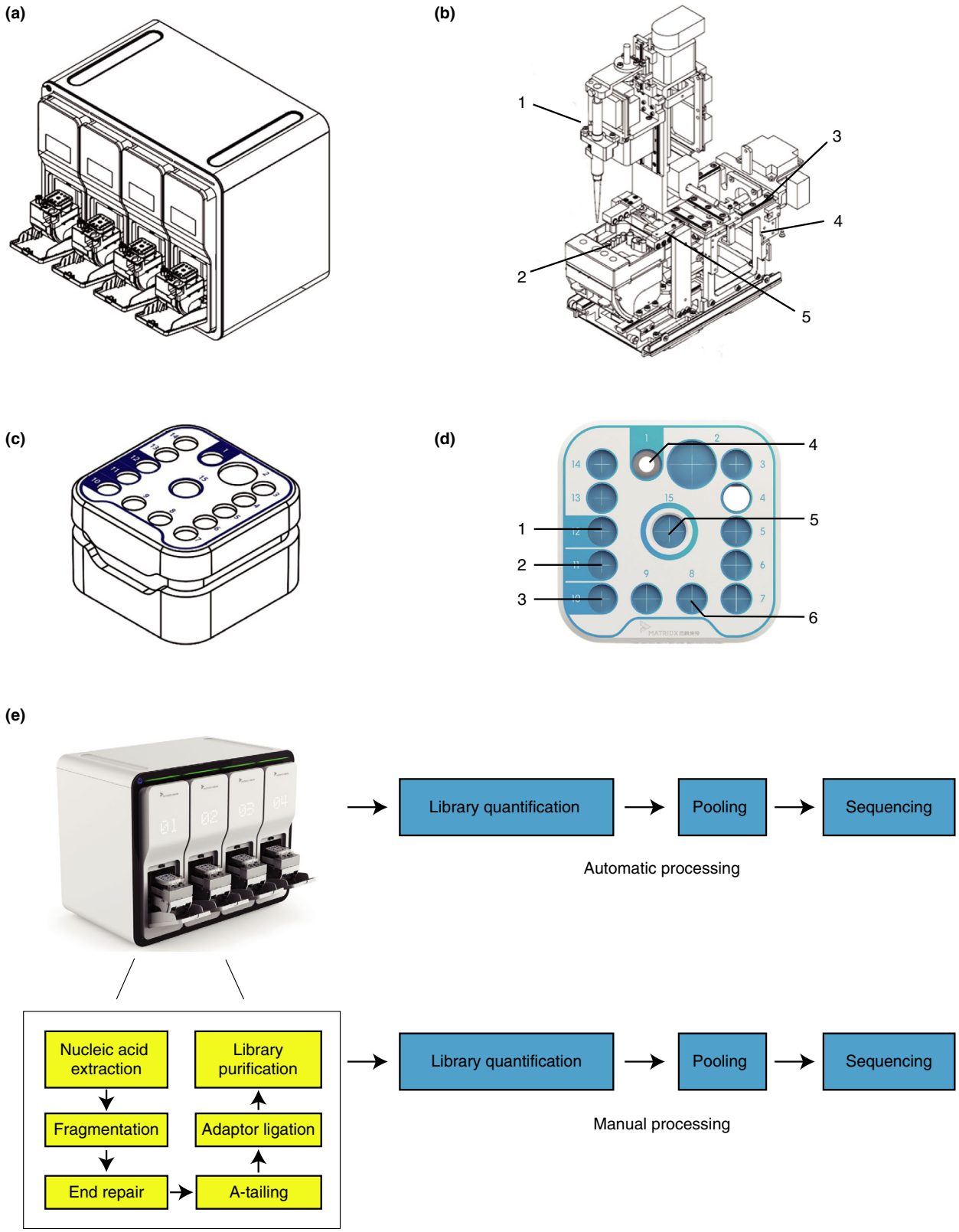
Materials and methods

Design and functionality of NGSmaster

The wet-lab workflow of mNGS consists of sample pretreatment (centrifugation, liquid transfer and handling, host depletion, etc.), nucleic acid extraction, library prep, purification and quantitation, pooling and sequencing. These processes are both time-consuming (approximately 7–9 h of library prep and 11 h of sequencing, depending on methods and sequencing platform; Gu *et al.* 2019) and labour-intensive, which typically require three skilled technicians (one person for sample handling and documentation, two for library prep and sequencing). Contaminating microbial reads derive from a variety of sources including sampling, transportation, reagents, environment and importantly, aerosols since PCR is used for library prep (Seitz *et al.* 2015). In addition, PCR is known to create bias towards sequences of different length and GC content (van Dijk *et al.* 2014). As a result, these issues make it challenging to carry out on-site mNGS in a health care facility.

In this study, we designed a device (NGSmaster) to automate the wet-lab procedures. It has four individual channels, each of which contains liquid handling, temperature control and magnetic separator modules (Fig. 1a,b). NGSmaster can perform cartridge-based PCR-free library preparation (Fig. 1c,d), which can handle up to four samples simultaneously. Each channel is independent and does not interfere with others, which is equipped with liquid handling, temperature control and magnetic separator modules. The liquid handling system consists of sampling pump and pipette tips, high-precision pipetting is achieved by the stepping motor, injector and software control. The temperature module is regulated by PID controller. The mobile guide platform consists of two parts: stepping motor and precision linear guide, which are operated by a single-chip microcomputer that moves the platform in three axes. The magnetic separator system is composed of direct-current geared-down motor and swing arm of a permanent magnet (Fig. 1b). Clinical samples can be directly loaded into the cartridge, which contains three layers: (i) reagents are pre-loaded in the bottom layer; (ii) middle layer is sealed with a thin

Figure 1 Design and functionality of NGSmaster. Schematics showing the exterior of NGSmaster with four channels (a). Each channel is independent and includes liquid handling, temperature control and magnetic separator modules. The designations are: 1. Liquid handling module; 2. Temperature control module; 3. Mobile guide platform; 4. Base support; 5. Magnetic separator (b). Reagents and samples are loaded into a cartridge and can be inserted into NGSmaster for automatic sample processing (c, d) that includes nucleic acid extraction, enzymatic fragmentation, end repair and A-tailing, adaptor ligation and library purification. The designations are: 1. End repair mix; 2. Sequencing adaptor; 3. Ligation reaction mix; 4. Magnetic beads; 5. Proteinase K and clinical sample; 6. Ethanol. (e) The illustration of mNGS workflow when conducted automatically or manually.


(a)
(b)
1
2
3
4
5
(c)
(d)
1
2
3
4
5
6
(e)
Library quantification
Pooling
Sequencing
Automatic processing
Nucleic acid extraction
Library purification
Fragmentation
Adaptor ligation
End repair
A-tailing
Library quantification
Pooling
Sequencing
Manual processing

aluminium film to ensure safety in storage; (iii) labels are printed on the top layer marking different loading wells (Fig. 1d). Once inserted, the bottom of cartridge closely contacts the temperature control module of NGSmaster. The operations are monitored and controlled by onboard software.

Additionally, multiple sets of devices can be connected in parallel, making them adaptable to different throughput needs. Reagents are pre-loaded in cartridges, which can be inserted into the device for processing. It can carry out the wet-lab procedures, including the extraction of nucleic acid, enzymatic fragmentation, end repair, A-tailing, ligation of sequencing adaptors (illumina) and library purification (Fig. 1e). Finished libraries can then be quantified and pooled for sequencing.

PCR-free library preparation and metagenomic sequencing

Whole blood was centrifuged at 1600 ***g*** for 10 min and supernatant was centrifuged at 16 000***g*** for 10 min to obtain plasma. One millilitre of plasma/cerebrospinal fluid (CSF) or 400 μl of bronchoalveolar lavage fluid (BALF) was pipetted into the cartridge. Sequencing library was prepared by reverse transcription (for RNA sequencing only), enzymatic fragmentation (except for plasma DNA sequencing, since cell-free DNA is intrinsically fragmented; Han *et al.* 2020), end repairing, terminal adenylation and adaptor ligation. Libraries were quantified by real-time PCR (KAPA) and pooled. Shotgun sequencing was carried out on illumina Nextseq. Approximately 20 million of 75 bp single-end reads were generated for each library. Bioinformatic analysis was conducted as described in a previous report (Miller *et al.* 2019). Sequences of human origin were filtered (GRCh38.p13) and the remaining reads were aligned to a reference database (NCBI GenBank and in-house curated microbial genomic data) to identify species, reads count and relative abundance. For each sequencing run, a negative control (NC; plasma from healthy donors) was included.

Artificial sample preparation

Cultured HeLa cells were re-suspended in PBS and cell concentrations were measured using a haemocytometer. Artificial CSF samples were prepared by mixing deactivated microbes with cell culture medium containing HeLa cells ($2{\cdot}0 \times 10^4$ cells per ml). The final concentrations of microbes were: Human Adenovirus Group B ($2{\cdot}0 \times 10^4$ copies per ml), *Staphylococcus aureus* ($1{\cdot}0 \times 10^5$ CFU per ml), *Pseudomonas aeruginosa* ($3{\cdot}3 \times 10^3$ CFU per ml) and *Cryptococcus gattii* ($1{\cdot}0 \times 10^4$ CFU per ml). Similarly, artificial BALF samples contained $2{\cdot}5 \times 10^5$ cells per ml of HeLa cells, Human Adenovirus Group B ($2{\cdot}0 \times 10^5$ copies per ml), Influenza A virus ($6{\cdot}0 \times 10^5$ copies per ml), *S. aureus* ($1{\cdot}0 \times 10^6$ CFU per ml), *P. aeruginosa* ($3{\cdot}3 \times 10^4$ CFU per ml) and *C. gattii* ($1{\cdot}0 \times 10^5$ CFU per ml). Artificial plasma samples were prepared by extracting the total DNA of artificial CSF samples, which was then sonicated to ~200 bp fragments and mixed with plasma from healthy donors (final conc. 40 ng ml^{-1}). The NC of each sample type consists of only HeLa cells and culture medium. The NC samples were repeated three times. The artificial BALF samples were repeated five times, while artificial CSF and plasma samples were each repeated six times.

Clinical samples

The use of clinical samples has been approved by the Ethics department at Sun Yat-Sen Memorial Hospital, Guangzhou, China. Plasma samples (~4 ml) were processed from peripheral blood collected from febrile inpatients at Hematology department of the Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health (Streck cfDNA Vacutainer). A total of three plasma samples were donated for research purpose and personal identifying information was omitted.

mNGS reporting criteria

Microbial species identified from a sample were reported if: (i) the results passed quality control filters (library concentration >50 pmol l^{-1}, Q20 > 85%, Q30 > 80%, GC-content < 45%, total reads > 10 million); (ii) NC in the same sequencing run does not contain the species or the $RPM_{NC} : RPM_{sample} < 5$.

Results

Validation of automated workflow in testing cultured bacteria

To evaluate the experimental and bioinformatic pipeline for microbial identification, we used NGSmaster to test 14 cultured bacterial samples. For each culture, three to five colonies were picked by sterile pipette and added into 1 ml of PBS. After vortexing, the solution was loaded into cartridges for DNA library prep. Analysed results are shown in Table 1. An average of 15·5 million reads were generated for each sample and all bacterial species were correctly identified. We used the total reads

Table 1 mNGS testing of cultured bacteria by NGSmaster

		mNGS results			
#	Culture results	Total reads (bp)	Genus and species	RPM	Coverage (%)
1	*Pseudomonas aeruginosa*	18 392 240	*Pseudomonas*	823 320·32	NA
			Pseudomonas aeruginosa	752 887·79	94·85
2	*Acinetobacter baumannii*	16 972 845	*Acinetobacter*	865 861·26	NA
			Acinetobacter baumannii	752 000·27	94·66
3	*Pseudomonas aeruginosa*	15 777 280	*Pseudomonas*	795 957·03	NA
			Pseudomonas aeruginosa	750 674·51	94·83
4	*Pseudomonas aeruginosa*	19 927 433	*Pseudomonas*	837 604·32	NA
			Pseudomonas aeruginosa	771 530·68	94·87
5	*Acinetobacter baumannii*	15 745 974	*Acinetobacter*	850 566·18	NA
			Acinetobacter baumannii	740 337·94	94·65
6	*Staphylococcus aureus*	19 277 195	*Staphylococcus*	814 405·67	NA
			Staphylococcus aureus	794 325·11	95·03
7	*Klebsiella aerogenes*	16 770 107	*Klebsiella*	626 027·07	NA
			Klebsiella aerogenes	607 137·21	94·55
8	*Serratia marcescens*	15 660 037	*Serratia*	661 122·77	NA
			Serratia marcescens	636 613·12	95·38
9	*Enterococcus faecalis*	15 463 713	*Enterococcus*	738 131·26	NA
			Enterococcus faecalis	728 834·27	94·98
10	*Staphylococcus haemolyticus*	9 182 090	*Staphylococcus*	797 665·35	NA
			Staphylococcus haemolyticus	761 609·94	95·57
11	*Enterococcus faecalis*	12 129 072	*Enterococcus*	815 176·71	NA
			Enterococcus faecalis	807276·02	95·84
12	*Staphylococcus haemolyticus*	12 078 171	*Staphylococcus*	711 868·05	NA
			Staphylococcus haemolyticus	669 375·6	95·68
13	*Staphylococcus haemolyticus*	14 991 232	*Staphylococcus*	843 156·19	NA
			Staphylococcus haemolyticus	806 221·46	94·99
14	*Stenotrophomonas maltophilia*	13 963 959	*Stenotrophomonas*	697 831·61	NA
			Stenotrophomonas maltophilia	632 600·32	95·48

mapped to each species to perform sequence assembly and obtained a whole-genome coverage of 95·25 ± 0·46%. These results showed that the automated mNGS could properly identify bacterial species.

Comparison between PCR-free and PCR-based library preparation

PCR is frequently used in the library preparation step for mNGS. However, we elected to utilize a PCR-free method to reduce TAT and possible aerosol contamination. We used cultured *Listeria monocytogenes* and *Enterobacter cloacae* to validate the performance of PCR-free method in comparison to PCR-based approach. Each bacterium was serially diluted in cell culture medium and each concentration was also tested in different levels of host cell background (10^4, 10^5, 10^6 and 10^7 HEK293 cells per ml). Sequencing libraries were generated by NGSmaster (PCR-free) and PCR-based method (TruSeq DNA Sample Preparation Kit, Illumina, Cat# FC-121-2001). We graphed the RPM (reads per million reads) of each species according to its concentrations (CFU per ml in log) (Fig. 2a,b). Using both methods, we discovered a linear correlation between RPM and the number of bacterial cells. As expected, microbial RPM was in a negative correlation with host cell background. In these experiments, PCR-free library prep was consistent with the PCR-based method across all tested concentrations.

Automation alleviates exogenous contamination in mNGS

The workflow of mNGS typically involves multiple purification (nucleic acid purification, library purification, etc.) and liquid transfer steps. When performed manually by a technician, exogenous microbial contaminations are inevitable, including air-borne micro-organisms and ones that dwell on the human skin. To assess whether automation is better than manual operation in reducing microbial contaminations, we compared the plasma results generated by NGSmaster and technicians. We analysed the frequency of occurrence for *Staphylococcus*

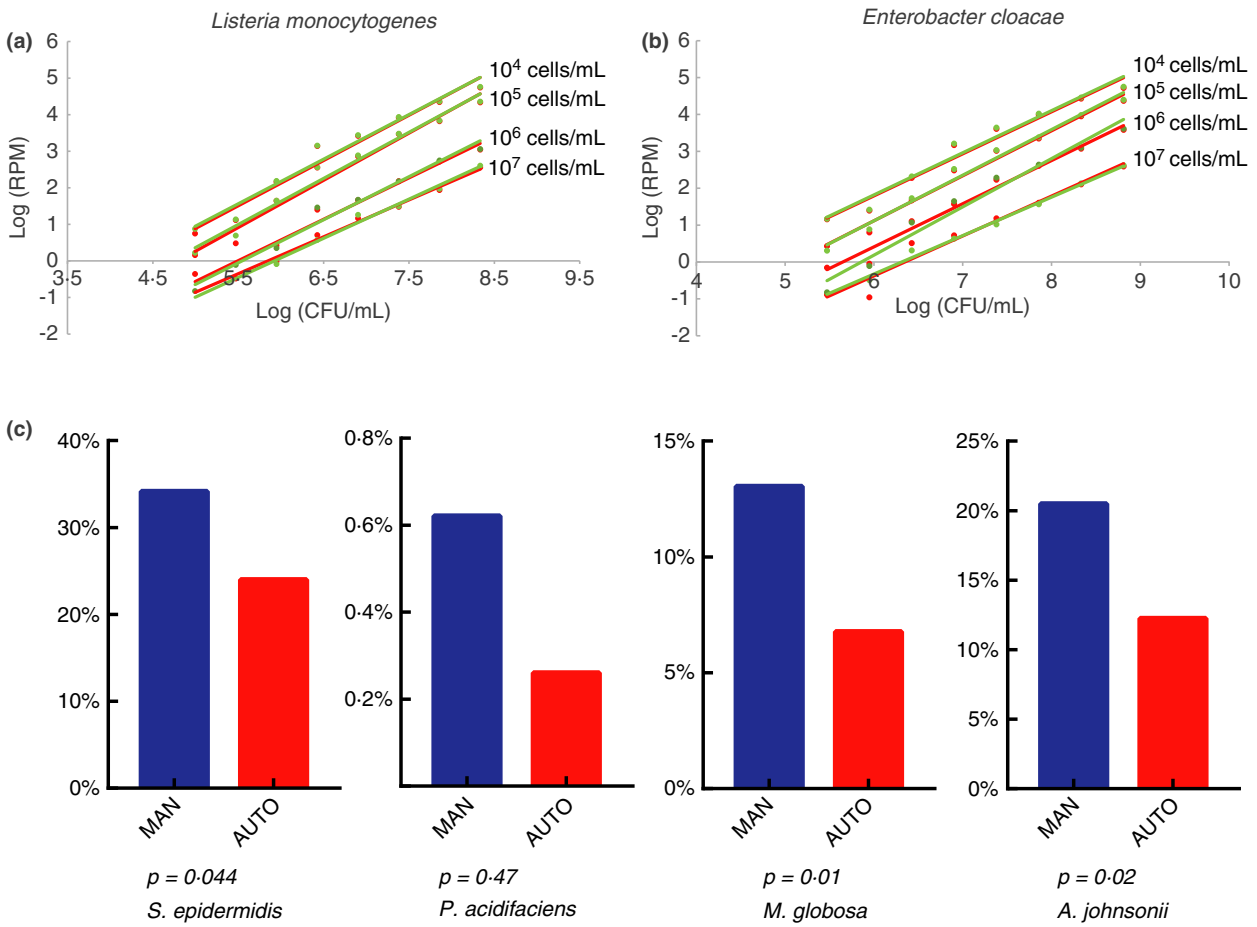


Figure 2 PCR-free automatic library preparation can identify microbial species and reduce contamination. Different dilutions of *Listeria monocytogenes* and *Enterobacter cloacae* were tested using both PCR and PCR-free library prep. Each concentration was also repeated in different levels of human cell background. The resulting microbial RPM (reads per million base) were graphed in (a) and (b). The occurrence rate of four common skin flora *Staphylococcus epidermidis*, *Propionibacterium acidifaciens*, *Malassezia globosa* and *Acinetobacter johnsonii* were compared and graphed (c) between 161 plasma samples that were tested manually (MAN) and 767 plasma samples that went through automatic processing (AUTO). (a) () PCR-based library preparation; (b) () PCR-free library preparation.

epidermidis, *Propionibacterium acidifaciens*, *Malassezia globosa* and *Acinetobacter johnsonii*, all of which were common components of human skin flora. A total of 161 plasma samples were tested manually (Manual group) and 767 were tested by automation (Automatic group). The sequencing and bioinformatic analyses were kept unchanged in both groups. The average reads for the manual group were 24·3 million, compared to 15·7 million for the automatic group (Table S1). In the manual group (Fig. 2c), *S. epidermidis* was detected (RPM > 0) in 34·2% of samples, relative to 24·0% in the automatic group ($P = 0{\cdot}044$, chi-squared test). *Propionibacterium acidifaciens* was detected in 0·62% of samples, relative to 0·26% in the automatic group ($P = 0{\cdot}47$, chi-squared test). *Malassezia globosa* was detected in 13·0% of samples, compared to 6·78% in the automatic group ($P = 0{\cdot}01$, chi-squared test). *Acinetobacter junii* was detected in 20·5% of samples, relative to 12·3% in the automatic group ($P = 0{\cdot}02$, chi-squared test). These results indicated that automation could indeed lower exogenous contaminations when compared to manual operation.

Validation of NGSmaster in pathogen detection

To evaluate the stability and reproducibility of NGSmaster in performing mNGS, we first used artificial samples containing deactivated Human adenovirus group B, influenza virus A, *C. gattii*, *P. aeruginosa* and *S. aureus*. These micro-organisms were chosen to represent disease-causing

DNA virus, RNA virus, fungi and two common bacterial pathogens respectively. We prepared three types of artificial samples: plasma, BALF and CSF (see Methods). Each sample was processed by NGSmaster and repeated for a minimum of three times. RNA libraries were generated for BALF while DNA libraries were prepared for plasma and CSF. An average of 28·8 million reads were produced for each library and the results were listed in Table 2. The coefficient of variation (CV) of microbial RPM were less than 10%, except for plasma, which had CVs of greater than 20%.

Next, we used NGSmaster to run plasma samples collected from febrile patients who were suspected of infections. Three patients had their peripheral blood drawn and sent for microbiological culture and mNGS testing. For these patients, culture identified *S. epidermidis*, *Klebsiella pneumoniae* and *P. aeruginosa* respectively. The corresponding mNGS discovered *S. epidermidis*, *K. pneumoniae*, *P. aeruginosa* and *Human parvovirus B19* respectively (Table 3). mNGS was concordant with culture in these samples and identified additional potential pathogen (Human parvovirus B19) in one patient.

Table 2 mNGS testing of artificial samples containing DNA virus, RNA virus, fungus, and bacteria by NGSmaster

Sample name	Library type	Total reads (bp)	Human adenovirus group B RPM	Influenza virus A RPM	*Cryptococcus gattii* RPM	*Pseudomonas aeruginosa* RPM	*Staphylococcus aureus* RPM
BALF-NC-1	RNA	26 647 566	0	0	0	0	0
BALF-NC-2	RNA	17 718 579	0	0	0	0	0
BALF-NC-3	RNA	21 947 900	0	0	0	0	0
BALF-1	RNA	40 864 828	98·28	49·77	27·63	64·06	17·57
BALF-2	RNA	28 891 156	95·67	53·37	27·14	63·96	14·54
BALF-3	RNA	35 648 400	81·38	47·88	25·78	57·25	14·19
BALF-4	RNA	21 179 707	99·86	57·04	30·69	64·54	16·62
BALF-5	RNA	19 225 934	101·43	43·53	29·08	62·47	15·97
Standard deviation (SD)			8·08	5·17	1·88	3·01	1·42
Arithmetic mean			95·32	50·32	28·06	62·46	15·78
Coefficient of variation (CV) (%)			8·5	10·3	6·7	4·8	9·0
CSF-NC-1	DNA	39 143 338	0	0	0	0·08	0·03
CSF-NC-2	DNA	31 534 190	0	0	0	0	0
CSF-NC-3	DNA	31 456 292	0	0	0	0	0
CSF-1	DNA	36 554 891	19·81	0	22·6	68·94	20·35
CSF-2	DNA	37 716 811	15·75	0	17·82	68·99	17·87
CSF-3	DNA	34 990 727	15·63	0	20·46	59·73	18·32
CSF-4	DNA	33 277 079	15·06	0	22·51	57·07	19·26
CSF-5	DNA	36 618 986	18·51	0	20·4	60·19	18·11
CSF-6	DNA	24 473 764	15·85	0	21·53	50·91	20·06
Standard deviation (SD)			1·92	0·00	1·78	7·02	1·05
Arithmetic mean			16·77	0·00	20·89	60·97	19·00
Coefficient of variation (CV) (%)			11·4	0·0	8·5	11·5	5·5
Plasma-NC-1	DNA	21 202 952	0	0	0	0	0·05
Plasma-NC-2	DNA	37 676 160	0	0	0	0·05	0·03
Plasma-NC-3	DNA	40 987 899	0	0	0	0	0·02
Plasma-1	DNA	22 150 245	31·65	0·05	15·49	39·77	28·13
Plasma-2	DNA	20 609 139	39·98	0	19·65	50·95	16·93
Plasma-3	DNA	13 463 541	32·09	0·07	18·57	30·08	27·56
Plasma-4	DNA	22 166 276	24·45	0·09	14·93	31·76	30·99
Plasma-5	DNA	24 177 783	16·96	0	9·02	30·36	34·33
Plasma-6	DNA	28 375 739	27·73	0	18·75	34·99	28·58
Standard deviation (SD)			7·80	0	3·94	8·04	5·86
Arithmetic mean			28·81	0	16·07	36·32	27·75
Coefficient of variation (CV) (%)			27·1	0·00	24·5	22·1	21·1

Table 3 mNGS testing of clinical plasma samples by NGSmaster

#	Gender	Age	Symptoms	mNGS results	Blood culture	WBC (10^9/l)	Neutrophil (10^9/l)	PCT (ng ml^{-1})	CRP (mg l^{-1})
1	Male	4 years 4 months	Recurrent fever	*Staphylococcus epidermidis*	*Staphylococcus epidermidis*	4·85	1·64	0·05	12·37
2	Male	9 months	Fever, abdominal pain, vomit, hematochezia	*Klebsiella pneumoniae*	*Klebsiella pneumoniae*	18·59	0·26	1·83	229·90
3	Female	11 years 6 months	Fever, abdominal pain, shock	*Human parvovirus B19* *Pseudomonas aeruginosa*	*Pseudomonas aeruginosa*	3·90	3·55	0·18	10·93

Discussion

In-house mNGS for rapid diagnosis of critical infections

The diagnostic value of mNGS has been recognized by both researchers and clinicians. However, this test comprises of laborious experimental steps and require expertise in sample handling. Fast TAT is essential for the diagnosis and treatment of critically ill patients, but the complexity of mNGS made it challenging for a healthcare facility to perform in-house testing. In addition, transporting clinical samples to a third-party laboratory is of biosafety concern, as specimens are likely to be contagious and require special handling such as heat deactivation and specialized packaging. Therefore, in-house mNGS solutions are urgently needed to achieve faster results, lower cost and more reliable quality control. In our study, we used a cartridge-based automation device to replace manual labour. This 'sample-in, library out' solution enables an easier and faster mNGS, shortening the TAT to 16 and 18·5 h for DNA and RNA sequencing respectively. Moreover the size of this device is relatively small (51·8 × 41·8 × 34·3 cm) and can be placed inside a biosafety cabinet, ensuring the safety in handling contagious samples and reduce health risks posed to lab personnel.

PCR-free library preparation prevents aerosol contamination

PCR is commonly used in the library prep for mNGS, including multiplex PCR amplification of targeted nucleic acids as well as the shotgun sequencing approach (Jovel *et al*. 2016). Notably, the sequencing output of mNGS drastically subsamples the original DNA and RNA content in a library (~0·02% of the original nucleic acids; Gu *et al*. 2019). Therefore, any bias introduced through PCR would inevitably modify the original information, such as the relative abundance of microbial species within a sample. Moreover PCR aerosols are problematic since mNGS can detect contaminating nucleic acids due to multiplexing of libraries in a single sequencing run, especially when certain samples with high levels of microbial sequences are processed together with other libraries. To determine whether PCR could introduce microbial contamination, we conducted an experiment by using two samples. We added *Stenotrophomonas maltophilia* into one sample (#2), but not the other (#1). Next, we performed PCR-based library prep and went through mNGS testing. We repeated the experiment twice. In the first sequencing run, adapters A and B were used for sample #1 while adapters C and D were used for sample #2. In the second run, we kept everything the same except that the adapters were swapped for these two samples. As shown in Table S2, the exchange of adapters has led to false positive detection of *S. maltophilia* in sample #1 (RPM of 1·15 and 0·6 respectively), which was most likely due to cross-contamination of PCR aerosols originating from PCR amplification of sample #2. This experiment indicated the possibility of PCR aerosol-mediated contamination in mNGS. Therefore, PCR-free library prep might be a proper choice since no nucleic acid amplification is involved.

Automation can reduce contaminations of human origin

Although sterile technique is applied throughout the experiment, contaminating microbial nucleic acids are difficult to eliminate since they exist virtually everywhere such as the skin, gloves, lab containers and reagents. Microbial sequences originating from the reagents can be filtered by setting up NCs in each sequencing run. However, monitoring exogenous contaminations are more challenging since they could vary considerably. For instance, skin flora are highly diverse from person to person and can change for the same individual depending on age, hygiene, nutritional levels and immune activities (Grice and Segre 2011). Therefore, replacing manual operation with automation could help alleviate exogenous contaminations in mNGS.

In summary, mNGS is of clinical value for the diagnosis of microbial infections in critically ill and immunocompromised patients. Short TAT, ease of operation and cost-effectiveness are important factors to consider before using mNGS as a first-line microbiological test. The 'sample-in, library-out' automation makes mNGS more convenient to perform in-house. More importantly, automation could reduce exogenous microbial contaminations and make the results more reliable and easier to interpret.

Comparison between NGSmaster and other techniques

When operated manually, most current sequencing platforms can complete mNGS testing from sample to results in 24–72 h (average 48 h; Han *et al*. 2019). To date, there has been no development of automatic solutions for mNGS that integrated all wet-lab procedures, though commercial devices for the automation of individual molecular biology applications are available, including nucleic acid extraction, liquid handling and PCR amplification. However, total laboratory automation such as NGSmaster has advantages over existing solutions due to the following: (i) it incorporates main wet-lab procedures including nucleic acid extraction, purification and library preparation. The finished library could be sequenced directly on an illumina sequencer; (ii) it eliminates the need for manually transferring products between different devices; (iii) it further reduces hands-on time such as transferring liquid from one microcentrifuge tube to another (i.e. pipetting extracted nucleic acid to set up a PCR reaction); (iv) it minimizes microbial contaminations since the majority of experimental procedures are carried out in a closed chamber without the need for PCR amplification; (v) the throughput of sample handling is more flexible since each chamber is independent and can initiate experiment whenever a clinical sample arrives at the laboratory, without the need to wait for a certain amount of samples before starting processing.

Acknowledgement

This project was supported by National Key Program of the Ministry of Science and Technology Foundation 2020YFC2004505, Guangdong Natural Science Foundation 2020A1515011467, and by Guangzhou City Science and Technology Project Plan Foundation 202002030149.

Author contributions

JW and CD conceived of the study. YL, HH, CL, BC, XL, YX and XL conducted the experiments, collected and analysed data. JC, BY and FH designed assay protocol and conducted bioinformatic analysis. CL, BC and JW wrote the manuscript with input from all authors. All authors discussed the results and commented on the manuscript.

Conflict of Interest

Chao Liu, Bin Chen, Jing Chen, Baochun Ye, Fei Huang and Jun Wang are employees at Matridx Biotechnology Co., Ltd. NGSmaster is a patented product in China (patent no. CN106404508A, CN106754339A).

References

Blauwkamp, T.A., Thair, S., Rosen, M.J., Blair, L., Lindner, M.S., Vilfan, I.D., Kawli, T., Christians, F.C. *et al.* (2019) Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol* **4**, 663–674.

Chiu, C.Y. and Miller, S.A. (2019) Clinical metagenomics. *Nat Rev Genet* **20**, 341–355.

van Dijk, E.L., Jaszczyszyn, Y. and Thermes, C. (2014) Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res* **322**, 12–20.

Forbes, J.D., Knox, N.C., Peterson, C.L. and Reimer, A.R. (2018) Highlighting clinical metagenomics for enhanced diagnostic decision-making: a step towards wider implementation. *Comput Struct Biotechnol J* **16**, 108–120.

Grice, E.A. and Segre, J.A. (2011) The skin microbiome. *Nat Rev Microbiol* **9**, 244–253.

Gu, W., Miller, S. and Chiu, C.Y. (2019) Clinical metagenomic next-generation sequencing for pathogen detection. *Annu Rev Pathol* **14**, 319–338.

Han, D., Li, R., Shi, J., Tan, P., Zhang, R. and Li, J. (2020) Liquid biopsy for infectious diseases: a focus on microbial cell-free DNA sequencing. *Theranostics* **10**, 5501–5513.

Han, D., Li, Z., Li, R., Tan, P., Zhang, R. and Li, J. (2019) mNGS in clinical microbiology laboratories: on the road to maturity. *Crit Rev Microbiol* **45**, 668–685.

Hogan, C.A., Yang, S., Garner, O.B., Green, D.A., Gomez, C.A., Dien Bard, J., Pinsky, B.A. and Banaei, N. (2020) Clinical impact of metagenomic next-generation sequencing of plasma cell-free DNA for the diagnosis of infectious diseases: a multicenter retrospective cohort study. *Clin Infect Dis*. https://doi.org/10.1093/cid/ciaa035

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D. *et al.* (2016) Characterization of the gut microbiome using 16s or shotgun metagenomics. *Front Microbiol* **7**, 459.

Kufner, V., Plate, A., Schmutz, S., Braun, D.L., Gunthard, H.F., Capaul, R., Zbinden, A., Mueller, N.J. *et al.* (2019) Two years of viral metagenomics in a tertiary diagnostics unit: evaluation of the first 105 cases. *Genes (Basel)* **10**.

Miao, Q., Ma, Y., Wang, Q., Pan, J., Zhang, Y., Jin, W., Yao, Y., Su, Y. *et al.* (2018) Microbiological diagnostic performance of metagenomic next-generation sequencing when applied to clinical practice. *Clin Infect Dis* **67**, S231–S240.

Miller, S., Naccache, S.N., Samayoa, E., Messacar, K., Arevalo, S., Federman, S., Stryke, D., Pham, E. *et al.* (2019) Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res* **29**, 831–842.

Pallen, M.J. (2014) Diagnostic metagenomics: potential applications to bacterial, viral and parasitic infections. *Parasitology* **141**, 1856–1862.

Qi, C., Hountras, P., Pickens, C.O., Walter, J.M., Kruser, J.M., Singer, B.D., Seed, P., Green, S.J. *et al.* (2019) Detection of respiratory pathogens in clinical samples using metagenomic shotgun sequencing. *J Med Microbiol* **68**, 996–1002.

Seitz, V., Schaper, S., Droge, A., Lenze, D., Hummel, M. and Hennig, S. (2015) A new method to prevent carry-over contaminations in two-step PCR NGS library preparations. *Nucleic Acids Res* **43**, e135.

Wilson, M.R., Naccache, S.N., Samayoa, E., Biagtan, M., Bashir, H., Yu, G., Salamat, S.M., Somasekar, S. *et al.* (2014) Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med* **370**, 2408–2417.

Young, J.C., Chehoud, C., Bittinger, K., Bailey, A., Diamond, J.M., Cantu, E., Haas, A.R., Abbas, A. *et al.* (2015) Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am J Transplant* **15**, 200–209.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Detection of normal skin flora in plasma samples by manual operation and automation.

Table S2. PCR-mediated aerosol contamination in mNGS.