

Isolation an Aldehyde Dehydrogenase Gene from Metagenomics Based on Semi-nest Touch-Down PCR

Rong Chen · Chenglu Li · Xiaolin Pei ·
Qiuyan Wang · Xiaopu Yin · Tian Xie

Received: 22 November 2012 / Accepted: 10 April 2013 / Published online: 10 May 2013
© Association of Microbiologists of India 2013

Abstract Culture-independent approaches to analyze metagenome are practical choices for rapid exploring useful genes. The *mg-MSDH* gene, acquired from the hot spring metagenomic, was retrieved full lengths of functional gene using semi-nest touch-down PCR. Two pairs of degenerate primers were used to separate seven conserve partial sequences by semi-nest touch-down PCR. One of them showed similarity with aldehyde dehydrogenase was used as a target fragment for isolating full-length sequence. The full-length *mg-MSDH* sequence contained a 1,473 bp coding sequence encoding a 490-amino-acid polypeptide and assigned an accession number JQ715422 in Genbank. The upstream sequences TAGGAG of the start codon (GTG), suggested that was a ribosome binding site. The coding sequence of *mg-MSDH* was ligated to pET-303 vector and the reconstructive plasmid was successfully overexpressed in *E. coli*. The purified recombinant *mg-MSDH* enzyme showed propionaldehyde oxidative activity of 3.0 U mg^{-1} at 37°C .

Keywords Metagenomic · Semi-nest touch-down PCR · Functional analysis

Introduction

Environment microorganisms contain a wide range of potentially useful genes of scientific and industrial interest.

More than 99 % of microorganisms are difficult to culture under laboratory condition, so it's limited to exploit by traditional culture-dependent methods. Recently, culture-independent approaches to analyze metagenome have been developed, including random shotgun sequencing and function-based screening. The former approach provides vast numbers of fragment sequences but with no phylogenetic or functional annotation leading to various pitfalls during analysis [1]. The latter approach is largely dependence on detecting of the enzymatic activity with a high-throughput screening technology, but it's mainly determined by sensitive and efficient screening methods and cost large work forces [2].

Another metagenomic approach, PCR-based techniques which based on the sequence information of known genes has been developed. Morimoto and Fujii retrieved benzoate 1,2-dioxygenase and chlorocatechol 1,2-dioxygenase genes from a 3-chlorobenzoate (3CB)-dosed soil sample using PCR denaturing gradient gel electrophoresis (DGGE-PCR) [3]. Insertion sequence-based cassette PCR (IS-PCR) and Cassette PCR were successful in the isolation of the complete genes from environmental DNA and tap into a vast pool of unexploited genetic diversity [4]. In our previous study, Wang et al. developed a strategy for collecting 23 lipase gene fragments for DNA shuffling from environmental samples by truncated metagenomic gene specific PCR (TMGS-PCR) [5]. Thus, PCR-based techniques are practical choices for rapid exploring useful genes from metagenomic DNA because it specifically amplifies the target gene [3].

Aldehyde dehydrogenases are NAD-linked enzymes (with some NADP-accepting examples) that act on a broad variety of aldehyde substrates, converting them to the corresponding carboxylic acid. Aldehyde dehydrogenases play an important role as a detoxifying enzyme to

R. Chen · C. Li · X. Pei · Q. Wang · X. Yin (✉) · T. Xie (✉)
Center for Biomedicine and Health, Hangzhou Normal
University, Hangzhou 310012, China
e-mail: yinxp.hznu@gmail.com

T. Xie
e-mail: tianxie.hznu@gmail.com

overcome the harmful effects of aldehydes that are widely present in the environment, especially in industrial wastewaters. [6]. Aldehyde dehydrogenases incorporate with other enzymes can convert cytotoxic aldehydes to organic acids and the bioprocess is environmental friendly, replacing current chemical process [7]. Moreover, the oxidation of aldehyde dehydrogenase in microbial fermentation to produce ethanol is a crucial step [8]. To our knowledge, most information about aldehyde dehydrogenase is based on cultured microorganisms. There is no an effective and rapid way to retrieve aldehyde dehydrogenase from metagenomic DNA. Here, we present a new approach to get out new aldehyde dehydrogenase genes from metagenomic DNA using culture-independent PCR and metagenome walking. The full-length sequence was cloning and expressed in *E. coli*. Its aldehyde oxidative activity was identified.

Materials and Methods

All standard molecular biology techniques were performed essentially as described previously [9]. The samples were collected from hot spring water and underwater soil located in Jiuqu, Guangxi, China. The temperature was range from 65 to 68 °C. The genomic DNA was directly isolated from samples according to the instruction using Power Max™ Soil DNA Isolation Kit. Metagenome DNA was stored at –80 °C until use. The quality of genome DNA was checked by pulsed field gel electrophoresis, PFGE). The migration was performed on 0.8 % agarose gel in 0.5× Tris Borate EDTA (TBE) buffer in a Bio-Rad CHEF-DRII apparatus at 14 °C for 16 h, and with switch times ramped from 10–25 s. The lambda ladder PFG marker and the low range PFG marker were used (BioLabs). The purify of genome DNA was analyze by UV260 and 280 nm.

Development of Degenerate Primers

Forty amino acid sequences typically representing aldehyde dehydrogenase subunits range in size from about 450 to 520 amino acids were obtained from GenBank. Those sequences include all invariant residues and conserved segments between Arg25 and Gly414 [10]. These sequences were aligned by the ClustalX module within Geneious Pro 3.0.5 (Biomatters Ltd.)

According to the sequences of ten most conserved motifs and the principle of primer designing, the motif 4, 6 and 10 were selected (Fig. 1). Examining the corresponding DNA sequences for evidence of codon bias and controlling 1,024-fold degeneracy, we designed degenerate primers (Table 1). The primers were synthesized by integrated DNA Technologies (Coralville, IA).

PCR Amplification and Partial Sequences Analysis

For improve the PCR products specificity, we used the semi-nested PCR technique similar to a nested PCR except that in the second PCR one of the primers is a primer that was used in the first PCR. Meanwhile, touch down PCR was used for optimizing PCRs to obtain the specificity and decreasing mismatches between primers and template [11]. The first PCR mixture contained 1.25 U of TaKaRa Ex Taq™ HS (Takara Bio Inc., Shiga, Japan), 1× Ex Taq buffer, 0.2 mM of each dNTP, 0.8 μM primer pair, 20 μg of bovine serum albumin, and 1 μl of soil DNA in a total volume of 50 μl. Conditions for the first PCR were as follows: initially 94 °C for 2 min; 25 cycles of 30 s each at 94, 48–55 °C, and 72 °C; and finally 72 °C for 1 min. the first PCR products diluted 50-fold and used for the second PCR as the template. Conditions for the second PCR were as follows: initially 94 °C for 2 min; 30 cycles of 30 s each at 94, 50 to 60 °C, and 72 °C; and finally 72 °C for 1 min.

The Target Bands Were Excised and Subcloned to PMD 18-T Vector for DNA Sequence Analysis

The homologues of the nucleotide and deduced amino acid sequence of partial sequences were identified using the BLASTn and BLASTx program with the default parameters [12]. Sequence alignments and similarity comparisons were initially conducted by the Clustal W method with default parameters [13] and the final alignment was manually performed with the multiple sequence editor MegAlign (DNAStar, Madison, WI, USA).

Metagenome Walking

One ALDH fragment sequence was selected and subsequently used to design the gene-specific primers(TSP) for metagenomic walking to clone of the 5' and 3' sequences. The most important principle of design TSP is that melting temperature(TM) of TSP1 was between 55 and 60 °C and that of TSP2-3 was between 60 and 65 °C. DNA Walking SpeedUp™ premix kits (Seegene, Inc. Seoul, Korea) were used for metagenome walking. The procedure is firstly based on one DNA walking PCR for amplifying target unknown sequence using DW-ACP™ and TSP1, following by two rounds of nested PCR using DW primer and rested TSP2-3 from previous PCR diluted products. One microliter of DNA was used as the first PCR template in each walk. Walking was performed both upstream and downstream from each target band according to the instructions of the manufacturer (Seegene, Inc.). Purified products were subjected to sequence analysis, and resultant sequences were used to design new primers for further walking or for retrieval of the full sequence of the target gene.

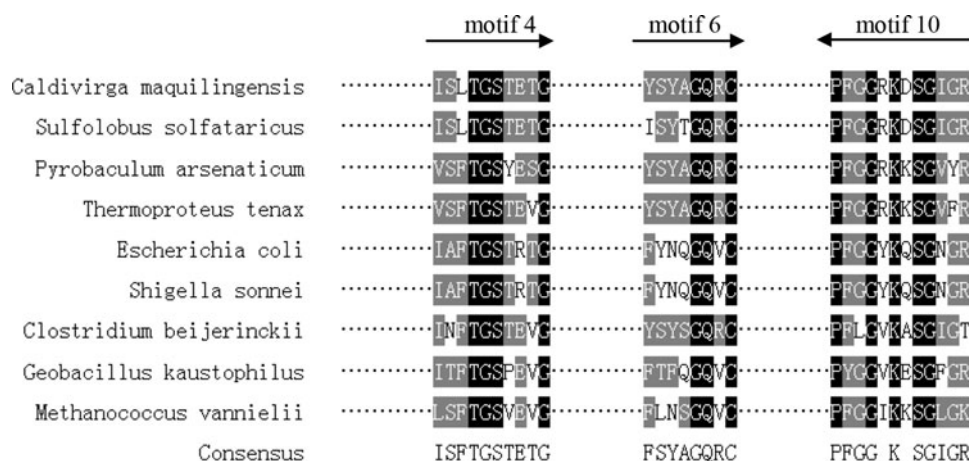


Fig. 1 Specific primers design for aldehyde dehydrogenase genes according to conservative motifs. The locations of the three primers and the direction of amplification are indicated by arrows. microbial source and GenBank number were given as follows: *Caldivirga maquilingsensis* (YP_001540801); *Sulfolobus solfataricus*

(NP_344502); *Pyrobaculum arsenaticum* (YP_001152801); *Thermoproteus tenax* (CAA71651); *Escherichia coli* (BAB35300); *Shigella sonnei* (AAZ88514); *Clostridium beijerinckii* (YP_001309399); *Geobacillus kaustophilus* (BAD76204); *Methanococcus vannielii* (ABR54702)

Table 1 The primers for PCR and walking

Primer name	For purpose	Primer sequence
F motif4	The first semi-nested PCR	5-TTHAGVGGHTGYBYBGW-3
F motif6	The second semi-nested PCR	5-GGBCARVKDTGYAWHK-3
R motif10	Semi-nested PCR	5-TTSHNGCCKCCRAARGG-3
DW-ACP 1	The first walking	5-ACP-AGGTC-3'
DW-ACP 2	The first walking	5-ACP-TGGTC-3'
DW-ACP 3	The first walking	5-ACP-GGGTC-3'
DW-ACP 4	The first walking	5-ACP-CGGTC-3'
DW-ACPN	The second walking	5-ACPN-GGTC-3'
Uni-primer	The first walking	5-TCACAGAAGTATGCCAAGCGA-3'
R-Tsp1	The first walking	5-TAAAAACGCCAGAGAGTTCT-3'
R-Tsp2	The second walking	5-CCTTCTTGATAGTGTGCCTGG-3'
R-Tsp3	The third walking	5-GCAATATAGGAGTCAACATAGG-3'
F-TSP1	The first walking	5-TCTATGTACCTCAGAACAGTGT-3'
F-TSP2	The second walking	5-AGGCTGCCTCGACAAGCG-3'
F-TSP3	The third walking	5-CTCGCCTGCGAGGCATCTCTG-3'
F-full	Full length sequence amplification	5-tctagaGTGTTAAATGGAGGATTAGAA-3'
R-full	Full length sequence amplification	5-ctcgagCCACCTGGATATGATGACCTTC-3'

By assembling the sequences of 3', 5' and the core partial sequences on ContigExpress (Vector NTI Suite 6.0), the full-length cDNA sequence of *ALDH* was deduced. The sequence comparison was conducted through database search using BLAST program including BLASTn, x, and p [11].

Expression and Functional Analysis

The full-length sequence was amplified using the primers (F-full + R-full) containing *Xba*I and *Xho*I sites, respectively. The amplified products and pET-303 vector were

digested with *Xba*I and *Xho*I restriction enzymes and then ligated. The constructed plasmid, pET303/TIADH, was introduced into *E. coli* BL21-codonplus. The resulting positive clone was induced with 0.5 mM isopropyl β -D-thiogalactopyranoside for 5 h at 30 °C. The recombinant protein was purified by Ni-NTA affinity chromatography. The purified enzymes were assayed by SDS-PAGE followed by Coomassie brilliant blue G-250 staining and the protein concentration was determined by the Bradford method. Band intensity was quantified by Glyko BandsScan software.

The ALDH activity was measured by following the oxidation of propionaldehyde in coupled reaction of

converting NAD^+ to NADH at 37 °C. The change in absorbance of NADH was monitored at 340 nm using an ultraviolet–visible spectrophotometer with a temperature-controlled cuvette holder (Shimadzu Co., Kyoto, Japan). One unit (U) activity was defined as the amount of ALDH required for the consumption or formation of 1 mol of NADH per min at 37 °C. The standard reaction mixture contained 100 μM of Tris buffer, pH 9.2, 100 μM of mercaptoethanol, 1 μM of NAD^+ , 25 μM of propionaldehyde, and enzyme in a total volume of 1 mL. Prior to activity measurement, all components except propionaldehyde were added and the cuvette carrier was placed in the cell compartment through which water was circulated at 37 °C. The reaction was initiated by the addition of the propionaldehyde. [14].

Results and Discussion

Metagenomic libraries herald the era of magnifying the microbial world and enhanced the rate of discovery of novel genes and pathways [15]. We isolated partial fragment of new aldehyde dehydrogenases gene from metagenome DNA of hot spring water and underwater soil using Power Max™ Soil DNA Isolation Kit. Pulsed Field Gel Electrophoresis was used to identify the quality of DNA, the result showed that the DNA fragments were ranging from 30 to 40 kD and can be used to any downstream application, for example, fosmid library construction or genomics walking. The isolated DNA had a high level of purity allowing for successful PCR amplification from the sample.

Using the metagenome DNA as a template, the semi-nest and touch down PCR was performed to amplify fragments, aim to enhance specificity. The first PCR had no visible production. further, we did the second and third semi-nest PCR with the diluted PCR production. The single bands were excised for sequencing analysis (Fig. 2). 7 relative fragments were isolated, meanwhile, some non-specific amplification products also were amplified, for example, DNA methylase, FAD linked oxidase domain-containing protein fragments according to homology analysis. Although the extremely high sensitivity of PCR can give rise to the non-specific amplification products, PCR technology was widely applied to metagenomics research [16]. The nested PCR that is useful in achieving the specificity by the amplification of an extended nucleotide sequence followed by the amplification of a region located within the first amplification. Meanwhile touchdown PCR can one-step optimize PCRs, increasing specificity, sensitivity and yield, without the need for lengthy optimizations and/or the redesigning of primers [17].

Based on semi-nest touch-down PCR, the possible ALDH fragments were identified using the BLASTn, x,

and p program, the results showed that they belonged to different kinds of aldehyde dehydrogenase subfamilies, such as ALDH-1 was homology with methylmalonyl semialdehyde DH (MMSALDH) and ALDH-2 was homology with γ -glutamyl semialdehyde DH, and glyceraldehyde-3-phosphate DH-nonphosphorylating.

Methylmalonate semialdehyde dehydrogenase (MSDH) is a NAD and coenzyme A (CoA) dependent enzyme which catalyses the reversible conversion of methylmalonate semialdehyde (MSA) into propionyl-CoA and of malonate semialdehyde into acetyl-CoA. MSDH is particular since it is the only known ALDH that requires CoA in the deacylation step instead of a water molecule. Therefore, we focus on the putative ALDH1 and named *mg-MSDH*.

Metagenome Walking and Isolation of *mg-MSDH* Genes

The sequence showing *MSDH* homology was used to design TSPs for metagenome walking to retrieve the flanking regions. Using DNA Walking SpeedUp™ premix kit, the 3'-end and 5'-end of *MSDH* were respectively obtained. By assembling the sequences of 3'-end, 5'-end and the core fragment on Contig Express (Vector NTI Suite 6.0), the full length sequence of *MSDH* was obtained that was 1,493 bps; finally the physical full-length *MSDH* was amplified and confirmed by sequencing.

The ORF finding analysis showed that the *MSDH* contained a 1,473 bp coding sequence encoding a 490-amino-acid polypeptide with a calculated molecular mass of 53.9 kDa and an isoelectric point of 5.02. The upstream sequences TAGGAG of the start codon (GTG), suggested that was a ribosome binding site. Then, the full-length *MSDH* sequence was submitted to GenBank and assigned an accession number JQ715422.

The deduced amino acid sequence of *MSDH* was submitted to NCBI for BLAST searching and the results showed that *MSDH* had high similarities with *MSDH* from bacterium Ellin514 (73 % similarities and 53 % identities) and *Bacillus subtilis* (63 % similarities and 46 % identities) [18], meanwhile, *mg-MSDH* had different levels of similarity with *MSDH* or aldehyde dehydrogenase from other microorganisms. Thus, the BLAST analysis results indicated that *MSDH* belonged to the aldehyde dehydrogenase superfamily. The multiple alignment analysis results demonstrated that ten conserved motifs and the conserved residues were found in *mg-MSDH* and selective aldehyde dehydrogenase [19, 20]. For example, motif 4 “VSVFGSTPVA” covered the essential NAD-binding turn of the Rossmann fold in the class 3 ALDH structure, in which the Gly187 is invariant in ALDHs and dehydrogenase families [21]. Motif 10 “SFSGWKESFFGD” was highly conserved at the C terminus in all aldehyde

Fig. 2 The agarose gel electrophoresis of walking products by touch down and semi-nested PCR technique. The first PCR showed no visible products (a), the second PCR showed small ones (b). The third PCR showed many bands and excised targets for sequencing (c). The number (1, 2, 3, 4) mean the four kinds of initiative template amplified by DW-ACP and uni-primer

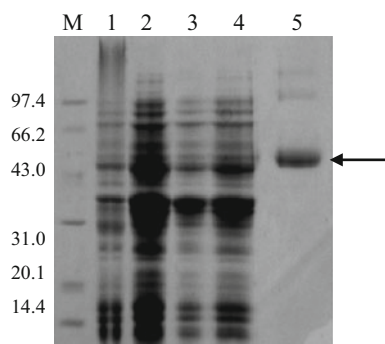
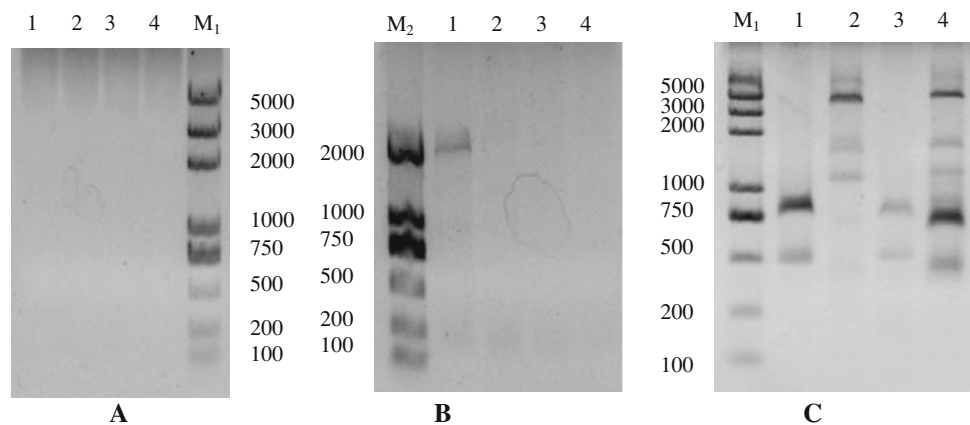


Fig. 3 SDS-PAGE Analysis of mg-MSDH. Lane M, codonplus; lane 2, BL 21-codonplus pET/mg-MSDH; BL 21-codonplus pET/mg-MSDH after IPTG induction; lane4, the supernatant of BL 21-codonplus pET/mg-MSDH; lane 5 the reconstruction mg-MSDH by Ni-NTA affinity chromatography

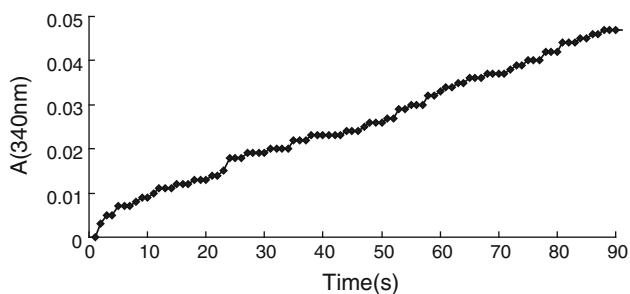


Fig. 4 Time course of enzymatic converting NAD^+ to NADH in coupled with oxidation of propionaldehyde by mg-MSDH monitored at 340 nm

dehydrogenase [22]. Finally, we have successfully isolated a new aldehyde dehydrogenase gene from the metagenome by PCR and metagenome walking.

Expression and Functional Analysis

The expression vector carrying *mg-MSDH* gene was constructed by ligating the PCR-amplified gene into

the plasmid pET-303/CT-His; the resulting plasmid, pET/*mg-MSDH*, was transformed into *E. coli* BL21-codonplus to yield a soluble expression of recombinant protein by IPTG-induction. The percentage of mg-MSDH in crude cell extract was approximately 10 % that was quantified by BandsScan analysis software. The recombinant protein was purified to homogeneity using a Ni-NTA affinity chromatography (Fig. 3). The recombinant mg-MSDH had the ability of oxidation of propionaldehyde in coupled reaction of converting NAD^+ to NADH and gave the activity of 3.0 U mg^{-1} at 37°C (Fig. 4). Although many new genes have been identified from the metagenomes, only a few of properties of their encoded proteins have been analyzed. In the present work, we not only isolate a new full length aldehyde dehydrogenase gene sequence, but also identify its aldehyde oxidative activity.

In conclusion, we have presented here a culture-independent approach to retrieve a target gene. And the gene showed propionaldehyde oxidative activity. However, the method based on PCR technology have its inhere drawback. The design of primers is dependent on existing sequence information and skews the search in favor of known sequence types. Functionally similar genes resulting from convergent evolution are not likely to be detected by a single gene-family-specific set of PCR primers.

Acknowledgments This research was financially supported by National Natural Science Foundation of China (21006018), Science and Technology Department of Zhejiang Province (2009C31086), Technology Research and Development Program for Institute of Hangzhou (20090331N03) and Young Nature Fund of Hangzhou Normal University (2010QN19).

References

1. Raes J, Foerstner KU, Bork P (2007) Get the most out of your metagenome: computational analysis of environmental sequence data. *Curr Opin Microbiol* 10:1–9
2. Nobutada K (2006) Metagenomics: access to unculturable microbes in the environment. *Microbes Environ* 21(4):201–215

3. Morimoto S, Fujii T (2009) A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Appl Microbiol Biotechnol* 83:389–396
4. Fuchu G, Ohtsubo Y, Ito M, Miyazaki R, Ono A, Tsuda M, Nagata Y (2008) Insertion sequence-based cassette PCR: cultivation-independent isolation of *g*-hexachlorocyclohexane-degrading genes from soil DNA and gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl Microbiol Biotechnol* 79:627–632
5. Wang Q, Wu H, Wang A, Du P, Pei X, Li H, Yin X, Huang L, Xiong X (2010) Prospecting metagenomic enzyme subfamily genes for DNA family shuffling by a novel PCR-based approach. *J Biol Chem* 285(53):41509–41516
6. Jaureguibeitia A, Saá L, Llama MJ, Serra JL (2007) Purification, characterization and cloning of aldehyde dehydrogenase from *Rhodococcus erythropolis* UPV-1. *Appl Microbiol Biotechnol* 73(5):1073–1086
7. Peng X, Shindo K, Kanoh K, Inomata Y, Choi SK, Misawa N (2005) Characterization of sphingomonas aldehyde dehydrogenase catalyzing the conversion of various aromatic aldehydes to their carboxylic acids. *Appl Microbiol Biotechnol* 69(2):141–150
8. Jing QQ, Wang JK, Wu GG (2012) Teth137, a conserved factor of unknown function from thermoanaerobacter ethanolicus JW200, represses the transcription of the *adhE* gene in vitro. *Indian J Microbiol*. doi:10.1007/s12088-012-0339-y
9. Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor press, Cold Spring Harbor
10. Hempela J, Lindahlb R, Perozich J, Wang BC, Kuoa I, Nicholas H (2001) Beyond the catalytic core of ALDH: a web of important residues begins to emerge. *Chem-Biol Interact* 130–132(30):39–46
11. Huang XQ, Cloutier S (2007) Hemi-nested touchdown PCR. *BMC Genet* 8:18
12. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
13. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
14. Talfournier F, Stines-Chaumeil C, Branlant G (2011) Methylmalonate-semialdehyde dehydrogenase from *Bacillus subtilis*: substrate specificity and coenzyme a binding. *J Biol Chem* 286(25):21971–21981
15. Sharma N, Tanksale H, Kapley A, Purohit HJ (2012) Mining the metagenome of activated biomass of an industrial wastewater treatment plant by a novel method. *Indian J Microbiol* 52(4):538–543
16. Sundarakrishnan B, Pushpanathan M, Jayashree M, Rajendhran J, Sakthivel N, Jayachandran S, Gunasekaran P (2012) Assessment of microbial richness in pelagic sediment of Andaman Sea by bacterial tag encoded FLX titanium amplicon pyrosequencing (bTEFAP). *Indian J Microbiol* 52(4):544–550
17. Korbie DJ, Mattick JS (2008) Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc* 3:1452–1456
18. Stines-Chaumeil C, Talfournier F, Branlant G (2006) Mechanistic characterization of the MSDH (methylmalonate semialdehyde dehydrogenase) from *Bacillus subtilis*. *Biochem J* 395:107–115
19. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34(2):W369–W373
20. Nakamura T, Ichinose H, Wariishi H (2010) Cloning and heterologous expression of two aryl-aldehyde dehydrogenases from the white-rot basidiomycete *Phanerochaete chrysosporium*. *Biochem Biophys Res Commun* 394(3):470–475
21. Bottoms CA, Smith PE, Tanner JJ (2009) A structurally conserved water molecule in Rossmann dinucleotide-binding domains. *Protein Sci* 11(9):2125–2137
22. Zhao TF, Lei MK, Wu YX, Wang CW, Zhang ZS, Deng F, Wang HB (2009) Molecular cloning and expression of the complete DNA sequence encoding NAD⁺-dependent acetaldehyde dehydrogenase from *Acinetobacter* sp. strain HBS-2. *Ann Microbiol* 59(1):97–104