Unraveling activities by functional-based approaches using metagenomic libraries from dry and rain forest soils in Puerto Rico

José M. Cruz1, Manuel A. Ortega1, Jean C. Cruz1, Pedro Ondina1, Rossivette Santiago2, and Carlos Ríos-Velázquez3

1Industrial Biotechnology Program, University of Puerto Rico at Mayagüez, 00681, Mayagüez, Puerto Rico
2Department of Bacteriology, University of Wisconsin, 53706, Madison, Wisconsin, United States
3Department of Biology, University of Puerto Rico at Mayagüez, 00681, Mayagüez, Puerto Rico

Studies have demonstrated that standard cultivable methods fail to access a representative amount of the microbes available in any environment, since only 1% of the whole population is cultivable. This leaves behind the other 99%, which represents an invaluable source of microorganisms with potentially novel metabolic activities that can be explored using culture independent strategies like Metagenomics. To exploit the biotechnological potential of such activities, two large (25-40 kb) fosmid libraries of environmental DNA (eDNA) from El Yunque rainforest and Guánica’s dry forest were generated. Functional selections and screenings, such as hydrolytic and colorimetric assays, were used to detect the presence of antibiotic resistance, lipolytic and urease activities in both libraries. The aim of this chapter is to introduce the concept of functional metagenomics as a tool to monitor and detect novel biologically active molecules in the total microbial population of tropical forest soil or ideally in any environment.

Keywords functional metagenomics; tropical forest

1. Introduction

Studies have demonstrated that one gram of soil may contain up to 4,000 different microbial species, from which only 1% is readily cultivable [1]. All the biologically active molecules such as restriction enzymes, antibiotics and metabolic enzymes discovered until recently had been isolated from this fraction of cultivable microorganisms [2-5]. Thus, the biotechnological and biomedical relevance of 99% of the uncultivable majority remains unknown. This demonstrates that soil is indeed a rich source for isolation of novel genes with a plausible wide range of biotechnological applications that has not been fully explored [2]. Metagenomics has proven to be a novel and effective strategy to unearth the functional capacity of the 99% of the uncultivable microorganisms, overcoming the limitations of traditional culture-dependent microbiological techniques.

Metagenomics is based on the genomic analysis of microbial DNA that is extracted directly from communities in environmental samples [2]. Metagenomic DNA can be analyzed by shotgun sequencing [6], PCR amplification of target genes [7], or by analyzing phenotypic evidence [8]. Functional metagenomics (FM) in particular has been used to detect biological molecules of industrial and biomedical relevance such as cellulases, lipases, antimicrobial and antifungal agents, among others [9-12].

FM relies on the expression of genes from foreign microbial DNA in a surrogate host through the generation of metagenomic libraries. This method grants access to discover novel genes, independent of prior sequence information. A simplified process can be seen in Fig. 1 and described as following according to Handelsman et al. (1998): isolation of environmental DNA directly from soil with gentle methods to preserve large pieces of DNA, cutting the DNA with restriction enzymes, and cloning the DNA into a readily cultivable organism such as Escherichia coli [3]. Metagenomics has yielded promising results, including the characterization of clones conferring indirubin and indigo production [13] and clones secreting the antibiotics turbomycin A and B [10].

Here we describe the generation and screening of metagenomic libraries from Puerto Rican soil with the purpose of discovering the soil’s biotechnological potential. Two large fosmid metagenomic libraries were generated from “El Yunque” National Forest Reserve and Guánica’s Dry Forest soils. The Guánica Dry Forest, located in the Southwest region of Puerto Rico, spans 11,000 acres including coastline. In 1981, the United NationESCO established the Guánica Dry Forest as the 2nd international biosphere reserve in Puerto Rico. The annual precipitation is approximately 10-32 inches, with a mean temperature of 34-39 °C [14]. The Yunque rainforest, a considerably different environment from Guánica’s Dry Forest, is located in the Northeast region of Puerto Rico. Officially named the Caribbean National Rainforest, it spans 28,000 acres. In 1980, UN-ESCO declared it the first international biosphere reserve in Puerto Rico. Differing from Guánica’s environmental factors, precipitation is approximately 120 inches of rain and temperature varies between 17-28 °C [15]. Our interest in studying these soils’ biotechnological potential comes from the different environmental conditions that each forest possesses.

Each of the metagenomic libraries was screened using various functional assays to detect enzymes of industrial and biomedical importance such as lipases, ureases and enzymes that confer antibiotic resistance. Lipases are enzymes that catalyze the hydrolysis and synthesis of acylglycerides and other fatty acids. Their biocatalytic importance in industrial
applications stems from their ability to catalyze reactions without the need of cofactors, their unique substrate specificity and their stability in organic solvents [16]. Ureases as well as lipases are hydrolytic enzymes. They catalyze the hydrolysis of urea to yield ammonia and a carbamate. Ureases are used in a wide range of applications such as in alcoholic beverages as a reducing agent [17], and as a biosensor for determining urea concentration in the environment [18]. While lipase and urease enzymes are both industrially and environmentally relevant, the selection and characterization of antibiotic resistance determinants addresses the important public health concern of global increases in multiple-drug resistant pathogens [19-20]. By studying antibiotic resistance reservoirs, from remote soil environments [21] to the human microbiome [22], methods can be developed to circumvent microbial resistance strategies.

Antibiotic resistance and urease activity have been detected via selection and liquid-based screening, respectively. Each of the activities is conferred by genes within the cloned metagenomic inserts from libraries of both forests. Though solid-based screening has been used to monitor lipase activity from cultivable isolates of the Guánica Dry Forest, no lipase activity attributable to a metagenomic insert has been detected yet under current experimental conditions.

2. Materials and Methods

2.1. Soil collection

Soil was collected from the Yunque Rainforest (18º13’38”N 65º47’29”W) and from the Guánica Dry Forest (17º57’56”N 66º52’45”W). Samples were transported at 4°C and stored at -20°C. All samples were thawed to room temperature before use.

2.2. Bacterial strains and vectors

The two libraries were generated using the fosmid vector pEpiFOS™-5 available in the EpiFOS™ fosmid library production kit (Epicentre Biotechnologies, Madison, WI). The bacterial host of the foreign eDNA was the EPI300™T1R Phage T1-Resistant Escherichia coli (Epicentre).

2.3. DNA extraction from soils

DNA extraction was carried out in the Metagenomic Library School at UW Madison in collaboration with UPR Mayagüez. DNA was extracted following the protocol detailed by Handelsman et al. in [23]. Briefly, soil (100 g/ca.) was sieved through a 2 mm mesh to eliminate particles larger than 2 mm. Samples were divided in half and 75 mL of Z buffer were added to each sample. Suspensions were treated with 2 freeze-thaw cycles consisting of 40 min. in dry ice-ethanol followed by 40 min. in a 65 °C water bath. After two cycles, suspensions were mixed gently with 9 mL of SDS (20 %), 4.5 mL of GITC (5 M) and incubated at 65 °C for 2 hrs. Following the incubation period, samples were centrifuged and DNA was collected from the supernatant by a standard phenol-chlorophorm extraction and ethanol precipitation. To purify and size soil DNA from impurities such as humic acids, 100 µg of DNA was pulsed-field gel electrophoresed in a low melting agarose gel (1 %) (Denville Scientific Inc., Metuchen, NJ) using a Chef Mapper (BioRad Hercules, CA). To visualize DNA for sizing, a small portion of the gel including the ladder and DNA was cut, stained with ethidium bromide and visualized using a Universal Hood II (BioRad). DNA with a 40 kb size was located in the stained gel, cut from the unstained gel and electroeluted from the agarose gel using the procedure previously described in [24].
Fig. 2 Steps used in the generation of the metagenomic libraries. After extracting eDNA from soil (a), 40kb fragments were obtained by shearing the eDNA by pipetting (b). The eDNA was blunt ended by end-repairing (c), and size selection was performed by purifying the eDNA from agarose gel after electrophoresis (d,e). (f) Finally, the chosen eDNA was ligated with the fosmid vector, (g) packaged into lambda phages, and transferred into *Escherichia coli* by transduction (h).

2.4. Library generation

The purified eDNA was ligated with the fosmid vector pEpiFOS™-5 as previously described in [20]. MaxPlax lambda packaging extracts were used to package the recombinant DNA into lambda phages following the manufacturer’s protocol. The recombinant packaged DNA was transduced into EPI300™-T1R, plated in Luria Bertani (LB) agar plates (Sigma St. Louis, MO) supplemented with chloramphenicol (Chl) (12.5 μg mL⁻¹), and incubated at 37 °C for 2 days, according to the manufacturer’s protocol. To identify the percentage of clones containing DNA inserts, random clones were picked and grown in LB broth containing Chl for plasmid extractions. Plasmid purification was carried out using a modified protocol from QIAprep spin miniprep kit (QIAGen, Valencia, Ca). The insert DNA was then analyzed by gel electrophoresis (1 %) after a *Not*I (New England Bio Labs, Ipswich, MA) digest. The libraries were divided in subpools and stored in 1.5 mL microtubes, containing LB with Chl and 20 % glycerol at -80 °C. A schematization of the whole process can be seen in Fig. 2.

2.5. Library screening

2.5.1 Lipolytic screening

To test for lipolytic activity, library subpools were grown at 37 °C in LB broth supplemented with Chl for 40 min. After growth, subpools were diluted with physiological buffer (0.85 % NaCl) to adjust the number of colonies to 300-450 colonies/plate. Using the appropriate dilution factor, clones were spread in LB agar supplemented with Chl and 1.5 % emulsified tributyrin (Sigma) as the indicator substrate. Plates were incubated at 37 °C and scored after a 48 hr incubation period for clear halos surrounding the colonies, which indicates presence of lipolytic activity. *Staphylococcus aureus* (ATCC 6538) and *Escherichia coli* EPI300™-T1R were used as positive and negative controls respectively to verify hydrolysis in the tributyrin agar base.

2.5.2. Selecting for antibiotic resistance.

All selections were carried out in LB agar with different antibiotic classes at the following concentrations: 50 μg mL⁻¹ ampicillin (Amp) (Sigma) 25 μg mL⁻¹ kanamycin (Kan) (Sigma), and 10 μg mL⁻¹ tetracycline (Tet) (Sigma) following the method described in [21]. Negative control *Escherichia coli* EPI300™-T1R presented no antibiotic resistance activity at these concentrations throughout the selections. Metagenomic libraries were inoculated on the day of selection in 5 mL LB broth plus 15 μg mL⁻¹ of Chl. Cultures were incubated for 60 min at 37 °C with shaking. Cultures were spread on LB agar plates containing the antibiotics described above, and incubated at 37 °C for up to 3 days. To confirm the resistance phenotype was conferred by the eDNA within the tested library fosmid, resulting colonies were transferred onto LB agar plus Chl and the corresponding antibiotic and incubated overnight. All resistant clones were evaluated by a *Not*I restriction digest and retransformation to confirm the phenotypes.

2.5.3 Urease screening

Metagenomic library sub-pools were inoculated in 5 mL of LB with Chl for 60 min and then in 5 mL of urea broth (UB) (Sigma). Positive subpools for the urea test were serially diluted and inoculated with UB containing Chl in 96 well
plates to reduce the number of bacterial cells/subpool. The wells showing urease activity were streaked on LB agar with Chi to obtain single colonies. Colonies were further re-assessed for urease activity. To confirm that the cloned eDNA conferred the putative urease activity, the cloned fosmid fragment was extracted, analyzed with a NotI digest and retransformed into an isogenic strain. The transformants were grown in LB with Chi, and re-tested for the urease activity.

2.6 Identifying and sequencing active tetracycline resistance and urease genes

To facilitate finding the gene(s) of interest within the 40 kb DNA insert size, an in vitro transposon mutagenesis was carried out using the Genome Priming System (GPS-1, New England Biolabs) according to the manufacturer’s protocol. Insertion mutants that either failed to grow on tetracycline or failed the urease test were identified as having transpositions within the active gene. Clones presenting loss of function were sequenced using the manufacturer’s GPS-1 primers. The sequencing reactions were carried out using the Macrogen sequencing system (http://dna.macrogen.com/eng/order/order_seq.jsp). Sequences were annotated using available Internet databases such as the Basic Local Alignment Search Tool (BLAST http://blast.ncbi.nlm.nih.gov/) [25].

3. Results

3.1. Generation of fosmid libraries

In order to characterize, describe and compare the microbial community and possible phenotypes found in two different environments, fosmid metagenomic libraries were generated from “El Yunque” National Forest Reserve and Guánica’s Dry Forest. The two libraries were designated RF (Rainy Forest) and DF (Dry Forest), respectively.

3.2. Characterization of both libraries

The total library is composed of approximately 32 Gbp of eDNA. The RF library consists of 14,631 clones. Approximately 0.06% (n=10) of the clones were examined for insert by restriction digest analysis; 100% contained insert DNA. The DF library consists of 781,199 clones. Approximately 0.003% (n=30) of the clones were examined for insert; 87% contained insert DNA. Based on the restriction digest analysis of both libraries, and given that the NotI recognition sequence is 5’-GCGGCGGC-3’, the data qualitatively suggests that our libraries contain eDNA that differ in GC content. Clones from DF contain more NotI recognition sites within the insert than clones from RF. Both libraries contain inserts where the maximum size is approximately 40 kb as seen in Fig. 3.

3.3. Library screening

3.3.1. Lipolytic screening

Subpools of clones were screened no less than 3-fold times their original number. No presence of lipases has been identified in the RF library under current experimental conditions. As for the DF clones, approximately 17% of the metagenomic library has been screened. To date, no lipase presence has been identified in this metagenomic library.

3.3.2. Selecting for antibiotic resistance

DF and RF library subpools were assayed for antibiotic resistance to Kan, Tet, and Amp as described above. Approximately 6%, 5%, and 11% of the RF clones resistant to Kan, Tet, Amp, respectively, were verified by NotI restriction analysis. Likewise, approximately 5%, 25%, and 2% of the DF clones resistant to the antibiotics mentioned above were also verified by NotI restriction analysis. Restriction analysis of tetracycline resistant clones revealed low diversity in the RF library. Values can be seen in Table 1.

<table>
<thead>
<tr>
<th>Library</th>
<th>Kanamycin (25 μg mL⁻¹)</th>
<th>Tetracycline (10 μg mL⁻¹)</th>
<th>Ampicillin (50 μg mL⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TAC</td>
<td>NCV</td>
<td>NDC</td>
</tr>
<tr>
<td>DF</td>
<td>213</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>RF</td>
<td>169</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

DF-Dry Forest, RF-Rainy Forest, TAC-Total Antibiotic Resistant Clones, NCV-Number of Clones Verified, NDC-Number of Different Clones
Fig. 3 Restriction digest analysis of clones from (a) Guánica Dry Forest and (b) Yunque Rain Forest libraries, digested with NotI. The line indicates the fosmid vector. First lane on the left contains the 1 Kb ladder. Comparing the number of internal NotI recognition sites, DF clones have DNA inserts higher in GC content than RF.

3.3.3. Urease Screening

DF and RF library subpools were assayed for urease activity as described above. To date no urease activity has been detected in the DF library. However, 37 clones were identified with urease activity in the RF library. Approximately 81% were verified by a NotI restriction analysis, from which 73% differ in their restriction pattern. This suggests that the eDNA insert among the tested clones is different.

3.4. Transposon mutagenesis and sequencing

The genes responsible for tetracycline resistance phenotypes and urease activity were interrupted by transposon insertion to facilitate their identification. Upon transformation of an in vitro transposition reaction of an RF insert conferring tetracycline resistance, 758 clones were observed. Because the selection for antibiotic resistance can be done using solid media, patching was used to identify loss of function as seen in Fig 4. A total of 200 transformants were transferred sequentially in a fixed order to 3 plates containing the appropriate markers. Loss of function of the original phenotype suggests that a transposition has occurred within the gene responsible for the activity. This loss of function was observed on approximately 154 clones. The transposon’s primers were then used to sequence outward of the transposon and over the gene of interest.

For the urease activity, UB was used to identify gene inactivation. Upon transformation of an in vitro transposition reaction of 10 inserts conferring urease activity, 4 different transformants were observed to have a loss of function phenotype. Preliminary sequencing data alignment with Blast X showed that fragments had similarity with DNA fragments from Acidobacterium capsulatum (E=2e-16), Koribacter versatillis (E=2e-95) and Solibacter usitatus (E=8e-51).

Fig. 4 The mutation of interest caused by transposition results in loss of antibiotic resistance function. Transposon mutagenized clones were individually transferred (patched) onto: media containing the library (Chl) and transposon (Kan) selection markers (A and C), and to media containing both markers from the first plate and tetracycline (Tet) (B). The chosen clone shows a loss of Tet resistance (no growth on B) due to transposition.
4. Discussion

4.1 Generation and molecular characterization of fosmid libraries

Before the advent of Metagenomics, the isolation and characterization of novel molecules was limited and biased by the dependence on culturing pure organisms, leaving behind the uncultivable. Here, as in previous works, we demonstrate a straightforward procedure of cloning the soil metagenome to attempt unbiased expression of total microbial genetic content, regardless the type of soil.

Two fosmid libraries with 100 g of forest topsoil from a tropical dry and rainforest were generated and screened for different activities. Approximately 14,631 clones were generated from El Yunque Rain forest, with approximately a 40 kb eDNA insert size which represents a total of ~585 Mbp of eDNA. On the other hand, for the Guánica Dry forest, we were able to generate approximately 781,199 clones. The total DF library represents ~3.12 Gbp of eDNA with an average insert size of ~40 kb. Though the same procedure was used for both samples, the DF library surpassed the total library clones generated when compared to the RF library. The reason for this goes beyond the scope of this work, but one possible explanation could be that the different environmental factors and type of soil plays an important role in cloning efficiency.

Molecular differences were observed between both libraries. According to the restriction digest analysis with NotI, DF clones contained DNA higher in GC content than RF clones as seen in Fig. 3. The temperature and precipitation discrepancies between both environments may contribute to more stable DNA in their respective environments. Microorganisms inhabiting the Guánica Dry Forest, a hotter and more arid environment when compared to the Yunque Rainforest, have likely undergone genome-wide adaptations, which have resulted in comparatively higher GC content and, subsequently, a higher number of hydrogen bonds between DNA strands to prevent denaturation.

4.2 Library screening strategy

Various methods have been employed to find these biomolecules activities: selecting for antibiotic resistance and screening for lipases using solid media, and screening for ureases using liquid media. Of these strategies, potential eDNA genes have been found that confer antibiotic resistance or urease activity.

The lipolytic screen requires every individual clone in the library be observed on solid media to determine lipolytic activity. Thus, for a +700,000 clone library and the capability of only being able to screen approximately 450 clones per plate, this strategy is the most resource consuming and least effective. Lipolytic clones had been characterized in libraries containing between 33,000 to 100,000 clones with an average 40 kb eDNA insert size [16]. Due to the small number of clones obtained in the RF library, results suggest that indeed the RF library does not necessarily represent the entire microbial community present in the rainforest. Thus, it is of no surprise that no lipolytic clone has been identified in this library. As for the DF library, only 17 % has been presently screened. Previous screenings of cultivable isolates from Guánica’s dry forest soil were identified to have lipase activity (data not shown). This suggests that indeed lipolytic genes may be present in the DF library, but require further thorough screening efforts to be detected.

The urease screens consisted of liquid media in which metagenomic library subpool fractions were assayed at a time. Urea broth media contains phenol red as a pH indicator. Under alkaline conditions, the media turns fuchsia allowing detection of urease activity due to ammonia accumulation. Serial dilutions were performed to narrow down any possible urease-positive candidates. This strategy allowed us to find putative urease-positive clones within the DF library, and was comparably quicker and more effective than the lipolytic screen. The restriction analysis done showed that 22 of them differ in their restriction pattern. However, due to the insert DNA size, whether these 22 clones hold different urease genes without sequencing cannot be stated. Preliminary in silico analysis done for putative urease-positive clones showed that the most prevailing sequence in the DNA inserts came from members of the Acidobacteria class. Previous molecular studies have shown that the most dominant microorganism present in soils belongs to this class [26]. Due to the limited number of sequences obtained, further transposition reactions and sequencing are required to understand and unveil the identity of the putative urease genes.

Though 22 different putative urease-positive clones were isolated, hundreds of antibiotic resistant clones were isolated within days, because of the type of monitoring strategy used. Selection strategies eliminate the need to observe all library clones by allowing only the growth of clones with inserts conferring the desired phenotype, in this case antibiotic resistance. Thus, provided that the activity being assayed for is present within the DNA fragments of the metagenomic clones and readily expressed by its surrogate host, a selection assay should be the quickest and most effective strategy, followed by screening on liquid media, and finally by screening on solid media.

Coupling transposon mutagenesis with a functional assay facilitated the search for the activities detected. Given the average size of the library inserts and the average size of tetracycline resistance genes (~< 1 kb), it would not have been cost-effective to employ a primer-walking strategy to find our gene of interest. As seen in Fig. 4, if a clone is able to grow on plates (A) and (C) but not (B), this may suggest the clone contains a transposition within the tetracycline resistance gene, or perhaps any other genetic element such as promoters or upstream genes (if the gene of interest is co-transcribed) which interrupt the resistance phenotype. Ongoing work is being done to increase the number of sequences
obtained to characterize the putative tetracycline resistance genes. Once the genes conferring the respective activity are fully described, PCR cloning can be used to isolate the gene products and continue with biochemical characterization and comparison with other antibiotic resistance projects. For example, Allen et al. (2009) found both ancient and novel β-lactamases in Alaskan soil, which provides evidence that resistance may be evolving in soil due both to human and natural selective pressures, even in remote locations [21]. Further comparison of antibiotic resistance reservoirs from different environments, including from the human distal gut [22], may provide clues to find the origin of antibiotic resistance and their transfer to clinical settings, and mechanisms to diminish the prevalence of multi-drug resistant pathogens.

4.3 Soil metagenomics as a tool

Cloning the soil metagenome of two different forests in Puerto Rico has allowed access to activities with biomedical and biotechnological applications present in microbes that cannot grow by standard methods. The growing field of Metagenomics casts a new perspective on soils, by providing tools to better understand the role microorganisms play in their ecosystem, and to unearth previously hidden microbial genes and their products. These generated libraries will remain both as a source to continue exploring other biological and functional traits, including the diversity of the microbial communities present, and as an educational tool to increase awareness in the general public towards the importance of soil and its conservation as a valuable natural resource.

Acknowledgements

We would like to give our thanks to the Handelsman Lab for their technical and intellectual contribution in the generation of the metagenomic libraries. We also thank the USDA Forest Service and the Department of Natural and Environmental Resources of Puerto Rico for kindly providing access to samples. Funding for this project was provided by USDA-CSREES: 2007-02386, PR-LSAMP, Bio-MINDS, and MARC programs at UPRM.

References


