Alternative genome sequencing approaches of SARS-CoV-2 using Ion AmpliSeq Technology

Md. Shazid Hasan [a], M. Shaminur Rahman [a], Prosanto Kumar Das [a],
A.S.M. Rubayet Ul Alam [a], Ovinu Kibria Islam [a], Hassan M. Al-Emran [b],
M. Anwar Hossain [c,1], Iqbal Kabir Jahid [a,c,1,*]

[a] Department of Microbiology, Jashore University of Science and Technology, Jashore 7408, Bangladesh
[b] Department of Biomedical Engineering, Jashore University of Science and Technology, Jashore 7408, Bangladesh
[c] Genome Centre, Jashore University of Science and Technology, Jashore 7408, Bangladesh

ARTICLE INFO

Method name:
Alternative approaches of SARS-CoV-2 complete genome sequencing

Keywords:
Ion AmpliSeq SARS-CoV-2 Research Panel
Alternative approaches
NEBNext® Fast DNA Library Prep Set
Ion torrent technology

ABSTRACT

A thorough understanding of SARS-CoV-2 genetic features is compulsory to track the ongoing pandemic across multiple geographical locations of the world. Thermo Fisher Scientific USA has developed the Ion AmpliSeq SARS-CoV-2 Research Panel for the targeted sequencing of SARS-CoV-2 complete genome with high coverage and lower error rate. In this study an alternative approach of complete genome sequencing has been validated using different commercial sequencing kits to sequence the SARS-CoV-2. Amplification of cDNA with the SARS-CoV-2 primer pool was performed separately using two different master mixes: 2X environmental master mix (EM) and Platinum™ PCR SuperMix High Fidelity master mix (PM) instead of 5X Ion AmpliSeq™ HiFi Mix whereas NEBNext® Fast DNA Library Prep Set for Ion Torrent™ kit was used as an alternative to Ion AmpliSeq Library Kit Plus for other reagents. This study demonstrated a successful procedure to sequence the SARS-CoV-2 whole genome with average ~2351 depth and 98.1% of total the reads aligned against the reference sequence (SARS-CoV-2, isolate Wuhan-Hu-1, complete genome). Although genome coverage varied, complete genomes were retrieved for both reagent sets with a reduced cost. This study proposed an alternative approach of high throughput sequencing using Ion torrent technology for the sequencing of SARS-CoV-2 in developing countries where sequencing facilities are low. This blended sequencing technique also offers a low cost protocol in developing countries like Bangladesh.

Specifications table

Subject area:	Biochemistry, Genetics and Molecular Biology
More specific subject area:	Complete Genome Sequencing
Name of your protocol:	Alternative approaches of SARS-CoV-2 complete genome sequencing

* Corresponding author at: Department of Microbiology, Jashore University of Science and Technology, Jashore 7408, Bangladesh.
  E-mail address: ikjahid_mb@just.edu.bd (I.K. Jahid).
1 These authors share the senior authorship.

Reagents/tools:	Reagents
	• Invitrogen purelink viral RNA/DNA mini kit (Thermo Fisher Scientific, USA) • Qubit RNA HS Assay Kit (Thermo Fisher Scientific, USA) • novel coronavirus (2019-nCoV) nucleic acid diagnostic kit (Sansure Biotech Inc., China) • SuperScript® III First-Strand cDNA Synthesis System (Thermo Fisher Scientific, Waltham, MA, USA) • 2X environmental master mix from Ion 16S™ Metagenomics Kit (Thermo Fisher Scientific, USA) • Platinum™ PCR SuperMix High Fidelity master mix from Ion Xpress™ Plus gDNA Fragment Library kit (Thermo Fisher Scientific, Waltham, MA, USA) • Ion AmpliSeq™ SARS-CoV-2 Research Panel (Thermo Fisher Scientific, USA) • NEBNext® Fast DNA Library Prep Set for Ion Torrent™ (New England Biolabs, MA, USA, NEB#E6270L, Version 9.1_8/20) • Agencourt™ AMPure™ XP Reagent (Beckman Coulter, Brea, CA, USA) • Ion Xpress™ Barcode Adapters Kits (Thermo Fisher Scientific, Waltham, MA, USA) • Ion Library TaqMan® Quantitation Kit (Thermo Fisher Scientific,Cat. No. 4,468,802, MAN0015802 rev A.0) • Ion 520™ & Ion 530™ Kit – OT2 (Thermo Fisher Scientific, Cat. No. A27751, MAN0010844 rev E.0)
	Equipment
	• Qubit 4 Fluorometer (Thermo Fisher Scientific, USA) • Applied Biosystems™QuantStudio™ 3 Real-Time PCR System(Thermo Fisher Scientific, USA) • Applied Biosystems™ SimpliAmp™ Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA) • DynaMag™−2Magnet(Thermo Fisher Scientific, USA) • Ion One Touch™ 2 System(Thermo Fisher Scientific, USA) • Ion One Touch™ ES system(Thermo Fisher Scientific, USA) • Ion S5™ System (Thermo Fisher Scientific, USA)
Experimental design:	We validated the low-cost approach to retrieve the SARS-CoV-2 complete genome sequencing.
Trial registration:	Not applicable.
Ethics:	The study was approved by the Ethical Review Committee of Faculty of Biological Sciences and Technology, Jashore University of Science and Technology (ERC no: ERC/FBST/JUST/2020–51) and performed all experiments according to the relevant guidelines and regulations. The participants were informed about the study and consent was obtained from them.
Value of the Protocol:	• An alternative approach of complete genome sequencing using different commercial sequencing kits to sequence the SARS-CoV-2. • Cost-effective. • Complete genomes can be retrieved.

Description of protocol

Background information

The unprecedented development in molecular biology in recent years makes genome sequencing the most important tool to interpret viral transmission, disease dynamics, evolution history, mutation dynamics, vaccine and therapeutic efficacy, and identification of variants of concern (VOCs) [1–5]. As genome sequencing technology has become relatively available recently, most low- and middle-income countries like Bangladesh have been unable to afford frequent sequencing of the virus due to high throughput costs.

In this article, we will describe a modified protocol for SARS-CoV-2 sequencing using Ion AmpliSeq SARS-CoV-2 Research Panel, which can be a low-cost alternative to available SARS-CoV-2 whole genome sequencing methods (Thermo Fisher Scientific, USA, Pub. No. MAN0019277 Rev. A.0).

Methodology

Sample selection

All positive samples were randomly selected from the continuous surveillance at the Genome centre, Jashore University of Science and Technology, which is part of national surveillance covering four districts (Jashore, Magura, Narail, and Jhenaidha) in the Southwest part of Bangladesh, authorized and approved by the Directorate General of Health Services (DGHS), Bangladesh, for the screening of COVID-19 (https://mofa.portal.gov.bd/sites/default/files/files/mofa.portal.gov.bd/page/ad1f289c_47cf_4f6c_8dee_887957be3176/Govt%20lab%20list%20for%20COVID%20test.pdf) [6]. Among all the samples from 1st May to 15th June 2020, only three positive samples from three distinct districts (Jhenaidah, Narail, and Bagerhat) were selected based on the low cycle threshold (Ct) values (Ct<22) and used the left-over nasopharyngeal samples from the routine surveillance.

RNA extraction and RT-PCR

- 300 μL of nasopharyngeal samples were used to extract the total nucleic acids (NA) using Invitrogen purelink viral RNA/DNA mini kit (Thermo Fisher Scientific, USA) and eluted with 60 μL sterile RNase-free water according to the manufacturer's protocol.
- Total NA concentration was assayed by Qubit RNA HS Assay Kit (Thermo Fisher Scientific, USA) with Qubit 4 Fluorometer (Thermo Fisher Scientific, USA).

- Real-time fluorescence quantitative PCR (qRT-PCR) technology targeting the *ORF1ab* gene and *N* gene of SARS-CoV-2 was applied to detect the viral RNA within clinical samples using the novel coronavirus (2019-nCoV) nucleic acid diagnostic kit (Sansure Biotech Inc., China) in Applied Biosystems^{TM}QuantStudio^{TM} 3 Real-Time PCR System (Thermo Fisher Scientific, USA).

Amplicon-based library preparation

Libraries were prepared using the Ion AmpliSeq^{TM} SARS-CoV-2 Research Panel (Thermo Fisher Scientific, USA, Pub. No. MAN0019277 Rev. A.0) comprised of two 5X primer pools for whole viral genome sequencing. This approach followed an amplicon method that targeted 237 amplicons specific to the SARS-CoV-2 with 5 human expression controls targeting five regions of the human genome. The panel provides >99% coverage of the SARS-CoV-2 genome (~30 kb), with amplicon length ranging from 125 to 275 bp and covers all potential serotypes. Among the two 5X primer pools, primer pool 1 contains 125 primer pairs of 250 nM (120 specific to the SARS-CoV-2 sequences) and primer pool 2 contains 122 primer pairs of 250 nM (117 specific to the SARS-COV-2) [7]. The primer panel covered positions of 43 to 29,842 in the SARS-CoV-2 sequence and missed the first 42 and last 61 nucleotides because of primer placements [7].

The total protocol was followed by the instruction of five different protocols: Ion AmpliSeq Library Kit Plus user guide (Thermo Fisher Scientific, Publication Number MAN0017003 rev C.0), Ion AmpliSeq^{TM} SARS-CoV-2 Research Panel (Thermo Fisher Scientific, Publication Number MAN0019277 rev B.0), Ion Xpress^{TM} Plus gDNA Fragment Library Preparation user guide (Thermo Fisher Scientific, Publication Number MAN0009847, rev H.0), Ion 16S^{TM} Metagenomics Kit (Thermo Fisher Scientific, Publication Number MAN0010799, rev C.0) and NEBNext® Fast DNA Library Prep Set for Ion Torrent^{TM} (New England Biolabs, MA, USA, NEB#E6270L, Version 9.1_8/20) with some major modifications. All PCR reactions were performed with Applied Biosystems^{TM} SimpliAmp^{TM} Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA).

- cDNAs were prepared using SuperScript® III First-Strand cDNA Synthesis System (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instruction.
- Amplification of cDNA with the Ion AmpliSeq^{TM} SARS-CoV-2 Research Panel primer panel was performed separately using two different master mixes available in the lab: 2X environmental master mix (EM) from Ion 16S^{TM} Metagenomics Kit (Thermo Fisher Scientific, USA) and Platinum^{TM} PCR SuperMix High Fidelity master mix (PM) from Ion Xpress^{TM} Plus gDNA Fragment Library kit (Thermo Fisher Scientific, Waltham, MA, USA) instead of 5X Ion AmpliSeq^{TM} HiFi Mix. Hereafter, we will use EM and PM instead of their full meaning.
- Although 4.5 μL of 5X Ion AmpliSeq^{TM} HiFi Mix was recommended in the standard protocol, we optimized the protocol with EM and PM. 15 μL of EM was added with 8 μL of cDNA and 30 μL of PM was added with 7 μL of cDNA to perform the amplification.
- Both the mixtures were divided into two different tubes and 2 μL of each of the 5X Ion AmpliSeq^{TM} Primer Pool 1 and 2 were added to the corresponding tubes.
- The amplification condition was modified to be set for EM (activation of the enzyme: 95 °C, 7:30 min; denaturation 95 °C, 30 s; annealing 60 °C, 3 min; extension 72 °C, 30 s and final extension 72 °C, 3 min) and for PM (activation of the enzyme: 95 °C, 5 min; denaturation 95 °C, 30 s; annealing 60 °C, 3 min; extension 70 °C, 45 s and final extension 70 °C, 3 min). The number of amplification cycles of both PCR was set to 16, which was fixed based on the viral copy number of the samples and following the manufacturer's protocol of Ion AmpliSeq^{TM} SARS-CoV-2 Research Panel (Thermo Fisher Scientific, Publication Number MAN0019277 rev B.0).
- Following the target amplification of cDNA and combining these target amplified reactions of two distinct primer pools, the standard protocol suggests adding 2 μL of FuPa reagent for partial digestion of primers.
- Instead of partial digestion using FuPa reagent, in our modified protocol the pooled target amplified reactions were subjected to purify with the Agencourt^{TM} AMPure^{TM} XP Reagent (Beckman Coulter, Brea, CA, USA) according to the manufacturer's protocol of Ion 16S^{TM} Metagenomics Kit (Thermo Fisher Scientific, Publication Number MAN0010799, rev C.0, page#13) with minor modificaton. Here we used 80%ethanol instead of 70% as we used NEBNext® Fast DNA Library Prep Set for Ion Torrent^{TM} as an alternative for other reagents of Ion AmpliSeq Library Kit Plus where protocol recommend 80% ethanol.
- End-repair of amplicons were performed according to the NEBNext® Fast DNA Library Prep Set for Ion Torrent^{TM} protocol. 51 μL(adjust with Nuclease-free Water / NFW) of 100 ng amplified cDNA was added with 6 μL NEBNext End Repair Reaction Buffer and 3 μL NEBNext End Repair Enzyme Mix and the mix was incubated at 25 °C for 20 min, followed by 10 min at 70 °C, hold at 4 °C.
- Subsequently, Ion P1 adapter and Ion Xpress^{TM} Barcode (Thermo Fisher Scientific, Waltham, MA, USA) were ligated with the amplicons following Ion AmpliSeq Library Kit Plus user guide and NEBNext® Fast DNA Library Prep Set for Ion Torrent^{TM} protocol with minor modification. Briefly, 1 μL Ion P1 Adapter and 1 μL Ion Xpress^{TM} Barcode diluted with 2 μL of NFW. Afterwards, instead of 5 μL NEBNext DNA Library for Ion Torrent this 4 μL diluted adapter was mixed with 10 μL T4 DNA Ligase Buffer for Ion Torrent, 1 μL *Bst* 2.0 WarmStart DNA Polymerase, 6 μL DNA Ligase, and 60 μL of end repaired DNA. To adjust the final volume we added 19 μL of NFW instead of 18 μL. The mix was then incubated at 25 °C for 15 min, followed by 5 min at 65 °C, hold at 4 °C.
- The adapter ligated Barcoded libraries were subjected to purify again with the Agencourt^{TM} AMPure^{TM} XP Reagent (Beckman Coulter, Brea, CA, USA) following the manufacturer's protocol of NEBNext® Fast DNA Library Prep Set for Ion Torrent^{TM}.
- Afterward, the purified adapter ligated Barcoded libraries were PCR amplified and then AMPure^{TM} XP Reagent cleanup was applied again to purify the final libraries according to the NEBNext® Fast DNA Library Prep Set for Ion Torrent^{TM} protocol.

Table 1
Run and alignment summary results of the Ion 520 chips.

Addressable Wells	12,530,194		
With ISPs	4671,902	37.3%	
Live	4560,929	97.6%	
Test Fragment	271,948	06.0%	
Library	4288,981	94.0%	
Library ISPs	4288,981		
Filtered: Polyclonal	754,507	17.6%	
Filtered: Low Quality	467,243	10.9%	
Filtered: Adapter Dimer	187,613	04.4%	
Final Library ISPs	**2879,618**	**67.1%**	
Read Length			
Mean	236bp		
Median	261 bp		
Mode	275 bp		
Alignment Summary			
Total Reads	2677,770		
Aligned Reads	2628,132	98.1%	
Unaligned Reads	49,638	1.9%	
Alignment Quality	AQ17	AQ20	Perfect
Total Number of Bases [Mbp]	604 M	585 M	477 M
Mean Length [bp]	233	228	188
Longest Alignment [bp]	332	329	327
Mean Coverage Depth	20,194.6	19,564.1	15,942.2

Final concentration of each barcoded library was determined by qPCR with the Ion Library TaqMan® Quantitation Kit (Thermo Fisher Scientific, Cat. No. 4,468,802, MAN0015802 rev A.0) and diluted each library into 100 picomolar (pM) concentration recommended in manufacturer's protocol of Ion 520™ & Ion 530™ Kit – OT2 (Thermo Fisher Scientific, Cat. No. A27751, MAN0010844 rev E.0)

Preparation of enriched, template-positive Ion Sphere™ Particles (ISPs) and sequencing

- All the libraries were pooled in equimolar concentration to prepare the template-positive Ion Sphere™ Particles (ISPs) using the Ion 520™ & Ion 530™ Kit – OT2 user guide on the Ion OneTouch™ 2 System.
- Ion OneTouch™ ES system was used to enrich the template-positive ISPs. This ES system was cleaned with 70% alcohol prior to its running and maintained within biosafety cabinet.
- The enriched template-positive ISPs along with control ISPs were loaded in Ion 520™ chip and sequenced in Ion S5™ System.

Genome assembly, annotation, and analysis

- All data was retrieved as Binary Alignment Map (bam) files from Torrent Suite Software (version 5.10.0) (Thermo Fisher Scientific, USA) and analyzed with an in-house pipeline.
- Short reads of Binary Alignment Map (bam) files were converted to Fastq files by using SAMtools v1.12 [8] and the quality was checked by using FastQC 0.11.9 [9].
- Trimmomatic v0.39 [10] was used to trimmed low quality read (LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:36) and consensus were generated by using Burrows-Wheeler Aligner (BWA v0.7.17) [11], SAMtools v1.12 [8] and BEDTools v2.30.0 [12] by aligning a reference genome (hCoV-19/Wuhan/WIV04 (GenBank accession no. MN996528.1)).
- Further indel and area-wise mutation coverage were checked by using snippy [13] and corrected the genome accordingly. The variant caller of snippy is Freebays (https://github.com/freebayes/freebayes) with the minimum number of reads covering a site to be considered (default=10) and the minimum VCF variant call "quality" (default=100)

Phylogenetic reconstruction and mutation analysis

Nextstrain SARS-CoV-2 phylogetetic tree generation pipeline (https://nextstrain.org/sars-cov-2) has been used to reconstruct the phylogenetic tree with nextstrain clade of the sequenced representing this study [14]. Nextstrain uses MAFFT [15] as an alignment tools, IQ-TREE [16] and TreeTime [17] to reconstruct tree and Auspice web server (https://auspice.us/) to visualize the tree. Mutations and clades have assigned from CoVserver mutation app (https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/)

Table 2
Overview library preparation and read number, coverage of three samples with two different cDNA amplification methods.

Master Mix Criteria	Sample ID	cDNA concentration ng/ μL	[d]Chip loading PicoMolar (pM)	Mean Read Length	Total read count	Total read count after QC[a] (%)	Mapped reads (%)	Genome Coverage (×)	Complete genome (bp)	Clade	GenBank Accession No.	SRA Accession No
EM[b]	GC15.09-env11	962	100	247 bp	962,670	96.17	98.90	4162.62	29,870	20A	MZ749909	SRR15412770
PM[c]	GC15.09-pla8	962	100	243 bp	248,167	95.70	97.41	1661.93	29,879	20A	MZ749906	SRR15412769
EM	GC40.38-env12	1010	100	244 bp	784,868	96.35	99.91	3982.4	29,866	20A	MZ749905	SRR15412768
PM	GC40.38-pla9	1010	100	240 bp	248,269	95.64	99.56	1654.28	29,866	20A	MZ749907	SRR15412767
EM	GC40.86-env13	975	100	235 bp	349,023	95.85	99.42	2114.76	29,872	20A	MZ749904	SRR15412766
PM	GC40.86-pla10	975	100	231 bp	84,773	93.88	92.69	530.45	29,873	20A	MZ749908	SRR15412765

[a] QC = Quality control with trimmomatic (Details in methods section),

[b] EM = Environmental Master Mix,

[c] PM = Platinum Master Mix,

[d] Final Library Concentration for Chip loading PicoMolar (pM).

of Global initiative on sharing all influenza data (GISAID) where hCoV-19/Wuhan/WIV04 (GenBank accession no. MN996528.1) was used as a reference genome.

Protocol validation

Integrated epidemiological and virological surveillance of SARS-CoV-2 is playing an important role in monitoring the global spread and evolution of variants. High throughput sequencing data has provided the most vital information for these analyses. However, NGS using Ion torrent technology is expensive and complex procedure compared with Illumina, Oxford Nanopore, and Pacbio sequencing procedures [18]. Short read (about 300 bp) sequencing is used by Ion torrent and Illumina technology, whereas Oxford Nanopore and Pacbio employ approximately 100–1000 kb albeit with high error rates [19].

Thermo Fisher Scientific developed Ion AmpliSeq SARS-CoV-2 Research Panel for the sequencing of SARS-CoV-2 which amplify 237 target regions by two primers pools of 474 different primers sites. Several researchers have successfully used the technology and performed the sequencing [7,20,21]. All the researchers successfully sequenced the whole genome of SARS-CoV-2 and identified the different clades and mutations. Using the Ion Torrent Genexus Integrated Sequencer (Thermo Fisher Scientific, San Diego, CA), even low viral copy (less than 20/microliter) have been successfully sequenced almost 100% genome and different synonymous and non-synonymous mutations were obtained [21]. In another study, Ion AmpliSeq SARS-CoV-2 Research Panel successfully sequenced the whole genome of SARS-CoV-2 from 10 ng of isolates and 1 ng of nasopharyngeal swabs [7]. Other authors also compare the mutations of SARS-CoV-2 by both Ion torrent and MinION sequencing platform and found that Ion torrent is better compared to MinION technology [22]. However, all of the researchers followed the standard protocol and reagents as suggested by Thermo Fisher Scientific.

Our study demonstrated a successful procedure to sequence SARS-CoV-2 whole genome using Ion AmpliSeq SARS-CoV-2 Research Panel with average ~2351 depth. A total of 6 libraries (for 3 pairs of samples) were sequenced on one Ion 520 chip (Table 1). Ion Sphere Particles (ISP) loading was 37.3% where 97.6% of the ISP was represented live by libraries. About 17.6% of the ISPs were polyclonal, while low quality products and adapter dimers represented 10.9% and 04.4% of the sequences, respectively. The final libraries' ISPs were 67.1% of the total. The alignment summary showed that aligned reads against the reference sequence (Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome) represented 98.1% of total the reads, with a very low amount of reads (1.9%) not aligned to the reference genome.

This study used an alternative sequencing protocol to limit the expense of the project and modified the standard protocol according to the immediate availability of the sequencing reagents. We used two master mixes: 2X environmental master mix (EM) and Platinum™ PCR SuperMix High Fidelity master mix (PM) for cDNA amplification. Afterwards, an alternative library preparation regents: NEBNext® Fast DNA Library Prep Set for Ion Torrent™ were used instead of Ion AmpliSeq Library Kit Plus. Three samples were sequenced and for each sample, the procedure was duplicated as we used PM and EM separately for cDNA amplification. Among the two master mixes, The EM-generated library resulted in higher mean read length and total read count than PM (Table 2). Mapped reads and genome coverage of libraries also gave similar results for both EM and PM. Mapped reads from all libraries prepared by EM
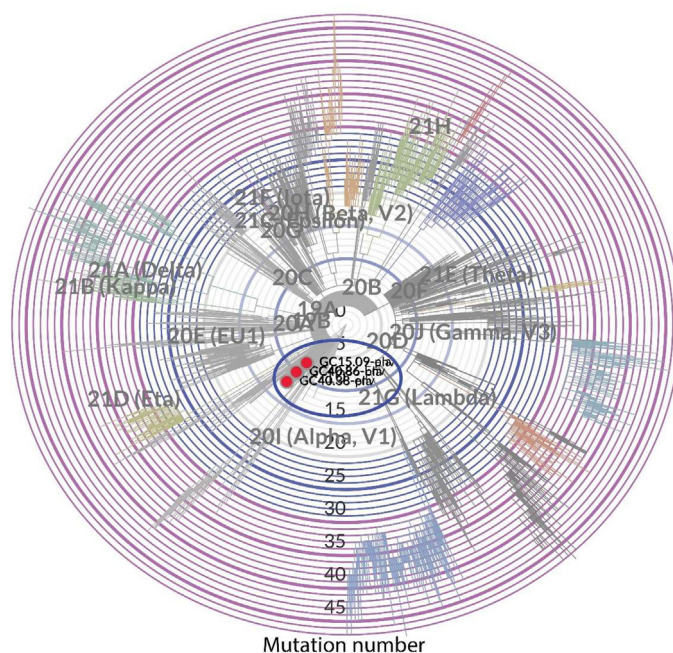

Fig. 1. SARS-CoV-2 genomes were visualized in world reference nextclade phylogenetic tree. All the samples were under 20A clade (Blue circle).

Table 3
Mutational spectra of 3 samples of SARS-CoV-2 genome with two different cDNA amplification methods (*PM = Platinum Master Mix, **EM = Environmental Master Mix) Evidence means the area wise coverage of that mutation.

Library type Sample ID	POS	TYPE	REF	ALT	EVIDENCE *PM	EVIDENCE *EM	NT_POS	AA_POS	EFFECT	GENE	PRODUCT
GC15.09	1163	snp	A	T	T:31 A:0	T:29 A:0	898/21,290	300/7095	missense_variant c.898A>T p.Ile300Phe	orf1ab	orf1ab polyprotein
	3037	snp	C	T	T:12 C:0	T:18 C:0	2772/21,290	924/7095	synonymous_variant c.2772C>T p.Phe924Phe	orf1ab	orf1ab polyprotein
	14,408	snp	C	T	T:23 C:0	T:26 C:0	14,143/21,290	4715/7095	synonymous_variant c.14143C>T p.Leu4715Leu	orf1ab	orf1ab polyprotein
	15,372	snp	G	T	T:23 G:0	T:26 G:0	15,107/21,290	5036/7095	missense_variant c.15107G>T p.Arg5036Leu	orf1ab	orf1ab polyprotein
	23,403	snp	A	G	G:27 A:0	G:34 A:0	1841/3822	614/1273	missense_variant c.1841A>G p.Asp614Gly	S	spike glycoprotein
	25,465	snp	C	T	T:29 C:0	T:35 C:0	73/828	25/275	missense_variant c.73C>T p.Pro25Ser	NS3	nonstructural protein NS3
GC40.38	1163	snp	A	T	T:51 A:1	T:40 A:0	898/21,290	300/7095	missense_variant c.898A>T p.Ile300Phe	orf1ab	orf1ab polyprotein
	3037	snp	C	T	T:14 C:0	T:14 C:0	2772/21,290	924/7095	synonymous_variant c.2772C>T p.Phe924Phe	orf1ab	orf1ab polyprotein
	3961	snp	C	T	T:30 C:0	T:38 C:0	3696/21,290	1232/7095	synonymous_variant c.3696C>T p.Ile1232Ile	orf1ab	orf1ab polyprotein
	4897	snp	C	T	T:31 C:0	T:43 C:0	4632/21,290	1544/7095	synonymous_variant c.4632C>T p.Phe1544Phe	orf1ab	orf1ab polyprotein
	12,170	snp	G	A	A:14 G:1	A:21 G:0	11,905/21,290	3969/7095	missense_variant c.11905G>A p.Ala3969Thr	orf1ab	orf1ab polyprotein
	14,408	snp	C	T	T:37 C:0	T:29 C:0	14,143/21,290	4715/7095	synonymous_variant c.14143C>T p.Leu4715Leu	orf1ab	orf1ab polyprotein
	14,973	snp	G	A	A:36 G:0	A:26 G:0	14,708/21,290	4903/7095	missense_variant c.14708G>A p.Arg4903Lys	orf1ab	orf1ab polyprotein
	20,274	snp	G	T	T:54 G:0	T:44 G:0	20,009/21,290	6670/7095	missense_variant c.20009G>T p.Trp6670Leu	orf1ab	orf1ab polyprotein
	23,108	snp	G	C	C:28 G:0	C:23 G:0	1546/3822	516/1273	missense_variant c.1546G>C p.Glu516Gln	S	spike glycoprotein
	23,403	snp	A	G	G:50 A:0	G:41 A:0	1841/3822	614/1273	missense_variant c.1841A>G p.Asp614Gly	S	spike glycoprotein
GC40.86	1163	snp	A	T	T:23 A:0	T:35 A:0	898/21,290	300/7095	missense_variant c.898A>T p.Ile300Phe	orf1ab	orf1ab polyprotein
	1191	snp	C	T	T:25 C:0	T:34 C:0	926/21,290	309/7095	missense_variant c.926C>T p.Pro309Leu	orf1ab	orf1ab polyprotein
	3037	snp	C	T	T:9 C:0	T:16 C:0	2772/21,290	924/7095	synonymous_variant c.2772C>T p.Phe924Phe	orf1ab	orf1ab polyprotein
	14,408	snp	C	T	T:13 C:0	T:25 C:0	14,143/21,290	4715/7095	synonymous_variant c.14143C>T p.Leu4715Leu	orf1ab	orf1ab polyprotein
	15,222	snp	C	T	T:23 C:0	T:33 C:0	14,957/21,290	4986/7095	missense_variant c.14957C>T p.Ser4986Phe	orf1ab	orf1ab polyprotein
	18,555	snp	C	T	T:11 C:0	T:22 C:0	18,290/21,290	6097/7095	missense_variant c.18290C>T p.Thr6097Ile	orf1ab	orf1ab polyprotein
	23,403	snp	A	G	G:12 A:0	G:19 A:0	1841/3822	614/1273	missense_variant c.1841A>G p.Asp614Gly	S	spike glycoprotein
	25,617	snp	G	T	T:33 G:0	T:45 G:0	225/828	75/275	missense_variant c.225G>T p.Lys75Asn	NS3	nonstructural protein NS3

Table 4
Cost comparison between Ion AmpliSeq Method and Modified Method of SARS-CoV-2 sequencing.

SI No	Name	Ion AmpliSeq Method			Modified Method		
		USD	Reaction	Per/sample USD	USD	Reaction	Per/sample USD
1	Ion 16 s Metagenomics kit	–	–	–	1755.81	100	17.56
2	Ion AmpliSeq Library kit	3488.37	24	145.35			
3	NEBNext® Fast DNA Library Prep Set for Ion Torrent™ (New England Biolabs)	–	–	–	1511.63	50	30.23
4	Ion Xpress™ Barcode Adapters	3720.93	20	11.63	3720.93	20	11.63
5	Qubit DNA Assay Kit	139.53	100	1.39	139.53	100	1.39
6	Agencourt AMPure XP, 60 Ml	1976.74		1.5	1976.74		1.5
7	OT2 kit	8372	4	261.63	8372	4	261.63
8	520 chip kit	4418.6	8	69.77	4418.6	8	69.77
9	Dynabeads™ MyOne™ Streptavidin C1	813.95	200	4	813.95	200	4
	Total			**~ $495.27**			**~ $397.71**

was 99.4% on average whereas the libraries derived from PM was 96.6%. We got more than twice genome coverage in EM-derived sequences compared to PM-derived sequences. Due to limited funding, our analysis was constrained by a small sample size and the absence of statistical analysis. Nevertheless, for both cases, the complete genomes were retrieved (Table 2). Nextclade analysis showed that all three pairs of genomes possessed clade 20A (Fig. 1). Mutation analysis from short-read sequences (raw read) and from the complete genome sequences of three samples showed no difference for EM and PM. Area-wise coverage of each mutation (Table 3) also showed no significant difference between the obtained sequences using these two master mixtures. Though the unavailability of certain reagents limited our ability to conduct a direct comparison, we are confident in the efficacy and cost-effectiveness of the alternative method presented as we found similar results when utilizing the standard library preparation method in a separate sequencing run [Data not presented]. Our method worked well with PM and EM master mix and thus we could reduce the total cost of sequencing. This approach might be convenient for low-income countries where the reagents are costly. The cost comparison showed that this modified approach needed around $397.71 whereas the Ion Ampliseq method required around $495.27 (all costs included import Tax and VAT, etc. applicable in Bangladesh) (Table 4).

The main limitation of our study is, we did not test the efficacy of our method for different variants which emerged later. But other researchers successfully sequenced different variants using Ion AmpliSeq SARS-CoV-2 Research Panel [23]. There were some unaligned reads, which might arise from the dimerization of primers during the PCR amplification step. These unaligned reads were not further investigated. Nevertheless, we found this protocol accurate and cost-effective.

Conclusion

The largest contributor to SARS-CoV-2 sequence data are Europe (63%) and North America (27%), while Asia (6%) and Africa (1%) remain low. Our study proposed an alternative approach of high throughput sequencing using Ion torrent technology for the sequencing of SARS-CoV-2 in developing countries where sequencing facilities are low. This blended sequencing approach offers low-cost operational technique in low-facilitated regions to generate the high throughput sequencing data.

Additional information

Coronavirus disease (COVID-19) caused by SARS-CoV-2 was first detected on December 8, 2019, in Wuhan city of Hubei Province, China [24]. Within the first couple of months, the virus had transmitted to different regions of the world [1,4] due to its high transmissibility and World Health Organization declared this disease as a global pandemic on March 11, 2020 [25]. Bangladesh reported its first COVID-19 case on March 8, 2020. As of March 7, 2022, in Bangladesh, a total of 1947,266 COVID-19 cases were confirmed through RT-PCR and rapid antigen test (14.4% detection rate), and 29,085 people died from this disease. (http://dashboard.dghs.gov.bd/webportal/pages/covid19.php). The death rate in Bangladesh increased during the second wave of infection which was caused initially by B.1351 [26] and B1.617.2 [27] variant.

SARS-CoV-2 virus is a positive sense single-stranded RNA virus having genome size of 29 kb [28] with high-frequency mutations to genes more vulnerable to RBD regions. The high mutation rates enable the virus to track the virus compared with DNA virus [29]. Because of the unprecedented development in molecular biology in recent years, genome sequencing became the most important tool to interpret the viral transmission, diseases dynamics, evolution history, mutations dynamics, vaccine and therapeutic efficacy, identification of variants of concern (VOCs) [5]. From January 2020 to February 2022, 8688, 321 complete genome sequence of SARS-CoV-2 were submitted to the GISAID database (gisaid.org), which is by far the highest number of sequences for a single virus. On the other hand, Genome sequencing technology have become relatively available in recent days, due to high throughput cost, most low- and middle-income countries like Bangladesh could not afford frequent sequencing of the virus. GISAID Data suggests that, 89% of the SARS-CoV-2 sequences (more than 90 million sequences) during corona pandemic have been submitted from European and North American countries, whereas only 1% (88,706) of the total submitted sequences were from Africa (gisaid.org).

For SARS-CoV-2 sequencing, Ion AmpliSeq SARS-CoV-2 Research Panel have been developed by Thermo Fisher Scientific for use with Ion S5 system. This system is relatively inexpensive at low throughput and lower error rate, accurate, quick run make this attractive to some researchers. The Pubmed search by the words "Ion Torrent" retrieved only 385 publications (https://pubmed.ncbi.nlm.nih.gov/?term=ion±ampliseq) and several authors have used the Ion Ampliseq Technology to sequence the SARS-CoV-2 [7,21,30]. But none of the authors modified the basic protocol of Ion AmpliSeq developed by Thermo Fisher Scietific (https://www.thermofisher.com/bd/en/home/life-science/sequencing/dna-sequencing/microbial-sequencing/microbial-identification-ion-torrent-next-generation-sequencing/viral-typing/coronavirus-research.html). Our study describe low-cost modified protocol for SARS-CoV-2 sequencing using Ion AmpliSeq SARS-CoV-2 Research Panel.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Md. Shazid Hasan: Conceptualization, Methodology, Validation, Investigation, Writing – original draft. **M. Shaminur Rahman:** Software, Data curation, Formal analysis. **Prosanto Kumar Das:** Writing – original draft, Writing – review & editing, Visualization. **A.S.M. Rubayet Ul Alam:** Writing – review & editing. **Ovinu Kibria Islam:** Writing – review & editing. **Hassan M. Al-Emran:** Investigation, Writing – review & editing. **M. Anwar Hossain:** Supervision, Project administration, Writing – review & editing. **Iqbal Kabir Jahid:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Data availability

The raw reads were submitted to the NCBI under BioProject accession number PRJNA753679 and SRA accession SRR15412765 - SRR15412770. The sequences of 6 SARS-CoV-2 genomes were submitted to the NCBI GenBank database under the identifiers MZ749904 - MZ749909 (Table 1).

Data will be made available on request.

Acknowledgments

References

[1] M.S. Hasan, M.T. Islam, A.S.M.R.U. Alam, S.L. Sarkar, M.S. Rahman, O.K. Islam, M.A.A. Setu, T. Chakrovarty, H.M. Al-Emran, I.K. Jahid, M.A. Hossain, Initial reports of the SARS-CoV-2 Delta variant (B.1.617.2 lineage) in Bangladeshi patients: risks of cross-border transmission from India, Health Sci. Rep. 4 (2021) e366, doi:10.1002/hsr2.366.

[2] H.M. Al-Emran, S. Rahman, M.S. Hasan, R. Ul Alam, O.K. Islam, A. Anwar, M.I.K. Jahid, A. Hossain, Microbiome analysis revealing microbial interactions and secondary bacterial infections in COVID-19 patients comorbidly affected by Type 2 diabetes, J. Med. Virol. (2022), doi:10.1002/jmv.28234.

[3] A.S.M.R.U. Alam, O.K. Islam, M.S. Hasan, M.R. Islam, S. Mahmud, H.M. Al-Emran, I.K. Jahid, K.A. Crandall, M.A. Hossain, Dominant clade-featured SARS-CoV-2 co-occurring mutations reveal plausible epistasis: an in silico based hypothetical model, J. Med. Virol. 94 (2022) 1035–1049, doi:10.1002/jmv.27416.

[4] O.K. Islam, H.M. Al-Emran, M.S. Hasan, A. Anwar, M.I.K. Jahid, M.A. Hossain, Emergence of European and North American mutant variants of SARS-CoV-2 in South-East Asia, Transbound. Emerg. Dis. 68 (2021) 824–832, doi:10.1111/tbed.13748.

[5] H.M. Al-Emran, M.S. Hasan, M.A. Ahasan Setu, M.S. Rahman, A.R.U. Alam, S.L. Sarkar, M.T. Islam, M.R. Islam, M.M. Rahman, O.K. Islam, I.K. Jahid, M.A. Hossain, Genomic analysis of SARS-CoV-2 variants of concern identified from the ChAdOx1 nCoV-19 immunized patients from Southwest part of Bangladesh, J. Infect. Public Health 15 (2022) 156–163, doi:10.1016/j.jiph.2021.12.002.

[6] DGHSCOVID-19 Dynamic Dashboard for Bangladesh, DGHS, 2022 http://dashboard.dghs.gov.bd/webportal/pages/covid19.php accessed February 6, 2022.

[7] F. Alessandrini, S. Caucci, V. Onofri, F. Melchionda, A. Tagliabracci, P. Bagnarelli, L. Di Sante, C. Turchi, S. Menzo, Evaluation of the Ion AmpliSeq SARS-CoV-2 Research Panel by Massive Parallel Sequencing, Genes 11 (2020) (Basel), doi:10.3390/genes11080929.

[8] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079, doi:10.1093/bioinformatics/btp352.

[9] S. Andrews, FastQC: a quality control tool for high throughput sequence data, (2010).

[10] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.

[11] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics 25 (2009) 1754–1760, doi:10.1093/bioinformatics/btp324.

[12] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (2010) 841–842, doi:10.1093/bioinformatics/btq033.

[13] T. Seemann, Snippy: Rapid Haploid Variant Calling and Core SNP Phylogeny, GitHub, 2015. https://github.com/tseemann/snippy.

[14] J. Hadfield, C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R.A. Neher, Nextstrain: real-time tracking of pathogen evolution, Bioinformatics 34 (2018) 4121–4123, doi:10.1093/bioinformatics/bty407.

[15] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, doi:10.1093/molbev/mst010.

[16] L.T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274, doi:10.1093/molbev/msu300.

[17] P. Sagulenko, V. Puller, R.A. Neher, TreeTime: maximum-likelihood phylodynamic analysis, Virus Evol. 4 (2018) vex042, doi:10.1093/ve/vex042.

[18] S.E. Levy, R.M. Myers, Advancements in next-generation sequencing, Annu. Rev. Genom. Hum. Genet. 17 (2016) 95–115, doi:10.1146/annurev-genom-083115-022413.

[19] V. Bansal, C. Boucher, Sequencing technologies and analyses: where have we been and where are we going? iScience 18 (2019) 37–41.

[20] G.S. Campos, S.I. Sardi, M.B. Falcao, E.M.M.A. Belitardo, D.J.P.G. Rocha, C.A. Rolo, A.D. Menezes, C.S. Pinheiro, R.H. Carvalho, J.P.P. Almeida, Ion torrent-based nasopharyngeal swab metatranscriptomics in COVID-19, J. Virol. Methods 282 (2020) 113888.

[21] A.M. Rachiglio, L. De Sabato, C. Roma, M. Cennamo, M. Fiorenza, D. Terracciano, R. Pasquale, F. Bergantino, E. Cavalcanti, G. Botti, G. Vaccari, G. Portella, N. Normanno, SARS-CoV-2 complete genome sequencing from the Italian Campania region using a highly automated next generation sequencing system, J. Transl. Med. 19 (2021) 246, doi:10.1186/s12967-021-02912-4.

[22] C.P. Sjaarda, N. Rustom, G.A. Evans, D. Huang, S. Perez-Patrigeon, M.L. Hudson, H. Wong, Z. Sun, T.H. Guan, M. Ayub, C.N. Soares, R.I. Colautti, P.M. Sheth, Phylogenomics reveals viral sources, transmission, and potential superinfection in early-stage COVID-19 patients in Ontario, Canada, Sci. Rep. 11 (2021) 3697, doi:10.1038/s41598-021-83355-1.

[23] A. Lai, A. Bergna, S. Caucci, N. Clementi, I. Vicenti, F. Dragoni, A.M. Cattelan, S. Menzo, A. Pan, A. Callegaro, A. Tagliabracci, A. Caruso, F. Caccuri, S. Ronchiadin, C. Balotta, M. Zazzi, E. Vaccher, M. Clementi, M. Galli, G. Zehender, Molecular tracing of SARS-CoV-2 in Italy in the first three months of the epidemic, Viruses 12 (2020), doi:10.3390/v12080798.

[24] Y. Wu, X. Xu, Z. Chen, J. Duan, K. Hashimoto, L. Yang, C. Liu, C. Yang, Nervous system involvement after infection with COVID-19 and other coronaviruses, Brain Behav. Immun. 87 (2020) 18–22, doi:10.1016/j.bbi.2020.03.031.

[25] WHOWHO Director-General's Opening Remarks at the Media Briefing on COVID-19–11 March 2020, WHO, 2020 https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020 accessed February 6, 2022.

[26] M. Rahman, T. Shirin, S. Rahman, M.M. Rahman, M.E. Hossain, M.H. Khan, M.Z. Rahman, S. El Arifeen, T. Ahmed, The emergence of SARS-CoV-2 variants in Dhaka city, Bangladesh, Transbound. Emerg. Dis. 68 (2021) 3000–3001, doi:10.1111/tbed.14203.

[27] A.A. Moona, S. Daria, M. Asaduzzaman, M.R. Islam, Bangladesh reported delta variant of coronavirus among its citizen: actionable items to tackle the potential massive third wave, Infect. Prev. Pract. 3 (2021) 100159, doi:10.1016/j.infpip.2021.100159.

[28] L. Mousavizadeh, S. Ghasemi, Genotype and phenotype of COVID-19: their roles in pathogenesis, J. Microbiol. Immunol. Infect. 54 (2021) 159–163, doi:10.1016/j.jmii.2020.03.022.

[29] N.D. Grubaugh, J.T. Ladner, P. Lemey, O.G. Pybus, A. Rambaut, E.C. Holmes, K.G. Andersen, Tracking virus outbreaks in the twenty-first century, Nat. Microbiol. 4 (2019) 10–19, doi:10.1038/s41564-018-0296-2.

[30] H. Gu, Q. Chen, G. Yang, L. He, H. Fan, Y.Q. Deng, Y. Wang, Y. Teng, Z. Zhao, Y. Cui, Y. Li, X.F. Li, J. Li, N.N. Zhang, X. Yang, S. Chen, Y. Guo, G. Zhao, X. Wang, D.Y. Luo, H. Wang, X. Yang, Y. Li, G. Han, Y. He, X. Zhou, S. Geng, X. Sheng, S. Jiang, S. Sun, C.F. Qin, Y. Zhou, Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy, Science 369 (2020) 1603–1607, doi:10.1126/science.abc4730.