De-MetaST-BLAST: A Tool for the Validation of Degenerate Primer Sets and Data Mining of Publicly Available Metagenomes

Christopher A. Gulvik1, T. Chad Effler1,2, Steven W. Wilhelm1, Alison Buchan1*

1 Department of Microbiology, University of Tennessee, Knoxville, Tennessee, United States of America, 2 Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, Tennessee, United States of America

Abstract

Development and use of primer sets to amplify nucleic acid sequences of interest is fundamental to studies spanning many life science disciplines. As such, the validation of primer sets is essential. Several computer programs have been created to aid in the initial selection of primer sequences that may or may not require multiple nucleotide combinations (i.e., degeneracies). Conversely, validation of primer specificity has remained largely unchanged for several decades, and there are currently few available programs that allows for an evaluation of primers containing degenerate nucleotide bases. To alleviate this gap, we developed the program De-MetaST that performs an in silico amplification using user defined nucleotide sequence dataset(s) and primer sequences that may contain degenerate bases. The program returns an output file that contains the in silico amplicons. When De-MetaST is paired with NCBI’s BLAST (De-MetaST-BLAST), the program also returns the top 10 nr NCBI database hits for each recovered in silico amplicon. While the original motivation for development of this search tool was degenerate primer validation using the wealth of nucleotide sequences available in environmental metagenome and metatranscriptome databases, this search tool has potential utility in many data mining applications.

Introduction

PCR is one of the most fundamental and powerful molecular biology tools available. PCR primer sets that contain degenerate bases allow for the amplification of homologous sequences and have been used in various applications, including genetic diversity analyses (e.g., [1–12]). Several software packages that use a nucleotide or amino acid alignment of the genetic target are available to aid in the initial development of degenerate primer sets (e.g., Amplicon [13], CODEHOP [14–16], DEFOG [17], DePict [18], HYDEN [19], MAD-DPD [20], PhiSiGns [21], and Primaclade [22]). In addition, manual identification of conserved regions from aligned sequences generated using software such as ARB [23], GlastalX [24], and MEGA [25] is also common practice (e.g., [26–31]). Once candidate primers are developed, thermodynamic properties and self-complementarity tests can be obtained using online tools (e.g., OligoCalc [32]).

Despite the utility and common use of degenerate primers, there are no software programs specifically designed to facilitate validation of their specificity. The most common practice for initial validation of degenerate primers is by direct sequence analysis of PCR amplicons (e.g., [33–37]). This can be both laborious and costly, and does not take advantage of the ever-increasing publicly available nucleotide data, including that derived from natural samples. In fact, environmental metagenomes and metatranscriptomes are especially attractive reference databases (e.g., CAMERA [38] [http://camera.calit2.net/] and MG-RAST [http://metagenomics.anl.gov/]) to perform in silico tests en masse to identify sequences a degenerate primer set might amplify.

To address this gap in available bioinformatic tools, we have developed a program termed De-MetaST. This program accepts primers that are degenerate using a meta-genome and – transcriptome search tool to retrieve in silico PCR amplicons. When paired with BLAST [39], the output provides the most homologous sequences in GenBank for each recovered in silico amplicon. In this report, we provide an overview of the program and outline its utility as a tool to validate the specificity of degenerate primer sets. This program is designed to be user-friendly for non-bioinformatics specialists and is publicly available; as are screencaast video tutorials demonstrating installation and implementation.

Design and Program Overview

De-MetaST is written in C++ and is provided as an executable wrapper to include BLAST (De-MetaST-BLAST) as well as an independent executable (De-MetaST). The function of De-MetaST is to implement a search routine based on bitwise comparisons. Initial steps translate the degenerate nucleotide sequences of each primer, as well as their reverse complement sequences, into unique and specific binary
representations. This approach facilitates rapid searches of large
databases that are also transformed into binary representations.
The specific computational steps of De-MetaST are outlined in
Figure S1.

How De-MetaST Works
The De-MetaST program initially converts the inputted
primer sequences into 4-digit binary code, where the 16 possible
combinations of nucleotides include: A, T, C, G, B, D, H, K,
M, N (or X), R, S, V, W, and Y (Figure 1). Then, each
sequence read within a user defined, FASTA formatted
database is converted to 4-digit binary codes and scanned
using a bitwise searching operation for the presence of both
primer sequences in the appropriate orientation. Limited
memory is necessary for this action because each sequence
read is individually transformed to binary and immediately
scanned for the presence of the primer sequences. The program
searches using both the original user inputted primers as well as
the reverse and complement of those sequences. This latter
search is done to insure identification of target sequences
regardless of whether the sense or antisense strand is
represented by the database sequence read scanned. The search
feature also allows a single primer to serve as both the forward
and reverse primer. When primers identify their respective
target(s) within a sequence read, the nucleotide sequence
delimited by the two primers, termed the in silico amplicon, is
retrieved. The primer(s) yielding each amplicon are reported in
the output. De-MetaST is written to parse in silico amplicons
>5000 bp into a separate FASTA formatted file that is not
subject to BLASTx; users can modify this length restriction by
ingrating the code. All in silico amplicons provided in the output
represent the sense strand in a 5' to 3' orientation. Thus, when
positive hits are made to reads representing antisense strands,
the complement and reverse of those reads are generated. Any
identifying features (e.g., unique read number) as well as the file
name for each predicted hit is recovered. Although developed to
accept degenerate primers, non-degenerate primers can also be
input into De-MetaST. Furthermore, the nucleotide query
database(s) themselves may contain sequence reads with de-
generate or ambiguous nucleotides (e.g., N). Finally, De-MetaST
accepts multiple primer sets as input; the in silico amplicons from
each set are output into separate FASTA files. As De-MetaST
accepts degeneracies in the input primer sequences, it requires
absolute conservation in the target sequences; it does not allow
for any mismatches between the primer sequence and target. In
this way, the user controls the level of primer specificity.

De-MetaST Paired with BLAST
Once the database sequence files have been queried for
predicted PCR amplicons, each in silico amplicon is subject to

---

**Figure 1. De-MetaST transformation of nucleotide sequences into a binary representation.** The binary representation for each of the 16 possible nucleotide character inputs is shown in the upper box. The lower box provides an example of the transformation using a mock primer sequence. Spaced gaps are shown for instructional purposes and do not occur in the De-MetaST search routine.
doi:10.1371/journal.pone.0050362.g001
a BLASTx analysis, which translates the nucleotide sequence in all six frames and performs queries for each translation against the non-redundant (nr) NCBI protein database. The top 10 BLASTx hits for each amplicon are formatted as an XML file. The final step of De-MetaST-BLAST compiles all of the meta-information of the BLASTx results for each amplicon retrieved (e.g., hit accession number, E-value, predicted function, nucleotide sequence, database file name, the primer combination that retrieved the amplicon, unique read number) into a single, tab-delimited TXT file. The BLASTx results file can also be exported as an XLS file format for direct use in Microsoft Excel or other suitable program. A graphical overview of the De-MetaST-BLAST workflow is shown in Figure 2.

Results and Discussion

We have developed a computational method to generate in silico amplifications from degenerate primer sets searched against user defined nucleotide databases. To illustrate the utility of De-MetaST-BLAST, we demonstrate its performance using a novel degenerate primer set designed for use on environmental samples. This primer set targets the bacterial boxB gene, which encodes the oxygenase component of a multi-enzyme epoxidase (EC 1.14.13) that is specific to a benzoate catabolic pathway [40]. Three metagenome libraries representing different environments, library size and DNA sequencing methods were searched and found to contain putative boxB amplicons of the appropriate size (300 bp) (Table 1). Figure 3 shows the typical output of De-MetaST-BLAST for one of those database searches, which includes for each in silico amplicon the top 10 BLASTx hits with their corresponding E-value and GenBank accession number.

To retrieve an in silico amplicon, the program requires both primers to match their respective targets in a single sequence read or sequence assembly (contig). Thus, an important consideration in terms of selection of appropriate searchable databases is the average length of the sequence read or assembly contained within it, as well as the desired amplicon size. This concern may be alleviated as longer read sequencing technologies are developed and/or as sequence coverage and assembly algorithms improve. Interestingly, our analysis demonstrates that in silico amplicons of
16S rRNA gene amplicons identified in representative metagenomes using De-MetaST-BLAST. In silico analysis, such as sequence alignment and subsequent phylogenetic analysis, is streamlined. Thus, in silico amplicons retrieved from existing sequence datasets can be readily compared to experimentally derived clone library sequences. Furthermore, as the nucleotide sequences targeted by the primers are returned in the De-MetaST output, users can draw on that information to further refine their primers according to a desired level of functional and/or phylogenetic specificity. The program also has utility beyond searches of environmental sequence databases. It can be used to query any nucleotide dataset, including those derived from single organisms. Thus, it has use in assessing the specificity of primers targeting multi-copy or homologous genes within a single organism or group of organisms.

### Benchmarks and System Requirements

De-MetaST-BLAST has been developed for the long-term support (LTS) Ubuntu operating systems 10.04 LTS and 12.04 LTS. While De-MetaST does not make use of multi-core processors, BLAST maintains that capability. Benchmarks were performed on an Intel i7-2600 processor (3.4 GHz quad-core, 8-thread) desktop using the developed degenerate boxB primer set against the Waseca Farm Soil metagenome (AAFX01000000) [41]. This search took approximately 11.7 s (Table 2). When the database size was artificially and incrementally increased up to five-fold (772 Mb) by replication of the original dataset, the processing time remained <1 min. Furthermore, to determine the effect of increased numbers of positive hits on run time, the libraries were seeded with additional sequences containing the target. Doubling of targets within the databases had no effect on run time (Table 2). In contrast to the relatively rapid processing speed of De-MetaST, implementation of the BLAST function can add significant processing time to the process, particularly if a local custom database is used. As an example, for the initial benchmark search against the locally installed Farm Soil metagenome that recovered two hits, the BLASTx function added 39.3 s using two threads. Thus, computational requirements and processing speed are primarily dictated by BLAST. When BLAST is performed remotely—the default setting [see below]—the return time is dependent upon availability and processing speeds of the NCBI servers.

Both De-MetaST and De-MetaST-BLAST can be run on any operating system with a C++ compiler [e.g., standard Windows and Mac OS]. However, users would need to ensure the BLAST installation is compatible with their processor.

### Availability of De-MetaST-BLAST

The De-MetaST package and the De-MetaST-BLAST wrapper are made freely available at [http://sourceforge.net/p/de-metast-blast/](http://sourceforge.net/p/de-metast-blast/) and [http://code.google.com/p/de-metast-blast/](http://code.google.com/p/de-metast-blast/). These files are also provided as supplemental information to this publication.
BLASTx results

| E-value | GenBank accession number | GenBank sequence record | queried sequence ("in silico amplicons") |
|---------|--------------------------|-------------------------|----------------------------------------|
| 1       | YP_00161149              | yfK gene product \(E. coli\) | CAM_PBO1_FarmSoil.read.1-1-1     |
| 2       | ZP_00246848              | benzoyl-CoA oxygenase B | CAM_PBO1_FarmSoil.read.1-1-1     |
| 3       | YP_00326460              | benzoyl-CoA oxygenase subunit \(B\) \(E. coli\) | CAM_PBO1_FarmSoil.read.1-1-1     |
| 4       | ZP_0028721               | benzoyl-CoA oxygenase component \(B\) \(P. shermanii\) | CAM_PBO1_FarmSoil.read.1-1-1     |
| 5       | ZO_0027992               | COG3396 Uncharacterized conserved protein | CAM_PBO1_FarmSoil.read.1-1-1     |
| 6       | YP_41593                 | hypothetical protein \(NM_0011315\) \(M. macedonii\) | CAM_PBO1_FarmSoil.read.1-1-1     |
| 7       | YP_0023932               | benzoyl-CoA oxygenase component \(B\) \(B. thuringiensis\) | CAM_PBO1_FarmSoil.read.1-1-1     |
| 8       | ZP_00212153              | benzoyl-CoA oxygenase component \(B\) \(B. thuringiensis\) | CAM_PBO1_FarmSoil.read.1-1-1     |
| 9       | ZO_00212153              | benzoyl-CoA oxygenase component \(B\) \(B. thuringiensis\) | CAM_PBO1_FarmSoil.read.1-1-1     |
| 10      | YP_0023932               | benzoyl-CoA oxygenase component \(B\) \(B. thuringiensis\) | CAM_PBO1_FarmSoil.read.1-1-1     |

De-MetaST output

| queried database | excision info | unique read identifier |
|------------------|--------------|------------------------|
| CAM_PBO1_FarmSoil.read.1-1-1 | 1-1-1 |CAM_PBO1_FarmSoil.read.1-1-1|
| CAM_PBO1_FarmSoil.read.1-1-1 | 1-1-1 |CAM_PBO1_FarmSoil.read.1-1-1|
| CAM_PBO1_FarmSoil.read.1-1-1 | 1-1-1 |CAM_PBO1_FarmSoil.read.1-1-1|
| CAM_PBO1_FarmSoil.read.1-1-1 | 1-1-1 |CAM_PBO1_FarmSoil.read.1-1-1|
| CAM_PBO1_FarmSoil.read.1-1-1 | 1-1-1 |CAM_PBO1_FarmSoil.read.1-1-1|

Figure 3. Example of De-MetaST-BLAST output. Text within the box denotes the spreadsheet output for a boxB primer set search against the WASECA Farm Soil Metagenome (AAFX01000000) [41] that recovers two in silico amplicons. Column descriptors are shown in color; select columns have been truncated due to space constraints. For the "excision info" column, the first alphanumeric character reports the "hit" number within a read (i.e. "1" indicates it is the first in silico amplicon found within a single read). The subsequent alphanumeric characters denote the primer orientation yielding the amplicon (F = forward, R = reverse). Whether a unique read identifier is returned is contingent upon the database itself.

Table 2. Runtime duration of De-MetaST.

| Files Input | Database size [Mbytes] | Sequences in database \(\times 10^9\) | Nucleotides in Database [Mbp] | Hits | Real Time [s] | User Time [s] | System Time [s] |
|-------------|-------------------------|--------------------------------------|-------------------------------|------|---------------|---------------|-----------------|
| 1           | 206.1                   | 1.4                                  | 154                           | 2    | 11.7          | 11.7          | 0.02            |
| 2           | 412.2                   | 2.8                                  | 309                           | 4    | 23.5          | 23.4          | 0.05            |
| 3           | 618.3                   | 4.2                                  | 463                           | 6    | 35.2          | 35.1          | 0.07            |
| 4           | 824.4                   | 5.5                                  | 618                           | 8    | 47.6          | 47.5          | 0.10            |
| 5           | 1030.5                  | 7.0                                  | 772                           | 10   | 58.6          | 58.5          | 0.12            |
| 1           | 206.1                   | 1.4                                  | 154                           | 4    | 11.9          | 11.9          | 0.02            |
| 2           | 412.2                   | 2.8                                  | 309                           | 8    | 23.3          | 23.3          | 0.05            |
| 3           | 618.3                   | 4.2                                  | 463                           | 12   | 35.6          | 35.5          | 0.08            |
| 4           | 824.4                   | 5.5                                  | 618                           | 16   | 47.3          | 47.1          | 0.10            |
| 5           | 1030.5                  | 7.0                                  | 772                           | 20   | 58.2          | 58.0          | 0.12            |

*The datasets used for benchmarking were manipulations of the Waseca Farm Soil metagenome (AAFX01000000); the average sequence read length in these datasets is 1117 bp.

doi:10.1371/journal.pone.0050362.t002

Conclusions

It was recently predicted that the increasing amounts of metagenome sequences will likely serve as a valuable resource in evaluation of the coverage and specificity of previously developed primer sets [42]. De-MetaST-BLAST will provide users with a useful tool in such evaluations. De-MetaST is designed to provide in silico amplicons generated by user defined degenerate
primers found within a user defined nucleotide database. When paired with BLAST, the program returns the most homologous GenBank hits, which are useful in assessing the specificity of degenerate primers. However, the program does not evaluate PCR kinetics and efficiencies with degenerate primers. Thus, users are encouraged to consult appropriate references on the use and design of degenerate primers (e.g., [43-44]), including those that discuss the merits of utilizing base analogs (e.g., inosine; [45]) that can reduce the overall degeneracy of primers.

Supporting Information

Figure S1  Computational procedures of De-MetaST are illustrated within the De-MetaST-BLAST wrapper.

File S1  Archiving the source code for De-MetaST.

References

1. Jarman SN, Deagle BE, Gales NJ (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. Molecular Ecology 13: 1313–1322.
2. Brown MV, Schwabach MS, Hesvon I, Fuhrman JA (2005) Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. Environmental Microbiology 7: 1466–1479.
3. Brennerova M, Josefová J, Brenner V, Pieper DH, Junca H (2009) Metagenomics reveals diversity and abundance of metazoan cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-fuel mixture. Applied and Environmental Microbiology 11: 2216–2224.
4. El Ashari N, Bru D, Sarr A, Martin-Laurent F (2006) Estimation of the density of the protociliate-degrading bacterial community in soil by real-time PCR. European Journal of Soil Science 59: 663–673.
5. Schmalenberger A, Kertesz MA (2007) Desulfurization of aromatic sulfonates by rhizosphere bacteria. Biochim Biophys Acta 9: 335–345.
6. Melia MP, Butterfield DA, Baross JA (2003) Phylogenetic diversity of nitrogenase [nifH] genes in deep-sea and hydrothermal vent environments of the Juan de Fuca ridge. Applied and Environmental Microbiology 69: 960–970.
7. Burgmann H, Widmer F, Von Süderl W, Zeyer J (2004) New molecular screening tools for analysis of free-living diazotrophs in soil. Applied and Environmental Microbiology 70: 240–247.
8. Luton PE, Wayne JM, Sharp RJ, Riley PW (2002) The mcrA gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogens populations in landfills. Molecular Microbiology 48: 3321–3350.
9. Chaudhain SMN, Schafer JK, Crane S, Zylstra GJ, Barkay T (2006) Analysis of mercuric reductase (mcrA) gene diversity in an anerobic mercury-contaminated sediment enrichment. Environmental Microbiology 8: 1746–1752.
10. Wang GZ, Wang YR, Yang PL, Luo HY, Huang HQ, et al. (2010) Molecular detection and diversity of xylanase genes in alpine tundra soil. Applied Microbiology and Biotechnology 87: 1383–1395.
11. Wang LP, Wang WP, Lai QL, Shao ZZ (2010) Gene diversity of CYP153A and A1B1 alkane hydroxylases in oil-degrading bacteria isolated from the Atlantic Ocean. Environmental Microbiology 12: 1230–1242.
12. Matteson AR, Loar SN, Bourbonnaire RA, Wilhelm SW (2011) Molecular enumeration of an ecologically important cyanophyae in a laurentian great lake. Applied and Environmental Microbiology 77: 6772–6779.
13. Jarman SN (2004) Amplicon: software for designing PCR primers on aligned DNA sequences. Bioinformatics 20: 1644–1645.
14. Stafhead JP, Boyce R, Kovarik D, Rose TM (2011) CODEHOP PCR and CODEHOP PCR primer design. Methods in molecular biology (Clifton, NJ) 569: 57–73.
15. Rose TM, Henikoff JG, Henikoff S (2003) CODEHOP (COnsensus-DEgenerate hybrid oligonucleotide primer) PCR primer design. Nucleic Acids Research 31: 3763–3766.
16. Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, et al. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Research 26: 1620–1633.
17. Fuchs T, Madeva C, Linhart C, Sharman R, Klen M, et al. (2002) DEFOG: A practical scheme for deciphering families of genes. Genomics 89: 295–302.
18. Wei X, Kuhn D, Narasimham G (2003) Degenerate primer design via clustering. Proceedings of the IEEE Computer Society Bioinformatics Conference. Stanford, CA. 73–83.
19. Linhart C, Sharman R (2007) Degenerate primer design: theoretical analysis and the HYDEN program. Methods in molecular biology (Clifton, NJ) 402: 221–244.
20. Najafabadi HS, Saberi A, Torabi N, Chamankhah M (2008) MAD-DPD: designing highly degenerate primers with maximum amplification specificity. Biotechnology and Biotechnological Equipment 22: 319–4.
21. Dwivedi B, Schmieder R, Goldamin DB, Edwards RA, Breitbart M (2012) PhiSiGns: an online tool to identify signature genes in phages and design PCR primers for examining phage diversity. BMC Bioinformatics 13.
22. Gazdbery MD, Malcomber ST, Doust AN, Kellogg EA (2005) Primacle - a flexible tool to find conserved PCR primers across multiple species. Bioinformatics 21: 1263–1264.
23. Ludwig W, Strunk O, Westram R, Richter I, Meier H, et al. (2004) ARB: a software environment for sequence data. Nucleic Acids Research 32: 1303–1319.
24. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal x version 2.0. Bioinformatics 23: 2947–2948.
25. Takekura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731–2739.
26. Kwok S, Chang SY, Sninsky JJ, Wang A (1994) A guide to the design and use of mismatched and degenerate primers. Genome Research 3: 839–847.
27. Junca H, Pieper DH (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. Environmental Microbiology 6: 95–110.
28. Hendricks B, Junca H, Vosholova J, Lindner A, Ruegg I, et al. (2006) Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: Distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. Journal of Microbiological Methods 64: 250–263.
29. Lyons JJ, Newell SY, Buchan A, Moran MA (2003) Diversity of ascomycete laccase gene sequences in a southeastern US salt marsh. Microbial Ecology 45: 270–281.
30. Allen AE, Ward BB, Song BK (2005) Characterization of diatom (Bacillariophyceae) nitrate reductase genes and their detection in marine phytoplankton communities. Journal of Phycology 41: 95–104.
31. Malmstrom RR, Coe A, Ketterl GC, Martinsey AG, Frias-Lopez J, et al. (2010) Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific oceans. ISME Journal 4: 1262–1264.
32. Kibbe WA (2007) OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Research 35: W43–W46.
33. Kirkman DL, Yu LY, Cottrell MT (2003) Diversity and abundance of uncultured Cytophaga-like bacteria in the Delaware Estuary. Applied and Environmental Microbiology 69: 6587–6596.
34. Ochsenreiter T, Selezni D, Quaiser A, Bonch-Osmolovskaya L, Schleper C (2005) Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. Environmental Microbiology 7: 767–797.
35. Stach JM, Maldonado LA, Ward AC, Goodfield M, Bull AT (2003) New primers for the class Actinobacteria: application to marine and terrestrial environments. Environmental Microbiology 5: 828–841.
36. Rothmayer JB, Witzel KP, Liesack W (1997) The ammonia monooxygenase structural gene amoA as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations. Applied and Environmental Microbiology 63: 4704–4712.
37. Lopes-Lopez A, Barntal SG, Stal L, Onyshchenko O, Rodriguez-Valera F (2005) Genetic analysis of housekeeping genes reveals a deep-sea ecotype of Alteromonas macleodii in the Mediterranean Sea. Environmental Microbiology 7: 649–658.
38. Sun S, Chen J, Li W, Ahintas I, Lin A, et al. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. Nucleic Acids Research 39: D546–D551.
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ [1990] Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.
40. Rather LJ, Knapp B, Haehnel W, Fuchs G (2010) Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. Journal of Biological Chemistry 285: 20613–20624.
41. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, et al. (2005) Comparative metagenomics of microbial communities. Science 308: 554–557.
42. Iwai S, Johnson TA, Chai B, Hashsham SA, Tiedje JM (2011) Comparison of the specificities and efficacies of primers for aromatic dioxygenase gene analysis of environmental samples. Applied and Environmental Microbiology 77: 3351–3357.

43. Kwok S, Chang SY, Sninsky JJ, Wang A (1994) A guide to the design and use of mismatched and degenerate primers. PCR-Methods and Applications 3: S39–S47.
44. Preston GM (1997) Cloning gene family members using the polymerase chain reaction with degenerate oligonucleotide primers. Methods in Molecular Biology (Clifton, N.J.) 69: 97–113.
45. Liu H, Nichols R (1994) PCR amplification using deoxynosine to replace entire codon and at ambiguous positions. Biotechniques 16: 24–26.
46. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75: 7537–7541.
47. Muyzer G, DeWoald EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology 59: 695–700.