Sequencing the Microbial World One Bug at a Time

Authors: Jay Moore, Ryan Yang, Kyle Wheeler, Dr. Noha Youssef*
Department of Microbiology & Molecular Genetics, Oklahoma State University

Abstract

Most of the microbial world has never been observed, let alone isolated by mankind. Despite scientific advancements soaring in the past couple decades, there still remains many mysteries surrounding microbiology. For starters, there are an unimaginable amount of microbes that are on our earth that we have never observed, studied, or classified. For most people, they are just strange bacteria, but to us, they hold the key to today's problems. Of the 250 isolates obtained, 24 were identified to be unique. DNA extracted from these samples were sent for Illumina sequencing at the University of Georgia Genomics Facility. Assembled genomes are currently being analyzed and prepared for publication in hopes of helping the scientific community and microbiologists alike to broaden their understanding of different microorganisms and to facilitate further discoveries within the medical and scientific field surrounding microbes.

Keywords: Microbiology, Novel strains, Metagenomics, Novel microbes, and Environmental metagenomics

Introduction

The number of validly described species in a given genus is one of the ways that scientists can measure biodiversity. For instance, there are an estimated 1.2 million insect species that have been validly identified, yet there have only been 7,000 bacterial species that have been validly described (Stewart 2012). The researchers of the Microbial Diversity Project hope to identify more novel organisms to increase the scientific community's understanding of the microbial world using one of the most promising new methodologies in microbiology: metagenomics. Metagenomics is the genomic analysis of microbial DNA that is isolated from environmental samples. Out of the few thousand microbes that have been sequenced and identified, most of them are animal and/or human pathogens. The Microbial Diversity Project seeks to find novel microbial strains in samples taken from exotic environments, for example, the Oklahoma Great Salt Plains. We also took samples from two MICR classes taught at Oklahoma State University. The courses in question asked students to take samples from random environments and then culture the strains found there.

Scientists in the past have worked with nonpathogenic bacterial strains to power batteries, restore destroyed ecosystems, and create biofuels. Of course, when working with these microbes, they probably did not fully understand what they had in their hands, or the implications of their research. In New Scientist, Randerson (2003) states that a team of scientists found an obscure microbe called *Rhodoferax ferrireducens*, which was described for the first time in 2003, that can break down industrial waste and form an electrical current. The scientists who found *R. ferrireducens* couldn't have predicted that this iron-philic organism would inspire other scientists to power batteries with their ability to break down sugars. Scientists also understand now that the microbes that live in soil contribute to plant productivity and diversity in terrestrial ecosystems (van der Heijden 2008).

The same is true for our work. While in these preliminary stages, it's hard to imagine what purpose the results of this project will serve. Researchers could travel to the most remote corners of the earth and collect samples that only culture known microbes, or researchers could find a plethora of novel strains in these places. Even if researchers isolate many new strains, there is no guarantee that those species will be useful to the scientific community as a whole. That is the beauty and the most frustrating aspect of environmental shotgun sequencing all rolled into one. While environmental shotgun sequencing offers one of the best ways to find never-seen-before microorganisms, this method of sequencing produces a vast amount of confusing data that can stump even the most well equipped research team.

Last year, a student working on this project discovered a new species of *Chryseobacterium* on a light switch from Stillwater, OK (Couger *et al.* 2015). They went on to sequence the genome of *Chryseobacterium* sp. strain Hurlbut01 at the University of Georgia Genomics Facility using

Table 1 *- shows the interesting 24 isolates, the genus and species they are belonging to and whether they belong to a species with no sequenced representatives or with sequenced but unpublished representatives. Assembly results are also shown.*

Isolate Name	Mass Spectrometry/16S rRNA Identity	Genome Available	Publication Available	Number of Contigs	N50	Size of Largest Contig	Genome Size	Percent Similarity to Closest Sequenced Representative
Rutherford	Rhodococcus corynebacteroides	No	No	40	1361378	1783990	4053164	99
Laake	Curtobacterium oceanosedimentum	Yes	No	28	1266321	1705193	3737896	98
Montgomery	Macrococcus sp.	No	Publication on Macrococcus caseolyticus	38	1291070	1291070	2021599	98
Davis	Paenibacillus urinalis	No	No	107	1308153	2273583	5815617	97
Newton	Exiguobacterium acetylicum	Yes	Yes on acetylicum	81	2179978	2179978	3324777	99
Peters	Exiguobacterium indicum	No	Yes on 10 unidentified species within the genus	39	2145682	2145682	3132088	99
Yoes	Erwinia sp.	No		59	1515681	1558341	4892520	99
Ironside	Kocuria rosea	No	No	51	857460	1594640	3880193	98
Allameh	Paracoccus yeei	Yes	No	211	529783	1510178	4459809	98
Garren	Bacillus drentensis	No	No	157	647351	993984	6183344	99
Hollocomb	Bacillus nealsonii	Yes	No					
Wright	Paenibacillus lactis	Yes	No	112	2345011	2405936	6749542	100
Sutherland	Staphylococcus condementi	No	No	62	1690626	1690626	2690862	99
Chappell	Paracoccus yeei	Yes	No	230	262192	462966	4424659	100
Laird	Pseudomonas vulva	No	No	149	250526	642493	4832320	99
Burchfield	Leclercia adecarboxylata	Yes	No	80	1252018	1412696	4703139	99
Smith	Pantoea calida	No	No	135	260770	819060	4297010	100
Francis	Azorhizophilus paspali	No	No					
Omps	Bacillus muralis	Yes	No	126	1296012	1340904	5062189	99
Landers	87 identity with Bacterium 'ma		No					
Rudy	Microvirga subterranea	No	No	153	841120	2515035	6652794	95
Prendergast	Chryseobacterium camelliae	No	No	42	2241341	2241341	4274967	98
Mccann	Sphingobacterium sp. SCU-B140	No	No	69	450238	889017	4649484	92
Kotzamanis	Sphingomonas pseudosanguinis	No	No	97	247125	618540	4313967	97

2×300 paired-end chemistry and a library insert size of 700 bp. The goal of this year's research is to follow in Allison Hurlbut's footsteps and isolate new strains of bacteria - primarily halophilic and animal microflora - to later sequence and then publish our findings. Even if we find new strains, their usefulness (or lack thereof) might not be apparent at first. But that's alright; that is the purpose of basic research. We want to expand the scientific community's knowledge of known microorganisms so that one day scientists can see a pattern or make an observation that nobody had before. This spark of genius that leads to the finding of uses for microorganisms is fueled by the basic research that we are conducting.

Methods

Environmental samples, mainly plant material and soil samples, were collected from the Oklahoma Great Salt Plains. Additionally, samples collected by students enrolled in MICR 2133 and MICR 4012 were used. Our team had to develop a special, high salt concentration, growth media formula for the salt plains samples. After numerous trials and errors, we succeeded. The environmental samples were diluted and put into a liquid growth media to be incubated. We then transferred it onto plates; these would become the master plate for the sample. Individual isolated colonies from the master plate would be transferred into new plates, in which they will be cultured and send to the Oklahoma Animal Disease Diagnostics Lab (OADDL) along with the already isolated strains obtained from MICR

2133 & MICR 4012 students to be analyzed with MALDI-TOF mass spectrometry and 16s rRNA sequencing. From here, 24 isolates were determined to be potentially unique. DNA extraction was then performed and sent off to the University of Georgia Genomics Facility for full genome sequencing. Results were then analyzed in preparation of publishing genome announcement manuscripts for each of the 24 isolates.

Results

During the Microbial Diversity Research project, MICR 2132 and MICR 4022 obtained 250 isolates that were submitted for screening by Mass Spectrometry Analysis and 16S rRNA gene sequencing. The results of these analyses identified 24 interesting strains that exhibited signs that they belonged to species with unsequenced strains. Researchers extracted DNA from these isolates and sent the DNA for Illumina sequencing at the University of Georgia Genomics Facility. Drafts of the sequenced DNA were produced and are now being analyzed before they will be published. See Table 1 at the end of the report.

Table 2 *– This table shows the contents of the studies vitamin solution.*

KH_2PO_4 (Monobasic)	0.45 g/L
K_2HPO_4	0.9 g/L
$MgSO_4$	0.5 g/L
NaCl	145 g/L for 2.5 M prep, 58 g/L for 1.0 M agar plates
NH_4Cl	0.5 g/L
KCl	0.3 g/L
Yeast Extract	0.2 g/L
$FeCl_3$	0.05 g/L
Vitamin solution (see below)	1.0 ml
Trace metal solution (see below)	10.0 ml

Discussion

Through the Microbial Diversity Research Project, we have discovered that we can more efficiently discover novel strains of bacteria by using Mass Spectrometry Analysis and 16S rRNA to rule out any known strains instead of only sending random isolates for sequencing without previous screening. In future projects, we plan to use other exotic samples in order to access more diverse ecosystems to increase our chances of discovering novel strains of bacteria. It was worth noting that screening of the isolates was a huge bottleneck.

The reason we only found 24 new strains out of the 250 we isolated was because fast growing bacterium would take over the plate before other strains had the chance to grow. In future experiments, we will introduce new variables in the media like a lower concentration of food or unusual pH so we can screen out the fast growers. We also grew all of the plates at 30° to 36° C. The normal body temperature range in humans is around 36.8° C, so it's quite possible that we found so many known organisms, because these have already been grown and studied (Mackowiak 1992). We suspect that most research in microbiology and research into discovering new microbes has revolved around humans, so it's only natural that most microbes that grow at the human body temperature have already been discovered. If we grow the future isolates in more extreme temperatures, we believe we will see more variability in the strains.

Acknowledgements

This research was supported by funding from the Howard Hughes Medical Institute. We would also like to give a huge thank you to Dr. Gary Marley and Dr. Wouter Hoff. Without their help, this project would not have been possible. Furthermore, we would like to recognize the work of the MICR 2133 and MICR 4012 students for their hard work.

References

Couger M.B, A. Hurlbut, C.L. Murphy, C. Budd, D.P. French, W.D. Hoff, M.S. Elshaheda, and N.H. Youssefa. 2015. Draft Genome Sequence of the Environmental Isolate Chryseobacterium sp. Hurlbut01. *Genome Announcements*. 2015;3(5):e01071-15. doi:10.1128/genomeA.01071-15.

Randerson, J. 2003. Can microbes munch waste and power batteries? *New Scientist*, 179, 19.

Stewart, E..J., 2012. Growing Unculturable Bacteria, *Journal of Bacteriology*, August 2012 vol. 194 no. 16 4151-4160

Van der Heijden, M.,G.A., R.D. Bardgett, and N.M. van Straalen. 2008. The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology* Letters, 11(3), 296.