JOURNAL OF
EMERGING INVESTIGATORS

Article

Genomic Signature Analysis for the Strategic Bioremediation of Polycyclic Aromatic Hydrocarbons in Mangrove Ecosystems in the Gulf of Tonkin

Nam H. Dao[1], Quan M. Le[2], Thang X. Dao[3]
[1]The Hill School, USA
[2]The Loomis Chaffee School, USA
[3]Unaffiliated

SUMMARY

Current oil spill cleanup methods can be inefficient, costly, and have many more downsides. In an effort to create an efficient, cost-effective, and environmentally-friendly solution to cleaning oil spills, there has been a flurry of advancements in microbiological vectors. These plasmids confer oil-degrading genes that can be implemented into the native microbial population. For this strategy to work, it is critical that an appropriate naturally-found plasmid be picked as a base to ensure long-term plasmid survival and activation of its genes. To help realize this goal, this study relies on emerging computational technologies to measure genetic similarity between the host chromosome and the potential plasmid vector, utilizing an algorithm for 3-mer genomic signature analysis to determine the best naturally-found plasmids for later use as vectors, as well as detect local species or genera that are most compatible with these plasmids. Our hypothesis is that the resulting plasmid-species combinations will reflect existing literature regarding historical host-species interactions, further supporting genomic signature analysis as a tool to find historical plasmid-host interactions and thus predict plasmid compatibility. The plasmids we identify can then be chosen as a base vector to be equipped with hydrocarbon-degrading genes in future studies. Based on our results, for mangrove forests located in the Gulf of Tonkin, the combinations IncP-9 (*Pseudomonas lurida*, *Pseudomonas pelagia*, *Pseudomonas alkylphenolica*, *Pseudomonas antarctica*), IncA/C (*Marinobacter salinus*), Ri/Ti (*Erythrobacter atlanticus*) or Ri/Ti (*Pseudorhodoplanes sinuspersici*) should be the targets of future oil-degrading bioremediation strategies.

INTRODUCTION

Our consistent dependence on oil consumption is pushing the planet's ecological habitats to the brink of collapse. This progression has resulted in an unprecedented level of oil spilled into ecosystems, hurting the biodiversity within these communities (1). In particular, mangrove forests, a type of coastal ecosystem that is often the subject of excessive hydrocarbon intake from ocean waves, are especially at risk. Over the last 6 decades, at least 238 significant oil spills have been recorded in these ecosystems, directly impacting up to 1.94 million hectares (ha), or 19,400 square kilometers, of mangrove habitat (2). Oil is responsible for blocking the pores of cell membranes, leading to a failure in essential metabolic processes. Once oil has entered the roots and rhizosphere (the soil directly influenced by plant roots) of mangrove trees, they fail to participate in regular salt management and respiration (3). As mangrove trees are responsible for various important functions in this habitat such as protection of soil from erosion, improving water quality or helping provide nutrients for microbial communities, the death or failure of these trees to function can pose heavy consequences for local organisms (3). Upon direct contamination by an oil spill, mangrove trees usually face death within several months, and other organisms living in the area that depend on a host of benefits provided by mangrove trees are subsequently either impacted directly by the influx of oil or indirectly through the loss of mangrove trees (3).

Mangrove forests are highly productive beyond local impact; local biodiversity has different commercial uses, such as providing refuge for shrimp and oysters. This ecosystem also functions as a carbon container, reducing the proliferation of carbon in the atmosphere and global climate change (3). Therefore, it is critical that drastic measures be taken to best reduce the impact that oil spills may have on mangrove forests.

Several methods of oil clean-up have been attempted in mangrove forests ranging from vacuuming to chemical shoreline cleaners, but the sophisticated networks of mangrove trees, soil, and organisms make it difficult to effectively dispose of oil without also harming local species (3). Bioremediation, the addition of hydrocarbon-degrading bacteria and/or genes to the local environment, has enjoyed special attention due to its efficiency, effectiveness, and eco-friendly nature. When applied in mangrove forests, a reliance on the local microbial community as part of a bioremediation strategy promises to catalyze the oil quickly while providing little disturbance to the ecosystem (1). Interestingly, the presence of natural hydrocarbon-degrading bacteria has been observed in various environments that have faced frequent oil spills, and up to 300 genera of bacteria have been identified to be mainly reliant on hydrocarbons as a source of energy and nutrients (4). Although oil-degrading microbes have already naturally emerged, they only form in small numbers and emerge after some time has already passed since the initial spill, where a larger, immediate presence offered by bioremediation will be quicker to minimize the effects of hydrocarbons (1). This strategy would rely on the existence of known hydrocarbon-degrading catabolic pathways, thereby creating an effective

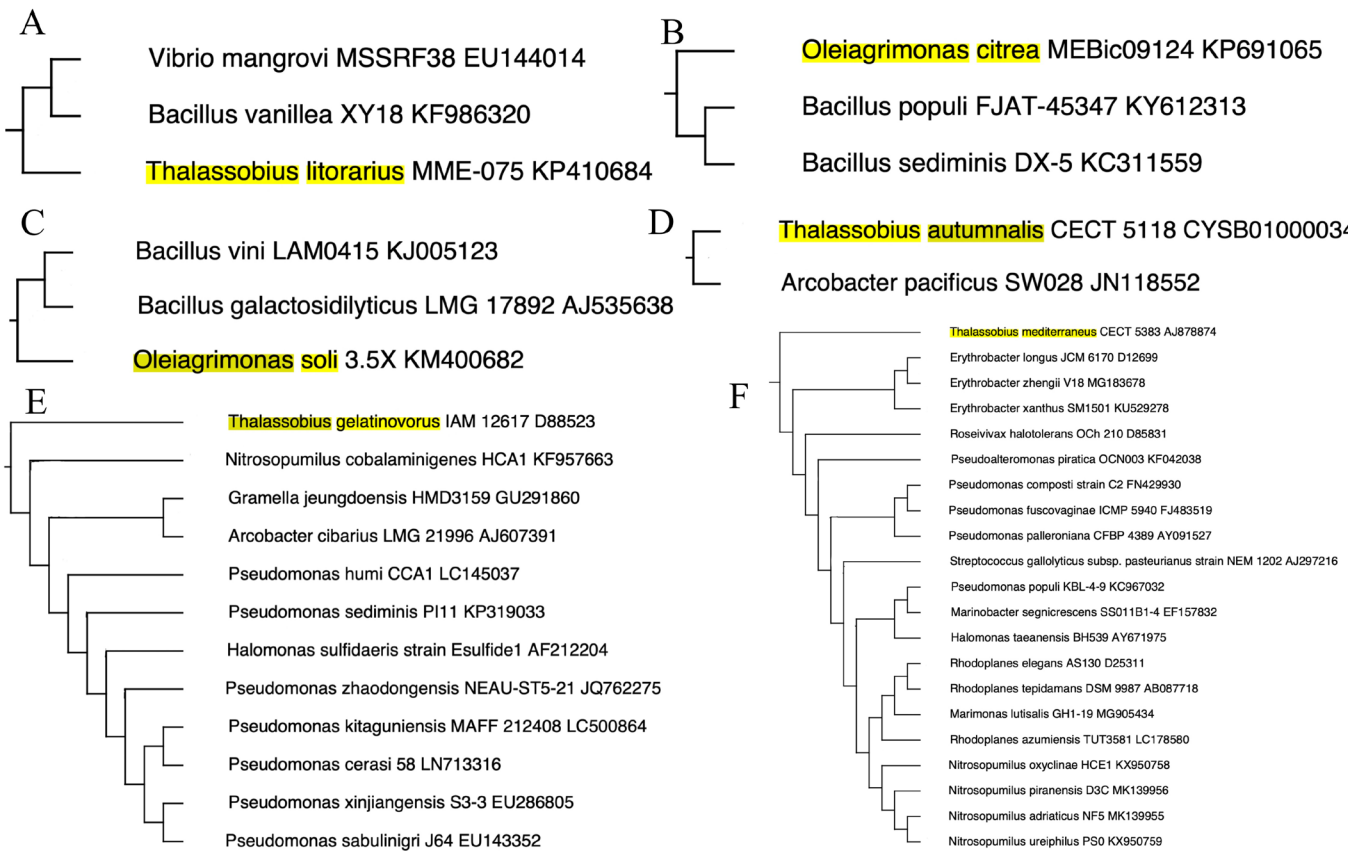

Figure 1. The 38 candidate species and their phylogenetic relations with *Thalassobius* or *Oleiagrimonas* species (highlighted in yellow) extracted from 16s rRNA-based phylogenetic tree. Each tree encapsulates the species from the focal ecosystem that are most genetically similar to a species belonging to the known PAH-degrading genera (*Thalassobius* and *Oleiagrimonas*). Each letter (**A** to **E**) depicts a set of branches on the phylogenetic tree, depicting the species from the Gulf of Tonkin most closely related to each PAH-degrading species. (**A**) The two local species that are most closely-related to *Thalassobius litorarius*. (**B**) The two local species that are most closely-related to *O. citrea*. (**C**) The two local species that are most closely related to *Oleiagrimonas* soli. (**D**) The species that is most closely related to *Thalassobius autumnalis*. (**E**) The 11 local species that are most closely related to *Thalassobius gelatinovorus*. (**F**) The 20 local species that are most closely related to *Thalassobius mediterraneus*.

solution that is both environmentally-friendly and financially-viable. This study utilizes an algorithm built upon evolutionary host-specific genetic ameliorations, which allows for the comparison of the chromosomal DNA with plasmid DNA to see if they have had historical interactions— a considerable measure of plasmid promiscuity (5, 6). A compatible plasmid, defined as being able to survive, reproduce and express its genes in the target bacterial species, can then be selected as a base for future gene recombinant designs. Due to the aforementioned risks that mangrove forests are subject to, various genera of bacteria from mangrove forests in the Gulf of Tonkin will be used as the subjects of this experiment. Our hypothesis is that the highest scoring host-plasmid combinations will successfully predict real-life historical host-plasmid interactions. These high scoring combinations solidify this technology as a tool for future predictions of plasmid compatibility and potential modalities for bioremediation in oil spill management.

RESULTS

Crude oil is composed of a variety of different components, such as aliphatics (alkanes, alkenes, alkynes, etc.) and aromatics (monoaromatic, polycyclic aromatic hydrocarbon). Although many species of bacteria are individually capable of the biotransformation of a small subset of components, the most efficient bioremediation effort of oil spills will need to comprise a variety of different species that can degrade various hydrocarbon components (1). Consequently, plasmids that are designed to confer hydrocarbon-degrading capabilities must be able to remain stable in a vast array of phylogenetic classifications, or evolutionary lineages of bacteria.

First, to identify any bacteria already present in the ecosystem that could degrade hydrocarbons, we compared local genera in the Gulf of Tonkin with genera that heavily depend on degrading hydrocarbons (4). Many more species have been reported to degrade hydrocarbons, but we focused specifically on this group because bacteria that rely more heavily on hydrocarbons are expected to have developed more efficient catabolic pathways for hydrocarbon degradation. Although 12

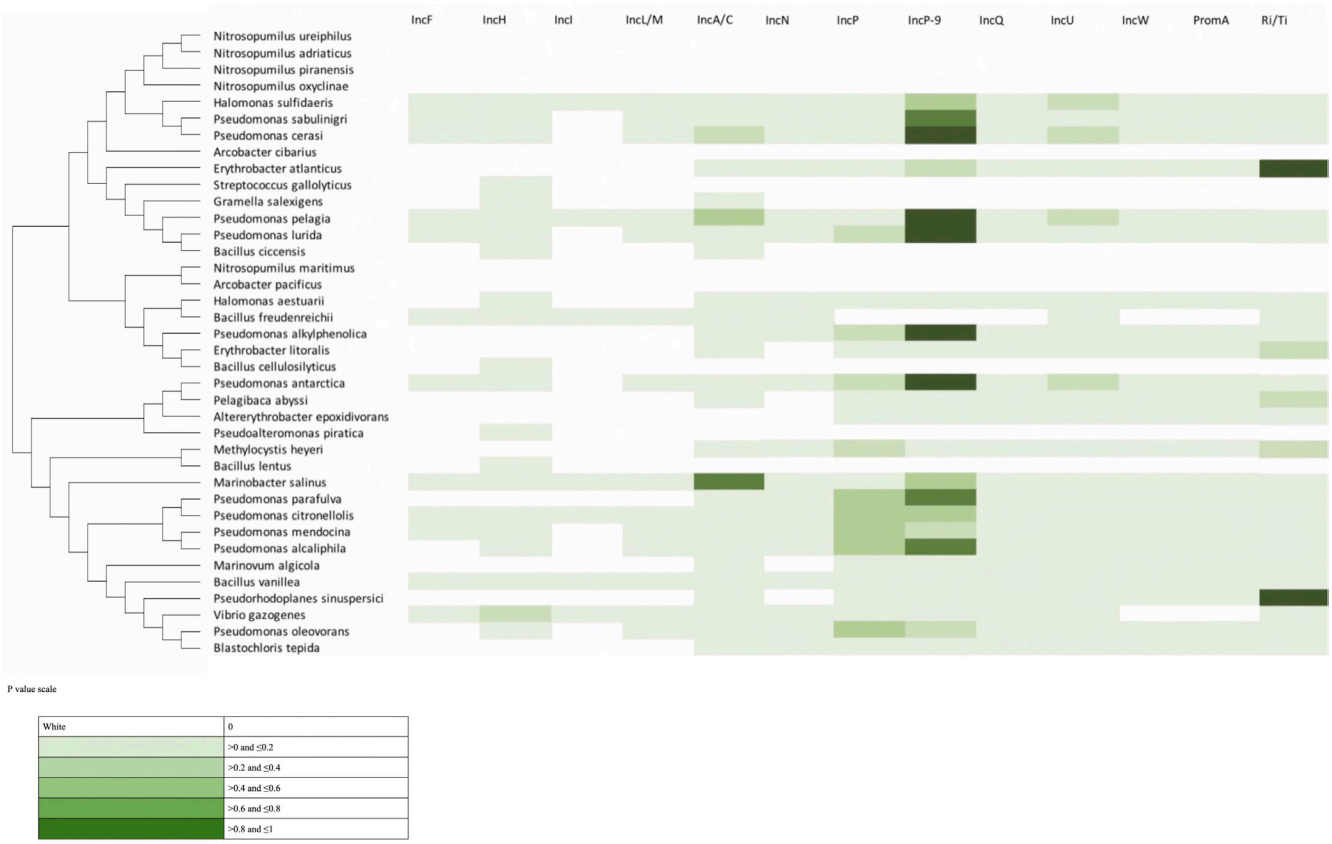

Figure 2: P values per incompatibility group of all 38 target species. Individual plasmid-species *P*-values were collected from R script algorithms and grouped/averaged based on plasmid incompatibility group. Inc groups differ in their replication and partition mechanisms, where plasmids from the same group will fail to concurrently exist in the same cell due to these similarities. The color for each species- Inc group pair darkens as its *P*-value approaches 1. MUSCLE and MEGA used for 16s rRNA phylogenetic tree creation.

local genera are known to degrade hydrocarbons, none of them heavily rely on the degradation of polycyclic aromatic hydrocarbons (PAHs), a type of hydrocarbon present in commercial oil that confers particular danger to mangrove trees and other local organisms. For this reason, we shifted our focus toward finding appropriate plasmids that can receive known PAH-degrading genes in future experiments.

Next, to identify the local species that are most genetically similar to PAH-degrading ones, we compared the genomes of the genera *Thalassobius* and *Oleiagrimonas*, two genera that heavily rely on catalyzing PAH, with species from the local genera by aligning and constructing a maximum parsimony phylogenetic tree using their 16s rRNA sequences and standard parameters on MEGA X (7, 8). The most genetically similar species across the two groups totaled 38 different species (**Figure 1**).

Then, to test the receptiveness of these species to various plasmids, we utilized a computational algorithm that contrasts 3-mer oligonucleotide sequences between plasmids and species to find the most compatible combinations. This yielded *P*-values measuring the degree of genetic similarity between a plasmid and species ranging from 0 to 1, with the highest P values indicating the most genetic similarity (*P*-values, defined as the results produced by the algorithm from Suzuki *et al.*, are not to be confused with *p*-values traditionally used in measuring statistical significance). With these values, we determined that several members of *Pseudomonas* genera and IncP-9 plasmids had the highest scoring species-plasmid combinations (**Figure 2**). Other notable species-plasmid group combinations include *M. Salinus* (Inc A/C); *E. atlanticus* (Ri/Ti) and *P. sinuspersici* (Ri/Ti).

Finally, to verify the results of this algorithm, the combinations were then juxtaposed with existing literature to confirm the existence of historical species-plasmid interactions. While the connection between IncP-9 plasmids and the several Pseudomonas bacterial species was documented, the other combinations did not have any historically substantial backing. No singular plasmid group ever consistently (≥19/38 species) scored above 0.6— the cutoff point determining whether a plasmid would be compatible with a species set by Suzuki, *et al* (6).

DISCUSSION

Our original hypothesis was partially supported by the results of the genomic signature comparison producing combinations already found in pre-existing literature, which confirms that IncP-9 plasmids did have historic interactions with the *Pseudomonas* genera of bacteria (9). However,

little is known about interactions between IncA/C plasmids with *M. salinus* or Ri/Ti plasmids with either *E. atlanticus* or *P. sinuspersici*, which cannot further confirm the reliability of oligonucleotide comparison as a technique for predicting historical plasmid-host interactions. Alternatively, this ambiguity can simply be attributed to lack of complete genomic sequencing and understanding of the aforementioned species of bacteria.

As for larger understanding of utilizing oil-degrading plasmids, two different conclusions as to which plasmid(s) should be picked for optimal chances of expressing PAH-degrading genes can be drawn from the results, depending on the choice of maintenance strategy: broad- or narrow-host range plasmids (**Figure 2**).

Assessing genomic signature is only part of the equation when it comes to measuring plasmid success. As it utilizes similarity of genomes, the algorithm can only detect genetically similar plasmids as potential candidates. This neglects broad-host-range (BHR) plasmids that have been shown to survive and have prolonged stability but whose own promiscuousness has prevented enough genomic amelioration to be picked up by the algorithm (5, 6). BHR plasmids such as pRA3 and pB10 have seen success in representatives across α-, β- and γ-proteobacteria, but their *P*-value scores from this study are not significant because of their promiscuous nature. Choosing a single BHR plasmid means that the range of the potential species is vast, but the probability of success is lower.

The tendency for BHR plasmids to have low *P*-value scores has been discussed by previous papers regarding genomic signature analysis (5, 6). There are currently few ways to differentiate BHR plasmids from plasmids that are simply incompatible, as an analysis of the genomic signatures of both can yield similar results. The matter is made even more complicated when analysis has shown that small genetic fragments may not be picked up by the algorithm due to the mosaic nature of plasmid genomes that may otherwise indicate a possible evolutionary host: each plasmid's genetic composition can be derived from several parental plasmids that come from phylogenetically distant bacteria (10). Still, other studies in correlating genomic signature performance with performance via physical experiments have yielded fairly positive results (5, 10, 11). As the catabolic pathways of PAH-degrading plasmids are well-studied, we hope genomic structure analysis can be optimized to the PAH-catabolic pathways for more effective predictions to use for vector in vivo applications (12).

In contrast, choosing a number of narrow-host-range-plasmids (NHR) to deploy PAH-degrading genes will promise a higher likelihood that the plasmid will remain stable even without positive selection. However, this promise comes with the caveat that plasmids will have a harder time transferring and succeeding across phylogenetic divisions, potentially reducing the process's efficiency (6). In such a case, the search for plasmids should go much further beyond the 92 plasmids used in this study, with a focus on plasmids that have been specifically found in target genera.

This study highlighted the plasmids that are most likely to be successful as a base for hydrocarbon-degrading genes, while classifying the possible approaches into two options. Still, comparative genomic signatures analysis is far from a perfect technology, as seen by the several seemingly dubious predictions regarding species-host historic interactions. Further improvements of the algorithm utilized will be needed to better predict a plasmid's host range based on its genome (5, 6, 10). Furthermore, more complete plasmid and bacterial genomes would provide a better picture of a specific plasmid's host range, as previous studies have highlighted that plasmids from the same incompatibility group can still vastly differ in their respective host ranges (13). Further study of PAH-catabolic pathways will also be needed to create a more comprehensive overall strategy in which a wider range of PAHs are degraded. These advancements can also be later utilized in bioremediation efforts for other ecosystems in the future.

METHODS

A list of the 35 most common genera of bacteria living in various mangrove trees spread throughout the Gulf of Tonkin was compiled (14). This list of genera was then compared to a list of 320 genera of bacteria that heavily rely on hydrocarbons as a source of nutrient and energy (4). From this, 12 of the 35 genera were found to feature such species (**Table 1**). However, as none of the 35 included genera specifically targeted PAHs, genomic comparisons were further completed between local genera and any that heavily rely on PAH degradation.

Genera of Hydrocarbon-Degrading Found in Mangrove Ecosystems	Phylum	Hydrocarbon Component Targeted
Bacillus	Firmicutes	Toluene
Roseivivax	Proteobacteria	Alkanes
Erythrobacter	Proteobacteria	Alkanes
Marinobacter	Proteobacteria	Crude Oil
Microbulbifer	Proteobacteria	Crude Oil
Pseudoalteromonas	Proteobacteria	Crude Oil
Serratia	Proteobacteria	Crude Oil
Halomonas	Proteobacteria	Crude Oil
Marinomonas	Proteobacteria	Phenanthrene
Vibrio	Proteobacteria	Phenanthrene
Pseudomonas	Proteobacteria	Gas Oil
Desulfococcus	Proteobacteria	Alkanes
Desulfosarcina	Proteobacteria	Xylene

Table 1. The 12 genera of bacteria local to the Gulf Tonkin that heavily rely on catalyzing hydrocarbons.

These were constructed using an algorithm and instructions provided online by Suzuki et al. and a standard computer command line (15).

From the aforementioned list of ~320 bacteria, 2 genera of hydrocarbon degrading bacteria were selected as the sole hydrocarbon degraders whose most typical target substrate was PAH. As Thalassobius and Oleiagrimonas are the only two genera found to mainly degrade PAHs, all local species whose 16s rRNA sequence most closely resembles those of species from these 2 genera are selected to be candidate species for future plasmid introduction based on the idea that more genetically similar species will be better suited for plasmid survival and reproduction. To this end, a maximum parsimony tree was created on the MEGA X program using the embedded MUSCLE multi-sequence alignment using the program's standard parameters to compare individual species present in mangrove forests and PAH-degrading bacteria.

The construction of this phylogenetic tree resulted in a list of 38 species whose genomes are the closest to those of the Thalassobius and Oleiagrimonas species, making them the best candidates to receive and utilize PAH-degrading genes. Next, a previously developed algorithm was utilized in order to calculate each species' potential receptiveness to 92 common plasmids, which were originally compiled by Suzuki, et al to similarly test the promiscuity of different plasmids (5, 6). The algorithm utilizes 3-mer genomic signature to determine the degree of similarity in oligonucleotide composition between the plasmids and host chromosomes by calculating the Mahalanobis distance, the difference in abundance of various trinucleotides, between their genomes. Evolutionary interactions and genomic amelioration between plasmids and species will lead to similar genetic composition, resulting in smaller Mahalanobis distances that will impact the final P-values; higher values mean that the oligonucleotide composition between a plasmid and host are more similar than not (10). The plasmid will thus have a higher degree of stability within the species and a higher probability that the genes it contains can be effectively utilized.

Utilizing 3-mer genomic signature to determine plasmid compatibility yielded P-values for individual plasmids for each species. Plasmid incompatibility groups are defined as plasmids that cannot be stably replicated and expressed together, so that plasmids from the same group will be incompatible over generations. The 92 plasmids were designated into groups IncF, IncH, IncI, IncL/M, IncA/C, IncN, IncP, IncP-9, IncQ, IncU, IncW, PromA and Ri/Ti (5). Thus, the scores of the plasmids were grouped together based on their incompatibility groups and averaged. This average was then chosen as the preferred indicator of plasmid success when comparing between different plasmids and incompatibility groups.

Received: January 14, 2021
Accepted: June 8th, 2021
Published: June 27, 2021

REFERENCES

1. Santos, Henrique F., *et al*. "Bioremediation of Mangroves Impacted by Petroleum." *Water, Air, & Soil Pollution*, vol. 216, no. 1–4, 2010, pp. 329–50., doi:10.1007/s11270-010-0536-4.
2. Duke, Norman C. "Oil Spill Impacts on Mangroves: Recommendations for Operational Planning and Action Based on a Global Review." *Marine Pollution Bulletin*, vol. 109, no. 2, 2016, pp. 700–715., doi:10.1016/j.marpolbul.2016.06.082.
3. Hoff, Rebecca, *et al*. "Oil Spills in Mangroves: Planning & Response Considerations." *NOAA/Office of Response and Restoration*, 2014., doi:10.13140/2.1.3588.5765.
4. Prince, Roger C., *et a*l. "Prokaryotic Hydrocarbon Degraders." *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes*, 2019, pp. 1–39., doi:10.1007/978-3-030-14796-9_15.
5. Suzuki, Haruo, *et al*. "Predicting Plasmid Promiscuity Based on Genomic Signature." *Journal of Bacteriology*, vol. 192, no. 22, 2010, pp. 6045–55., doi:10.1128/jb.00277-10.
6. Suzuki, Haruo, Masahiro Sota, *et al*. "Using Mahalanobis Distance to Compare Genomic Signatures between Bacterial Plasmids and Chromosomes." *Nucleic Acids Research*, vol. 36, no. 22, 2008, p. E147., doi:10.1093/nar/gkn753.
7. NCBI Resource Coordinators. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research*, vol. 45, no. D1, 2016., doi:10.1093/nar/gkw1071.
8. Kumar, Sudhir, *et al*. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." *Molecular Biology and Evolution*, vol. 35, no. 6, 2018, pp. 1547–1549., doi:10.1093/molbev/msy096.
9. Sevastsyanovich, Yanina R., *et al*. "Diversity of IncP-9 Plasmids of Pseudomonas." *Microbiology*, vol. 154, no. 10, 2008, pp. 2929–41., doi:10.1099/mic.0.2008/017939-0.
10. Suzuki, Haruo, Celeste J. Brown, *et al*. "Genomic Signature Analysis to Predict Plasmid Host Range." *Molecular Life Sciences*, 2018, pp. 458–64., doi:10.1007/978-1-4614-1531-2_574.
11. Tokuda, Maho, *et al*. "Determination of Plasmid PSN1216-29 Host Range and the Similarity in Oligonucleotide Composition Between Plasmid and Host Chromosomes." *Frontiers in Microbiology*, vol. 11, 2020., doi:10.3389/fmicb.2020.01187.
12. Seo, Jong-Su, *et al*. "Bacterial Degradation of Aromatic Compounds." *International Journal of Environmental Research and Public Health*, vol. 6, no. 1, 2009, pp. 278–309., doi:10.3390/ijerph6010278.
13. Yano, Hirokazu, *et al*. "Host Range Diversification within the IncP-1 Plasmid Group." *Microbiology*, vol. 159, no. Pt_11, 2013, pp. 2303–15., doi:10.1099/mic.0.068387-0.
14. Gong, Bin, *et al*. "High-Throughput Sequencing and Analysis of Microbial Communities in the Mangrove Swamps along the Coast of Beibu Gulf in Guangxi, China." *Scientific Reports*, vol. 9, no. 1, 2019., doi:10.1038/s41598-019-45804-w.
15. Suzuki, Haruo. "Measuring Mahalanobis Distances between Plasmid pSN1216-29 and Chromosomes." *GitHub*,

github.com/haruosuz/plasmids/tree/master/pSN1216-29.