

# Whole metagenome analysis of extreme environment samples by quasi-linear whole genome amplification

## Abstract

Microbiomes are gaining a lot of importance owing to its impact in the human health and other life processes. The significance of microbes in the deep sea is often undermined but it holds the key for the environmental changes that affects our planet. Recently scientists are investigating the diversity of microbial life in the sea to learn its role in the ocean processes, its interactions with other marine life, and its importance to humans. Emergence of cutting edge genomics and molecular biology tools has propelled the field of metagenomics to study the microbial populations without the need to culture them. In spite of these growth studies of extreme environmental samples with very low nucleic acid concentration remains still a major challenge. A whole genome amplification method with less bias is required to get over this problem. Commercially available single cell whole genome amplification kits don't perform well with the low concentration of samples. A minimum of 100ng of DNA is the input requirement for these kits and some of the extreme environment samples can only yield 1-10ng of DNA. Our aim was to devise an efficient and cost effective workflow that could work with very low concentration of metagenomic DNA. We have adopted a quasi-linear whole genome amplification methodology, Multiple Annealing and Looping Based Amplification Cycles (MALBAC) that was earlier used in the single cell genomic studies. MALBAC methodology was carried out using 4ng of the deep sea sediment DNA and it was compared against the NEB Next Ultra DNA Library prep kit using 125ng of deep sea sediment DNA. The data generated were analysed and compared for the microbial diversity and gene function annotations. The results obtained through MALBAC methodology were comparable to that of the commercial kit based procedure utilizing nearly 125ng of DNA. The MALBAC based whole metagenome analysis of deep sea sample from Arabian Sea revealed the immense microbial diversity of unexplored extreme environment. The study also identified the functional genes coding for valuable enzymes and proteins that could be used in the biotech industry. This study has showed that MALBAC is a cost effective procedure for extreme environment samples containing very low concentration of DNA.

**Keywords:** MALBAC, deep sea, amplification, metagenome, extreme environment

Volume 4 Issue 5 - 2016

Beena PS, Rajadurai CP, Aswathy S, Tintu Joy, Saju Michel, Sam Santhosh, Dhinoth Kumar Bangarusamy

SciGenom Labs Pvt Ltd, India

**Correspondence:** Beena PS, SciGenom Labs Pvt Ltd, 43A, SDF 3rd Floor, CSEZ, Kakkannad, Kerala, India, Tel +91 9446971288, Fax 04842413398, Email beenaps@scigenom.com, beena.santhosh@gmail.com

**Received:** October 25, 2016 | **Published:** December 20, 2016

**Abbreviations:** MALBAC, multiple annealing and looping based amplification cycles; DNA, deoxyribonucleic acid; MDA, multiple displacement amplification method; ngS, next generation sequencing

## Introduction

The deep sea is an extreme unique environment with low temperature and high pressure. Similar to the ocean surface, several microbes living in the deep sea have specialized genomic features associated with their adaptation to extreme conditions. Deep sea is such an ecosystem in which 99% remains unexplored compared to other ecosystems studied. This genomic divergence can lead to difficulty in categorization, characterization and in turn exploration of biological treasure of deep sea for the benefit of human kind. Difficulty in culturing majority of microbes in an ecosystem prevents the potential study of the specific system. There are many molecular methods to explore these diverse potential. Cultivable organisms and its potential could be studied without any fail if the samples are available. But uncultivable microbes remain as a question mark.

Culture independent techniques based on nucleic acid extraction from the ecosystem provide information on community structure and diversity. Hence microbial diversity in deep sea has the capability to tolerate nutrient poor environments, high concentration of salts and temperature variations. The potential enzymes or metabolites produced by these organisms would be much more stable and useful than those produced by terrestrial organisms.

The microbiome studies like Earth microbiome<sup>1</sup> and Human gut microbiome<sup>2</sup> the entire DNA of a sample is extracted and sequenced to determine its taxonomy and to annotate the genes present in sample and overall functionalities and pathways present in the specific environment. For a whole metagenome study the most critical aspect is isolation of quality and quantity DNA from the environmental sample. The sample features differ from place to place, as the salinity and other mineral content differs, and the extraction protocols could not be standardized for each and every location. The success of whole genome metagenome sequencing relies on the quality and amount of DNA obtained from the sample. Even though there are many extraction protocols that promises high yield of DNA, there are cases of extreme environments from which we fail to isolate quality and

quantity DNA for high through put sequencing. Here we come up with an amplification procedure which is generally used for single cell genomics which could get rid of the problem of low quality and quantity of DNA. The first metagenomic studies attempted using high-throughput sequencing method was massively parallel 454 pyrosequencing.<sup>3</sup> Three other technologies commonly applied to environmental sampling are the Ion Torrent Personal Genome Machine, the Illumina MiSeq or HiSeq and the Applied Biosystems SOLiD system.<sup>4</sup>

In the late 1970s, Carl Woese proposed the use of ribosomal RNA genes as molecular markers for life classification.<sup>5</sup> This idea in conjunction with the Sanger automated sequencing<sup>6</sup> method revolutionized the study and classification of microorganisms. Later, advances in molecular techniques were applied to microbial diversity studies to a “new uncultured world” of microbial communities. Some of these techniques, which had a remarkable impact, were the polymerase chain reaction (PCR), rRNA genes cloning and sequencing. Polymerase chain reaction (PCR) based amplification have sequence dependent bias due to the exponential amplification with random primers.<sup>7–9</sup> Multiple displacement w29 DNA polymerase with random primers under isothermal conditions is widely applied in single-cell genomics projects.<sup>10–12</sup> But amplification bias still exists in Multiple Displacement Amplification method (MDA). An alternative method of whole genome amplification, so-called multiple annealing and looping-based amplification cycles (MALBAC), emerged recently.<sup>13,14</sup>

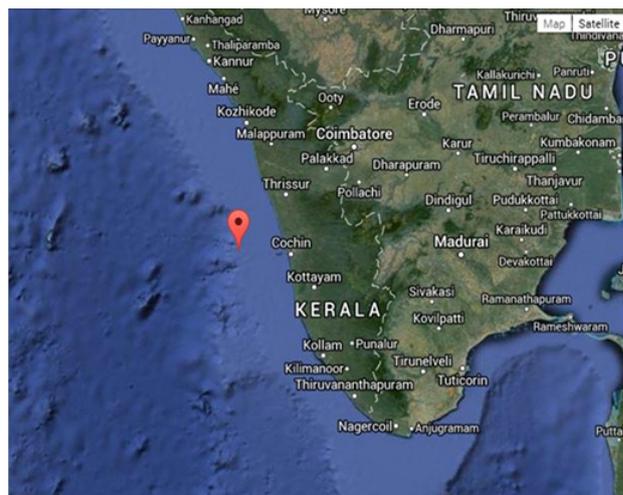
Combining advantage features of MDA and tweaked PCR, MALBAC substantially reduces experimental bias related to non-linear amplification. Amplification with MALBAC is initiated with a group of random primers that can evenly hybridize to the templates at 0°C. MALBAC primers are 35 nucleotides long<sup>13</sup> which have a common 27-nucleotide sequence and 8 variable nucleotides. The common nucleotide sequence is GTG AGT GAT GGT TGA GGT AGT GTG GAG. The 8 variable nucleotides anneal randomly to the single stranded genomic DNA molecule. After a 5-cycle initial reaction, specific DNA polymerase derived from *Bacillus stearothermophilus* (Bst polymerase) with strand-displacement activity were used to generate semi-amplicons at 65°C. The same primers having complementary ends were then used to generate full amplicons after the annealing at 94°C. As the common sequence of random primer on terminal can form loop-like amplicons, enough quantity of DNA for sequencing will be obtained after 20 cycle regular PCR amplification. By this approach, MALBAC method can prevent a lot of random amplification bias.

We report on assessment of the ability of MALBAC protocol for generating good quality and quantity DNA for whole metagenomic sequencing approach to produce DNA profiles that can give taxonomic classification and gene annotation as a whole genome metagenome sequencing approach from high quality and quantity DNA. This methodology could be adopted for extreme environmental samples from which DNA isolation is critical.

## Materials and methods

### Sample collection

Sediment sample collected onboard on the FORV (Fishery and Oceanographic Research Vessel) “Sagar Sampada” the cruise no. 321 (during 03/12/2013 to 14/12/2013) along southwest coast of India. Sediment samples collected using a Smith McIntyre grab (having a bite area of 0.2m<sup>2</sup>) from 9°57'N, 75°42'E, continental shelf region of Arabian Sea at a depth of 10<sup>2</sup>m depth on 5/12/2013 (Figure 1).



**Figure 1** Location in Arabian sea where the deep sea sediment samples were collected.

### DNA isolation and MALBAC PCR

DNA was extracted from 50mg sediment using Ultra clean Soil DNA Isolation Kit (MO-BIO, USA) following the manufacturer’s recommendations. The quality and quantity of the extracted DNA was checked using agarose gel (1%) stained with sybr safe and Qubit dsDNA HS assay kit (Invitrogen, USA) respectively followed by the MALBAC procedure.

50µl of a linear pre-amplification mix containing 36µl H<sub>2</sub>O, 5µl 10x standard Taq reaction buffer (NEB, USA), 2µl 10mM dNTP, 2µl 50mM MgSO<sub>4</sub>, and 2µl of 5mM of each MALBAC Primers was added to 4ng of deep sea sediment DNA. The reaction mixture was incubated at 94°C for 3min and immediately quenched on ice. 1µl of Bst polymerase (NEB,USA) was added to the reaction and run at 10°C for 45sec; 20°C for 45sec; 30°C for 45sec; 40°C for 45sec; 50°C for 45sec; 65°C for 2min; 94°C for 20sec, and immediately quenched on ice. Another 1µl of Bst polymerase (NEB, USA) was added to the reaction and run at 10°C for 45sec; 20°C for 45sec; 30°C for 45sec; 40°C for 45sec; 50°C for 45sec; 65°C for 2min; 94°C for 20sec; 58°C for 20sec and immediately quenched on ice. Thesecond pre-amplification step was repeated 5 times and an aliquot of 10µl from this mixture was further subjected to PCR amplification using 5µl 10X standard taq reaction buffer (NEB,USA), 1µl 10mM dNTP, 3.35µl 50mM MgSO<sub>4</sub>, 2µl Bst polymerase (NEB, USA), 3µl 10 uM MALBAC PCR primer (5’GTGAGTGTGGTTGAGGTAGTGTGGAG3’) and 25.65µl of H<sub>2</sub>O and incubated for 20 cycles at 94°C for 20sec; 59°C for 20sec; 65°C for 1min and 72°C for 2min. Final PCR reactions were then cleaned up using Invitrogen Pure Link custom PCR purification columns (Invitrogen, USA)and quantified using the Qubit HS DNA kit (Invitrogen, USA).

Simultaneously 125ng of the deep sea sediment DNA was processed using NEB Next Ultra DNA Library prep kit using the procedure recommended by the manufacturer. Libraries obtained from this procedure were compared against the MALBAC libraries in order to study the efficiency and sensitivity of MALBAC procedure on low concentration samples.

Primers used:

MALBAC1 5’GTG AGT GAT GGT TGA GGT AGT GTG GAG NNNNNTCG 3’

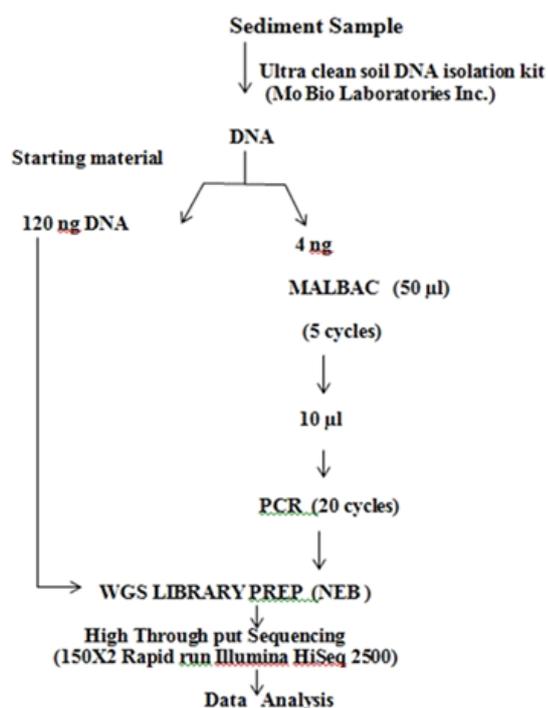
MALBAC2 5’GTG AGT GAT GGT TGA GGT AGT GTG GAG NNNNNAGT 3’

MALBAC3 5’GTG AGT GAT GGT TGA GGT AGT GTG GAG NNNNGAC 3’

MALBAC4 5'GTG AGT GAT GGT TGA GGT AGT GTG GAG NNNNNTCA 3'  
 MALBAC5 5'GTG AGT GAT GGT TGA GGT AGT GTG GAG NNNNNCAT 3'  
 MALBAC6 5'GTG AGT GAT GGT TGA GGT AGT GTG GAG NNNNNGAT 3'  
 MALBAC7 5'GTG AGT GAT GGT TGA GGT AGT GTG GAG NNNNNTAG 3'  
 MALBAC8 5'GTG AGT GAT GGT TGA GGT AGT GTG GAG NNNNNGCT 3'  
 MALBAC PCR 5'GTG AGT GAT GGT TGA GGT AGT GTG GAG 3'

## Library preparation

The DNA sample (DS) from deep sea sediment was divided into two parts. One part of the sample with ~120ng concentration (DS1) was sheared on Covaris220 (Covaris, USA) instrument using the following program: 2min 10% duty cycle, intensity 5, 200cycles/burst and under frequency sweep. Sequencing library was then made using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® according to the manufacturer's protocol. Another part of the sample (DS2) with 4ng concentration was subjected to MALBAC procedure and 120ng of the resulting product was used to prepare library using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® following the manufacturer's instruction. Final libraries were quantified with the Qubit HS DNA kit, visualized on an Agilent Bio analyzer 2100, and the libraries were sequenced from both ends for 150 cycles on Illumina HiSeq 2500 using rue Seq SBS V3 kit (Figure 2).

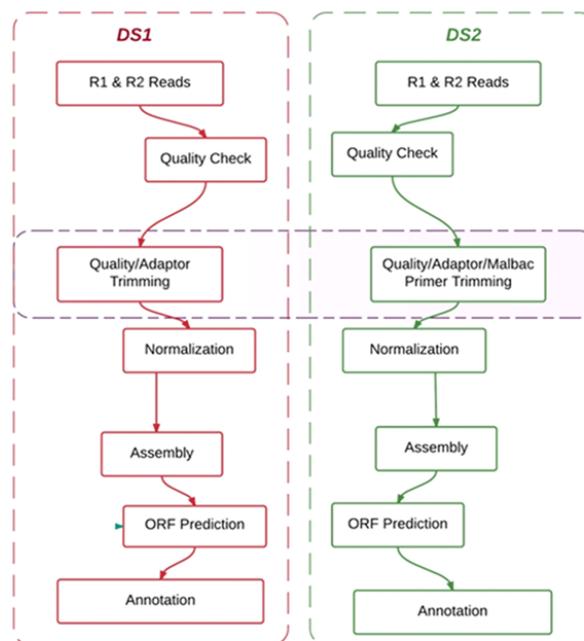


**Figure 2** Work flow from sample to Data.

## Quality trimming and alignment of sequence

The raw reads of samples (DS1 and DS2) were quality checked using FastQC tool<sup>15</sup> and the reads with Phred quality  $\geq 20$  were retained for further downstream analysis. The adapters were removed from the sample reads using cutadapt, version 1.8.<sup>16</sup> MALBAC primers were removed from raw reads of DS2 using an in-house perl program. Prior to assembly, all reads were normalized to the lowest read count (1.92million reads) among the two samples (DS1 & DS2), in order to avoid any assembly bias. Pre-processed sequences were the assembled with RayMeta<sup>17</sup> using a k-mer size of 31. The contigs

with less than 150bp length were filtered and protein coding regions (Figure 3) were predicted using Glimmer-MG v0.3.2<sup>18</sup> program. We performed BLASTX of all the predicted genes against the protein database using the BLAST version 2.2.29+<sup>19</sup> with an e value of 1e-5. The functional analysis of all hits were analysed using the KEGG<sup>20</sup> and SEED<sup>21</sup> options provided in the MEGAN software.<sup>22</sup>



**Figure 3** Overall bioinformatics workflow of metagenomics studies on DS1 and DS2 sample.

## Results and discussion

The focus of metagenomics is to assess the frequency of taxa and gene functions of microbes within an environment through DNA sequence data. The austerity of these evaluation lies on how accurately the frequency of the originating microbial population within the community can be captured. Next-generation sequencing (NGS) provides us exciting possibilities to achieve rapid and cost-effective identification of both cultivable and uncultivable microorganisms in the sample. The accuracy of thengS data relies on the quality and quantity of the Nucleic acid content derived from the samples.<sup>23</sup> The concentration of the nucleic acid in the sample plays a significant role in determining the specificity of the analysis and hence it is a challenge to analyse the extreme environment samples. In this study we have showed that MALBAC procedure accurately reflect the true composition and frequency of the microbial population and hence can be effectively used in the case of extreme environment samples with low nucleic acid content.

Sizeable number of studies exploring the microbial population of environments have focused on DNA profiling,<sup>24,25</sup> which requires efficient and unbiased DNA extraction and amplification procedures. These requirements are more critical in the case of ngs based metagenomic analysis of complex and extreme environments.<sup>23</sup>

In this study we wanted to explore the efficiency of MALBAC procedure and its compatibility with thengS applications to analyse the low concentration of nucleic acid obtained from the deep sea sediment sample (DS). In order to assess the specificity and sensitivity of the procedure we compared the MALBAC mediated library

preparation (DS2) with a direct procedure (DS1) that doesn't involve the genome amplification step. The 150 base paired reads from whole metagenome sequenced libraries were analyzed after filtering out low quality reads and trimming out the adaptors and primers. The reads were then normalized and assembled to contigs. The number of contigs and the N50 (Longest contig length of DS2 was 920 compared to 517 of DS1) was found to be more for MALBAC based procedure compared to the direct method indicating the quality of assembly and the efficient capture of the microbial population in the sample.<sup>26</sup>

The contig size length distribution differs for DS1 and DS2 (Figure 4) (Figure 5). MALBAC based sample has less contig length than the direct methodology, even though the longest contig was obtained in DS2. Microbial ecology has many tools for assessing species diversity. Rarefaction curves are used to estimate the coverage obtained from sampling (Figure 6). No significant difference observed between the number of leaves in taxonomy of DS1 and DS2.

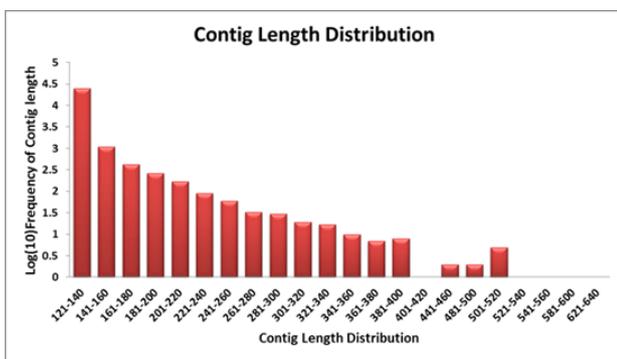


Figure 4 Length distribution of all contigs of sample DS1 after assembly.

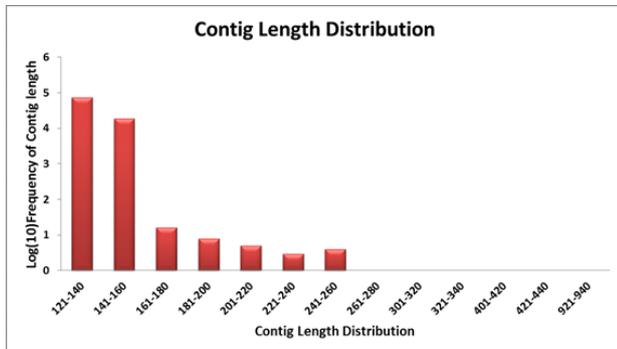


Figure 5 Length distribution of all contigs of sample DS2 after assembly.

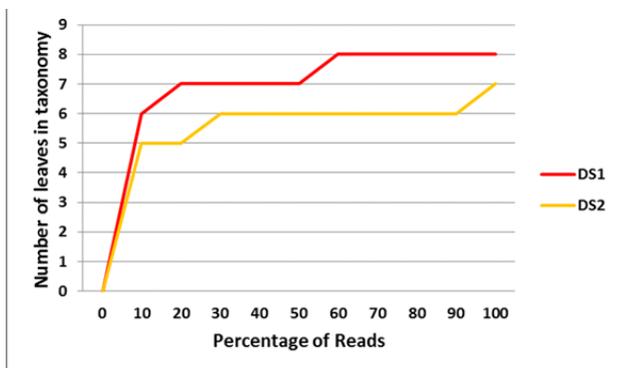


Figure 6 Rarefaction curve of sample DS1 and DS2.

The microbial community of the sediment was dominated by Proteobacteria, followed by Planctomycetes, Actinobacteria,

Firmicutes, Bacteroidetes, Cyanobacteria, Chloroflexi, Nitrospirae, Acidobacteria etc. The major representatives are found to be similar in both of the samples with slightly varied abundance (Figure 7) Microbial diversity in both of the methodology shows the presence of *Shewanella frigidimarina* Antarctic bacterial species which has the ability to produce Omega 3 fatty acid like Eicosapentaenoic acid which has medicinal properties. Other major representative in both methodologies is *Vibrio rumoiensis*, psychrophilic bacteria, which has high catalase activity.

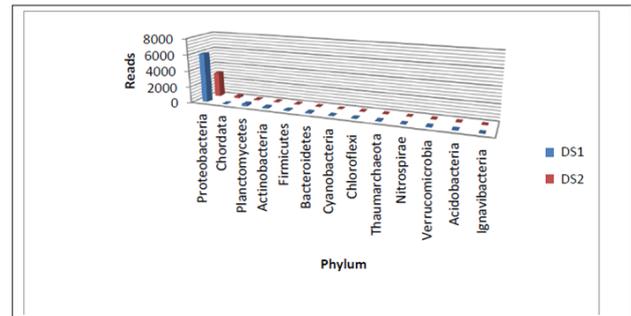


Figure 7 Phylum representations of DS1 and DS2.

One goal of sequencing-based metagenomic community analysis is the quantitative taxonomic assessment of microbial community compositions. In particular, relative quantification of taxons is of high relevance for metagenomic community comparison. Hence finding the relative abundance of species is very important in determining the metagenome of an extreme environment. Figure 8 shows the relative abundance of species in direct method and MALBAC approach. Even though quantitatively DS2 abundance is less, qualitatively all the major representations are present in DS2 which shows the usability of the approach in samples with low concentration.

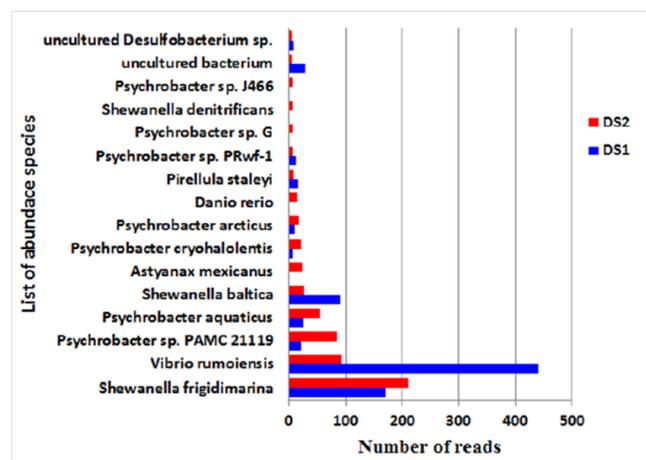
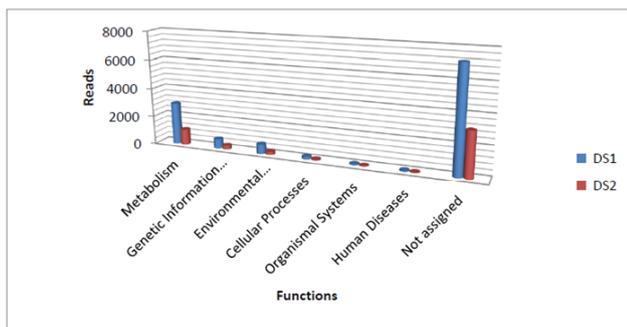


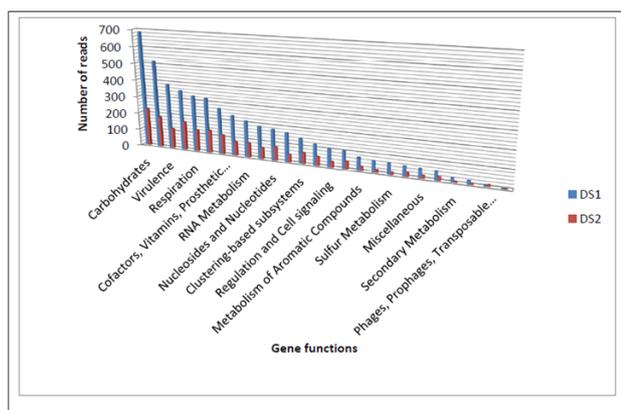
Figure 8 Relative abundance of species distributed among DS1 and DS2 Sample.

Comparison of KEGG functional classification results obtained from KEGG mapper (Figure 9) indicated that DS1 and DS2 were able to detect the prominent metabolic pathways in spite of the lower read abundance in DS2. The datasets were further annotated with SEED (Figure 10) indicating a similar trend for 12 metabolic pathways. Results showed that at the same sequencing depth, abundance of annotated genes in DS1 dataset were generally more than DS2. Similar to the findings about soil,<sup>27</sup> and marine microbial communities,<sup>28</sup> protein metabolism, carbohydrates, amino acids and derivatives, as well as miscellaneous were the four most abundant categories

in the global metabolism of microbial communities, suggesting the dominant roles of these categories in the microorganisms. These results indicated that MALBAC based procedure was able to detect almost all the metabolic pathways captured by the direct method, although there was a difference in the number of reads representing these pathways. BLASTX results were utilized for annotating *Vibrio sps* functional genes (camphor resistance protein CrcB, N- acetyl glutamate synthase, Valyl t RNA synthase, carboxyl phosphatase synthase, chitinase, recombinase RecA) and , *Shewanella sps* functional genes (heavy metal efflux pump, CRISPR associated CsyZ family protein etc) (Table 1).



**Figure 9** KEGG method based gene functional classification of DS1 and DS2 Sample. Functional classification with reference to number of reads.



**Figure 10** SEED method based gene functional classification of DS1 and DS2 Sample. Gene function with reference to number of reads

**Table 1** Assembly summary for Illumina data of DS1 and DS2 samples. It represents raw reads obtained, normalized read count, average read length obtained in bp, total number of contigs obtained from data, N50 and longest contig representation of DS1 and DS2 sample.

Assembly	DS1 sample	DS2 sample
Total raw reads	2,684,044	18,820,122
Normalized read count	964,712	964,712
Average read length (bp)	150	150
Total number of contigs	27639	93163
N50	517	920
Longest Contig	623	920

Whole-genome amplification strategy is indispensable for precious samples from extreme environments. Majority of the commercial kits in market focus on human samples requires template in excess

of 10ng and utilize a clonal amplification step via random primers that can result in a bias.<sup>29-31</sup> There are several multiple displacement amplification procedures (MDA) that has come out recently which uses random priming and the strand-displacing phi 29 polymerase under isothermal condition,<sup>32</sup> to provide improvements over the conventional PCR-based methods. These methods still exhibit considerable bias due to nonlinear amplification.

In order to overcome these obstacles we tried Multiple Annealing and Looping Based Amplification Cycles (MALBAC)<sup>13</sup> for the whole genome amplification of an extreme environmental sample and compared it with the high concentration DNA library preparation methodology to assess the uniformity, reproducibility and specificity. This methodology introduces quasi-linear preamplification to reduce the bias associated with nonlinear amplification. Picograms of DNA fragments (~10 to 100kb) from each individual in a population serve as templates for amplification with MALBAC. The amplification is initiated with a pool of random primers, each having a common 27-nucleotide sequence and 8 variable nucleotides that can evenly hybridize to the templates at 0°C. At an elevated temperature of 65°C, DNA polymerases with strand displacement activity are used to generate semi-amplicons with variable lengths (0.5-1.5kb), which are then melted off from the template at 94°C. Amplification of these semi-amplicons gives full amplicons which have complementary ends. The temperature is cycled to 58°C to allow the looping of full amplicons, which prevents further amplification and cross hybridizations. Five cycles of preamplification procedure are followed by exponential amplification of these full amplicons by PCR in order to generate micrograms of DNA required for next generation sequencing.<sup>13</sup>

## Conclusion

MALBAC strategy renders the specific amplification of target template and uniformity of coverage with very little bias even with extreme amplification. Except for the difference in read counts, all of the major phylum and species that are captured by the direct methodology were also represented by MALBAC based protocol. Gene functional classification study also indicates that the MALBAC protocol can perform well as the direct protocol and its reproducibility is reliable. This protocol is also cost effective and enables the use of standard reaction volumes. Hence MALBAC protocol can be utilized as a methodology for amplification of low quantity precious DNA from extreme environments for high through put sequencing and in turn help us to explore the uncultured potential of extreme environments for novel enzymes proteins and other metabolites.

## Acknowledgements

We thank Dr. Sarita G Bhat, PI, and Centre for Marine Living Resources & Ecology-ministry of Earth Sciences project (No. MOES/10-MLR/2/2007), Department of Biotechnology, Cochin University of Science and technology and Sagar Sampada Cruise no. 321 participants especially Dr. Raghu Zubin for providing samples for present study.

## Conflict of interest

“The study was funded by SciGenom Labs Pvt Ltd, Research and Development division. The authors declare that they have no competing interests financially and non-financially. All authors of manuscript are employed by SciGenom Labs Pvt Ltd and have worked from processing of sample until analysis of result. This does not alter the author’s adherence to all policies on sharing data and materials”.

## References

- Bravo A, Cases M, Queralt-Rosinach N, et al. A knowledge-driven approach to extract disease-related biomarkers from the literature. *BioMed research international*. 2014;2014:253128.
- Jensen K, Panagiotou G, Kouskoumvekaki I. Integrated text mining and cheminformatics analysis associates diet to health benefit at molecular level. *PLoS Comput Biol*. 2014;10(1):e1003432.
- Krallinger M, Valencia A. Text-mining and information-retrieval services for molecular biology. *Genome Biol*. 2005;6(7):224.
- Fleuren WW, Alkema W. Application of text mining in the biomedical domain. *Methods*. 2015;74:97-106.
- Abul Seoud RA, Mabrouk MS. TMT-HCC: a tool for text mining the biomedical literature for hepatocellular carcinoma (HCC) biomarkers identification. *Comput Methods Programs Biomed*. 2013;112(3):640-648.
- Donlan RM. Biofilms: microbial life on surfaces. *Emerg Infect Dis*. 2002;8(9):881-890.
- Doulgeraki AI, Di Ciccio P, Ianieri A, et al. Methicillin-resistant food-related *Staphylococcus aureus*: a review of current knowledge and biofilm formation for future studies and applications. *Res Microbiol*. 2016;S0923-2508(16)30083-3.
- Gupta P, Sarkar S, Das B, et al. Biofilm, pathogenesis and prevention—a journey to break the wall: a review. *Arch Microbiol*. 2016;198(1):1-15.
- Margarit y Ros I. *Streptococcus pyogenes* Pili. In: Ferretti JJ, Stevens DL, et al. editors. *Streptococcus pyogenes: Basic Biology to Clinical Manifestations*. Oklahoma City, USA; 2016.
- Vuong C, Otto M. *Staphylococcus epidermidis* infections. *Microbes Infect*. 2002;4(4):481-489.
- Vuong C, Voyich JM, Fischer ER, et al. Polysaccharide intercellular adhesin (PIA) protects *Staphylococcus epidermidis* against major components of the human innate immune system. *Cell Microbiol*. 2004;6(3):269-275.
- Otto M. Staphylococcal biofilms. *Curr Top Microbiol Immunol*. 2008;322:207-28.
- Goh KI, Cusick ME, Valle D, et al. The human disease network. *Proc Natl Acad Sci U S A*. 2007;104(21):8685-8690.
- Yildirim MA, Goh KI, Cusick ME, et al. Drug-target network. *Nat Biotechnol*. 2007;25(10):1119-1126.
- Durmus S, Cakir T, Ozgur A, et al. A review on computational systems biology of pathogen-host interactions. *Front Microbiol*. 2015;6:235.
- Spasic I, Ananiadou S, McNaught J, et al. Text mining and ontologies in biomedicine: making sense of raw text. *Brief Bioinform*. 2005;6(3):239-251.
- Roberts RJ. PubMed Central: The GenBank of the published literature. *Proc Natl Acad Sci U S A*. 2001;98(2):381-382.
- Rogers FB. Medical subject headings. *Bull Med Libr Assoc*. 1963;51:114-116.
- Jenssen TK, Laegreid A, Komorowski J, et al. A literature network of human genes for high-throughput analysis of gene expression. *Nat Genet*. 2001;28(1):21-28.
- Chavalarias D, Wallach JD, Li AH, et al. Evolution of Reporting P Values in the Biomedical Literature, 1990-2015. *JAMA*. 2016;315(11):1141-1148.
- Rebholz-Schuhmann D, Grabmuller C, Kavaliuskas S, et al. A case study: semantic integration of gene-disease associations for type 2 diabetes mellitus from literature and biomedical data resources. *Drug Discov Today*. 2014;19(7):882-889.
- Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504.
- Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25-29.
- Wang C, Zhang H, Zheng WM, et al. FALCON@home: a high-throughput protein structure prediction server based on remote homologue recognition. *Bioinformatics*. 2016;32(3):462-464.
- Mashiach E, Schneidman-Duhovny D, Andrusier N, et al. FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res*. 2008;36(Web Server issue):W229-W232.
- Schneidman-Duhovny D, Inbar Y, Nussinov R, et al. atchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res*. 2005;33(Web Server issue):W363-W367.
- Vuong C, Saenz HL, Gotz F, et al. Impact of the agr quorum-sensing system on adherence to polystyrene in *Staphylococcus aureus*. *J Infect Dis*. 2000;182(6):1688-1693.
- Li H, Liu C. Biomarker identification using text mining. *Computational and mathematical methods in medicine*. 2012;2012:135780.