Supplemental Materials

for

Development of the Organonitrogen Biodegradation Database: Teaching Bioinformatics and Collaborative Skills to Undergraduates during a Pandemic

Serina L. Robinson1, Troy Biernath2, Caleb Rosenthal3, Dean Young3, Lawrence P. Wackett4, and Betsy M. Martinez-Vaz3*

1Institute of Microbiology, ETH Zurich, Zürich 8093, Switzerland; 2Department of Chemistry and Biochemistry Program, Bethel University, Saint Paul, MN 55112, USA; 3Department of Biology and Biochemistry Program, Hamline University, Saint Paul, MN 55104, USA; 4Department of Biochemistry, Molecular Biology and Biophysics and Biotechnology Institute, University of Minnesota, Saint Paul, MN 55108, USA

Table of Contents
(Total pages 34)

Appendix 1: Project timeline
Appendix 2: Chemical data mining
Appendix 3: Literature searches
Appendix 4: Reading research papers
Appendix 5: Pathway prediction exercises with the EAWAG-BBD
Appendix 6: ONDB class activity

Appendix 7: ONDB pages
Appendix 8: ONDB compound information
Appendix 9: Examples of ONDB students’ presentations
Appendix 10: Examples of ONDB students’ posters
Appendix 11: ONDB sample tutorials

*Corresponding author. Mailing address: Department of Biology and Biochemistry Program, Hamline University, 1536 Hewitt Ave., Saint Paul, MN 55104. Phone: 651-523-2493. E-mail: bmartinezvaz01@hamline.edu.
Received: 29 September 2020, Accepted: 9 December 2020, Published: 31 March 2021

©2021 Author(s). Published by the American Society for Microbiology. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial-NoDerivatives 4.0 International license (https://creativecommons.org/licenses/by-nc-nd/4.0/ and https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode), which grants the public the nonexclusive right to copy, distribute, or display the published work.
## Appendix 1

### Timeline of activities for the ONDB Research Experience

| Week | Activity |
|------|----------|
| 1    | Compound selection  
Exploring PubChem and ChemSpider  
R tutorial: Downloading and installing R  
How import data and install packages in R  
Virtual lab meeting |
| 2    | Assessment of the biodegradation of compounds using the EAWAG-BBD  
R tutorial: Importing data into R  
Data types and data structures in R  
Progress meeting with mentors  
Virtual lab meeting |
| 3    | Literature searches to learn about degradation pathways  
R tutorial: Data types and data structures in R  
Virtual lab meeting |
| 4    | Discussion of papers related to the biodegradation of compounds of interest  
Discussions to select chemical reactions and narratives to include in the ONDB  
Progress meeting with mentors  
Virtual lab meetings  
Creating pages using Shiny and flexdashboard |
| 5    | Designing the ONDB Web interface  
Discussions to decide the organization of reaction pages  
Evaluating ONDB logo and page formats  
Creating pages using Shiny and flexdashboard  
Virtual lab meetings |
| 6    | Formatting structures and text for inclusion in the ONDB pages  
Evaluating web links and resources to connect to the ONDB  
Progress meeting with mentors  
Virtual lab meeting |
| 7    | Examination of ONDB Prototype I: Mission and Resource Pages  
Evaluating and reformatting images, narratives and references in the ONDB  
Testing ONDB’s links to other databases  
Virtual lab meeting |
| 8    | Examination and Evaluation of ONDB Prototype I: Database and Reaction Pages  
Work on posters  
Final progress meeting with mentors  
Virtual lab meeting |
| 9    | Examination and Evaluation of ONDB Prototype I: Finalize all pages, information and links  
Live demonstrations of the ONDB Web tool  
Work on posters and oral presentations |
| 10   | Final poster and oral presentations  
Final presentation at lab meeting |
Appendix 2  
Chemical Data Mining Using PubChem and ChemSpider

ChemSpider and PubChem are chemical repositories containing useful information about diverse compounds. The first part of this research experience consists of learning about the physical and chemical properties of the organonitrogen compounds of interest. Use the following URLs to access the ChemSpider and PubChem pages

Chemspider: http://www.chemspider.com  
PubChem: https://pubchem.ncbi.nlm.nih.gov

1. Compare and contrast the types of information presented in these databases. List three differences and three similarities between these tools.

2. Explore ChemSpider and PubChem. Using these databases, retrieve the following information for the organonitrogen compound of interest:
   a. Synonyms
   b. Smile structure
   c. Molecular weight
   d. Physical properties (boiling point, refractive index, density)

3. What chemical methods are used to detect the compound of interest? Discuss and provide representative spectra.

4. Locate the structural search feature in either PubChem or ChemSpider. Perform a structural search to identify three compounds sharing the same chemical formula and structural features of the organonitrogen compound of interest. Provide the names and structures of these compounds.

5. Briefly discuss the common uses of the compound if interest. Support your answer with documentation available in PubChem and ChemSpider. Examples of supportive information include but are limited to research articles, patents, and current news reports. Are the compounds with similar structure (identified in question #4) used for the same purposes?

6. Which chemical information about the compound of interest should be included in the ONDB? Why?
Appendix 3: Effective Literature Searches

Background

Searching the published literature effectively can be a challenging task. A good research project starts with a strong revision of the literature in the field of interest. To carry out an effective literature search, we need to consider five important factors:

1. Key words, operators and filters
2. Search tools
3. Types of literature
4. Evaluate Information
5. Organize Research

The selection of appropriate keywords can impact the effectiveness of any literature search. To select the right words, you should start by thinking about the goals of the project and the questions you want to address. This is the first step to gather information relevant to the topics you want to investigate. Another important aspect of a literature search is the use of operators and filters. These tools help cut off unnecessary information to focus the search and obtain the best sources and materials for a project. Common Boolean operators include “OR”, “And” and “NOT”. Additionally, we can continue to narrow down information gathered after a search by using filters to select articles by year, author, content, relevance, etc.

Given the vast amount of information available online, we need to carefully consider the search tools to carry out our literature research. We are interested in obtaining scholarly references to support biodegradation hypothesis, as well as the reactions and pathways shown in the ONDG. Ideally, we want to use a database that is easily accessible and provides full length articles free of cost. The NCBI PubMed database is ideal for this purpose given its accessibility, filter options connections to journals and other NCBI resources. This database also provides access to different type of literature including books, reviews, clinical trials and research papers.

Once we have selected a database and a set of keywords and operations, we are ready to start our literature search. To evaluate the information sources obtained after the search, we will use a modified version of the CRAAP developed by the Meriam Library at California State University, Chico. Finally, once we have selected the papers and resources we want to keep for our research, it is important to catalog and store them in an organized manner. Consider creating a document library with folders dedicated to various research topics. In addition, you should think of using a reference manager to create a list of citations in APA or MLA style. A reference list is valuable resource for future poster, papers and oral presentations.

References:

1) Webb, K. (2018). Conducting a successful literature search: A researcher’s guide to tools, terms and techniques. https://www.elsevier.com
2) Hopkins, J. (2015). Research Guides: Evaluating Sources: The CRAAP Test. https://researchguides.ben.edu/source-evaluation
Literature Search Guide for the ONDB

1. Use the following URL to access NCBI PubMed: https://pubmed.ncbi.nlm.nih.gov

2. You are interested in learning about the microbial degradation of methylenediurea. List up five keywords that could be used in your PubMed search: ____________________________________________

3. How many references did the search retrieve? __________

4. Select one or more operators to refine your search. List the operators selected: ____________________________

5. How many references did the search retrieve after using the different operators? ______________________

6. Which combination of keywords and operators produced the most relevant on the microbial degradation of methylenediurea? Explain.

7. What kind of sources did you find? Classify them as research papers, reviews, book, etc.

8. Evaluate two of the sources obtained after your searches using the CRAAP test provided on page #3.
CRAAP Test: https://researchguides.ben.edu/source-evaluation

Cite the source you are evaluating:

Currency: the timeliness of the information

- When was the information published or posted?
- Has the information been revised or updated?
- Is the information current or out-of date for your topic?
- Are the links functional?

Relevance: the importance of the information for your needs

- Does the information relate to your topic or answer your question?
- Who is the intended audience?
- Is the information at an appropriate level (i.e. not too elementary or advanced for your needs)?
- Have you looked at a variety of sources before determining this is one you will use?
- Would you be comfortable using this source for a research paper?

Authority: the source of the information

- Who is the author/publisher/source/sponsor?
- Are the author's credentials or organizational affiliations given?
- What are the author's credentials or organizational affiliations given?
- What are the author's qualifications to write on the topic?
- Is there contact information, such as a publisher or e-mail address?
- Does the URL reveal anything about the author or source? Examples:
  
  - .com (commercial),
  - .edu (educational),
  - .gov (U.S. government),
  - .org (nonprofit organization),
  - or .net (network)
Accuracy: the reliability, truthfulness, and correctness of the content

- Where does the information come from?
- Is the information supported by evidence?
- Has the information been reviewed or refereed?
- Can you verify any of the information in another source or from personal knowledge?
- Does the language or tone seem biased and free of emotion?
- Are there spelling, grammar, or other typographical errors?

Purpose: the reason the information exists

- What is the purpose of the information? to inform? teach? sell? entertain? persuade?
- Do the authors/sponsors make their intentions or purpose clear?
- Is the information fact? opinion? propaganda?
- Does the point of view appear objective and impartial?
- Are there political, ideological, cultural, religious, institutional, or personal biases?

After evaluating this source, do you think you will use it for your paper? Why or why not? If you are not sure, explain why.
Appendix 4
ONDB Database Research Project
Guide to Read and Interpret a Research Paper

Reading and interpreting research papers is a valuable skill but an overwhelming task. The best way to read a paper is to break it into sections and take notes about the main ideas presented in each portion of the manuscript. We will use this approach to read and discuss a paper about the microbial degradation of Cyanuric acid, an organonitrogen compound commonly used to clean swimming pools.

Start by locating the following research paper:

Aukema KG, Tassoulas LJ, Robinson SL, Konopatski JF, Bygd MD, Wackett LP. (2020). Cyanuric Acid Biodegradation via Biuret: Physiology, Taxonomy, and Geospatial Distribution. Appl Environ Microbiol. 7:86(2):e01964-19.

1. Read the abstract. In your own words, describe the “problem” being investigated or the main purpose of Aukema’s paper. Your answer should not be longer than one paragraph.

2. Following the abstract, we find the introduction of the paper. Read this section and answer the following questions.
   a. Why are the authors interested in studying the microbial degradation of cyanuric acid? Why is this a relevant research topic?
   b. What other compounds are structurally related to cyanuric acid? Why are these compounds relevant? Explain
   c. Based on the information presented in the introduction, how do bacteria degrade cyanuric acid? What enzymes are important to the degradation pathway?
   d. Compare and contrast the cyanuric acid degradation pathways presented on the paper. What are the main differences in between these pathways?

3. Results This is the section of the paper that supports the story the authors want to tell the reader. Let’s start with figure #1.
   a. What's the basic idea of Fig 1? Which paragraph does it seem to match? Did it get referenced in that paragraph?
   b. Repeat step “a” for figures 2-6. Explain the main experimental findings presented in each figure.
   c. In your opinion, what is the most important figure presented in Aukema’s paper? How does it support the hypothesis posed by the authors? Explain.
4. Discussion. This is section of the paper where the authors explain the meaning of their results. Sometimes this section is combined with the results portion of the paper.

   a. What do the authors say about their results. Did their data support the original hypothesis of the study?
   b. Discuss two main limitations of the work presented in Aukema’s manuscript. What experiments would you do to address these limitations?
   c. Do the authors identify the next logical steps for the research? What would you guess might be done next?

5. Material and Methods. Sometimes this section follows the introduction, and sometimes it’s tucked at the end of the paper. It’s very detailed, in case others want to think carefully about or even use the same techniques.

   a. What technique did the authors use to isolate cyanuric acid -degrading bacteria?
   b. Briefly describe the protein purification approach employed by the author to isolate cyanuric acid degrading enzymes.
   c. In this paper, the authors used both computational and wet-bench methods. How did the computational research support or complements the wet-bench experiment? Support your answer with at least two examples.

6. Write 2-3 sentences summarize the main findings of this paper and their relevance to the field of biodegradation of organonitrogen compounds.

7. Which of the information presented in this paper should be added to the ONDB? Explain your reasoning.
Appendix 5
Assessing Biodegradation Potential Using the EAWAG-BBD Pathway Prediction System

The goal of this exercise is to investigate the microbial degradation potential of various organonitrogen compounds using a pathway prediction tool.

Use the following URLs to access the EAWAG-BBD Pathway Prediction System: http://eawag-bbd.ethz.ch/predict/

1. Enter the structure of the compound of interest in SMILES format. An example of a SMILES string is illustrated in the figure below. Alternatively, you can draw the structure using the EAWAG-BBD drawing tool.
   \[ \text{C(=NC(=S)N)(N)N} \]

2. Write the predicted biodegradation reaction and state the level of likelihood provided by the EAWAG-BBD system. Is the degradation pathway predicted aerobic or anaerobic?

3. Click on the box by the arrow to investigate the enzymes and pathways associated with the predicted biodegradation route.

4. Examine the predicted biodegradation reaction. Label the enzymes, substrates and products of the predicted reactions.
   a. Which pathways is the reaction associated with? Provide 1-3 examples based on the information provided by the EAWAG-BBD system.
   b. Are there any microorganism capable of metabolizing the compound of interest?
5. Note the name of the enzymes listed as catalysts for the proposed biodegradation reactions. Search the Uniprot (https://www.uniprot.org/uniprot) database to learn about the protein sequence, structure and distribution of these enzymes. Write a short paragraph summarizing your findings.

6. Use the Uniprot database (https://www.uniprot.org) to retrieve the protein sequence of the enzyme involved in the first reaction of the biodegradation pathway. Conduct a BLASTp search using this sequence. Which proteins share the highest degree of homology with the protein sequence used for the search? Prepare a table reporting the names, species, % identity and E-values for the ten closest relatives of the enzyme you searched.
Appendix 6

ONDB Research Project
Investigating the Biodegradation of Cyanoguanidine

Introduction

Cyanoguanidine is an artificially synthesized compound heavily used in agriculture to prevent soil nitrogen loss and nitrate leaching into water bodies. In addition, cyanoguanidine is often used in the manufacturing of paint, lumber, rubber, metal and raw materials. Unfortunately, the compound poses a toxicity threat to certain mammals, as it has been shown in animal studies to cause reactions such as eosinophilia, a form of excessive immune response. Further, tests of intradermal cyanoguanidine injection have proven fatal for guinea pigs at a mere 200 parts per billion.

Therefore, cyanoguanidine is considered an emergent pollutant and its removal from soils and water environments is of growing concern. Attempts at degrading cyanoguanidine through chemical means have shown partial success, but another approach shows great promise: bioremediation. Bioremediation is the use of microorganisms such as bacteria to break down unwanted and harmful compounds into benign products. The concept of bioremediation is not new; in fact, it was used to clean the Deep Horizon oil spill, in a manner one study described as “rapid and robust”. However, the idea of bioremediation for cyanoguanidine, while showing great promise, has been hardly explored.

The purpose of this class exercise is to investigate the potential for microbial degradation of cyanoguanide using various computational resources such as the organonitrogen biodegradation database (ONDB), and the EAWAG-BBD Pathway Prediction System and the MetaCyc metabolism database.

Database Exploration Activity

1. Access the organonitrogen biodegradation database (https://z.umn.edu/ondb). In the database tab, locate cyanoguanidine. Write down the compound’s smile string, molecular weight and synonyms.

2. Go to the reaction page and click on cyanoguanidine. Examine the biodegradation pathways shown for this compound. What are the main products of cyanoguanidine biodegradation? Are any of the biodegradation products known to be toxic to humans, plants or animals? Support your answer with at least one scholarly reference.

3. Which type of nutrients would bacteria get from the degradation of cyanoguanidine? Explain.

4. Which enzymes have been linked to the degradation of cyanoguanidine? Based on the information shown in the database, what type of reactions do they catalyze (hydrolysis, condensation, oxidation-reduction, etc.)?

5. Access the MetaCyc metabolism database (https://metacyc.org/). How are the cyanoguanidine degradation products link to other metabolic pathways in bacteria? Select one of the degradation products and explain its connection to other metabolic pathways in bacteria in detail.

6. Go back to the organonitrogen biodegradation database (https://z.umn.edu/ondb). Examine the compounds listed in the reaction page. Do any of these compounds connect with the biodegradation pathway of cyanoguanidine? If yes, explain the metabolic connections in detail.
7. Access the EAWAG-BBD Pathway Prediction System (http://eawag-bbd.ethz.ch/predict). Enter the structure or the smile string of cyanoguanidine. Run the predicted biodegradation algorithm. Compare and contrast the predicted biodegradation pathway to the reactions reported in the ONDB. List the main differences and similarities amongst the pathways. Which metabolic routes do you think are mostly likely to be present in bacteria? Why?

8. Imagine you work at the biotechnology company and have been assigned to a project to develop a biodegradation strategy to clean up an environment contaminated with cyanoguanidine. Which enzyme(s) from the proposed pathways would you choose? Why?
Organonitrogen Degradation Database

Our Mission:

Cataloguing microbial transformations of organonitrogen inputs in soils

Organonitrogen compounds are a major soil input used for fertilizers, pesticides and other agricultural purposes. One such compound, urea, is predicted to reach 1.87 x 10e11 kg by 2021 (International Fertilizer Association, 2017), and recent findings suggest these compounds can increase productivity and decrease environmental impacts. The intent of the ON-DB is to aid in remediation efforts by providing information on nitrogen-rich compounds.

“Agriculture is the single largest source of nitrogen compounds entering the environment in the U.S.”

-USDA: Economic Research Service

Last updated: July 2020
| Compound         | Formula | MW   | SMILES            | Synonyms                                                                 | CN_Ratio | Cost_metric_ton |
|------------------|---------|------|-------------------|--------------------------------------------------------------------------|----------|-----------------|
| Allophanate      | C2H3N2O3- | 103.06 | NC(=O)NC([O-])=O | Allophanic acid, urea-1-carboxylate, Allophanate                        | 1:1      | NA              |
| Biuret           | C2H5N3O2 | 103.08 | NC(=O)NC(N)=O     | Allophanamide, Carbamylurea, Imidodicarbonic diamide, Biuret             | 2:3      | 440.00          |
| Cyanamide        | CH2N2   | 42.04 | NC#N              | Amidocyanogen, Carbamoniitrile, Carbimide, Carbodiimide, carbodiimide, Cyanogenamide, Dormex, hydrogen cyanamide, Cyanamide | 1:2      | 660.00          |
| Cyanoguanidine   | C2H4N4  | 84.08 | NC(N)=NC#N, tautomer NC(N)NC#N | Dicyandiamide, Cyanoguanidine                                       | 1:2      | 1540.00         |
| Cyanuric acid    | C3H3N3O3 | 129.08 | O=C1NC(O)NC(O)N1  | 1,3,5-Triazine-2,4,6(1H,3H,5H)-trione; 1,3,5-Triazine-2,4,6-triol; 2,4,6-Trihydroxy-1,3,5-triazine; Isocyanuric acid; Normal cyanuric acid; Pseudocyanuric acid; sym-triazine-2,4,6-triol; sym-triazinetril; Triazine-2,4,6(1H,3H,5H)-trione; Triazine-2,4,6-triol; Triazinetrione; Triazinetrilone; Tricarbimide; Tricyan acid; Trihydroxy-1,3,5-triazine; Trihydroxycyanidine, Cyanuric acid | 1:1      | 990.00          |
| Guanyl thiourea  | C2H6N4S | 118.16 | C(=NC(=S)N)N    | Amidinothiourea, Guthimine, Guanyl thiourea                              | 1:2      | NA              |
| Methyleneurea    | C3H8N4O2 | 132.12 | NC(=O)NCNC(N)=O | Diuridomethylene, Methyleneurea                                         | 3:4      | 1100.00         |
| Urea             | CH4N2O  | 60.06 | NC(N)=O           | Aquadrate, Carbamide, Carbamimidic acid, Carbonyl diamide, Carbonyldiamine, Isourea, Ureaphil, Ureophil, Urea | 1:2      | 220.00          |
Guanyl thiourea

Guanyl thiourea (1-amidino-2-thiourea, GTU) is a nitrification inhibitor. A nitrification inhibitor compound reduces the amount of nitrogen that is capable of being lost, known as leaching, and therefore unusable by surrounding plants (Amberger, et al 1990). Several benefits of using GTU in contrast to other inhibitors exist. GTU is not harmful to plants, organisms that reside in the soil, or other species that are more complex. This nitrification inhibitor is also non-volatile which means that it can be added to solid fertilizers (Amberger, et al 1990). Guanyl thiourea, in addition, reduces both nitrous oxide and methane making it a safer and more environmentally friendly alternative to other nitrogen inhibitors. At $5.20 per kg, Guanyl thiourea is competitively cheap as well. It uses vary from golf courses to farm fields (Trenkl, 1997, Hang-Wei, et al 2015). Guanyl thiourea’s degradation product, dicyandiamide (DCD), is widely used as a nitrogen inhibitor as well (Trenkl, 1997). This compound is an irritant and health hazard as classified by GHS.

Enzymes in Literature:
- in vivo (liver)- monooxygenase can catalyze S-oxygenation of thiourea resulting in formamidine sulfenic acid and formamidine sulfinic acid. (Sahu et al. 2011)
- in vivo (humans)- Peroxidase oxidizes thiourea when it is in the presence of iodine. This will result in the formation of formamidine disulfide which breaks down into cyanamide. (Sahu et al. 2011)

Pathway Prediction System:

Bacteria
- Pseudomonas sp. NCIMB (Wang et al. 2017)

Reactions
The Guanyl thiourea enzymes Thioacetamide S-oxygenase and Ethionamide monooxygenase are from the Biocatalysis and Biodegradation Database and reaction is from Amberger and Germann-Bauer 1990.
Resources, links, and references

Located on the right and found below are the resources and citations relevant to the compounds listed in the ON-DB. These resources can serve as a point of further investigation.

Last updated: July 2020

Web Resources

Agricultural and fertilizer information:
- International Fertilizer Association
- Urea Fertilizer – University of Minnesota Extension
- Nutrient Loss Database for Agricultural Fields and Forests in the United States
- USDA – Fertilizer Use in the United States

Pathway information:
- EAWAG - Biocatalysis and Biodegradation Database
- MetaCyc - Metabolic Pathway Database
- Uniprot - Universal Protein Resource

Enzyme and gene information:
- Expasy - Bioinformatics Resource Portal
- KEGG - Kyoto Encyclopedia of Genes and Genomes

Compound information:
- Pubchem - Chemistry Database
- ChemSpider - Royal Society for Chemistry
- NPIC - National Pesticide Information Center

Scientific Publications

1. Ahmed, Moddassi, et al. “Hazards of Nitrogen Fertilizers and Ways to Reduce Nitrate Accumulation in Crop Plants.” Environmental Science and Pollution Research International, vol. 27, no. 15, 2020, pp. 17661-17670. PubMed, https://www.ncbi.nlm.nih.gov/pubmed/32180142, doi:10.1007/s11356-020-08236-y.
ONDG Compound Information Form

* Required

This is a summary form to collect information about the organ-nitrogen compound studied. The data collected will be used to create and update pages in the Organonitrogen Biodegradation Database (ONDG).

1. Compound's name: *

2. Common uses and applications *

3. Synonyms: *

4. Chemical formula *
5. Molecular weight *

6. SMILES string structure *

7. C:N Ratio *

8. Cost per metric ton *

9. Predicted biodegradation products *

10. Reported biodegradation products: *
11. Enzymes involved in the biodegradation pathway *





12. Bacterial and fungal species linked to the biodegradation of this compound *





13. Upload a ChemDraw file illustrating the biodegradation pathway of the compound studied *

Files submitted:





14. References *
15. Additional resources. Add the links to other database or electronic resources that can help you learn about the compound studied. *

This content is neither created nor endorsed by Google.

Google Forms
Example 8-week R coding tutorial for undergraduate students
The course has been generalized to learn R coding skills with basic datasets

Week 1: Introduction to R and to the project

Reading material:
Chapter 1 - Introduction to Data Science: Getting started with R and RStudio
Chapter 2 - Introduction to Data Science: R Basics
Chapter 3 - Introduction to Data Science: R Programming

Coding challenges:
Complete the exercises in Chapter 2 and 3 of Introduction to Data Science

Bonus coding challenges:
Basics of programming in R - swirl course

Primary literature (to discuss together):
Biodegradation and Bioinformatics: Arora PK, Bae H. (0214). Integration of bioinformatics to biodegradation. Biol Proced Online. 16:8.
https://biologicalproceduresonline.biomedcentral.com/articles/10.1186/1480-9222-16-8

Review articles (bonus reading):
Pathway Prediction System: Gao J, Ellis LB, Wackett LP. (2011) "The University of Minnesota Pathway Prediction System: multi-level prediction and visualization." Nucleic Acids Research 39 Suppl 2: W406-11.
https://academic.oup.com/nar/article/39/suppl_2/W406/2505766

Week 2: Data wrangling

Reading material:
Chapter 4 - Introduction to Data Science: The tidyverse
Chapter 5 - Introduction to Data Science: Data import

Bonus reading material:
Data wrangling in R
Wrangling unruly data
Joining data tables

Coding challenges:
Complete the exercises in Chapter 4 and 5 of Introduction to Data Science
Data carpentry dplyr exercises
Data wrangling exercises
Bonus challenges:
Getting and cleaning data - swirl course

Week 3: Data visualization with ggplot2

Main package documentation:
ggplot2 - for customizable graphs

Reading material:
Chapter 6 - Introduction to Data Science: Introduction to data visualization
Chapter 7 - Introduction to Data Science: ggplot2
10 levels of ggplot, from basic to beautiful

Coding exercises:
Exercises from Chapter 7 of Introduction to Data Science
Basic graphics with ggplot exercises
How to plot with ggplot and patchwork exercises

Bonus material:
Browse the ggplot2 gallery for inspiration!
A ggplot2 cheatsheet

Now apply these skills to make graphs for your own research project!

Week 4: Interactive data visualization with plotly

Main package documentation:
plotly - interactive graphing library

Reading material:
Chapters 1-6 - Plotly R book

Coding challenges:
Getting started with plotly exercises
Advanced plots and features

Now apply these skills to make interactive visualizations in plotly for your own research project!
Week 5: Introduction to descriptive analytics and statistics

Main package documentation:
R stats package

Reading material:
Chapter 11 - Introduction to Data Science: Robust summaries
Chapter 12 - Introduction to Data Science: Statistics with R
Descriptive statistics in R (using your new ggplot2 skills!)

Coding challenges:
Exercises from Chapters 11 and 12 of Introduction to Data Science

Bonus coding challenges:
swirl course - Regression models
swirl course - Statistical inference

Now apply these skills to do some descriptive analysis and statistics for your own dataset!

Week 6: Basics of R Shiny web applications I

Main package documentation:
R Shiny - for building interactive web apps
flexdashboard - use Markdown syntax to build interactive dashboards

Reading material:
The basic parts of a Shiny app

Bonus:
For inspiration, see the R Shiny app gallery

Coding challenges:
Building a Shiny apps exercises
How to create a flexdashboard: exercises

Now build a basic Shiny app for your own research purposes!

Week 7: Basics of R Shiny web applications II

Main package documentation:
R Shiny - for building interactive web apps
flexdashboard - use Markdown syntax to build interactive dashboards

Reading/lecture material:
Reactivity 101
Reactivity in Shiny

Coding challenges:
Shiny app layouts exercises
Interactive data tables exercises

Make improvements to your Shiny app!

Week 8: Geospatial mapping (optional, if relevant for research project)

Main package documentation:
ggmap  - spatial visualization
leaflet  - interactive maps

Coding challenges:
Leaflet mapping exercise 1
Leaflet mapping exercises 2

Now apply these skills to make a map using data for your research project!
Or choose to learn an additional topic in R that is more relevant to your research.
ORGANONITROGEN DEGRADATION DATABASE

ORGNONITROGEN COMPOUNDS

- Used in agriculture
- Mobile in soil
- Emerging pollutants
- Cause environmental damage

WHY ORGANONITROGEN?

- 100x increase in past 4 decades (2006)
- Global warming: 300x more radiatively reactive than CO2
- Reduction in species
- Bioservices cost estimated $170 billion

SOLVING THE PROBLEM WITH NATURE’S BEST CHEMISTS

- How to remove?
  - Many methods not feasible
  - Good news: bacteria can break down (“degrade”) these compounds
  - Use this to our advantage

IN ORDER TO USE BIOREMEDIATION:

- Need to know how it degrades:
  - Bacterial species
  - Enzymes
  - Products
  - Conditions

We need a database of how these organonitrogen compounds degrade.
MOVING FORWARD

"Microbial processes do more to abate and actually, by nature’s design, counteract contamination of ground water systems than any other natural or human-contrived mechanism." (Chapelle 1995, emphasis added)

Why Bioremediation?
• Natural
• Low cost
• Demonstrated repeated success

Thank You
z.umn.edu/andsb
Development of the Organonitrogen Degradation Database: A Bioinformatics Approach to Study the Microbial Degradation of Cyanoguanidine and Related Compounds

Organonitrogen Compounds

The use of nitrogen-rich fertilizers, pesticides and similar agriculturally-applied chemicals has increased to billions of tons as agriculture consumes more land and demands greater yield. An average of 80% of these organonitrogen compounds are lost to the environment through mechanisms such as soil leaching, where these chemicals enter into waterways and cause extensive nitrogen pollution. As these chemicals break down, they create toxic ammonia blooms, and cause hypoxia, the “dead zones” of coastal waters that suffocate aquatic life. One such compound, cyanoguanidine, has been found entering waterways and even contaminating dry milk export. Cyanoguanidine has been found to disrupt multiple kingdoms of life, including plants, animals and bacteria (Salis et al. 2017). For these reasons, John Lawton of the Natural Environment Research Council points out that nitrogen pollution from compounds such as cyanoguanidine is “the third major threat to our planet after biodiversity loss and climate change.” (Giles 2005)

Fortunately, there is a way we can mitigate these threats - using microbes. Previous research has shown microbes have the ability to degrade organonitrogen compounds into harmless products that can enter the natural biogeochemical cycles: this approach is known as bioremediation and has been used in many scenarios with repeated success (Kimes et al. 2014). However, no central source of information on how to conduct remediation of these compounds exists. The Organonitrogen Degradation Database fills this gap of knowledge and assists bioremediation efforts.

Methods

- Literature Search
- Pathway Prediction System
- Chem Sketch
- Enzymes
- Bacteria

Results

- Figure 1. Flowchart of work process
- Figure 2. Pathway Prediction Tool
- Figure 3. Database Homepage
- Figure 4. Resources
- Figure 5. Menu for selecting compounds
- Figure 6. Multiple biodegradation pathways of cyanoguanidine (Halling-Bankert et al. 1996; Schwarzer et al. 1998)
- Figure 7. Biodegradation Pathways

Bioremediation

To reduce the environmental hazards caused by organonitrogen compounds, the Organonitrogen Degradation Database provides information that can be used for bioremediation.

- Bacteria are capable of degrading these compounds into harmless products
- Multiple bacteria may be able to degrade a given compound
- Different bacterial species may degrade compounds differently (see figure below)
- One mode of bioremediation is through utilizing bacteria already present in soil
- Protein application can also be used for biodegradation

Future Directions

The Database is rich with future possibilities. The following are some future goals we have identified:

- Monitor and update compound usage information
- Incorporate new information as it is discovered
- Expand the current list of molecules in the database
- Use the database as a classroom exercise
- Incorporate specific methods for bioremediation (such as hydrogel for cyanoguanidine as developed by Minet et al. 2013)

References

Giles, Jim. “Nitrogen Study Fertilizer Fears of Pollution” Nature, vol. 433, no. 791, 2005, https://doi.org/10.1038/433791a

Halling-Bankert, S. et al. “Several aspects of Bacterial Dicyandiamide Degradation” Nature Biotechnology, vol. 77, 1990, pp. 332-334

Kimes, Nikole E et al. “Microbial transformation of the Deepwater Horizon oil spill-post, present, and future perspectives.” Frontiers in Microbiology vol. 5, no. 605, 2014 : doi: 10.3389/fmicb.2014.0605

Martín-Hendrie, Daniel. “Factors affecting the degradation of the nitrification inhibitor in a lowland Canterbury stream.” Lincoln University, 2014.

Minet, E. P. et al. “Slow delivery of a nitrification inhibitor (dicyandiamide) to soil using a biodegradable hydrogel of chitosan.” Scientific Reports, vol. 93, no. 5, 2013, pp. 2854-2858

Salis, R K et al. “High-throughput amplicon sequencing and stream bacterial taxonomy: identifying the taxonomic level for multi-stressor research.” Scientific Reports, vol. 7, no. 44657, 2017 https://doi.org/10.1038/srep44657

Schwarzer et al. “Physiological and electron microscopical investigations on syntrophic dicyandiamide degradation by soil bacteria.” Soil Biology and Biochemistry, vol. 30, no. 3, 1998, pp. 385-391 https://doi.org/10.1016/s0038-0717(98)00127-2

Smith, I & Schallenberg, M. “Occurrence of the agricultural nitrification inhibitor, dicyandiamide, in surface waters and its effects on nitrogen dynamics in an experimental aquatic system.” Agriculture, Ecosystems and Environment, vol. 164, 2013, pp. 23-31 https://doi.org/10.1016/j.agee.2012.09.002

Acknowledgements

Thank you to Dr. Lawrence Wackett, Dr. Serina Robinson, Kathryn Malody, the Hamline Biology Department, the SCUR program, the Ridgeway Forum Fund and the USDA. A special thank you to Dr. Betsy Martinez-Vaz for her support and guidance throughout the project.
Example 8-week R coding tutorial for undergraduate students
The course has been generalized to learn R coding skills with basic datasets

Week 1: Introduction to R and to the project

Reading material:
Chapter 1 - Introduction to Data Science: Getting started with R and RStudio
Chapter 2 - Introduction to Data Science: R Basics
Chapter 3 - Introduction to Data Science: R Programming

Coding challenges:
Complete the exercises in Chapter 2 and 3 of Introduction to Data Science

Bonus coding challenges:
Basics of programming in R - swirl course

Primary literature (to discuss together):
Biodegradation and Bioinformatics: Arora PK, Bae H. (0214). Integration of bioinformatics to biodegradation. Biol Proced Online. 16:8.
https://biologicalproceduresonline.biomedcentral.com/articles/10.1186/1480-9222-16-8

Review articles (bonus reading):
Pathway Prediction System: Gao J, Ellis LB, Wackett LP. (2011) "The University of Minnesota Pathway Prediction System: multi-level prediction and visualization." Nucleic Acids Research 39 Suppl 2: W406-11.
https://academic.oup.com/nar/article/39/suppl_2/W406/2505766

Week 2: Data wrangling

Reading material:
Chapter 4 - Introduction to Data Science: The tidyverse
Chapter 5 - Introduction to Data Science: Data import

Bonus reading material:
Data wrangling in R
Wrangling unruly data
Joining data tables

Coding challenges:
Complete the exercises in Chapter 4 and 5 of Introduction to Data Science
Data carpentry dplyr exercises
Data wrangling exercises
Bonus challenges:
Getting and cleaning data - swirl course

Week 3: Data visualization with ggplot2

Main package documentation:
ggplot2 - for customizable graphs

Reading material:
Chapter 6 - Introduction to Data Science: Introduction to data visualization
Chapter 7 - Introduction to Data Science: ggplot
10 levels of ggplot, from basic to beautiful

Coding exercises:
Exercises from Chapter 7 of Introduction to Data Science
Basic graphics with ggplot exercises
How to plot with ggplot and patchwork exercises

Bonus material:
Browse the ggplot2 gallery for inspiration!
A ggplot2 cheatsheet

Now apply these skills to make graphs for your own research project!

Week 4: Interactive data visualization with plotly

Main package documentation:
plotly - interactive graphing library

Reading material:
Chapters 1-6 - Plotly R book

Coding challenges:
Getting started with plotly exercises
Advanced plots and features

Now apply these skills to make interactive visualizations in plotly for your own research project!
Week 5: Introduction to descriptive analytics and statistics

Main package documentation:
R stats package

Reading material:
Chapter 11 - Introduction to Data Science: Robust summaries
Chapter 12 - Introduction to Data Science: Statistics with R
Descriptive statistics in R (using your new ggplot2 skills!)

Coding challenges:
Exercises from Chapters 11 and 12 of Introduction to Data Science

Bonus coding challenges:
swirl course - Regression models
swirl course - Statistical inference

Now apply these skills to do some descriptive analysis and statistics for your own dataset!

Week 6: Basics of R Shiny web applications I

Main package documentation:
R Shiny - for building interactive web apps
flexdashboard - use Markdown syntax to build interactive dashboards

Reading material:
The basic parts of a Shiny app

Bonus:
For inspiration, see the R Shiny app gallery

Coding challenges:
Building a Shiny apps exercises
How to create a flexdashboard: exercises

Now build a basic Shiny app for your own research purposes!

Week 7: Basics of R Shiny web applications II

Main package documentation:
R Shiny - for building interactive web apps
flexdashboard - use Markdown syntax to build interactive dashboards

Reading/lecture material:
Reactivity 101
Reactivity in Shiny

Coding challenges:
Shiny app layouts exercises
Interactive data tables exercises

Make improvements to your Shiny app!

Week 8: Geospatial mapping (optional, if relevant for research project)

Main package documentation:
ggmap  - spatial visualization
leaflet  - interactive maps

Coding challenges:
Leaflet mapping exercise 1
Leaflet mapping exercises 2

Now apply these skills to make a map using data for your research project!
Or choose to learn an additional topic in R that is more relevant to your research.