2024 Resource Sustainability Project Review Meeting

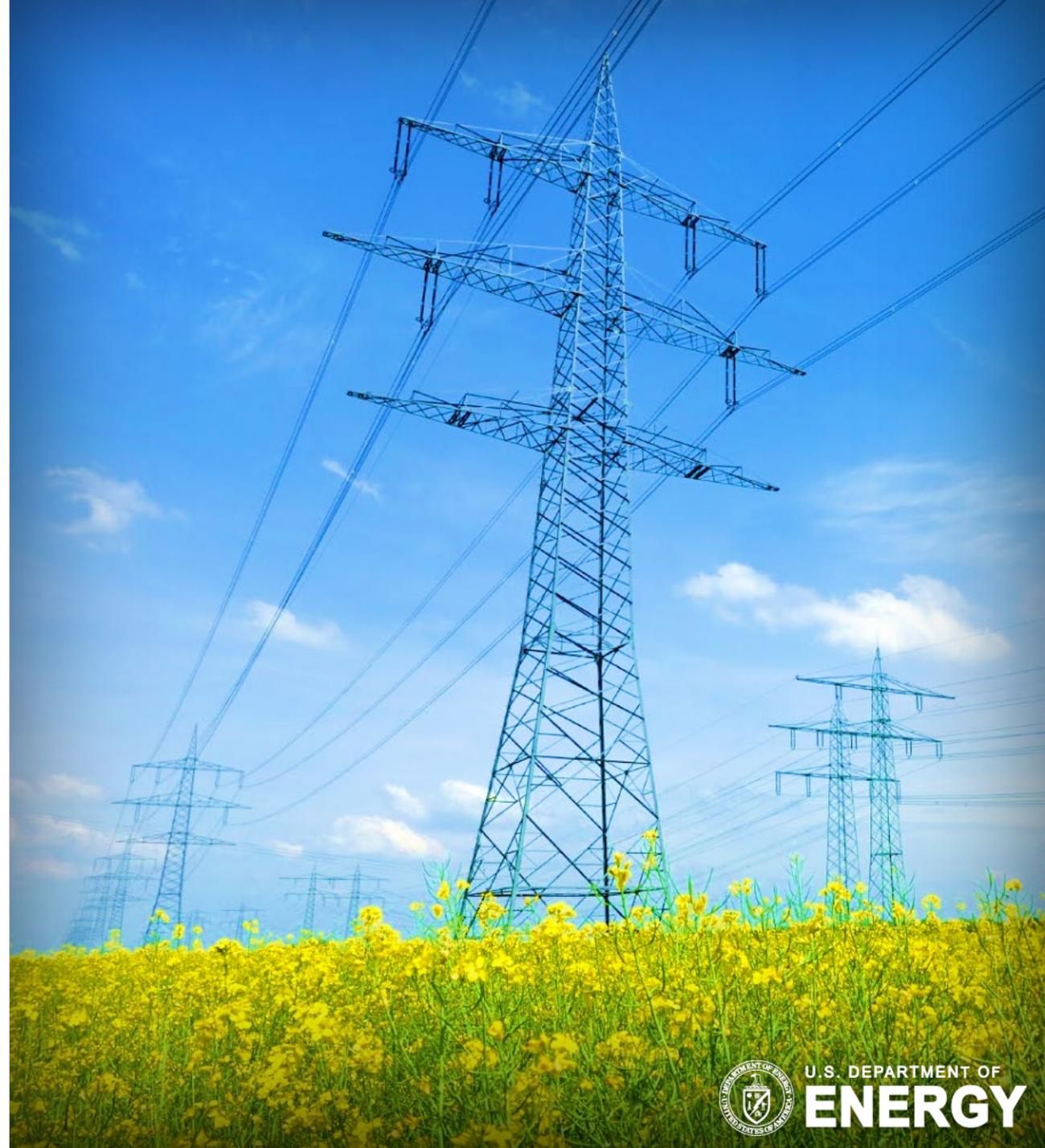# Microbial DNA Database and Demonstration

NETL Researchers: Dr. Hannah Schweitzer, Dr. Djuna Gulliver, Dr. Kara Tinker

USGS Researchers: Dr. Elliot Barnhart, Dr. Denise Akob

Technical Task Lead: Djuna Gulliver

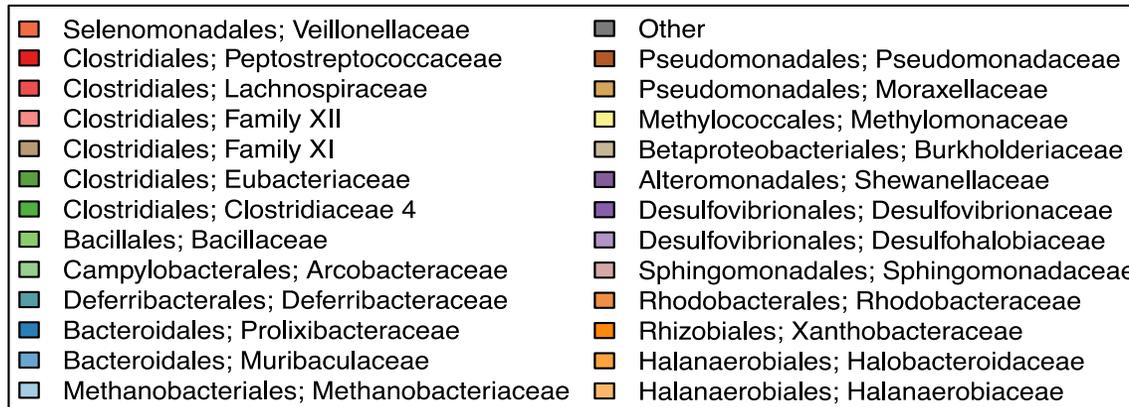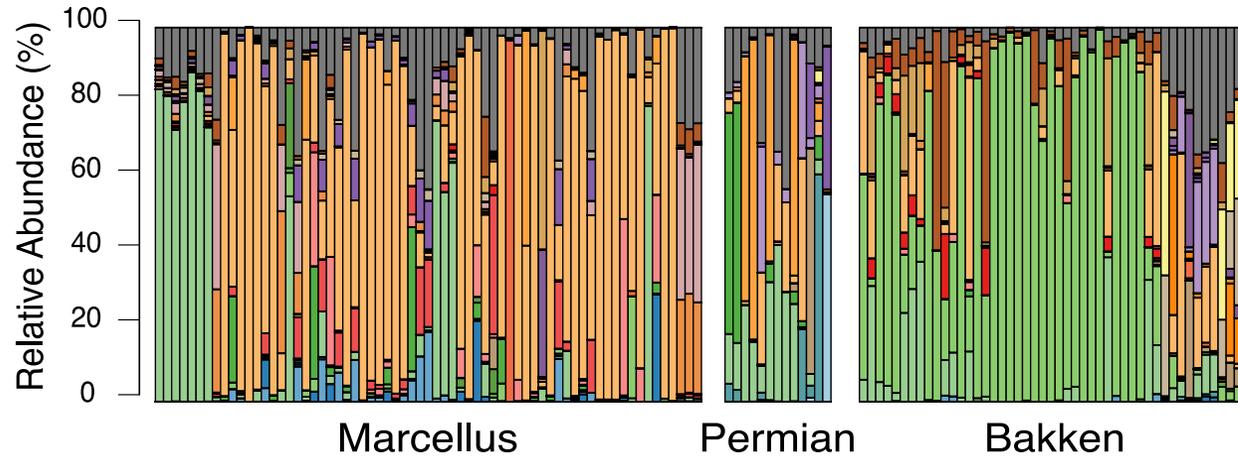Technical Portfolio Lead: Nicholas Siefert

April 3, 2024

NETL NATIONAL ENERGY TECHNOLOGY LABORATORY

U.S. DEPARTMENT OF ENERGY

Previous Work in Produced Water Microbiology

Geomicrobiology

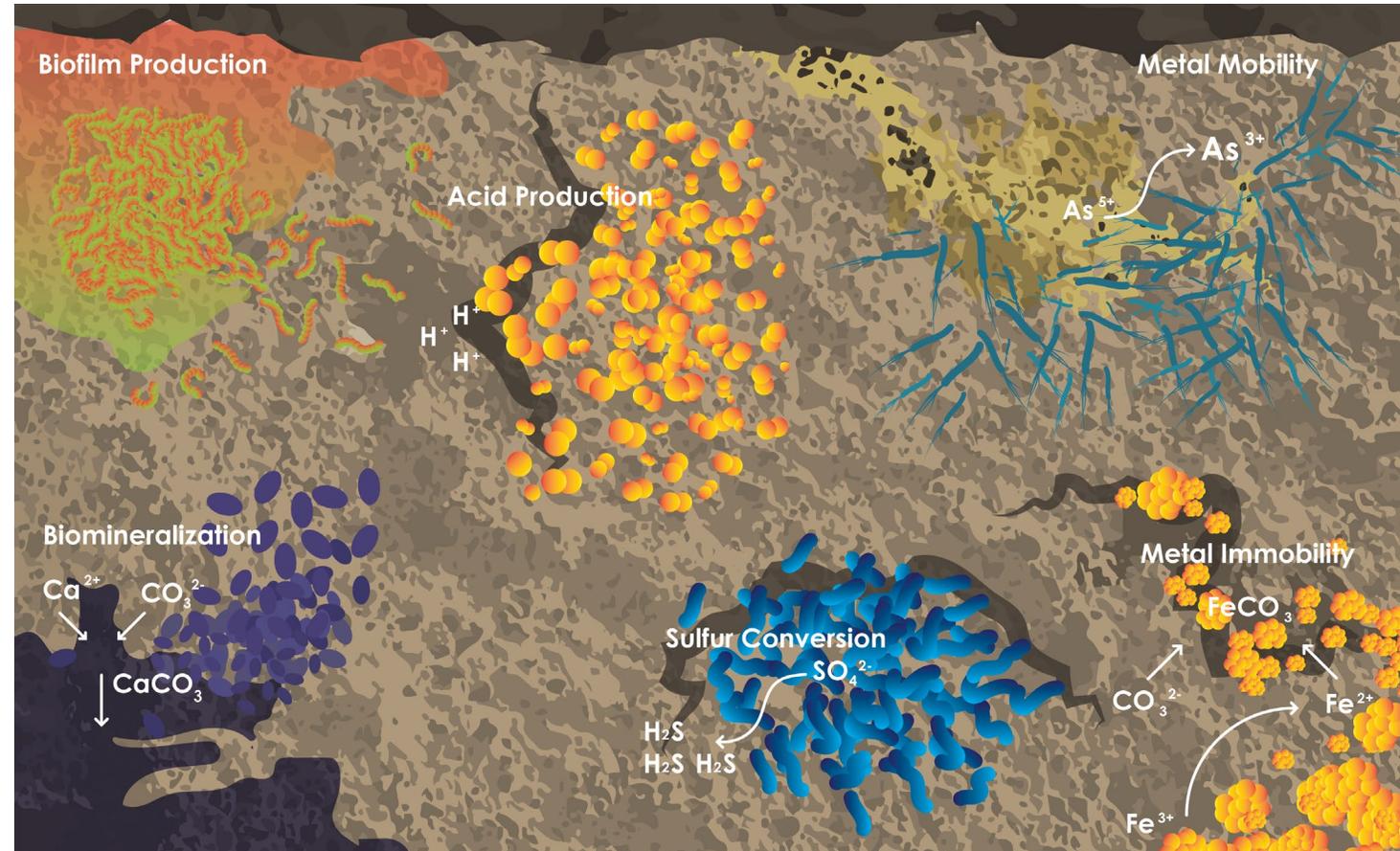**NETL has 8 years of experience characterizing the biogeochemistry of produced water.**

- Currently have hundreds of DNA datasets with relevant metadata

- USGS additionally has hundreds of DNA datasets

- Several Universities have hundreds of DNA datasets

- Data exists, just not in a usable form

Legend:
- Selenomonadales; Veillonellaceae
- Clostridiales; Peptostreptococcaceae
- Clostridiales; Lachnospiraceae
- Clostridiales; Family XII
- Clostridiales; Family XI
- Clostridiales; Eubacteriaceae
- Clostridiales; Clostridiaceae 4
- Bacillales; Bacillaceae
- Campylobacterales; Arcobacteraceae
- Deferribacterales; Deferribacteraceae
- Bacteroidales; Prolixibacteraceae
- Bacteroidales; Muribaculaceae
- Methanobacteriales; Methanobacteriaceae
- Other
- Pseudomonadales; Pseudomonadaceae
- Pseudomonadales; Moraxellaceae
- Methylococcales; Methylomonaceae
- Betaproteobacteriales; Burkholderiaceae
- Alteromonadales; Shewanellaceae
- Desulfovibrionales; Desulfovibrionaceae
- Desulfovibrionales; Desulfohalobiaceae
- Sphingomonadales; Sphingomonadaceae
- Rhodobacterales; Rhodobacteraceae
- Rhizobiales; Xanthobacteraceae
- Halanaerobiales; Halobacteroidaceae
- Halanaerobiales; Halanaerobiaceae

Importance of microbiology

- Assessment of microbiology can better predict chemical reactions:

  - Mineralization

  - Biofilm production

  - Acid production

  - H_2S production

- Assessment of microbiology can result in better targeted biological control treatments

- Assessment of microbiology can signal subsurface behavior

- Data exists, but need to make it more useable

Bakken Shale Correlation Example

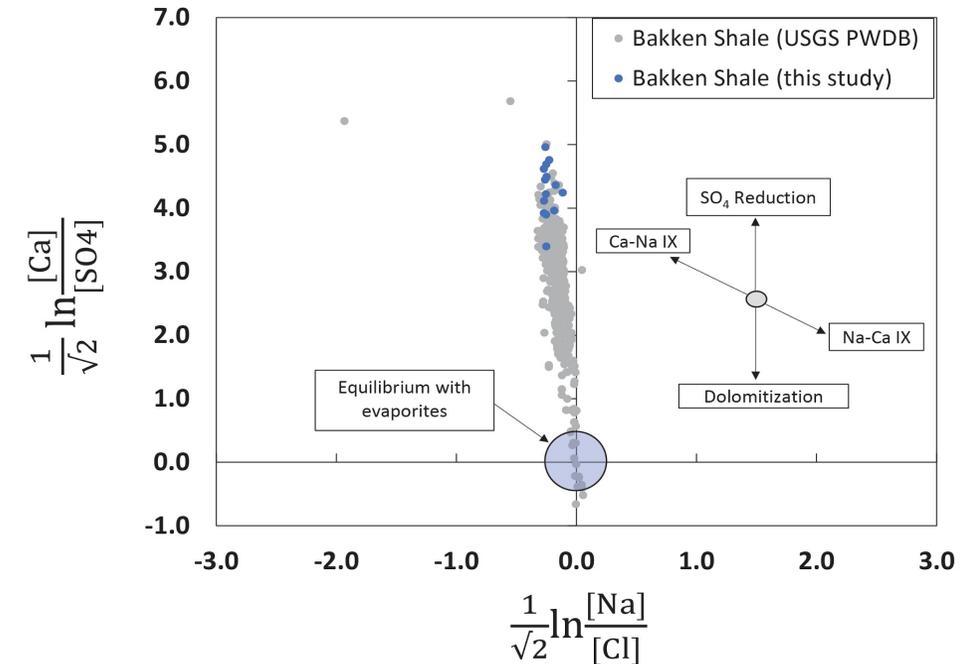| Well #	Well Age* (months)	Formation**	TDS (mg/L)	pH	Alkalinity (mg/L)	TIC (mg/L)	NPOC (mg/L)	16S rRNA gene copes/mL	
1	52	BKN	215484	6.5	330	23.5	291	223	
3	49	BKN	277081	6.3	500	29.8	76	83	
4	49	TF1	229002	6.5	450	47.2	139	149	
5	79	BKN	330664	6.3	1440	13.2	116	102	
6	79	TF1	326055	5.8	2000	20.9	140	66	
7	49	BKN	343563	6.1	1600	12.3	342	63	
8	49	BKN	354667	6.1	740	11.7	63	180	
10	61	BKN	294780	6.4	1750	16.4	407	65	
11	55	BKN	134561	6.9	440	38.2	543	305800	Mostly sulfate reducers
12	55	TF1	303176	6.7	560	13.8	71	22	
13	55	TF2	311195	6.4	398	12.2	90	404	
14	55	BKN	270392	6.8	426	19.1	333	325	
17	53	TF1	356018	6.4	186	4.1	66	127	
18	1	U	290236	6.4	310	24.8	326	30	

*Well age at May 2018 sampling event
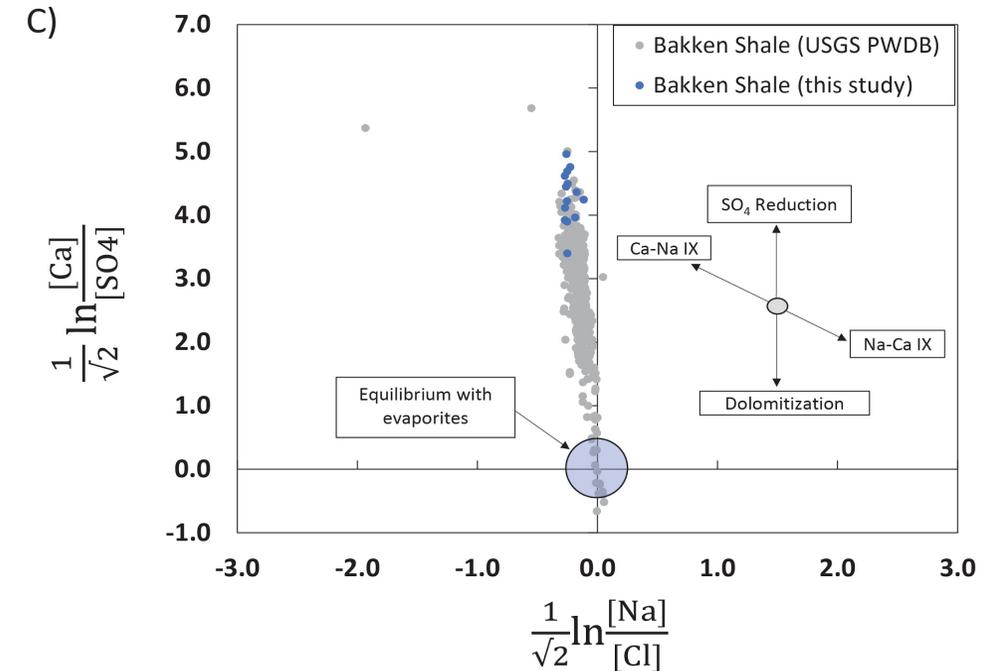
**BKN = Bakken formation; TF1 = First bench of Three Forks formation; TF2 = Second bench of Three Forks formation; U = Unavailable Data

U.S. DEPARTMENT OF ENERGY

Previous Work in Produced Water Microbiology

Bakken Shale Correlation Example

Well #	Well Age* (months)	Formation**	TDS (mg/L)	pH	Alkalinity (mg/L)	TIC (mg/L)	NPOC (mg/L)	16S rRNA gene copes/mL
1	52	BKN	215484	6.5	330	23.5	291	223
3	49	BKN	277081	6.3	500	29.8	76	83
4	49	TF1	229002	6.5	450	47.2	139	149
5	79	BKN	330664	6.3	1440	13.2	116	102
6	79	TF1	326055	5.8	2000	20.9	140	66
7	49	BKN	343563	6.1	1600	12.3	342	63
8	49	BKN	354667	6.1	740	11.7	63	180
10	61	BKN	294780	6.4	1750	16.4	407	65
11	55	BKN	134561	6.9	440	38.2	543	305800
12	55	TF1	303176	6.7	560	13.8	71	22
13	55	TF2	311195	6.4	398	12.2	90	404
14	55	BKN	270392	6.8	426	19.1	333	325
17	53	TF1	356018	6.4	186	4.1	66	127
18	1	U	290236	6.4	310	24.8	326	30

Mostly sulfate reducers (annotation at Well 11, 305800)

*Well age at May 2018 sampling event

**BKN = Bakken formation; TF1 = First bench of Three Forks formation; TF2 = Second bench of Three Forks formation; U = Unavailable Data

Plot axes: x-axis $\frac{1}{\sqrt{2}}\ln\frac{[Na]}{[Cl]}$, y-axis $\frac{1}{\sqrt{2}}\ln\frac{[Ca]}{[SO4]}$

Legend: Bakken Shale (USGS PWDB); Bakken Shale (this study)

Annotations: SO$_4$ Reduction; Ca-Na IX; Na-Ca IX; Dolomitization; Equilibrium with evaporites

Bakken Shale Correlation Example

Well #	Well Age* (months)	Formation**	TDS (mg/L)	pH	Alkalinity (mg/L)	TIC (mg/L)	NPOC (mg/L)	16S rRNA gene copes/mL
1	52	BKN	215484	6.5	330	23.5	291	223
3	49	BKN	277081	6.3	500	29.8	76	83
4	49	TF1	229002	6.5	450	47.2	139	149
5	79	BKN	330664	6.3	1440	13.2	116	102
6	79	TF1	326055	5.8	2000	20.9	140	66
7	49	BKN	343563	6.1	1600	12.3	342	63
8	49	BKN	354667	6.1	740	11.7	63	180
10	61	BKN	294780	6.4	1750	16.4	407	65
11	55	BKN	134561	6.9	440	38.2	543	305800 Mostly sulfate reducers
12	55	TF1	303176	6.7	560	13.8	71	22
13	55	TF2	311195	6.4	398	12.2	90	404
14	55	BKN	270392	6.8	426	19.1	333	325
17	53	TF1	356018	6.4	186	4.1	66	127
18	1	U	290236	6.4	310	24.8	326	30

*Well age at May 2018 sampling event

**BKN = Bakken formation; TF1 = First bench of Three Forks formation; TF2 = Second bench of Three Forks formation; U = Unavailable Data

C)

16S rRNA gene copies and geochemistry are not deposited in NCBI

Tinker, Gardiner, J., K., Lipus, D., Sarkar, P., Stuckman, M., **Gulliver, D. M.**, 2020. Geochemistry and microbiology predict environmental niches with conditions favoring potential activity in the Bakken Shale, *Frontiers in Microbiology*. doi: 10.3389/fmicb.2020.01781

Bakken Shale Correlation Example

Well #	Well Age* (months)	Formation**	TDS (mg/L)	pH	A			
1	52	BKN	215484	6.5				
3	49	BKN	277081	6.3				
4	49	TF1	229002	6.5				
5	79	BKN	330664	6.3				
6	79	TF1	326055	5.8				
7	49	BKN	343563	6.1				
8	49	BKN	354667	6.1				
10	61	BKN	294780	6.4	1750	16.4	407	65
11	55	BKN	134561	6.9	440	38.2	543	305800 Mostly sulfate reducers
12	55	TF1	303176	6.7	560	13.8	71	22
13	55	TF2	311195	6.4	398	12.2	90	404
14	55	BKN	270392	6.8	426	19.1	333	325
17	53	TF1	356018	6.4	186	4.1	66	127
18	1	U	290236	6.4	310	24.8	326	30

*Well age at May 2018 sampling event

**BKN = Bakken formation; TF1 = First bench of Three Forks formation; TF2 = Second bench of Three Forks formation; U = Unavailable Data

3 Reason why good example
- Use of big database
- Use of metadata
- Use of multiple produced waters

The holistic approach was required to understand that well 11 wasn't an outlier

- Bakken Shale (USGS PWDB)
- Bakken Shale (this study)

SO_4 Reduction

Ca-Na IX

Na-Ca IX

Equilibrium with evaporites

Dolomitization

0.0

-1.0

-3.0   -2.0   -1.0   0.0   1.0   2.0   3.0

$$\frac{1}{\sqrt{2}}\ln\frac{[Na]}{[Cl]}$$

16S rRNA gene copies and geochemistry are not deposited in NCBI

Curated microbial DNA Database

16 GB

Curated microbial DNA Database

Build a DNA Database for PW systems
- Curated to make it usable
- Metadata (geochemistry, qPCR) available for each dataset
- Categories for better visualization/filtering

Produced Water
DNA Database

Welcome to the PW-DNA database build in collaboration with

The goal of the PW-DNA database is to improve the ability for researchers to characterize produced water to help determine options for treatment, reuse, and recovery of valuable resources. These samples are found across North America from Shale, Coal Bed Methane, and Crude Oil resources. This narrative includes 6 files that you can either interact with inside of KBase or you can download to use on your own hardware. The files available are:

File	Number of Samples Included	Where to find Narrative
Amplicon	47	https://narrative.kbase.us/narrative/156895
MetaG Full Assembly	26	https://narrative.kbase.us/narrative/156896
MetaG Bins and Genome Isolates	XXX	https://narrative.kbase.us/narrative/156897
MetaT Assembly	4	https://narrative.kbase.us/narrative/156898
qPCR Analysis	17	https://narrative.kbase.us/narrative/156899
Clone Libraries	13	https://narrative.kbase.us/narrative/156900

These files were generated within KBase using publically available data from databases such as: NCBI, JGI-IMG, JGI-Gold, NMDC, and MR-RAST. The DOIs for each of the work and the KBase narrative for the data is available:

Database

| PW-DNA ID	DOI	Author	Sample Type	Samples Inlcuded (with KBase Narrative links)
PWDNA-C1	https://doi.org/10.1111/j.1574-6941.2011.01272.x	Wawrik et al. (2012)	Coal	Amplicon and qPCR
PWDNA-C2	https://doi.org/10.1016/j.coal.2013.03.006	Barnhart et al. (2013)	Coal	Clone Library
PWDNA-C3	https://doi.org/10.1128/aem.01737-15	Lawson et al. (2015)	Coal	MetaG and Amplicon
PWDNA-C4	https://doi.org/10.1016/j.coal.2016.05.001	Barnhart et al. (2016)	Coal	Clone Library
PWDNA-C5	https://doi.org/10.1128/genomea.01024-16	Ross et al (2016)	Coal	Genome Isolate
PWDNA-C6	https://doi.org/10.1016/j.fuel.2017.09.074	Davis et al. (2018)	Coal	Amplicon
PWDNA-C7	https://doi.org/10.1016/j.gca.2018.11.009	Schweitzer et al. (2019)	Coal	Amplicon and qPCR
PWDNA-C8	https://doi.org/10.1038/s41396-021-01139-x	McKay et al. (2021)	Coal	MetaG
PWDNA-C9	https://doi.org/10.1016/j.coal.2021.103860	Smith et al. (2021)	Coal	Amplicon and qPCR
PWDNA-C10	https://doi.org/10.1038/s41522-022-00267-2	Schweitzer et al. (2022)	Coal	MetaG
PWDNA-C11	https://doi.org/10.3389%2Ffmicb.2023.1097500	Platt et al. (2023)	Coal	Amplicon
PWDNA-C12	unpublished	Schweitzer et al.	Coal	Amplicon
PWDNA-C13	unpublished	Schweitzer et al.	Coal	MetaG and MetaT
PWDNA-CO1	https://doi.org/10.1021/es4020184	An et al. (2013)	Coal, Oil	MetaG and Amplicon
PWDNA-O1	https://doi.org/10.1016/j.resmic.2005.03.009	Grabowski et al. (2005)	Oil	Clone Library
PWDNA-O2	https://doi.org/10.1111/j.1462-2920.2008.01751.x	Pham et al. (2009)	Oil	Clone Library
PWDNA-O3	https://doi.org/10.1371/journal.pone.0023258	Ren et al. (2011)	Oil	Clone Library
PWDNA-O5	https://doi.org/10.3389/fmicb.2014.00197	Berdugo-Clavijo (2014)	Oil	Amplicon
PWDNA-O4	https://doi.org/10.3389%2Ffmicb.2014.00409	Piceno et al (2014)	Oil	Clone Library
PWDNA-O7	https://doi.org/10.1016/j.jes.2015.04.030	Wang et al. (2015)	Oil	Clone Library
PWDNA-O6	https://doi-org.proxybz.lib.montana.edu/10.1007/s11356-015-4947-z	Yeung et al. (2015)	Oil	Clone Library and qPCR
PWDNA-O8	https://doi.org/10.1128/mbio.01669-15	Hu et al. (2016)	Oil	MetaG
PWDNA-O9	https://doi.org/10.1038/ismej.2017.78	Vigneron et al. (2017)	Oil	MetaG, Amplicon and qPCR
PWDNA-O10	https://doi.org/10.3389/fmicb.2017.00099	Bonifay et al. (2017)	Oil	MetaG, Amplicon and qPCR
PWDNA-O11	https://doi.org/10.1016/j.apenergy.2016.10.057	Sheikhyousefi et al. (2017)	Oil	Amplicon
PWDNA-O12	https://doi.org/10.1186/s40168-017-0392-1	Liu et al. (2018)	Oil	MetaG and MetaT
PWDNA-O13	https://doi.org/10.1128/AEM.01819-20	Lahme et al. (2021)	Oil	MetaG, Amplicon and qPCR
PWDNA-O14	https://doi.org/10.1128/spectrum.00770-22	Campa et al. (2022)	Oil	MetaT and Amplicon
PWDNA-S1	https://doi.org/10.1128/AEM.00233-11	Struchtemeyer et al. (2011)	Shale	Clone Library and qPCR
PWDNA-S2	https://doi.org/10.1007/s00248-012-0073-3	Davis et al. (2012)	Shale	Amplicon
PWDNA-S3	https://doi.org/10.1111/j.1574-6941.2011.01196.x	Struchtemeyer et al. (2012)	Shale	Amplicon
PWDNA-S4	https://doi.org/10.1021/es402928b	Mohan et al. (2013)	Shale	Clone Library and qPCR
PWDNA-S5	https://doi.org/10.1111/1574-6941.12183	Mohan et al. (2013)	Shale	Clone Library and qPCR
PWDNA-S6	https://doi.org/10.3389/fmicb.2013.00367	Wuchter et al. (2013)	Shale	Amplicon
PWDNA-S7	https://doi.org/10.1021/es501173p	Cluff et al. (2014)	Shale	Amplicon
PWDNA-S8	https://doi.org/10.1371/journal.pone.0107682	Mohan et al. (2014)	Shale	Amplicon and qPCR
PWDNA-S9	https://doi.org/10.1061/(ASCE)EE.1943-7870.0000792	Strong et al. (2014)	Shale	Amplicon
PWDNA-S10	https://doi.org/10.1016/j.apgeochem.2015.04.011	Akob et al. (2015)	Shale	Clone Library
PWDNA-S11	https://doi.org/10.1038/nmicrobiol.2016.146	Daly et al. (2016)	Shale	MetaG
PWDNA-S12	https://doi.org/10.3389/fmicb.2016.00988	Liang et al. (2016)	Shale	Amplicon
PWDNA-S13	https://doi.org/10.1128/genomea.00899-16	Lipus et al. (2016)	Shale	Genome Isolate
PWDNA-S14	https://doi.org/10.1007/s00248-016-0811-z	Vikram et al. (2016)	Shale	MetaT, Amplicon and qPCR
PWDNA-S15	https://doi.org/10.3389/fmicb.2017.01164	An et al. (2017)	Shale	Amplicon
PWDNA-S16	https://doi.org/10.1128/mspheredirect.00257-17	Booker et al. (2017)	Shale	MetaG
PWDNA-S17	https://doi.org/10.1128/AEM.02659-16	Lipus et al. (2017)	Shale	Amplicon and qPCR
PWDNA-S18	https://doi.org/10.1039/C6EW00286B	Zhang et al. (2017)	Shale	Amplicon
PWDNA-S19	https://doi.org/10.1073/pnas.1800155115	Borton et al. (2018)	Shale	MetaG
PWDNA-S20	https://doi.org/10.1111/1462-2920.14467	Borton et al. (2018)	Shale	Genome Isolate
PWDNA-S21	https://doi.org/10.3389/fmicb.2018.02646	Evans et al. (2018)	Shale	MetaG and Genome Isolate
PWDNA-S22	https://doi.org/10.1016/j.scitotenv.2018.06.067	Hull et al. (2018)	Shale	Amplicon
PWDNA-S23	https://doi.org/10.1093/femsle/fny107	Lipus et al. (2018)	Shale	Amplicon and qPCR
PWDNA-S24	https://doi.org/10.1128/genomea.00159-18	Panescu et al. (2018)	Shale	Genome Isolate
PWDNA-S25	https://doi.org/10.1039/C8EM00291F	Roger et al. (2018)	Shale	Amplicon
PWDNA-S26	https://doi.org/10.1128/genomea.00155-18	Tummings et al. (2018)	Shale	Genome Isolate
PWDNA-S27	https://doi.org/10.1186/s13568-019-0766-9	Berdugo-Clavijo et al. (2019)	Shale	Amplicon
PWDNA-S28	https://doi.org/10.1128/AEM.00018-19	Booker et al. (2019)	Shale	Genome Isolate
PWDNA-S29	https://doi.org/10.1038/s41564-018-0312-6	Daly et al. (2019)	Shale	MetaG
PWDNA-S30	https://doi.org/10.1021/acs.estlett.9b00473	Evans et al. (2019)	Shale	MetaG
PWDNA-S31	https://doi.org/10.1038/s41396-019-0466-0	Evans et al. (2019)	Shale	MetaG and Genome Isolate
PWDNA-S32	https://doi.org/10.1080/01490451.2019.1599470	Lipus et al. (2019)	Shale	MetaG
PWDNA-S33	https://doi.org/10.3389/fmicb.2019.00376	Morono et al. (2019)	Shale	Amplicon
PWDNA-S34	https://doi.org/10.1128/mSphere.00613-19	Nixon et al. (2019)	Shale	MetaG
PWDNA-S35	https://doi.org/10.1016/j.watres.2019.114942	Wang et al. (2019)	Shale	Amplicon
PWDNA-S36	https://doi.org/10.1039/C6EW00286B	Zhong et al. (2019)	Shale	Amplicon
PWDNA-S37	https://doi.org/10.3389/fmicb.2020.00286	Cliffe et al. (2020)	Shale	MetaG and Amplicon
PWDNA-S38	https://doi.org/10.1128/mSystems.00098-20	Danczak et al. (2020)	Shale	MetaG
PWDNA-S39	https://doi.org/10.3389/fmicb.2020.536978	Shelton et al. (2020)	Shale	Amplicon and qPCR
PWDNA-S40	https://doi.org/10.3389/fmicb.2020.01781	Tinker et al. (2020)	Shale	Amplicon and qPCR
PWDNA-S41	https://doi.org/10.15530/urtec-2020-2979	Tinker et al. (2020)	Shale	Amplicon
PWDNA-S42	https://doi.org/10.1038/s41598-020-73010-6	Zhang et al. (2020)	Shale	MetaG and Amplicon
PWDNA-S43	https://doi.org/10.1021/acsearthspacechem.1c00087	Stemple et al. (2021)	Shale	MetaG, Amplicon and qPCR
PWDNA-S44	https://doi.org/10.1186/s40168-021-01194-8	Amundson et al. (2022)	Shale	MetaG and Amplicon
PWDNA-S45	https://doi.org/10.1128/spectrum.03640-22	Cliffe et al. (2022)	Shale	Amplicon
PWDNA-S46	https://doi.org/10.1128/spectrum.00049-22	Tinker et al. (2022)	Shale	MetaG and Amplicon
PWDNA-S47	https://doi.org/10.1016/j.jhazmat.2021.127649	Zhou et al. (2022)	Shale	MetaG and Amplicon
PWDNA-S48	https://doi.org/10.1186/s40793-023-00465-1	Hernandez-Becerra et al. (2023)	Shale	Amplicon
PWDNA-S49	unpublished	Schweitzer et al.	Shale	Amplicon

Database

MetaData and Geochemical Data

All the data available throughout PW-DNA database is connected by a unique identifier (PWDNA-###). This identifier also connects to other databases such as the

Produced Waters Geochemical Database v2.3

The Produced Waters Geochemical Database v2.3 contains other published geochemical data that may not contain molecular data like the Produced Water DNA Database. If you are interested in more geochemical data the USGS Produced Water Geochemical Database may be of interest to you. All data can be filtered via the metadata and the geochemical data using any of the narratives. The samples found in the PW-DNA have been collect across the United States and also from China, Australia, and Norway. Data can be visualized and filtered below

Database

Database
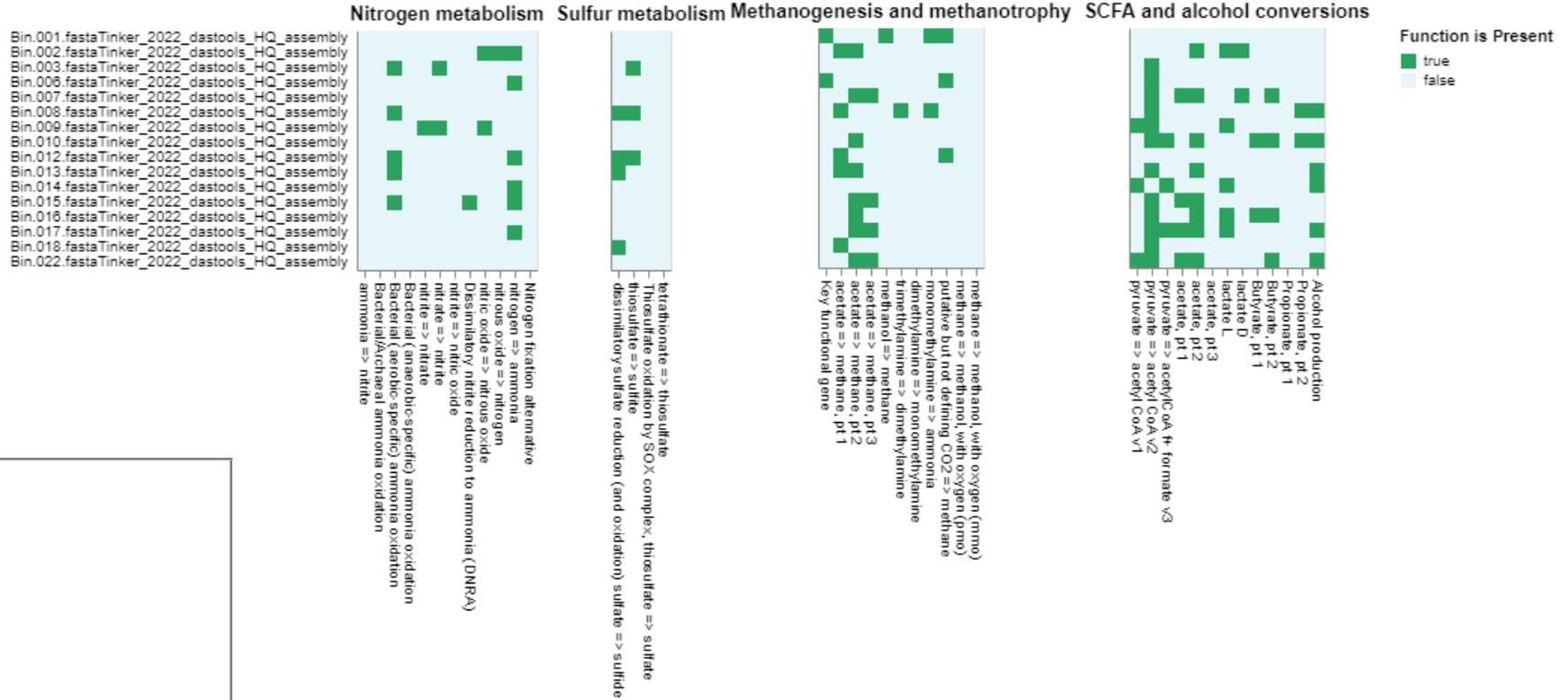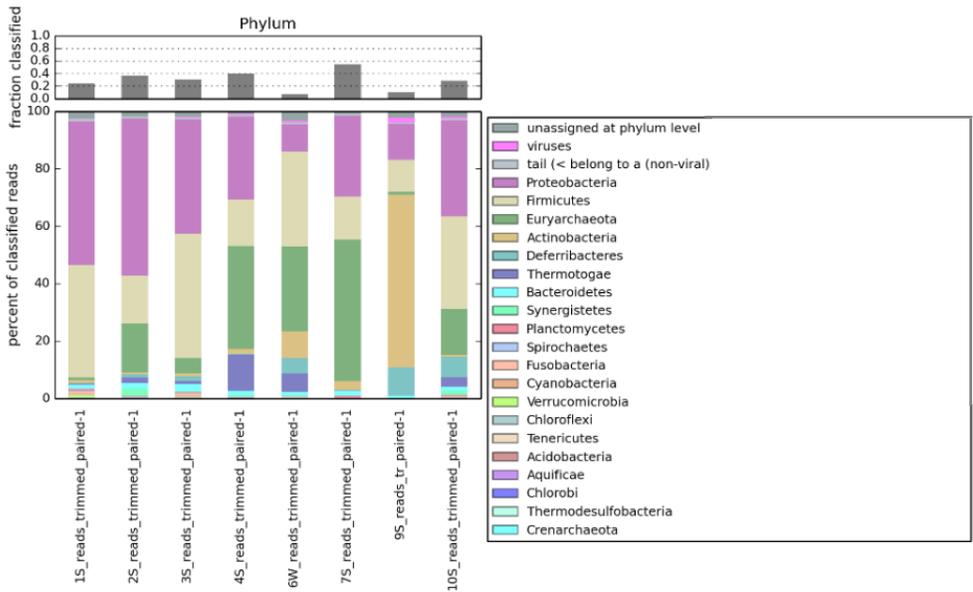
Progress

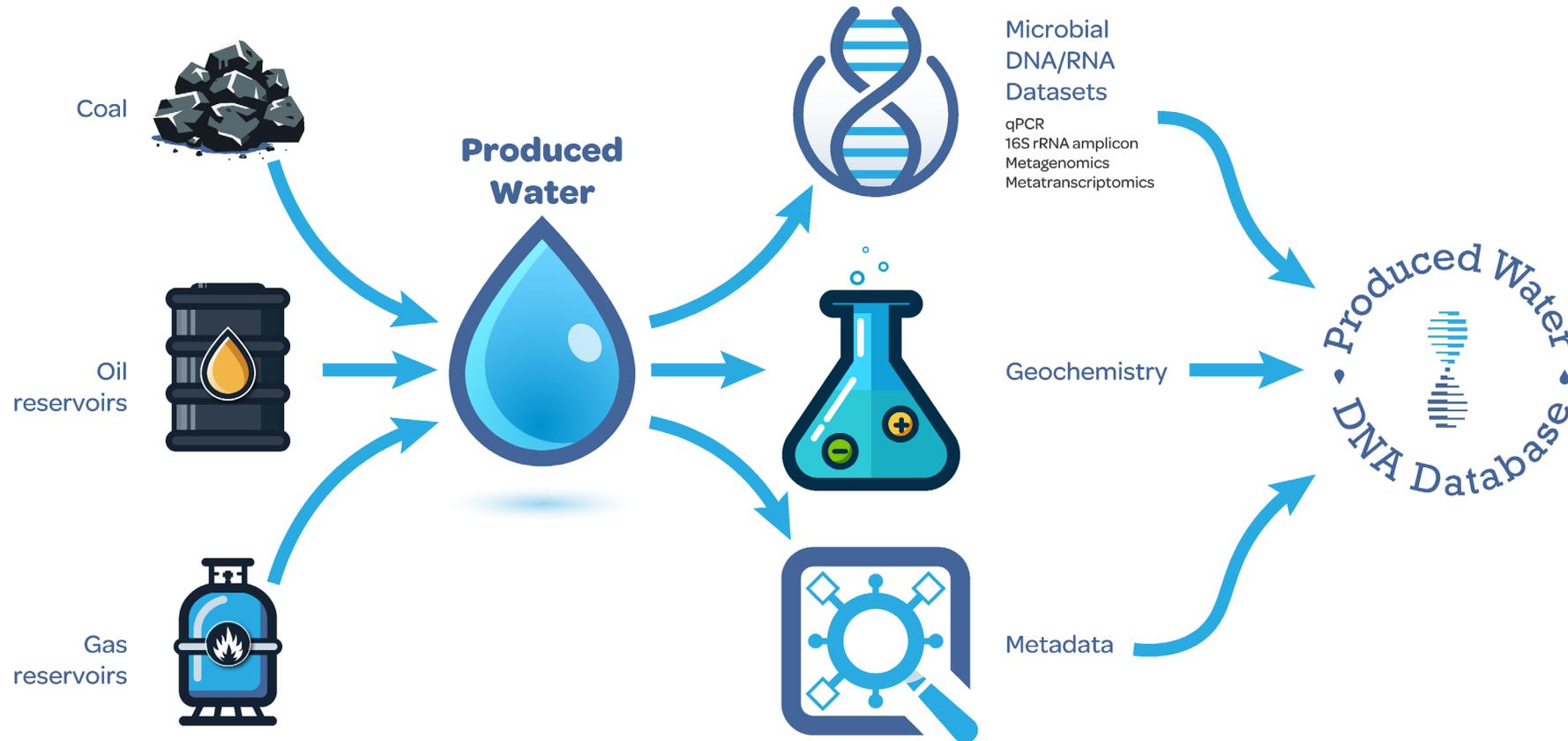| Identifier	Type[1]	Expected Completion Date	Description (What, How, Who, Where)
EY23.2.A	Major	12/31/2023	Completed. Compilation of relevant produced water DNA datasets and the relevant metadata.
EY24.2.B	Major	12/31/2024	On Track. Launch of public-facing DNA database for produced water management.
EY24.2.C	Major	03/31/2025	Network analysis and identified keystone species using produced water DNA datasets compiled in EY22–EY23.
EY25.2.D	Project	03/31/2026	Correlation of network analysis with geochemistry metadata to determine biogeochemical trends of produced water systems.
EY25.2.E	Major	03/31/2026	Management guidance to ensure the database remains current and curated.

Completed:
- Comprehensive literature search
- Downloaded DNA sequencing data from institutions and NCBI
- Designed KBase tutorials to guide users through KBase pipelines
- Designed KBase narrative page
- Applied to be part of the Powell Center program

Currently in process of:
- Formatting Geochemistry data and metadata
- Running all DNA Sequences through same bioinformatic platforms
- Determining best app for geospatial visualization
- Drafting a FEMS Perspective

Questions

Contact: PW-DNA@netl.doe.gov

NATIONAL ENERGY TECHNOLOGY LABORATORY

Coal

Oil reservoirs

Gas reservoirs

Produced Water

Microbial DNA/RNA Datasets
qPCR
16S rRNA amplicon
Metagenomics
Metatranscriptomics

Geochemistry

Metadata

Produced Water · DNA Database ·
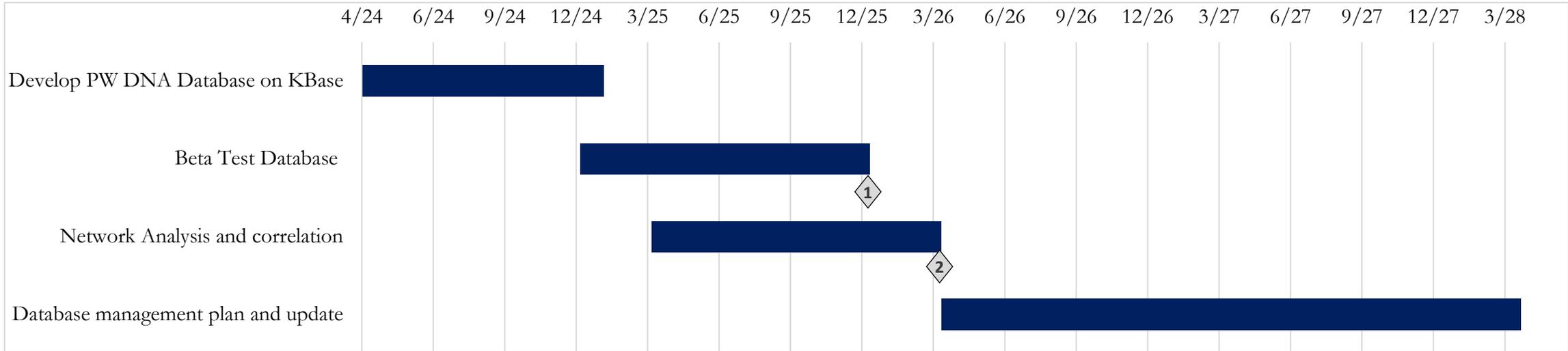
Organization Chart

NETL

- Djuna Gulliver, NETL-RIC (PI)
- Kara Tinker, Leidos-NETL-RIC
- Hannah Schweitzer, ORISE-NETL-RIC
- Preom Sarkar, ORISE-NETL-RIC

USGS

- Denise Akob
- Elliott Barnhart

Gantt Chart

	4/24	6/24	9/24	12/24	3/25	6/25	9/25	12/25	3/26	6/26	9/26	12/26	3/27	6/27	9/27	12/27	3/28

Tasks:
- Develop PW DNA Database on KBase
- Beta Test Database
- Network Analysis and correlation
- Database management plan and update

◇ **Milestones**

1. Public launch of PW DNA Database.(Q4, January 2025)
2. Demonstration of PW DNA Database through completion of network analysis.(Q4, March 2026)