Pacific Northwest
NATIONAL LABORATORY

PNNL-33795

Redox Sensitive Transcriptional Regulators in Model Cyanobacteria to form a Carbon Reducing Microbial Community for Bioproduction

December 2022

Christopher E. Cotillo
Pavlo Bohutskyi

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-33795

Redox Sensitive Transcriptional Regulators in Model Cyanobacteria to form a Carbon Reducing Microbial Community for Bioproduction

Enter Subtitle Here (or delete)

December 2022

Christopher E. Cotillo
Pavlo Bohutskyi

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Redox Sensitive Transcriptional Regulators in Model Cyanobacteria to form a Carbon Reducing Microbial Community for Bioproduction

Christopher Cotillo

Office of Science, Science Undergraduate Laboratory Internship Program

University of California, Berkeley, CA

Pacific Northwest National Laboratory

Richland, Washington

December 23, 2022

Student: Christopher Cotillo

Research Mentor: Pavlo Bohutskyi

ABSTRACT

Cyanobacteria can utilize sunlight to capture CO2 and convert it into sugars. By pairing the sucrose synthesizing *Synechococcus elongatus* PCC7942 with a yeast cell bio-factory, there is potential for generation of carbon-neutral bioproducts, evading the traditional demand for obtaining sugars from grains and competing with food production. Traditional methods for bioproduction revolves around burning fossil fuels for the essential nitrogen fixation in yeast metabolic processes, a method that pollutes the atmosphere with excess CO2. A solution this project looks towards is microbial communities. However, assembly of effective engineered microbial community requires exhaustive understanding of the cyanobacteria metabolism including revealing intrinsic insights on activation or suppression of function of transcriptional regulators, which control metabolism by expression of most genes. This project was focused on comparative investigation of transcriptional regulators in three model cyanobacteria including S. *elongatus PCC 7942*, *Synechococcus* sp. PCC 7002, and *Synechocystis* sp. PCC 6803. Specifically, we examined the importance of cysteine oxidative post-translational modifications to function as redox-switches that modulate activity of transcriptional factors by inducing protein structural changes in response to oxidative stimuli. Results demonstrated several redox sensitive regulators consistent within all three organisms, though more than half of the genes in the proteomic analysis showed to be specific to its microbe. Furthermore, the overlapping transcriptional regulators mentioned earlier were recorded for

our datasets, as these regulators were proving to be essential for the engineering of the intended microbial community. Additional methods in comparisons are being looked towards for a more thorough analysis with these regulators, and different methods for perturbation tests are being examined by members of the team. Better comprehension of the mechanisms governing activity of these essential regulatory factors can be used to engineer novel and more efficient biomanufacturing platforms that do not pollute the atmosphere, but rather utilize atmospheric CO2 as a resource.

I. INTRODUCTION

Microbial communities are demonstrating promise for bioproduction output, reducing the dependence on natural uncontrollable factors that may inhibit maximum potential of production, all while reducing excess atmospheric carbon dioxide in the process. However, there is lack of information on microbial communities and their ability to be properly engineered. The quantity of microbes in already pre-existing communities makes it near impossible to determine any communicative patterns amongst the different genes. This project's goal is to engineer a microbial community starting with two microbes: a cyanobacteria and a diazotroph. These two microbes would metabolically dependent on each other, and would be paired with a yeast to form carbon based product. Not only will the formation of this community provide a new basis of idea revolving around the potential of microbial community engineering, but the community would in-turn reduce total carbon dioxide from the atmosphere.

The first microbe, *Synechococcus elongatus* PCC 7942 is a cyanobacteria that has been recognized for its metabolic capabilities. Specifically, this cyanobacteria shows ability to replace the sugars produced by grains in the fermentation process achieved by a yeast to form alcohol, bread, and other carbon based products. Essentially, the cyanobacteria *Synechococcus elongatus* PCC would be enacted as the heterotroph, synthesizing compatible sugar to the yeast in the form of sucrose. The second microbe would be a diazotroph – *Azotobacter vinelandii* DJ. This microbe would act as the nitrogen fixating organism for the yeast, using the same sucrose produced by *Synechococcus elongatus* PCC 7942, a process illustrated in **Figure 1**. Through this new process of nitrogen fixation performed by the diazotroph, the harmful traditional process of nitrogen fixation is being avoided. Traditional methods revolving around this nitrogen fixation involved burning fossil fuels, which is a process that has been adding continuous excess carbon dioxide into the atmosphere.

Using transcriptoanalysis methods, my role in the project revolved around focused research on two specific model microbes in order to obtain more information on how this microbial community would function. The research analyzed overall genome structure and gene biofunctionality. This collected data from the two model microbes would then be compared to that of the two intended microbes for the community. The lack of knowledge revolving around microbial communities fronts the project as a primary task in order to co-culture the two microbes effectively.

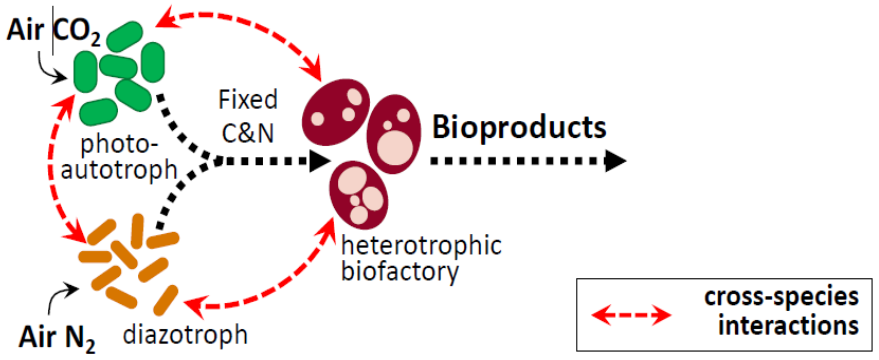


FIG. 1. Metabolic process representing the diazotroph and heterotroph process for bioproduct output. This microbial community represents the yeast, *Synechococcus elongatus* PCC 7942, *Azotobacter vinelandii* DJ in a co cultured setting. This microbial community allows for bioproduct output while reducing the overall carbon footprint in our atmosphere.

II. METHODS

A. Transcriptoanalysis

The methods of approach to form the community started with data collection amongst all of the targeted microbes. In this case, *Synechococcus elongatus* PCC 7942, was a microbe that was the first target for genomic analysis to determine functionality in specific conditions. Genomic analysis in itself can take some time to fully complete, as the amount of genes per organism can vary. In this instance, the organism had approximately a total of 2800 genes. My contributions to the project revolved around identifying transcriptional factor genes amongst the genome, using different computational databases including: NCBI, P2TF, DEEPTF, ENTRAF. By identifying the transcriptional factor genes, studies can be specified to these specific governing genes, rather than the entire sequence. The illustration of total transcriptional factors identified per database for *Synechococcus elongatus* PCC 7942 is shown in figure 2.

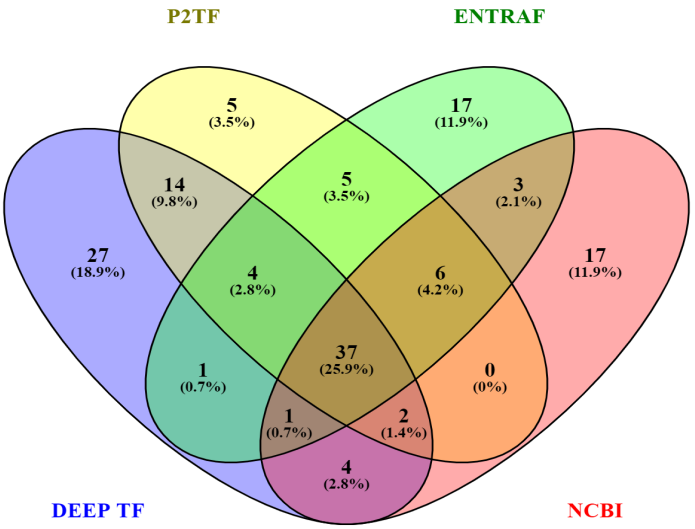


FIG. 2. *Synechococcus elongatus* PCC 7942- transcriptional factor overlap across the four different databases listed. Each database cited their own annotation for gene description for the microbe, the vienn diagram reflects those that were identified as transcriptional regulators.

These transcriptional factor genes were then compared to other homologically similar organisms in order to determine biological functionality. Two microbes that were used for this comparison were Synechococcus sp. PCC 7002 and Synechocystis sp. PCC 6803. The illustration of these regulator categorization per database are shown in **figures 3** and **4**. In order to make this comparison, the same analysis of overall transcriptional regulator gene identification was done on these two microbes. Once the regulators were determined, the consistency of the specific regulators were compared through the three different organisms. The comparisons amongst the three different list of governing genes were made to distinguish the genes with the highest importance in overall functionality.
After the comparisons between the three organisms were made in terms of their transcriptional factors, the next question to address was what genes controlled these regulators. In other words, what has control over these governing genes we have listed? A hypothesis that my project focused on was the redox state of the microbe. The redox state of the microbe suggested gene activity in light-dark states of the organism.

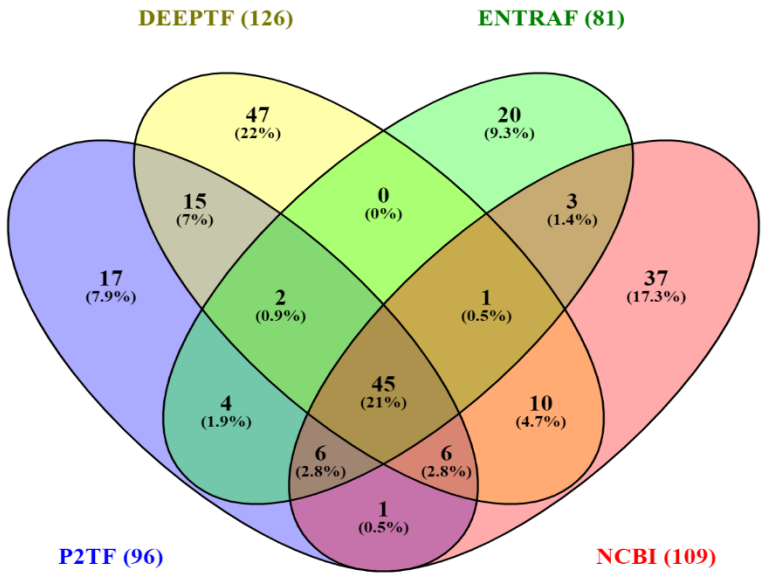


FIG. 3. Synechococcus sp. PCC 7002 - transcriptional factor overlap across the four different databases listed. Each database cited their own annotation for gene description for the microbe, the vienn diagram reflects those that were identified as transcriptional regulators.

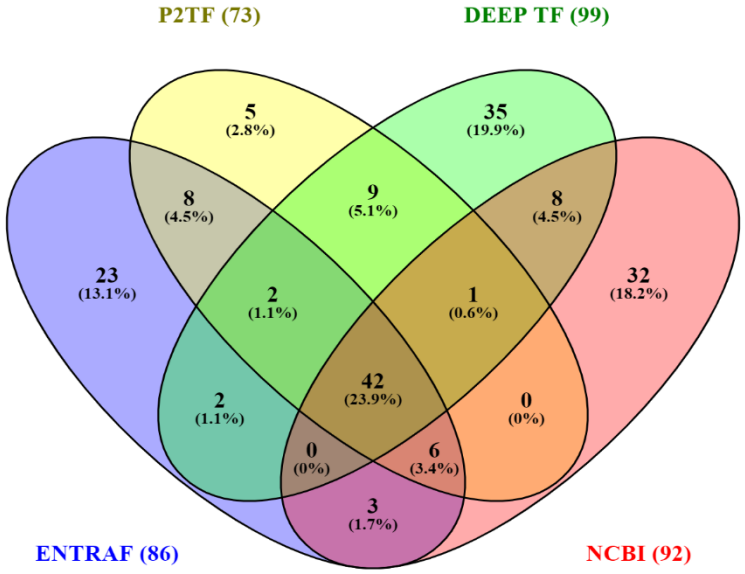


FIG. 4. Synechocystis sp. PCC 6803 - transcriptional factor overlap across the four different databases listed. Each database cited their own annotation for gene description for the microbe, the vienn diagram reflects those that were identified as transcriptional regulators.

B. Proteomics

Proteomic data is data that shows expression within an entire set of proteins. Proteomic data was cross referenced with the three different lists of transcriptional regulators- corresponding to the three different organisms. Synechocystis sp. PCC 6803 and Synechococcus sp. PCC 7002 had data that was already documented that was easy to cross reference. For this study, the team used proteomic data that was specific to cysteine sites with significant redox change. In other words, there were genes highlighted within the proteomic data that indicated greater activation than the rest of the recorded genes. These genes were cross referenced with the list of transcriptional regulators that were mentioned earlier.

The results for the cross referenced Synechocystis sp. PCC 6803 indicated that there were only a few genes that were both in the proteomic data and transcriptional regulatory category. This is shown in the **supplementary data**. The results for Synechococcus sp. PCC 7002 also showed similar results within the proteomic data. As of now, data shows that there is no significant correlation between redox state activation and regulator categorization. However, the hypothesis cannot be completely overruled as the proteomic data for the redox state in the cysteine site is still being recorded for one of our main organisms, *Synechococcus elongatus* PCC 7942.

I was able to compare the datasets for each model organism's proteomic data. In other words, the overlap of transcriptional factor and proteomic gene activation between the two organisms. The small list of genes that each microbe consisted of was compared through the NCBI database. Specifically, I was looking for any high base percentage of homological similarity. This would indicate that there are important genes that are consistent upon organisms that have taxonomically different family categorization. This comparison is shown in **figure 5**.

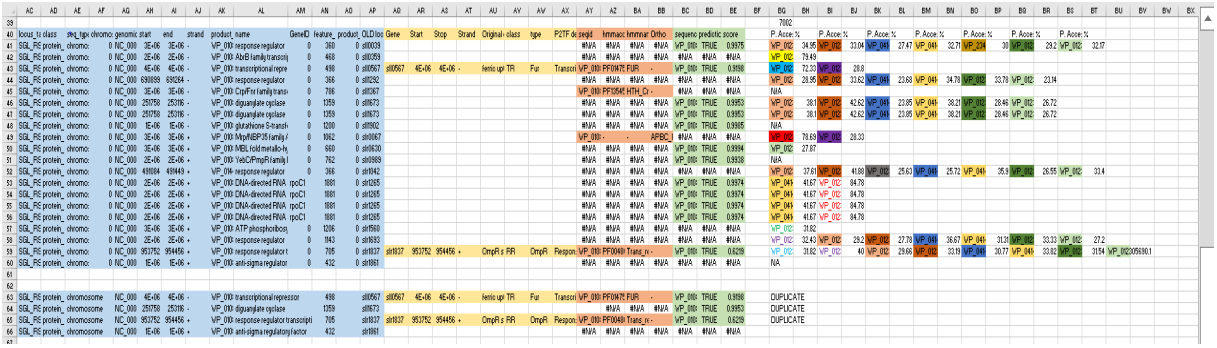


FIG. 5. Proteomic Comparisons- The list up on the left (shown columns AC-BE) represents the list gathered from Synechocystis sp. PCC 6803. The list on the right (shown in columns BG-BT) represented Synechococcus sp. PCC 7002's dataset. Each highlighted color in columns BG-BT represented a gene on their original list. This data shows the percentage of similarity of that of the gene on the list on the left. If there was not any significance in similarity, the box was left empty.

The data indicates that there were approximately six genes amongst the entire list that shared a homological similarity profile above 70%. This high percentage demonstrates that there is a a high resemblance in gene structure. The factors that these high percentage genes consists of has not yet been looked into, but is included in my team's schedule for analysis. The rest of the genes shared an average similarity around 35%. The consistency of this low percentage, does

indicate that there are some consistent factors amongst the genes compared, even if it is a small fraction. The next step with this specific comparison would be to identify the consistent factors that are present in both microbes within these specific genes of high percentage resemblance.

III. DESCRIPTION OF THE DATA

A. Genome Annotations

With the analysis described above, my team looked towards the newer found annotations of each genomic sequence. This was done due to the amount of newer found information that pertained to the description of each gene, some of which were transcriptional regulators. It is recognized that, although these annotations might be less cited publically, these annotations were primarily sourced from the old annotation. The only difference is the newer found information on a few more genes in regards to their biological function.

B. Proteomic Data

In terms of the proteomic data, both reference model organisms used were obtained from already existing publications[1&2]. This data was cross referenced with the list of transcriptional regulators obtained from the the two model microbes. Although thorough gene descriptions of *Synechococcus elongatus* PCC 7942 has already been documented, our team is currently waiting on the results of the proteomic data with this organism.

III. INCONSISTENCIES & SOLUTIONS

There were a number of inconsistencies that my team was unable to address, and these were due to pre-existing factors that aren't directly able to be controlled.

A. Annotation Labeling

1. The annotation used for this model microbe were based on the most up to date found genes in the NCBI database. The newer genome annotation included description of genes that the previous annotation had considered as "hypothetical" or "unknown". However, using these newer annotations resulted in some conflicts. These conflicts revolved around the fact that newer annotations commonly used different locus tags. These new locus tags were not recognized by many publications, and it was often times difficult to cross reference these genes. The reasoning behind using the newer annotations were strictly due to updated description specifying gene biological function.
2. Another conflict that was present in some of these datasets were genes that were reverted to a description labeled as "unknown". For instance, a gene description for the old annotation had a description of a gene that was more in depth than that of the newer annotated description. In this instance, my team decided to fill in these otherwise described genes with the older annotation description- to get the best possible data in our dataset. These inconsistencies and old annotated referral genes are highlighted in pink.

IV. CONCLUSIONS AND FUTURE GOALS

Moving forward, the next steps underway involve making the comparisons between *Synechococcus elongatus* PCC 7942 and the two model microbes mentioned earlier. Because these comparisons are strictly limited to the cysteine site activation due to redox change, this project would look towards other possible perturbation tests that may return more activated genes than what is being shown by the model organisms. The data collected is well documented, so now it is only a matter of identifying factors that control these transcriptional regulators. The direction of our project afterwards is a little unclear, as the project is in the

early stages of data collection and analysis to gain as much of an understanding with microbial communities as possible.

Throughout this internship, I gained appreciation for working within model organisms. I did not have any experience to microbial data analysis in the past, but I was able to learn fairly quickly thanks to my mentor and project team members. My prior experience revolved around work in a wet lab, a setting where most of my challenges revolved around measurement precision. However, I enjoyed reading the papers and studies that discussed the model organisms I was researching, and enjoyed learning about the different projects they were being used for. I also had never been exposed to any computational database in the past, which was extremely useful throughout my time here as an intern and my contributions to the project. Learning to ask questions and not soley depend on myself for these answers was something that I struggled with in the beginning of my internship, but gradually grew out of as I started being more communicative in my project's team meetings. I also had a small learning curve with the supplementary data that each research paper had, as I was not too familiar with the different figures that were being discussed. I never thought that I would go from learning about what the supplementary excel sheets mean to making my own data for other project team members to use. Although my pathway in terms of my career has always been in the trajectory of a physicans assistant, doing research here along with the quality of team members that I had has opened me to the idea of continuing this genre of work after I graduate in the spring of 2023.

V. ACKNOWLEDGMENTS

I would like to sincerely thank my mentor, Pavlo Bohutskyi, for assisting and guiding me throughout this entire project. Pavlo Bohutskyi would always go above and beyond what I asked for in terms of answering my questions, and always made time to meet with me amidst his busy schedule. I would also like to thank Sarah Bogart, another member of the Synthetic Biology Team. Not only did Sarah aid me with making figures for my data, she also provided career guidance with any questions I had throughout the last 16 weeks. Furthermore, she also introduced me to the concept of a regulatory network as a whole. I would also like to thank Zachardy D. Johnson, for aiding me with any ongoing questions I had regarding any research papers I encountered. His knowledge and experience with different types of data allowed me to identify the factors I was looking for in the different publications. I would like to thank Nicole Castilleja-Bentley, Juan Lopez, and Cassandra Clark for overall support within the Science Undergraduate Laboratory Internships (SULI) Program. Funding for the overall project is credited to Office of Biological and Environmental Research, Genomic Science Program, Award Number DE-SC0019181. *This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program.*

VI. INDEX FOR SUPPLEMENTARY DATA

A. *Synechococcus elongatus* PCC 7942 Highlights

1. Green & Yellow: Data from previous author of the annotation, David Anderson.
2. Pink: Researched information manually inputted into the annotation.
3. Blue (light): Manual "tRna" input for the priorly left blank cells in the data.
4. Blue (dark): Researched Cog Category categorization, otherwise priorly noted as having two or more categories. This was simplified to one.

5. Teal: Manually shortened gene name, aswell as manually edited cog category. Pre-existing cog category defined the gene as having more than categories, there was less public data to be as certain as the cells highlighted in blue (dark).
6. Orange: There were additions made to the description, manual edits are indicated after the "/".

B. Synechocystis sp. PCC 6803 Highlights

1. All highlights under the column T describe manual additions to the priorly blank gene name. These were researched and looked for in the previous annotation.

C. Synechococcus sp. PCC 7002 Higlights

1. Gray: Product accessions with this color highlight represent old accession numbers. Although there is a more up to date product accession in the annotation, new annotation searches returned no data on the gene. The old accession (gray) was needed to capture this information.
2. Red: Old product accession, there is no new product accession found. This was recorded from the old annotation data.

VII. REFERENCES

[1] Proteomic Data for Synechocystis sp. PCC 6803: Jia Guo, Amelia Y Nguyen, Ziyu Dai, Dian Su, Matthew J Gaffrey, Ronald J Moore, Jon M Jacobs, Matthew E Monroe, Richard D Smith, David W Koppenaal, Himadri B Pakrasi, Wei-Jun Qian. 13. "Proteome-wide light/dark modulation of thiol oxidation in cyanobacteria revealed by quantitative site-specific redox proteomics". 2014.

[2] Proteomic Data for Synechococcus sp. PCC 7002: Natalie C. Sadler, Matthew R. Melnicki, Margrethe H. Serres, Eric D. Merkley, William B. Chrisler, Eric A. Hill, Margaret F. Romine, Sangtae Kim, Erika M. Zink, Suchitra Datta, Richard D. Smith, Alexander S. Beliaev, Allan Konopka, and Aaron T. Wright. "Live Cell Chemical Profiling of Temporal Redox Dynamics in a Photoautotrophic Cyanobacterium". 2014.

[3] *Synechococcus elongatus* PCC 7942 Annotation: South China Normal University. " *Genome assembly ASM1469890v*". 2020.

[4.] Synechococcus sp. PCC 7002 Annotation: Penn State University. "Genome assembly ASM1948v1". 2008.

[5.] Synechocystis sp. PCC 6803 Annotation: Kazusa. "Genome assembly ASM972v1". 2004.

[6] Paper describing Transcriptional Factors: Kevin Rychel 1 , Anand V. Sastry 1 & Bernhard O. Palsson. "Machine learning uncovers independently regulated modules in the Bacillus subtilis transcriptome". 2020.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov