Molecular Biology Reports (2018) 45:297–308
https://doi.org/10.1007/s11033-018-4162-3

ORIGINAL ARTICLE

Statistical assessment of DNA extraction methodology for culture-independent analysis of microbial community associated with diverse environmental samples

Rishi Mahajan[1] · Sampan Attri[1] · Kavita Sharma[1] · Niharika Singh[1] · Deepika Sharma[1] · Gunjan Goel[1]

Received: 2 May 2017 / Accepted: 12 February 2018 / Published online: 16 February 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

Cost-effectiveness, quality, time-effectiveness and ease of the methodology are the most crucial factors in isolating quality DNA from wide variety of samples. Thus, research efforts focusing on the development of an efficient DNA extraction protocol is the need of the hour. The present study therefore, focuses on development of an efficient, rapid and free of inhibitory substances based methodology for extracting metagenomic DNA from diverse environmental samples viz. anaerobic biogas digesta, ruminant stomach, human feces, soil, and microbial starter cultures used for preparation of fermented food. PCR–DGGE based analysis and quality metagenomic library preparation, using DNA extraction methodology, validates the developed protocol. The developed protocol is cost effective, capable of isolating DNA from small sample size (100–1000 µl), time efficient (1.5–2.0 h protocol) and results in significantly higher DNA yield (4–8 times increased yield) when compared to previously available DNA extraction method and a commercial DNA extraction kit. The DNA extracted from the samples using different protocols was evaluated based on its ability to identify diverse microbial species using PCR–DGGE profiles targeting variable region within the 16S rRNA gene. The results of microbial community analysis revealed comparability of the developed protocol to commercial kits, in effectively identifying dominant representatives of the microbial community in different samples. Using the DNA extracted from the presented methodology, metagenomic libraries were prepared, which were found suitable for sequencing on Illumina platform.

Keywords DNA extraction · PCR–DGGE · Metagenomic library · Environmental samples · Microbial community

Introduction

The molecular approaches are reported for their significance in getting an insight into the complex microbial community structure in diverse environments. To an estimate, < 1% of the total microbial community in diverse environments is culturable [1, 2], therefore, the application of culture independent techniques to study complex microbial communities is on an all time increase [3]. Irrespective of the recent developments in high-throughput metagenomic technologies used in analysis of diverse microbial communities, the basic and foremost step of primary importance is the extraction of high-quality metagenomic environmental DNA. The DNA extraction protocols have been modified extensively, based on the sample requirement [4–8], however, reports of a single DNA extraction procedure, that could be used for diverse environmental samples, whether aerobic of anaerobic are scanty.

A review of the recent literature indicates that several researchers have correlated cost-dynamics with environmental DNA sampling procedures [9–11]. However, importance of budget requirements and the time invested in the actual procedure of environmental DNA extraction cannot be ignored. Though, several commercial kits are widely available [12], but the costs and time involved in the procedure of DNA extraction per sample can vary and certainly influence the overall research plan, especially if considering the implementation of next-generation sequencing (NGS)

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11033-018-4162-3) contains supplementary material, which is available to authorized users.

✉ Gunjan Goel
gunjanmicro@gmail.com

[1] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 173234, India

analysis. The involvement of interfering contaminates, while extracting DNA from environment samples poses other hurdles in characterization of diverse and specific microbial communities. The use of carcinogenic chemicals and cost intensive methodologies to mitigate the effect of environmental contaminants in DNA extraction procedure adds to the total cost of the culture independent molecular analysis of the samples [11, 12].

The present study is therefore aimed at developing a highly efficient, cost effective and time efficient DNA extraction methodology, which could be used for samples irrespective of their environmental source. Another aim of the study, while developing the DNA extraction methodology is to mitigate the negative influence of environmental contaminants that interfere with the downstream processing of DNA extraction and hinder the evaluation of complex microbial communities. Based on the results, a statistical model has been proposed, taking into account the factors like, actual chemical costs and time invested in the environmental DNA extraction procedure, without compromising the microbial species detection rate.

Materials and methods

Collection of samples

The samples of anaerobic digesta were collected from two anaerobic digesters (Table 1), ruminal fluid from freshly slaughtered cattle, from a butchery located in Solan, Himachal Pradesh. In order to validate the protocol, samples were also collected from other diverse environmental sources viz. (a) Samples were collected from infant's feces, at different time intervals, over a period of 4 months, until the infant's diet was altered from breast feed to formula feed. (b) Samples from Chir pine tree's rhizosphere soil. (c) Pine forest needle forest litterfall from sub-tropical Kandaghat pine forest range (31°00′57″N; 77°04′17″E). (d) Microbial starter cultures viz. *malera* and *phab* which are used for preparation of fermented beverages in Himachal Pradesh. (e) pure cultures of bacteria (*Lactobacillus* sp.) and fungi (*Cotylidia pannosa; Aspergillus* sp.; *Penicillium chrysogenum*).

Metagenomic environmental DNA extraction protocol

Test samples (Table 1) were mixed with 1000 µl of 'Buffer A' (50 mM EDTA; 200 mM Tris HCL; 1.5% SDS; 12 mg/ml Lysozyme) in a sterile 2 ml tubes, which were vortexed for 1 min followed by incubation at 70 °C for 35–40 min with gentle inversions, every 5 min. During the this duration, sample tubes were subjected to repeated cycles of incubation at − 20 °C for 2–5 min, followed by immediate immersion in water bath, set at 70 °C with simultaneous gentle inversions. Following this, the tubes were supplemented with 'Solution A' (60% PEG 400; 20 mM KOH; 2% $CaCl_2$) and mixed with gentle inversions for 2–3 min. The supernatant was collected after centrifugation at 6000×*g* for 2 min and an equal volume of 'Buffer B' (equal volume of phenol:chloroform in 1:1 ratio) and 0.2 volume 20% PEG 400 was added (Solution B). Further, the tubes were centrifuged at 12,000×*g* for 3 min for the recovery of aqueous phase. This step of aqueous phase separation was repeated employing similar conditions as mentioned above, with addition of equal volume of chloroform. The separated phase was transferred to a fresh tube containing 500 µl of 'Buffer D' (equal volume of pre-chilled absolute isopropanol) and supplemented with 'Solution C' (0.5 volume 50% PEG 400; 0.1 volume 5 M Nacl; 2% $CaCl_2$). The suspension was incubated at 20 °C for 30 min. The tubes were then centrifuged at maximum speed and the DNA pellet obtained was washed with 70% ethanol. The DNA pellet was air dried and suspended in 1000 µl nuclease free water.

For comparison purposes, metagenomic environmental DNA was also extracted by using (A) commercially available kit HPurA™ Stool DNA Purification Kit (Himedia Laboratories, Mumbai, India) (B) Protocol of previously used DNA extraction by Holben et al. [13] and Dempster et al. [14].

Purity and yield

The purity and concentration of extracted metagenomic environmental DNA was determined using spectrophotometric analysis in µDrop™ plate (Thermo Scientific MULTISCAN GO Spectrophotometer, USA). The quality of DNA and protein contamination was determined by spectrophotometric measurements at 260 and 280 nm. The DNA concentration was calculated using standard formula. Finally, the extracted DNA was analyzed for its integrity in 1.5% agarose gel (containing 0.05% ethidium bromide) using UV transilluminator.

PCR amplification, DGGE analysis and sequencing, phylogenetic analysis

For PCR amplification, *16S* ribosaomal RNA genes for total bacterial and archaeal; 18S–28S ITS genes *Eucarya* (Fungi) were targeted (Table S1). The PCR amplification was carried out as per details in Table S1. The amplified products were separated on denaturing gradient gel electrophoresis (DGGE: Bio-Rad, Hercules, CA, USA) and analysed using methodology as described previously Mahajan et al. [15].

The purified PCR products were sequenced and analysed. Multiple sequence alignments were used to carry out phylogenetic analysis and Unweighted Pair Group

Table 1 Description of samples along with DNA concentration and purity using spectrophotometric methods based on DNA extraction methodology used in the present study

Samples	Sampling source description	Sample pretreatment	Sample volume	DNA concentration (μg/ml)	DNA purity (A260/280)
Anaerobic digested sludge	Anaerobic digester at Shimla, Himachal Pradesh (Site I: 31°11′23″N, 77°10′17″E)	No pretreatment	500 μl	232 ± 7.71	1.65 ± 0.03
	Anaerobic digester at Shimla, Himachal Pradesh (Site II: 31°11′13.1″N 77°24′18.1″E)	No pretreatment	500 μl	215 ± 5.57	1.60 ± 0.04
Rumen Sample	Sampling from ruminant stomach of freshly slaughtered cattle, from a butchery located in Solan, Himachal Pradesh	No pretreatment	100 μl	277 ± 4.90	1.34 ± 0.03
Human infant feces	Sampling after 1st month	Slurry was prepared in nuclease free water	500 μl	285 ± 5.15	1.45 ± 0.03
	Sampling after 2nd month			355 ± 4.64	1.30 ± 0.04
	Sampling after 3rd month			476 ± 4.00	1.55 ± 0.02
	Sampling after 4th month			555 ± 7.38	1.60 ± 0.03
Soil samples	Chir pine tree's rhizosphere soil	No pretreatment	1000 mg	270 ± 5.43	1.30 ± 0.04
Forest litterfall	Pine forest needle forest litterfall	Pine needles powdered and slurry was prepared in nuclease free	500 μl	190 ± 6.48	1.45 ± 0.03
Microbial fermentation starter cultures	*Malera*	Samples powdered and mixed with minimum amount nuclease free water to obtain slurry	100 μl	719 ± 4.74	1.45 ± 0.03
	Phab			329 ± 5.79	1.9 ± 0.05
Bacterial pure cultures	*Brevibacillus aydinogluensis* strain BTM9	Overnight grown cell pellet	100 μl	625 ± 3.15	1.9 ± 0.03
	Brevibacillus thermoruber strain CD13		100 μl	600 ± 4.55	1.75 ± 0.03
	Enterococcus sp. strain GTM14		100 μl	710 ± 3.46	1.70 ± 0.05
	Brevibacillus thermoruber strain HM29		100 μl	586 ± 3.29	1.79 ± 0.04
	Brevibacillus thermoruber strain HM34		100 μl	620 ± 4.96	1.80 ± 0.03
	Weissella confuse strain C1		100 μl	589 ± 5.00	1.77 ± 0.04
	Lactobacillus paracasei strain CD4		100 μl	564 ± 4.47	1.80 ± 0.03
	Lactobacillus gastricus strain BTM7		100 μl	550 ± 5.15	1.83 ± 0.04
Fungal pure cultures	*Cotylidia pannosa* strain F6	Mycelia/fungal spores	1000 μl	620 ± 4.80	1.75 ± 0.04
	Penicillium chrysogenum DST-RFBR1		1000 μl	575 ± 4.06	1.80 ± 0.06
	Aspergillus sp. DST-RFBR2		1000 μl	600 ± 7.97	1.79 ± 0.03

Method with Arithmetic Mean hierarchical clustering method was used to construct phylogenetic trees. Statistical validation of the phylogenetic tree was also carried out (bootstrap confidence values: 100 permutations). The identified sequences of all the microbial cultures were submitted in the database of National Center for Biotechnology Information (NCBI). The accession numbers provided by NCBI have been provided in Supplementary Table S2.

Metagenome library preparation and sequencing

Based on the methodology developed for DNA extraction in the present study, 25 ng of nano drop quantified DNA was used for amplification of 16S rRNA region (variable region of V3–V4). The primers used were “taged” with a complimentary sequence to Illumina sequence adapter and also to the index primers from the Nextera XT “Index kit V2”. An amplicon of 500–550 bp was obtained in this round of PCR, and the results were confirmed on agarose gel. In the following indexing PCR, ‘Illumina sequencing adapters’ and ‘dual indexing barcodes’ were added and a product size of ~600–650 bp was obtained. An aliquot (1:10 dilution) was run on High Sensitivity Bioanalyzer Chip (Agilent Technologies, USA), in order to validate the quality of the libraries.

Data analysis

All the data are expressed as the average of five replications along with standard error of the mean (±SEM). The statistical model developed in the present study is based on slight modifications in the previously described statistical models as proposed by Smart et al. [11] and Moore et al. [16]

Results

The use of metagenomic environmental DNA extraction methodology, developed in the present study resulted in significantly higher yields of DNA from the two diverse anaerobic environment sources viz. anaerobic digested sludge samples and from ruminant stomach, when compared with yields of DNA extracted using commercially available kit and DNA extraction protocols evaluated from previous research studies (as shown in Fig. 1). Using protocol developed in the present study, the samples from anaerobic digested sludge resulted in maximum DNA yield of 232 µg/ml with corresponding A260/280 ratios of 1.65. The use of commercial kit resulted in significantly lower DNA yields (47 µg/ml) with corresponding A260/280 ratios of 1.0. The lowest DNA yield of 18.5 µg/ml was observed when DNA extraction protocol from previous studies was used. A similar trend was observed in case of samples from ruminant stomach. DNA yields of 277, 70 and 52.40 µg/ml with corresponding A260/280 ratios of 1.34, 0.70 and 1.24 were observed when DNA extraction was carried out using methodology developed in present study, using commercial kit and protocol from previous studies, respectively.

Samples collected from infant’s feces, at different time intervals, over a period of 4 months, resulted in DNA yield ranging from 285 to 555 µg/ml with corresponding A260/280 ratios ranging from 1.3 to 1.6 (Table 1). The DNA extraction protocol also resulted in effective DNA yields from samples of Chir pine tree’s rhizosphere soil (270 µg/ml; A260/280:1.30) and Pine forest needle forest litterfall (190 µg/ml; A260/280:1.45). Extraction beverages resulted in DNA yields ranging from 255 to 320 µg/ml with corresponding A260/280of DNA from *malera* and *phab* samples (starter cultures used for preparation of fermented ratios ranging from 1.45 to 1.55). The DNA extraction methodology was also validated using pure bacterial and fungal cultures (*Brevibacillus aydinogluensis; Brevibacillus thermoruber; Enterococcus* sp.; *Weissella confuse; Lactobacillus paracasei; Lactobacillus gastricus; Cotylidia pannosa; Aspergillus* sp.; *Penicillium chrysogenum*) that resulted in DNA yields ranging from 575 to 710 µg/ml with

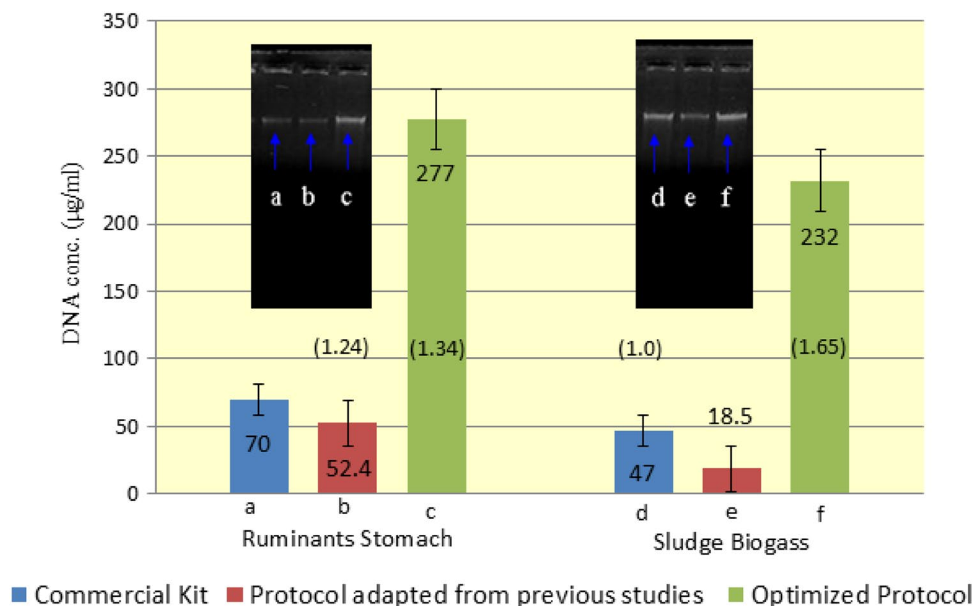

Fig. 1 Comparative analysis of DNA extraction methodologies viz. (a, d) commercial kit (b, e) protocol adapted from previous studies (c, f) optimized protocol developed in the present study, using samples from ruminant stomach and anaerobic biogas digested sludge. Bars represent DNA concentration (µg/ml) and figures in parenthesis represent DNA purity A260/280

corresponding A260/280 ratios ranging from 1.70 to 1.90 (as detailed Table 1).

In an attempt to validate the DNA extraction methodology, the microbial communities in extracted environmental DNA from diverse samples were identified using PCR–DGGE experiments, which was followed by nucleotide sequencing and their phylogenetic analysis. A diverse number of microbes were identified in each environmental sample. Environmental samples from anaerobic digester resulted in identification of a number of diverse microbial species which included eubacterial and archaeon species. Phylogenetic analysis, as depicted in Fig. 2a revealed occurrence of *Anaerobaculum* sp.; *Clostridium* sp.; *Coprothermobacter* sp.; *Prosthecochloris* sp.; *Methanomicrobiales* archaeon; *Conexibacter* sp.; *Xylanimonas* sp.; *Acidothermus* sp.; *Methanocalculus* sp.; *Methanogenic* archaeon; *Crenarchaeote* sp.; *Metallosphae; Pyrobaculum* sp. ; *Acidianus* sp. as most dominant microbial species.

The pine forest litter was associated with the presence of *Agrococcus* sp.; *Blastocatella* sp.; *Leifsonia* sp.; *Microbacterium* sp.; *Paenibacillus* sp.; *Psychrobacillus* sp.; *Saccharomonospora* sp.; *Glycomyces* sp. Their phylogentic relatedness is presented in a circular cladogram in Fig. 2b.

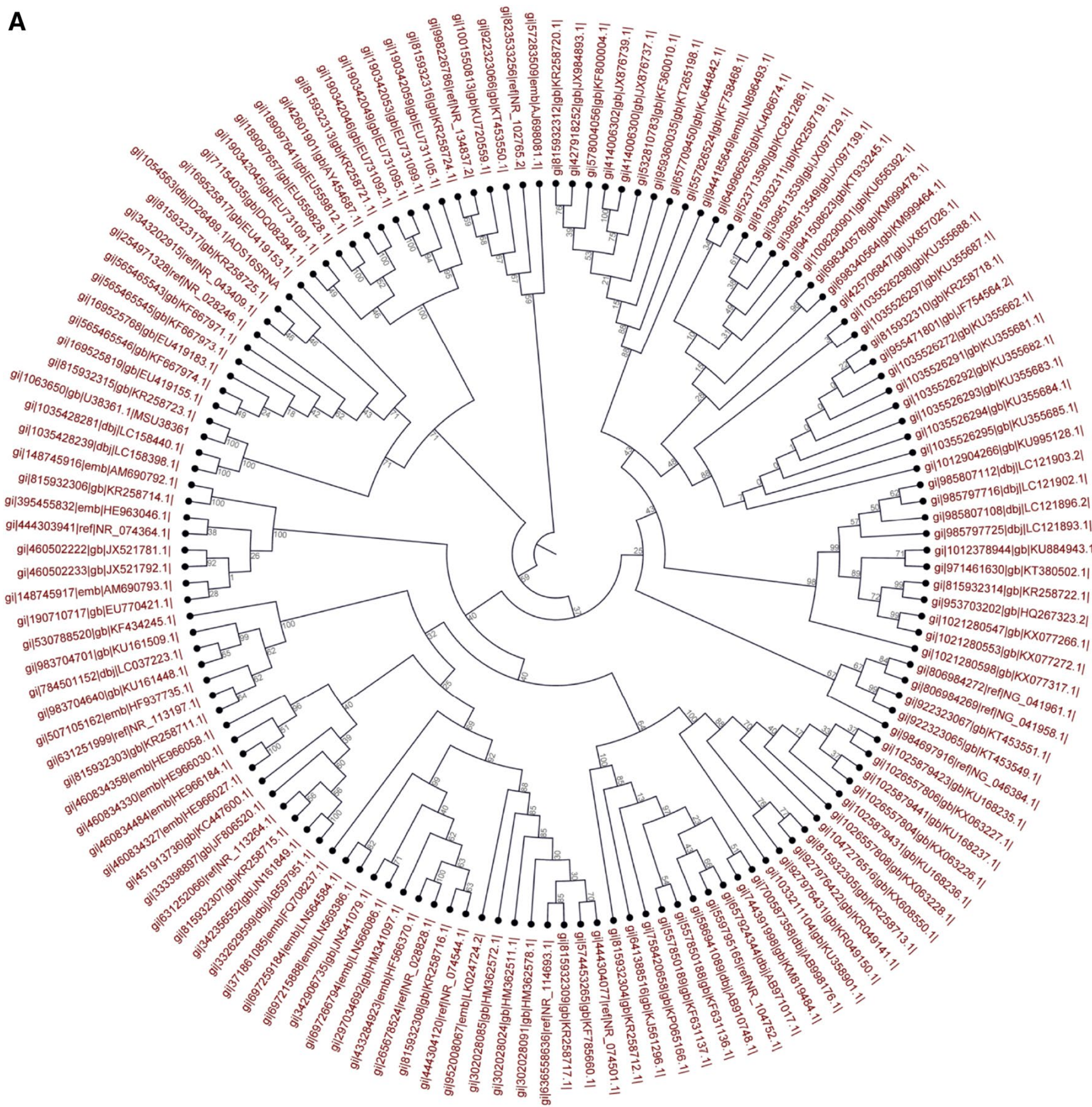

Fig. 2 **a** Circular cladogram depicting phylogenetic relatedness of microbial community in anaerobic digester. **b** Circular cladogram depicting phylogenetic relatedness of microbial community in pine forest litter. **c** Circular cladogram depicting phylogenetic relatedness of fecal microflora in an infant. **d** Circular cladogram depicting phylogenetic relatedness of microbial communities in traditional starter used in food fermentation. **e** Circular cladogram depicting phylogenetic relatedness of Pure culture (bacterial and fungal) isolates

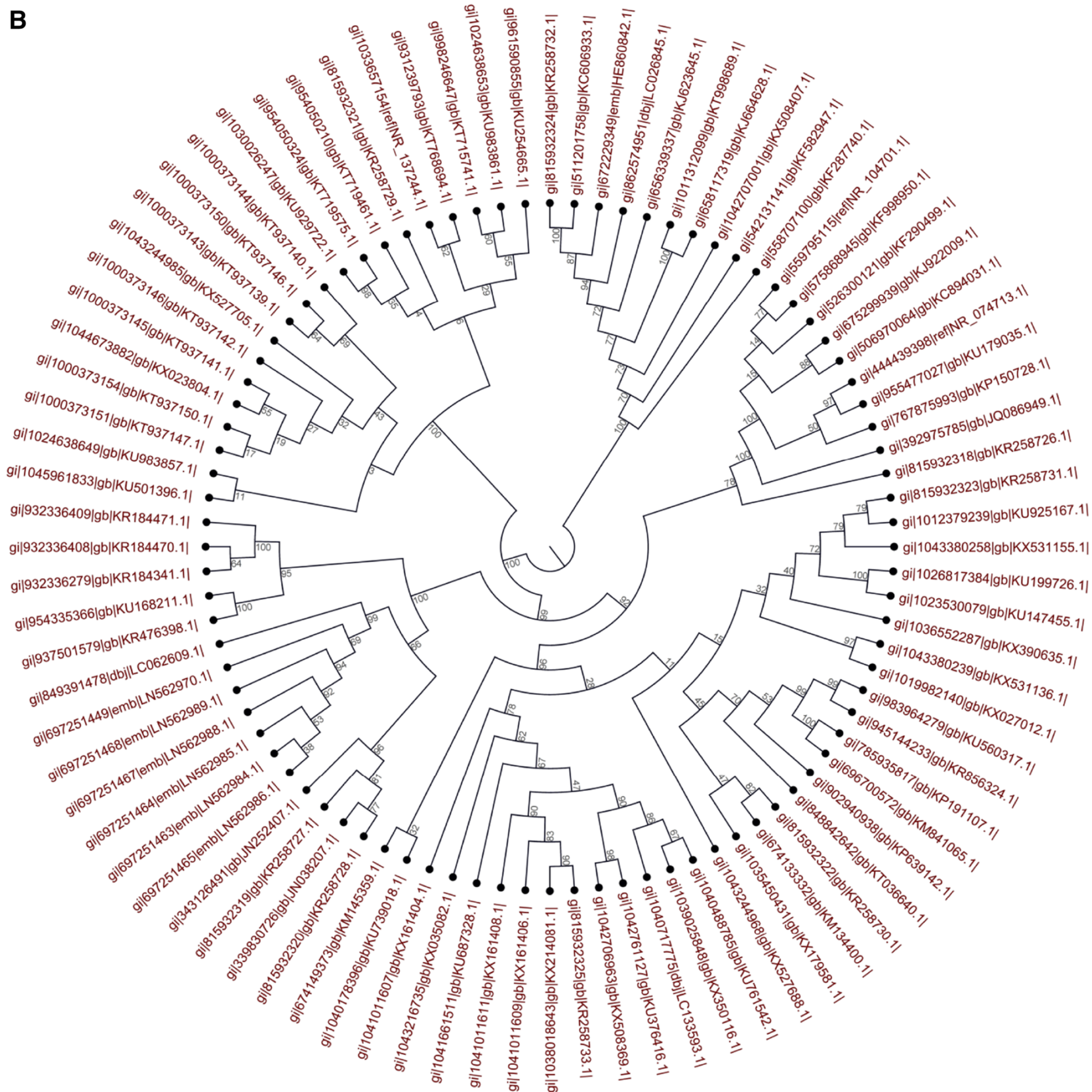

Fig. 2 (continued)

Using the improved DNA extraction methodology, fecal microflora of an infant revealed the presence of *Enterococcus* sp.; *Streptococcus* sp.; *Bifidobacterium* sp.; *Eubacterium* sp.; *Dialister* sp.; *Gemmiger* sp.; *Streptococcus* sp.; *Veillonella* sp.; *Brevundimonas* sp.; *Lactobacillus* sp.; *Streptococcus* sp.; *Clostridium* sp.; *Enterobacteriaceae* bacterium isolate and *Escherichia* sp. (Table S2) with phylogenetic relatedness is depicted in Fig. 2c.

The PCR–DGGE analysis of microbial communities in traditional starter used in food fermentation indicated the presence of *Klebsiella alba; Bifidobacterium pseudocatenulatum; Bifidobacterium kashiwanohense; Bifidobacterium catenulatum; Bifidobacterium angulatum; Alloscardovia omnicolens ;Bifidobacterium callitrichos; Bifidobacterium pseudocatenulatum; Bifidobacterium pseudocatenulatum; Bifidobacterium longum; Bifidobacterium breve; Streptomyces argenteolus; Dickeya chrysanthemi; Klebsiella* sp.; *Bifidobacterium* sp.; *Propionivibrio militaris* (Table S2).

Using DNA extracted from the methodology developed in the present study, we were able to prepare metagenomic library using 25 ng of nano drop quantified DNA. All the libraries showed an expected size of ~580–620 bp for V3–V4 region with an effective insert size of ~440–480 bp flanked on each size by adapters with a combined size of ~140 bp.

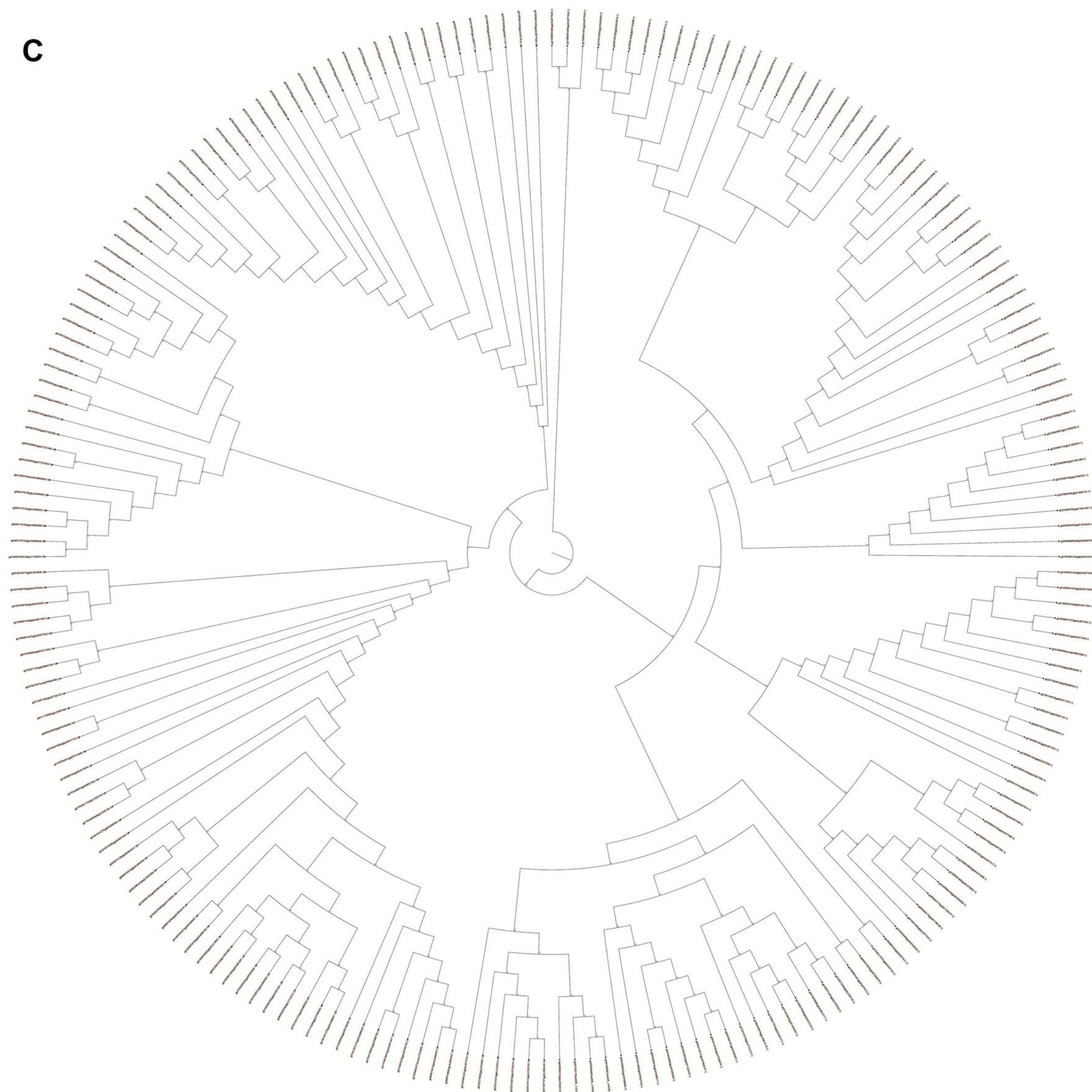

Fig. 2 (continued)

Discussion

The present study was performed to develop a rapid, cost effective methodology for extracting metagenomic DNA from anaerobic/aerobic environmental samples. The robustness of the developed methodology is exemplified by its ability to extract DNA not only from anaerobic samples but also from diverse environmental samples viz. fecal, soil, and starter cultures used for preparation of fermented food. Comparing metagenomic environmental DNA extraction performance from different environmental sources within the method developed (in the present study), ruminant stomach samples generated greater yields whereas, anaerobic digested sludge samples with higher purity (with values closest to ideal A260/280 values). Thus, indicating the effectiveness of the developed protocol in satisfactory removal of inhibitory contaminants viz. humic compounds and phenolics. The literature studies have highlighted that contaminating substances (humic compounds and phenolics) interfere with downstream processes, when isolating DNA from environmental samples, especially soil samples as these contaminates sediment and co-precipitate with the isolated DNA [17]. Different pre and post treatments of extracted DNA with aluminium sulphate, aluminium ammonium sulphate, calcium carbonate and polyvinylpolypyrrolidone have been reported for effective removal of humic acids [6, 18]. In the present study, the most best method for extraction of metagenomic DNA from anaerobic environmental samples was by supplementing "Buffer A" with 60% PEG 400; 20 mM

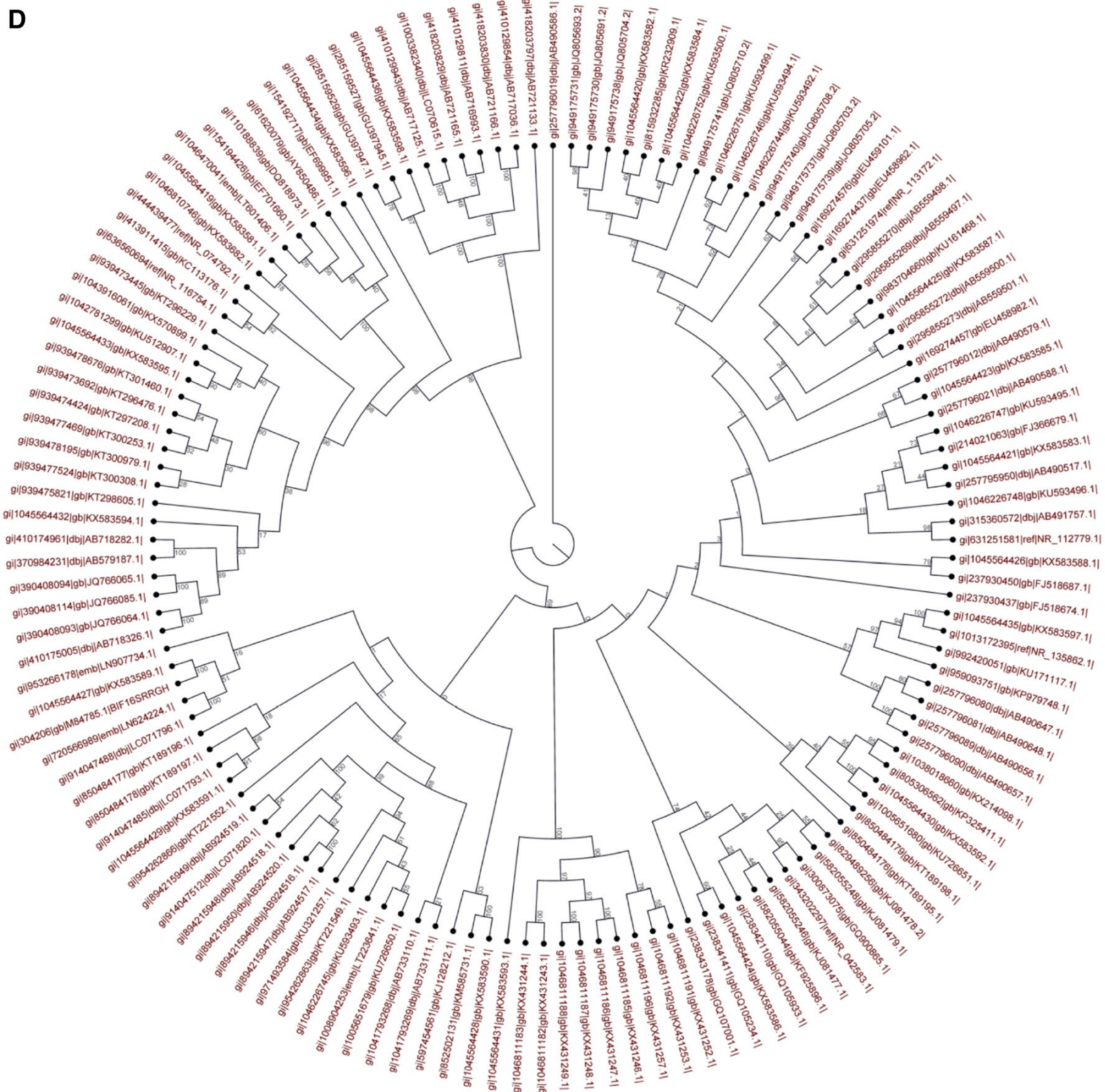

Fig. 2 (continued)

KOH that had a pH of 13.3–13.5; and 2% $CaCl_2$. “Buffer D” was supplemented with 0.5 volumes of 50% PEG 400 along with 0.1 volume of 5 M NaCl and 2% $CaCl_2$. As a result of use of alkaline PEG reagent (as indicated in the present study) the environmental DNA remained in solution without precipitation, which allowed optimal sample processing. Under saturating conditions, the addition of PEG at saturation level allowed the formation of large random coils in water whereas the accurate concentration of salt allowed the aggregation and precipitation of DNA [19, 20]. Purdy et al. [21] and LaMontagne et al. [22] also reported that PEG precipitation of DNA can also results in effective removal of humate contamination of DNA. The role of $CaCl_2$ in purification of the extracted environmental DNA is that it prevents the humic substances, to undergo oxidation forming quinones, which covalently bind to the DNA, thus hampering the DNA and Taq polymerase interaction [23].

Explaining the high recovery efficiency for DNA, using the methodology cited in here, the possibility of co-extraction of host DNA was ruled out by use of PCR application targeting microbe-specific primers. Spectrophotometric and PCR based assay were used as two endpoints to determine the quality and quantity of the extracted DNA. Generally, extraction of high quality of metagenomic environmental DNA for PCR assay from human faecal samples is challenging because of the simultaneous co-extraction of the interfering compounds (such as glycoproteins and phenolics), which often results in unproductive DNA yield [24, 25]. However, using the DNA extraction protocol developed, successful PCR amplification

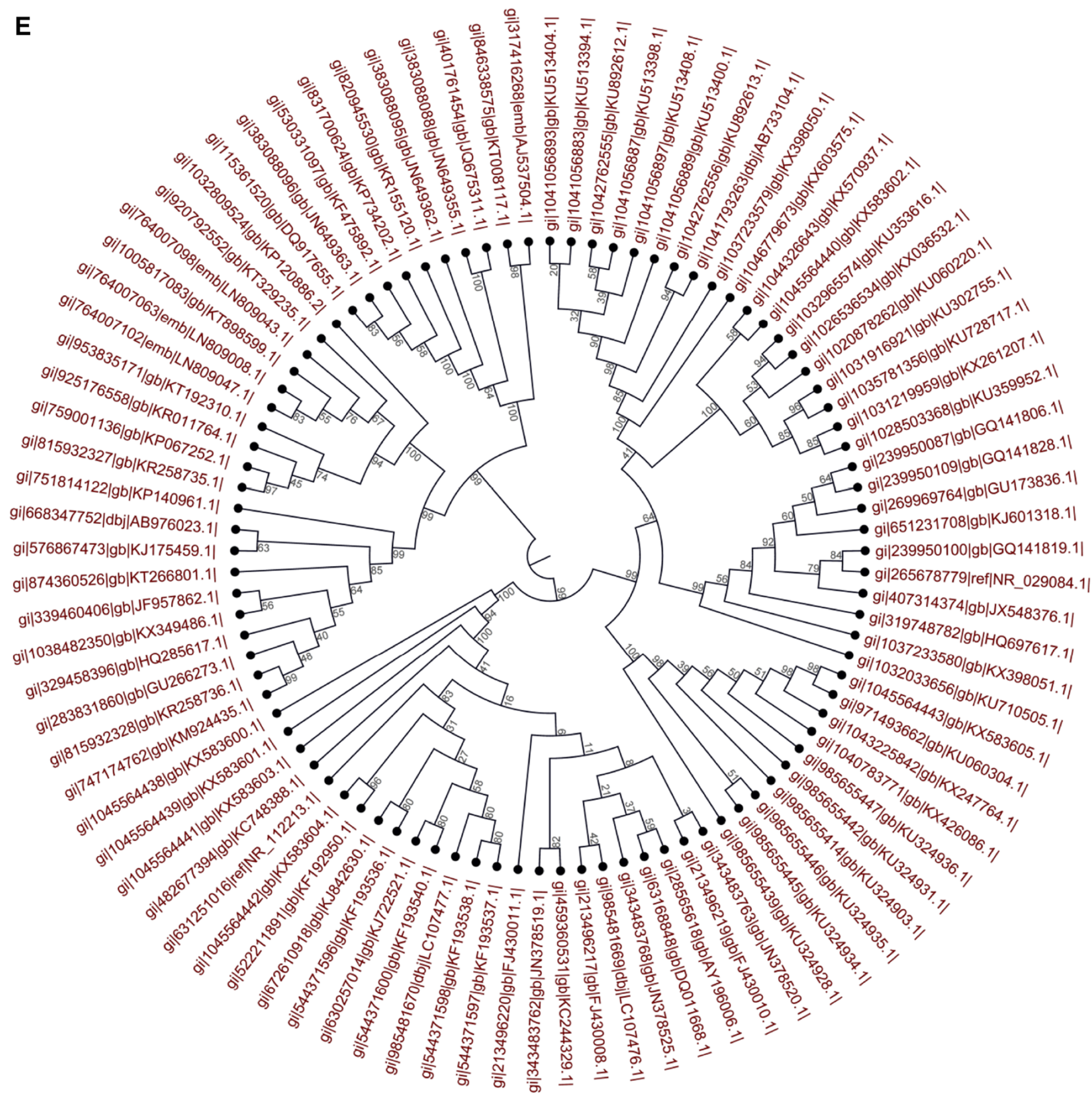

Fig. 2 (continued)

of faecal samples was achieved without any interference of the above said contaminants. In other environment samples tested such as soils and biogas digesta, humic acid is a major component that inhibits DNA extraction and further molecular analysis such as PCR based molecular techniques [26], and membrane hybridization reactions [27]. The humic acid is reported to form quinones which covalently binds to DNA and interferes with the amplification protocols [28]. Inhibition of PCR is often sourced to DNA contaminated with humic acids, co-extracted from the soil [26, 29].The DNA extraction methodology employed in the present study resulted in effective removal of interfering components as evident from the amplification of 16S rRNA and 18S–28S ITS genes, since no interference was observed in PCR based assay. The sequencing of the genes resulted in quality sequences based on which reliable phylogenetic analysis could be performed (as shown in Fig. 2a–e). DNA extracted using different methodologies viz. (A) commercially available kit; (B) DNA extraction protocols evaluated from previous research studies; (C) protocol developed in the present study, were used for generation of PCR–DGGE profiles for each anaerobic sample set (samples from anaerobic digested sludge samples and from ruminant stomach). The DGGE gel bands represent the most representative microbial species in the samples. The comparison of the DGGE profiles, within each sample set revealed identical banding patterns irrespective of the DNA extraction methodology used. The findings, thus clearly demonstrate that the quality of the DNA, extracted using methodology developed in the present study, is comparable to DNA extraction using

commercial kit, when identifying dominant microbial populations in environmental samples.

The protocol developed can be used for NGS of different samples as the NGS analysis allows deep sequence analysis of the mixed amplicon pools, with increased feasibility to incorporate NGS into ongoing research programmes [30]. Using DNA extracted from the methodology developed in the present study, we were able to prepare metagenomic library which were found to be suitable for sequencing on Illumina platform, as described in "Results" section. The findings, thus, further suggest that the DNA used for NGS library preparation was free from any inhibitory components and therefore, further authenticated the quality of the developed DNA extraction protocol.

The modified DNA extraction methodology for analysis of microbial community, associated with diverse environmental samples could be presented as a statistical model. The statistical model presented in the present manuscript has been adopted, with some modifications from the statistical models previously developed by Smart et al. [11] and Moore et al. [16].

In the presented statistical model, if the number of microbial species is detected using DGGE run (i), by DNA extraction methodology (j) at a detection rate of ($\lambda_{i,j}$), and probability of detecting only a single microbial species in an environmental sample ($k_{i,j}$) is:

$$\exp\left(-\lambda_{i,j}\, k_{i,j}\right) \tag{1}$$

Since the purity of extracted DNA from environmental samples (A_{260}/A_{280}) and the time duration involved in its extraction is critical when analyzing microbial communities associated with diverse environmental samples, the Eq. (1) should also consider two more factors viz. DNA purity ($Dp_{i,j}$) and time involved in ($t_{i,j}$) while assessing the probability of detecting only a single microbial species in an environmental sample ($k_{i,j}$) therefore, we propose modification in Eq. (1):

$$exp\left(-\lambda_{i,j}\, k_{i,j}\, Dp_{i,j}\, t_{i,j}\right) \tag{2}$$

Assuming that microbial species detection rate ($\lambda_{i,j}$) is independent of DNA extraction methodology, then the probability of not being able to detect even a single microbial species in (n) environmental samples using method (j) is:

$$\begin{aligned} Q_j &= \prod_{i=1}^{n} \exp\left(-\lambda_{i,j}\, k_{i,j}\, Dp_{i,j}\, t_{i,j}\right) \\ Q_j &= exp\left(-\prod_{i=1}^{n} -\lambda_{i,j}\, k_{i,j}\, Dp_{i,j}\, t_{i,j}\right) \\ Q_j &= exp\left(-A_j\right) \end{aligned} \tag{3}$$

where $A_j = \sum_{i=1}^{n} \lambda_{i,j}\, k_{i,j}\, Dp_{i,j}\, t_{i,j}$ is the expected number of microbial species detected in an environmental sample using DNA extraction methodology (j).

Our objective in the present study was to develop a DNA extraction methodology for diverse environmental samples that can result in maximum number of microbial species detection using limited budget inputs and minimum time involved in the entire DNA extraction procedure.

Therefore, firstly we found the minimum budgetary requirements (chemicals) that would result in good quality and pure environmental DNA (free from DNA and PCR inhibitory components). Using the extracted DNA, we expected to achieve DGGE banding patterns (i.e. number of microbial species, since each band in DGGE profile is assumed to represent an individual microbial species) similar to the ones observed in DGGE profiles, when DNA extraction method using commercially available kit was used. This would result in minimizing the expected probability of failed detection E[Q] of microbial species using modified DNA extraction procedure, developed in the present study. Assuming (X) to be a normal random variable, m as mean; v as variance; the cumulative density function of $Q_j = \exp\left(-A_j\right) = \exp\left(-\exp\left[X\right]\right)$ can be given by:

$$\begin{aligned} F_{(qj)} &= P_r(Q_j < q_j) \\ &= 1 - \frac{1}{2}\left(1 + erf\left[\frac{-m + \ln\left[-\ln\left[q\right]\right]}{\sqrt{2}\sqrt{v}}\right]\right) \end{aligned}$$

where erf (), is an error function.

Assuming that $F_{(qj)}$ denotes the probability density function of q_j;

$$F_{(qj)} = \frac{dF\ (qj)}{dqj}$$

The expected value of Q_j is obtained is:

$$\begin{aligned} E\left[Q_j\right] &= \int_0^1 qj \frac{dF\ (qj)}{dqj} dqj \\ E\left[Q_j\right] &= \int_0^1 -\frac{e^{-\frac{(-m+\ln(-\ln(q)))^2}{2\,v}}}{\sqrt{2\pi}\sqrt{v}\ \ln(qj)}\, dqj \end{aligned} \tag{4}$$

Substituting the values of m and v will give expected value of Q_j. The resulting expression would be in terms of mean (μ) and variance (σ^2) of the microbial species detection rate (λ), number of environmental samples (n) and the time involved in the DNA extraction procedure (t).

As previously discussed, one of the objectives of developing the modified DNA extraction methodology involved, the use of minimized budget *(B)* and minimum time *(t)* spent on the entire procedure of DNA extraction, without compromising the microbial detection rate ($\lambda_{i,j}$). The total budget in the environmental DNA extraction greatly depends on chemical costs *(C)*, which is fixed cost. The time duration in which DNA extraction procedure is completed determines (*Kt*) signifies the variable cost, since it is critical for further experimentation. The equation for the expected value ofQ_j, therefore need to be modified in a manner, that it encompasses the above cited factors, which can be presented as follows:

$$B = n\,(C + K\,\mathrm{t}) \tag{5}$$

Solving Eq. (5) for value of *K*

$$K = \left(\frac{B}{n} - C\right)\left(\frac{1}{t}\right) \tag{6}$$

Substituting the value of K, results in an expression for the expected value of Q_j [11, 16], with inclusion of new factors (as reported in the previous studies) like purity of the extracted environmental DNA and the time involved in DNA extraction procedure.

For extraction of high quality metagenomic environmental DNA from diverse environmental samples, the study developed an extraction protocol using chemical and mechanical steps for cell lysis. The increased freeze–thaw cycles may have impacted success in high efficiency of DNA extraction. The combination of mechanical and chemical approaches for cell lysis has been reported as highly effective relative to chemical lysis only [25, 31–33]. Complexity in the nature of environmental sample due to proteins and non-nucleic organic compounds plays an important role in obtaining a high quality DNA, which necessitates extra measures to remove co-extracted contaminants [34]. The removal of these components is required for ecological and molecular PCR-based studies. The developed protocol is an efficient, rapid and free of inhibitory substances based methodology for extracting total metagenomic environmental DNA from diverse samples. The recent studies have also highlighted the importance of environmental DNA extraction especially considering the time and cost issues in the context of biodiversity conservation [35]. The use of simple, readily available reagents and supplies needed for the DNA based extraction protocol reduced the total cost than the commercial kit (1 USD/sample or 0.88 Euro/sample for present method versus 2.83 USD/sample or 2.48 Euro/sample for the commercial kit).

In conclusion, the developed protocol is cost effective, capable of extracting DNA from small sample size, time efficient (1.5–2.0 h protocol) and results in significantly higher DNA yield (4–8 times increased yield), when compared to other existing protocols and therefore, be used for high quality extraction of metagenomic environmental DNA from diverse environmental samples, which can be used to unearth previously unexplored diverse microbial communities.

Author contributions RM and GG designed the research, assessed and interpreted the results and prepared the manuscript. RM, SA, NS, KS, DS analyzed samples using the developed protocols. All authors read and approved the final manuscript.

Funding The research is funded by Department of Science and Technology, Government of India for Indo-Russian collaborative project “Elucidating the linkage between key limiting processes and microorganisms during anaerobic degradation of lignocellulosic waste” INT/RUS/RFBR/P-175.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interest to declare.

Ethical approval All experimental protocols used in the study were carried out in accordance with the guidelines and regulations of Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, India. The stool sample collection, storage and analysis were approved by Jaypee University institutional ethical and biosafety committee (IEC No. 22/2015).

Informed consent Informed consent was obtained from the parents of all the infants, for using their stool samples.

References

1. Schleper C, Jurgens G, Jonuscheit M (2005) Genomic studies of uncultivated archaea. Nat Rev Microbiol 36:479–488
2. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC et al (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. Proc Natl Acad Sci USA 10940:16213–16216
3. Zhou J, He Z, Yang Y, Deng Y, Tringe SG, Alvarez-Cohen L (2015) High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. MBio 61:e02288-14
4. Howeler M, Ghiorse WC, Walker LP (2003) A quantitative analysis of DNA extraction and purification from compost. J Microbiol Methods 54:37–45
5. Fortin N, Beaumier D, Lee K, Greer CW (2004) Soil washing improves the recovery of total community DNA from polluted and high organic content sediments. J Microbiol Methods 56:181–191
6. Dong D, Yan A, Liu H, Zhang X, Xu Y (2006) Removal of humic substances from soil DNA using aluminium sulfate. J Microbiol Methods 66:217–222
7. Zinger L, Chave J, Coissac E, Iribar A, Louisanna E et al (2016) Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. Soil Biol Biochem 96:16–19
8. Gerasimidis K, Bertz M, Quince C, Brunner K, Bruce A et al (2016) The effect of DNA extraction methodology on gut microbiota research applications. BMC Res Notes 9(1):365
9. Biggs J, Ewald N, Valentini A, Gaboriaud C, Dejean T et al (2015) Using eDNA to develop a national citizen science-based monitoring programme for the great crested new *Triturus cristatus*. Biol Conserv 183:19–28

10. Sigsgaard EE, Carl H, Møller PR, Thomsen PF (2015) Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. Biol Conserv 183:46–52
11. Smart AS, Weeks AR, van Rooyen AR, Moore A, McCarthy MA, Tingley R (2016) Assessing the cost-efficiency of environmental DNA sampling. Methods Ecol Evol 7(11):1291–1298
12. Whitehouse CA, Hottel HE (2007) Comparison of five commercial DNA extraction kits for the recovery of *Francisella tularensis* DNA from spiked soil samples. Mol Cell Probe 21:92–96
13. Holben WE (1994) Microbiological and biochemical properties: isolation and purification of bacterial DNA from soil. In: Weaver RW, Angle S, Bottomley PD (eds) Methods of soil analysis. Soil Science Society of America, Madison, pp 727–752
14. Dempster EL, Pryor KV, Francis D, Young JE, Rogers HJ (1999) Rapid DNA extraction from ferns for PCR-based analyses. BioTechniques 27:66–68
15. Mahajan R, Nikitina A, Nozhevnikova A, Goel G (2016) Microbial diversity in an anaerobic digester with biogeographical proximity to geothermally active region. Environ Technol 37(21):2694–2702
16. Moore AL, McCarthy MA, Parris KM, Moore JL (2014) The optimal number of surveys when detectability varies. PLoS ONE 912:e115345
17. Tsai Y, Olson BH (1992) Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. Appl Environ Microbiol 58:2292–2295
18. Sagova-Mareckova M, Cermak L, Novotna J, Plhackova K et al (2008) Innovative methods for soil DNA purification tested in soils with widely differing characteristics. Appl Environ Microbiol 74:2902–2907
19. Lerman LS (1971) A transition to a compact form of DNA in polymer solutions. Proc Natl Acad Sci USA 688:1886–1890
20. Bloomfield VA (1997) DNA condensation by multivalent cations. Biopolymers 443:269–282
21. Purdy KJ, Embley TM, Takii S, Nedwell DB (1996) Rapid extraction of DNA and rRNA from sediments by a novel hydroxyapatite spin-column method. Appl Environ Microbiol 62:3905–3907
22. LaMontagne MG, Michel FC, Holden PA, Reddy CA (2002) Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. J Microbiol Methods 493:255–264
23. Siddhapura PK, Vanparia S, Purohit MK, Singh SP (2010) Comparative studies on the extraction of metagenomic DNA from the saline habitats of Coastal Gujarat and Sambhar Lake, Rajasthan India in prospect of molecular diversity and search for novel biocatalysts. Int J Biol Macromol 473:375–379
24. Monteiro L, Bonnemaison D, Vekris A, Petry KG, et al (1997) Complex polysaccharides as PCR inhibitors in feces: *Helicobacter pylori* model. J Clin Microbiol 354:995–998
25. Vo AT, Jedlicka JA (2014) Protocols for metagenomic DNA extraction and Illumina amplicon library preparation for faecal and swab samples. Mol Ecol Res 146:3–1197
26. Tebbe CC, Vahjen W (1993) Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and yeast. Appl Environ Microbiol 59:2657–2665
27. Alm EW, Zheng D, Raskin L (2000) The presence of humic substances and DNA in RNA extracts affects hybridization results. Appl Environ Microbiol 66:4547–4554
28. Young CC, Burghoff RL, Keim JG, Minak-Berbero V et al (1993) Polyvinylpyrrolidone–agarose gel electrophoresis purification of polymerase chain reaction amplifiable DNA from soils. Appl Environ Microbiol 59:1972–1974
29. Rondon MR, Goodman RM, Handelsman J (1999) The earth's bounty: assessing and accessing soil microbial diversity. Trends Biotechnol 17:403–409
30. Neiman M, Lundin S, Savolainen P, Ahmadian A (2011) Decoding a substantial set of samples in parallel by massive sequencing. PLoS ONE 6:17785–17792
31. Ariefdjohan MW, Savaiano DA, Nakatsu CH (2001) Comparison of DNA extraction kits for PCR-DGGE analysis of human intestinal microbial communities from fecal specimens. Nutr J 9:23–30
32. Salonen A, Nikkilä J, Jalanka-Tuovinen J, Immonen O et al (2010) Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. J Microbiol Methods 812:127–134
33. Smith B, Li N, Andersen AS, Slotved HC, Krogfelt KA (2011) Optimising bacterial DNA extraction from faecal samples: comparison of three methods. Open Microbio J 5:14–17
34. Jedlicka JA, Sharma AM, Almeida RPP (2013) Molecular tools reveal diets of insectivorous birds from predator fecal matter. Cons Gen Res 5:879–885
35. Stewart K, Ma H, Zheng J, Zhao J (2017) Using environmental DNA to assess population wide spatiotemporal reserve use. Conserv Biol 31(5):1173–1182. https://doi.org/10.1111/cobi.12910