

NEMC Conference: Current Topics in Microbiology • Wednesday, August 7, 2019

Next Generation Sequencing to Detect Environmental Stress

Krista Ternus, PhD

Signature Science, LLC

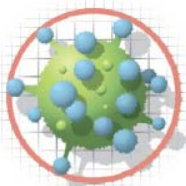
kternus@signaturescience.com



signature
science[®] LLC

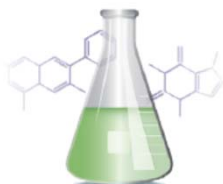
www.signaturescience.com

Signature Science, LLC



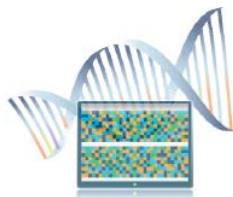
Life & Physical Sciences

- Environmental genomics
- Proteogenomics
- Applied Materials Science



Laboratory Support Services

- Quality Assurance
- Proficiency Testing
- Third Party Test and Evaluation



DNA Forensics

- Human DNA Casework
- Next Generation Sequencing
- Advanced Analytics



Data Science

- Applied Statistics
- Bioinformatics and Chemometrics
- Natural Language Processing



Intel & Operations

- Training, Exercise, and Mission Rehearsal



Systems Integration/Engineering

- Chemical Sensor Technology Design, Integration, Engineering, and Support
- Environmental Collection and Sampling

More About Us:

2001

- Founded in March 2001
- A Subsidiary of Southwest Research Institute

40

- ~ \$40 M Annual Revenue in FY18
- Federal Government Clientele

215

- 215 Employees
- Four Locations Across the U.S.

100

- 100+ Business and Academic Partner Network

Austin, TX
(HQ)



Atlantic City, NJ
Arlington, VA
Charlottesville, VA

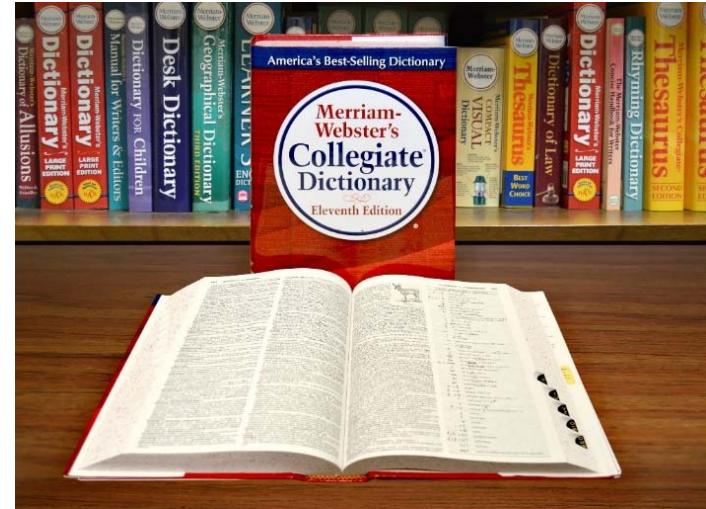
signature[®]
science LLC

Environmental Metagenomics



<https://www.nps.gov/subjects/geology/sri.htm>

Great for unbiased,
exploratory analyses, as
well as when standard
assays provide
ambiguous or no results



signature
science[®] LLC

Metagenomics: Powerful Investigative Tool and Regulatory Challenge

The New York Times

In a First, Test of DNA Finds Root of Illness



Joshua Osborn outside his home in Cottage Grove, Wis., with his father, Clark Osborn. Joshua suffered from swelling in the brain, but tests, a spinal tap and a biopsy were inconclusive. John Maniaci.

www.nytimes.com/2014/06/05/health/in-first-quick-dna-test-diagnoses-a-boys-illness.html

Contains Nonbinding Recommendations

Draft - Not for Implementation

Infectious Disease Next Generation Sequencing Based Diagnostic Devices: Microbial Identification and Detection of Antimicrobial Resistance and Virulence Markers

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This draft guidance document is being distributed for comment purposes only.

Document issued on: May 13, 2016

You should submit comments and suggestions regarding this draft document within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to <http://www.regulations.gov>. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Identify all comments with the docket number listed in the notice of availability that publishes in the *Federal Register*.

Finding Real World Data

SRA Search Results

Search: copper mine illumina metagenomics soil

Access: Public (22)
Source: DNA (22)
Type: genome (1)

Summary: 20 per page

View results as an expanded interactive table using the RunSelector. [Send results to Run selector](#)

Search results
Items: 1 to 20 of 22

1. microbial diversity of waste areas in Dexin copper mine : Sample 2#DG03
1 ILLUMINA (Illumina MiSeq) run: 39,304 spots, 11.8M bases, 6.3Mb downloads
Accession: SRX1615817
2. microbial diversity of waste areas in D...
1 ILLUMINA (Illumina MiSeq) run: 43,643 spots
Accession: SRX1615816
3. microbial diversity of waste areas in D...
1 ILLUMINA (Illumina MiSeq) run: 35,855 spots
Accession: SRX1615815
4. microbial diversity of waste areas in D...
1 ILLUMINA (Illumina MiSeq) run: 27,185 spots
Accession: SRX1615814
5. microbial diversity of waste areas in D...
1 ILLUMINA (Illumina MiSeq) run: 25,000 spots
Accession: SRX1615813

BioProject Details

Search: Copper Stressed Treatment and Controls

Accession: PRJEB24797 ID: 436543

Comparison of copper contaminated soils, with similar uncontaminated controls.
Two sites of long-term copper polluted soil were sampled, along with nearby uncontaminated controls. [More...](#)

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	14
OTHER DATASETS	
BioSample	14

Project Data:

Parameter	Value
Data volume, Gbases	15
Data volume, Mbytes	8311

Related information
BioSample
SRA

Recent activity
Turn Off Clear

- Copper Stressed Treatment and Controls (BioProject)
- 4 (BioSample)
- ERR2282012 (1) (SRA)
- copper mine illumina metagenomics soil (22) (SRA)
- copper illumina metagenomics soil (69) (SRA)

[See more...](#)

www.ncbi.nlm.nih.gov/sra

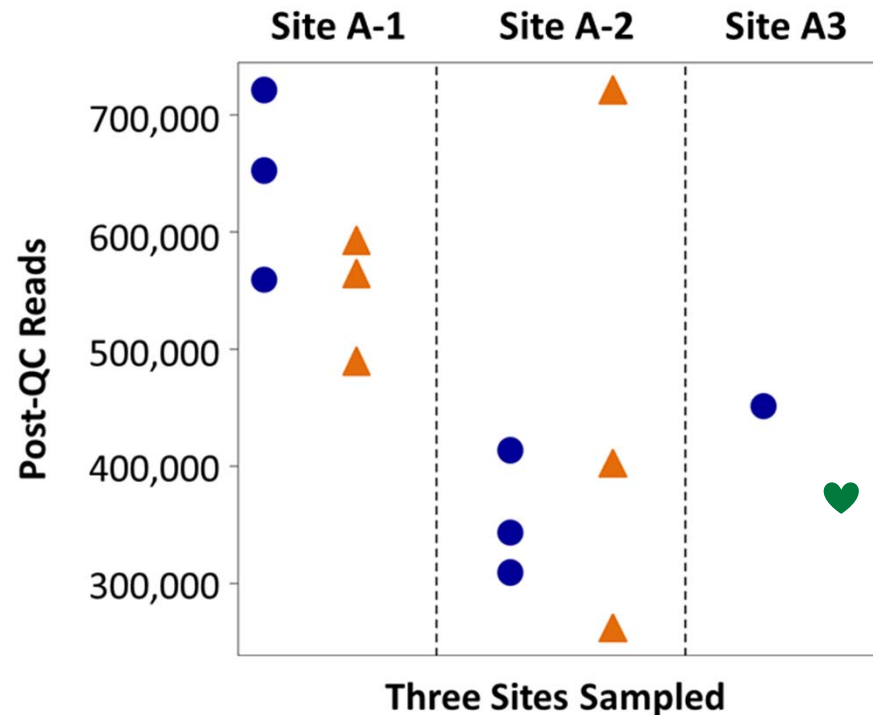
www.ncbi.nlm.nih.gov/bioproject/

Acid Mine Drainage



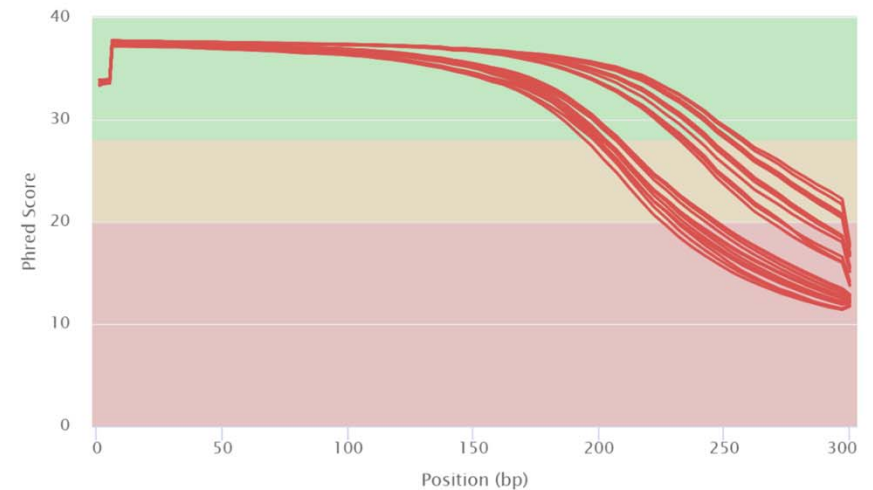
<https://www.fws.gov/mountain-prairie/contaminants/acidMineDrainage.php>

Quality Control

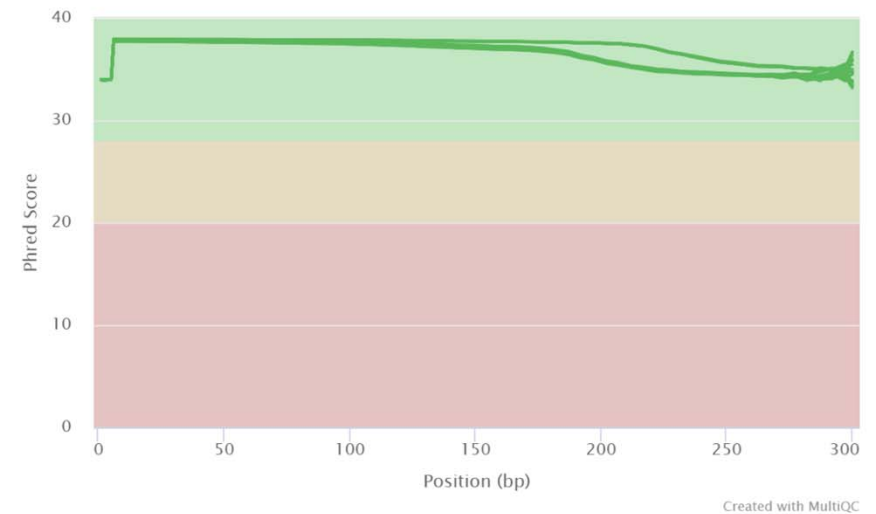


▲	Controls
●	Copper
♥	Copper and Arsenic

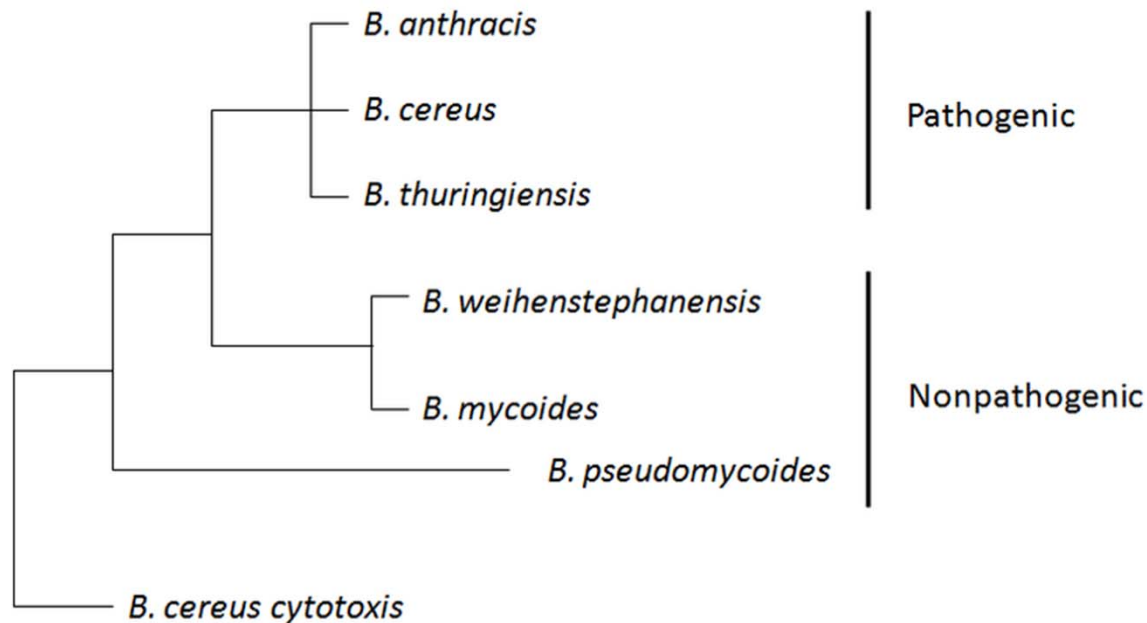
FastQC: Mean Quality Scores



FastQC: Mean Quality Scores



Methods for Taxonomic Classification



RLA0_PYRAE	-MMLAIGKRRYVRTQYPAKRVKIVSEATELLQKYPYVFLFDLHGLSSRLHEVRYRLRRY-GVIKLIKPTLFKIAFTKVYGG---TPAE	85
RLA0_METAC	-----MAEERHHTTEHIPQWKKDEIENIKELIQSHKVFGMVIEGILATKMKIRRDLDKV-AVLKVSRLTLTERALNQLG-----ETIP	78
RLA0_METMA	-----MAEERHHTTEHIPQWKKDEIENIKELIQSHKVFGMVIEGILATKMKIRRDLDKV-AVLKVSRLTLTERALNQLG-----ESIP	78
RLA0_ARCFU	-----MAAVRGS-----PPEYKVRVVEEIKRMISSKPVVAIVSFRNVFAGQOMKIRREFRGK-AEIKVVKNLTLERALDAG-----GDYL	75
RLA0_METKA	MAVKAKGQPPSGYEPKVAEWKRREVKELKLMDEYENVGLVDLEGIPAPLOEIRAKLRERDIIIRMSRNTLMRIALEEKLDER--PELE	88
RLA0_METTH	-----MAHVAEWKKKEVQELHDLIKGYEVVGIANLADIPARQLOKMRQTLRDS-ALIRMSKKLLISLALAKAGREL--ENVV	74
RLA0_METTL	-----MITAESEHKIAPWKIEEVNKLKLLKNGQIVALVDMMEVPARQLOEIRDKIR-GTMTLKMSRNTLIERAIKEVAEETGNPEFA	82
RLA0_METVA	-----MIDAKSEHKIAPWKIEEVNKLKLLKNSANVIALIDMMEVPARQLOEIRDKIR-DQMTLKMSRNTLIERAIKEVAEETGNPEFA	82
RLA0_METJA	-----METKVKAHVAPWKIEEVKTLKGLIKSKPVVAIVDMMDVFPAPLOEIRDKIR-DKVVKLRMSRNTLIERALKEAAEELNNPKLA	81
RLA0_PYRAB	-----MAHVAEWKKKEVEELANLIKSPYVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSRNTLIELAIKKAAGELGKPELE	77
RLA0_PYRHO	-----MAHVAEWKKKEVEELAKLIKSPYVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSRNTLIELAIKKAAGELGKPELE	77
RLA0_PYRFU	-----MAHVAEWKKKEVEELANLIKSPYVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSRNTLIELAIKKAAGELGKPELE	77
RLA0_PYRKO	-----MAHVAEWKKKEVEELANLIKSPYVIALVDVAGVPAYPLSKMRDKLR-GKALLRVSRNTLIELAIKKAAGELGKPELE	76
RLA0_HALMA	-----MSAESERKTETIPWKQEEVDIVFMIESYESVGVVNIAGIPSRLODMRRDLHGT-AELRVSRNTLIERALDDVD-----DGLE	79
RLA0_HALVO	-----MSESEVRQTEVIPQWKREEVDLVDFIESYESVGVVGVAGIPSRLODMRRDLHGS-AAVRMSRNTLVNRALDEVN-----DGFE	79
RLA0_HALSA	-----MSAEQRTTEEVPEWKRQEVAEVDLLETYDSVGVVNVVTGIPSKLODMRRGLHGO-AALRMSRNTLLVRALEAG-----DGLD	79
RLA0_THEAC	-----MKEVSQOKKELVNEITRIKASRSVAIVDTAGIRTRQIODIRGKNRGK-INLKVIKKTLLFKALENLGD-----EKLS	72
RLA0_THEVO	-----MRKINPKKKEIVSELAQDITKSKAVAIVDIKGVTRMODIRAKNRDK-VKIKVVKKTLLFKALDSIND-----EKLT	72
RLA0_PICTO	-----MTEPAQWKIDFVKNLENEINSRKVAIAVISIKSLRNNFQKIRNSIRDK-ARIKVSRRALLRLAIENTGK-----NNIV	72
ruler	1.....10.....20.....30.....40.....50.....60.....70.....80.....90	

Scrabble Analogy: One Possibility



Scrabble Analogy: Thousands of Possibilities!



spanks
spanless
spanned
spanner
spanners
spanning
spanokopita
spanokopitas

spanceling
spancelled
spancelling
spancels
spandex
spandexes
spandrel
spandrels

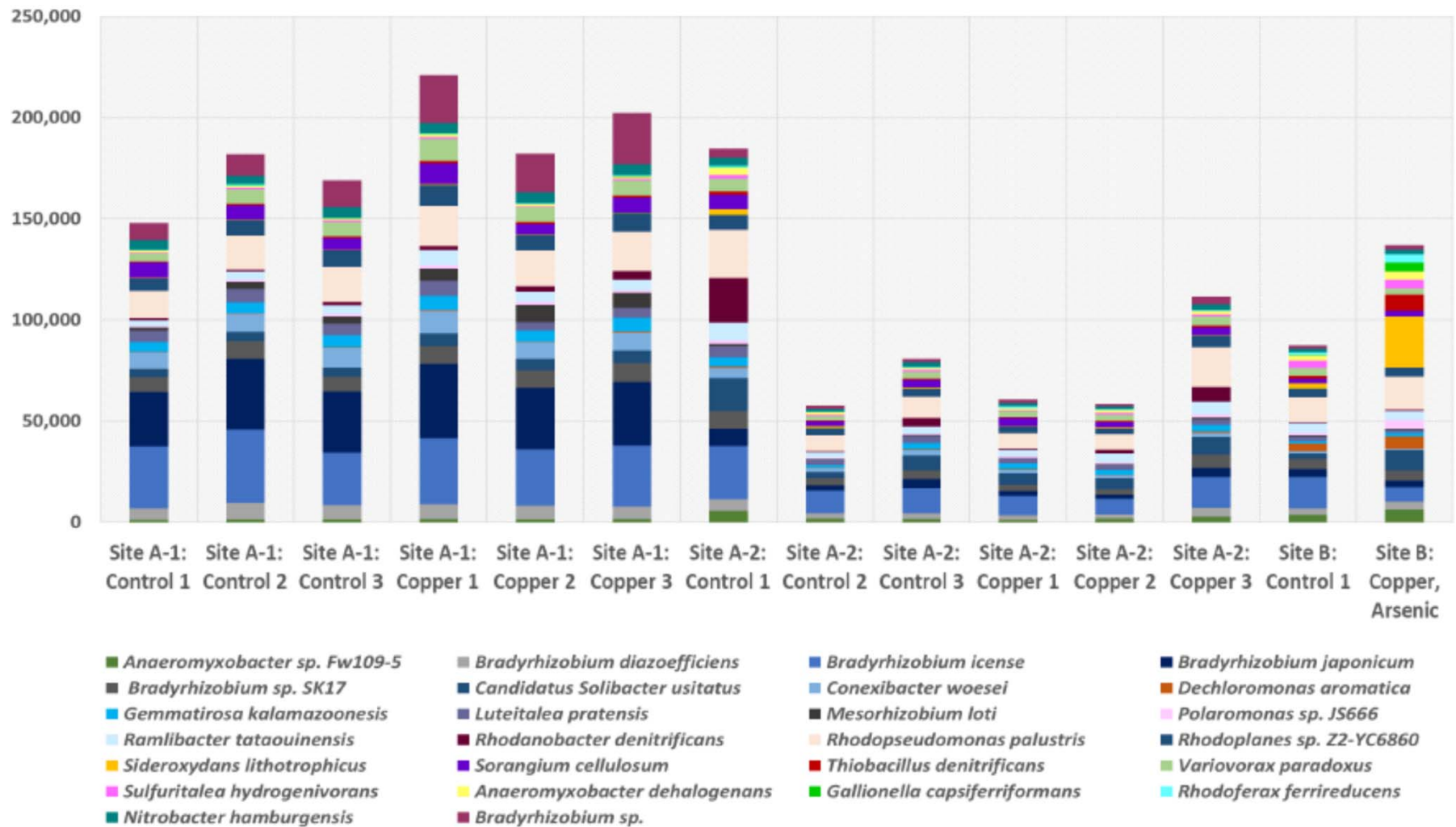
spaniel
spaniels
spank
spanked
spanker
spankers
spanking
spankings

spaes
spaetzle
spaetzles
spaghetti
spaghettilike
spaghetтини
spaghetтинis
spaghetтis

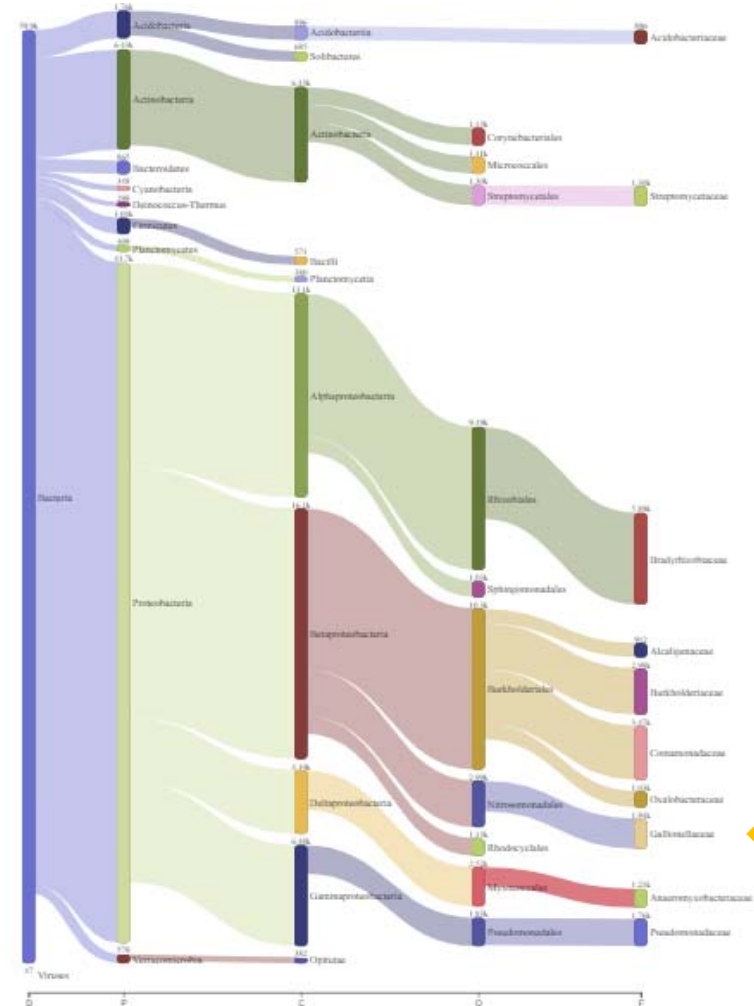
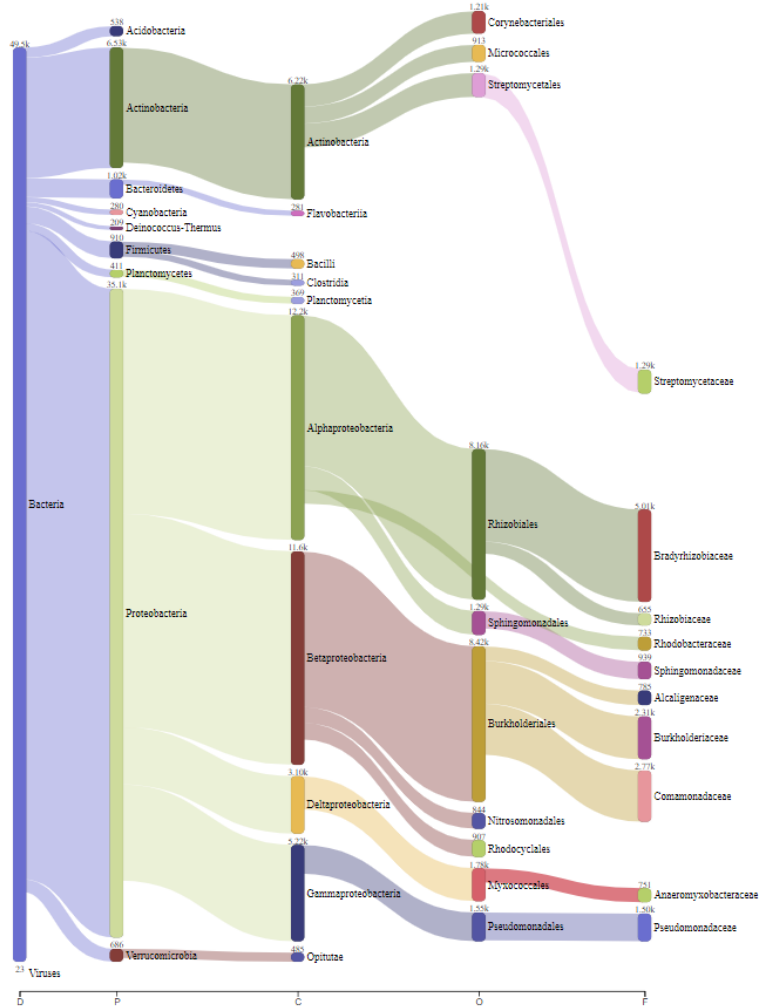
spandril
spandrils
spang
spangle
spangled
spangles
spangler
spangliest
spadework
spadeworks
spadices
spadille
spadilles
spading
spadix
spadixes

spado
spadones
spae
spaed
spaeing
spaeings
spaes
spaetzle
spagyric
spagyrics
spahee
spahees
spahi
spahis
spail
spails

Taxonomic Classification



Visualizing Taxonomies



Sideroxydans lithotrophicus

NCBI Resources ▾ How To ▾

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed ▾ Sideroxydans lithotrophicus arsenic
[Create alert](#) [Advanced](#)

Article types
Clinical Trial
Review
Customize ...

Text availability
Abstract
Free full text
Full text

Search results
Items: 0

No documents match your search terms

Limited publications are available for this species, especially with the words “arsenic” or “copper”

Sideroxydans lithotrophicus copper
[Create RSS](#) [Create alert](#) [Advanced](#)

[Send to ▾](#)

While this may not have been an obvious species ahead of time, it was successfully identified with an unbiased sequencing approach

Microb Ecol. 2017 Aug;74(2):264-277. doi: 10.1007/s00248-017-0950-x. Epub 2017 Feb 18.

Attached and Suspended Denitrifier Communities in Pristine Limestone Aquifers Harbor High Fractions of Potential Autotrophs Oxidizing Reduced Iron and Sulfur Compounds.

Herrmann M^{1,2}, Oplitz S³, Harzer R³, Totsche KU⁴, Küsel K^{3,5}.

Author information

Abstract

Oxygen and nitrate availability as well as the presence of suitable organic or inorganic electron donors are strong drivers of denitrification; however, the factors influencing denitrifier abundance and community composition in pristine aquifers are not well understood. We explored the denitrifier community structure of suspended and attached groundwater microorganisms in two superimposed limestone aquifer assemblages with contrasting oxygen regime in the Hainich Critical Zone Exploratory (Germany). Attached communities were retrieved from freshly crushed parent rock material which had been exposed for colonization in two groundwater wells (12.7 and 48 m depth). Quantitative PCR and amplicon pyrosequencing of nirK and nirS genes encoding copper-containing or cytochrome cd1 heme-type nitrite reductase, respectively, and of bacterial 16S ribosomal RNA genes showed a numerical predominance of nirS-type denitrifiers in both attached and suspended groundwater communities and a dominance of nirS-type denitrifiers closely related to the autotrophic thiosulfate- and hydrogen-oxidizing Sulfuritalea hydrogenivorans and the iron- and sulfide-oxidizing Sideroxydans lithotrophicus ES-1. Potential rates of nitrate reduction in association with exposed crushed rock material were higher with an inorganic electron donor (thiosulfate) compared to an organic electron donor (fumarate/acetate) in the upper aquifer assemblage but similar in the lower, oxic aquifer. Our results have clearly demonstrated that groundwater from pristine limestone aquifers harbors diverse denitrifier communities which appear to selectively attach to rock surfaces and harbor a high potential for nitrate reduction. Our findings suggest that the availability of suitable inorganic versus organic electron donors rather than oxygen availability shapes denitrifier communities and their potential activity in these limestone aquifers.

KEYWORDS: Denitrification; Groundwater; nirK; nirS

Sideroxydans lithotrophicus

Organism Overview

ID: 1862

Sideroxydans lithotrophicus

Sideroxydans lithotrophicus overview

Lineage: Bacteria[24883]; Proteobacteria[7898]; Betaproteobacteria[1124]; Nitrosomonadales[141]; Gallionellaceae[26]; Sideroxydans[2]; Sideroxydans lithotrophicus[1]

Sideroxydans lithotrophicus. *Sideroxydans lithotrophicus* is an iron-oxidizing freshwater bacterium that has been isolated from iron contaminated groundwater.

gene

978988..980019

/locus_tag="Slit_0997"

CDS

978988..980019

/locus_tag="Slit_0997"

/inference="protein motif:TFAM:TIGR00832"

/note="KEGG: lch:Lcho_3118 arsenical-resistance protein;

TIGRFAM: arsenical-resistance protein;

PFAM: Bile acid:sodium symporter"

/codon_start=1

/transl_table=11

/product="arsenical-resistance protein"

/protein_id="ADE11235.1"

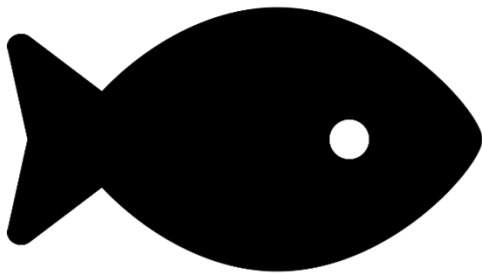
/db_xref="InterPro:IPR002657"

/db_xref="InterPro:IPR004706"

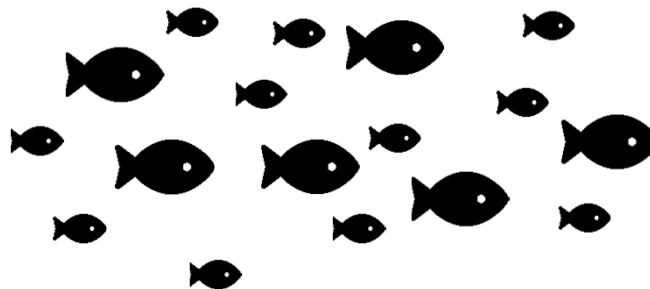
/translation="MSIFERYLTLWVFLCIIIGVVLGQTIPAPFHWLGGLEIAKVNLP
VGLLIWMIIPMLLRIDFGALHEVRKHWRGIGVTLFINWAVKPFSMALLGWLFIRHLF
AAYLPAEQLDSYIAGLVLLAAAPCTAMVFVWSRLVNGEPLFTLSQVALNDTIMVFFAFA
PLVALLGISSIVVPWNTLLVSVVLYIVLPVLLSQVWRRLLLARGEDALQRTLHALGP
ISIAALLTLVLLFAFQGAIVEQPLIILMLAVPITIQVYFTSGLAYWLNRRFGEAHC
VAGPSALIGASNFFELAVAAAISLFGFESGAALATVVGVLIEVPVMLSVVHIVKRSQG
WYQTGGAAN"

Its reference
genome contains
genes involved in
arsenic resistance

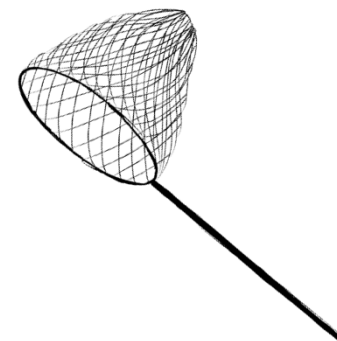
Complex Environmental Samples



Genome

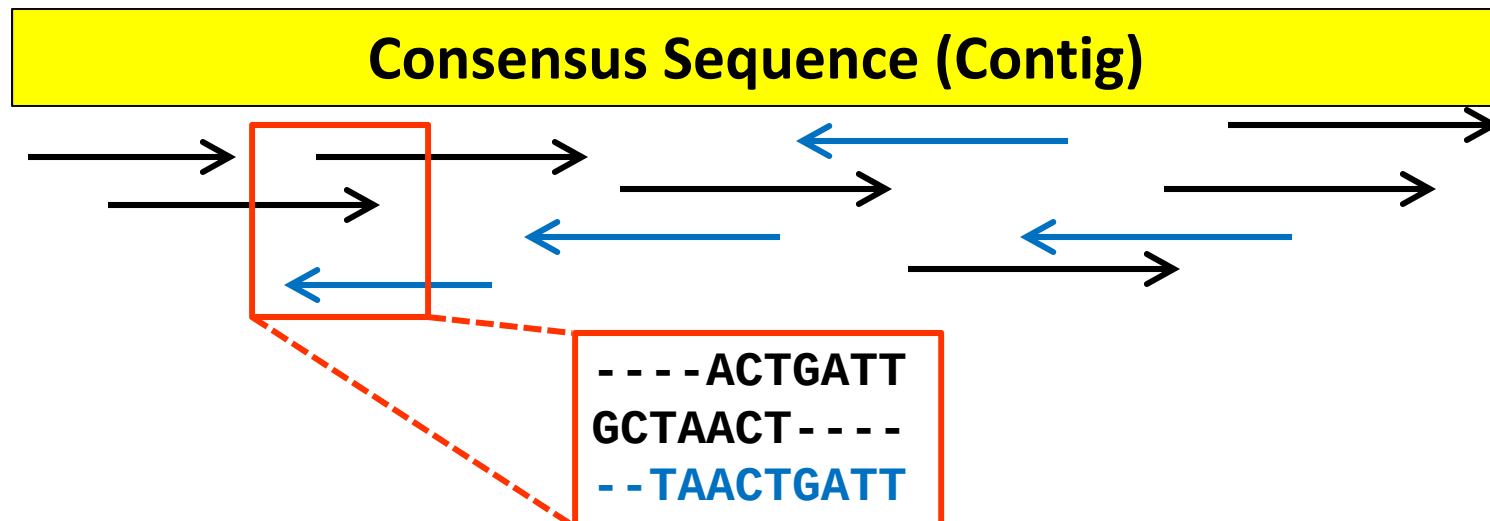


Genes



Metagenome Assembly

Sequence Name	@HWI-EAS367_0010:3:1:2380:6567#0/1
Sequence	TAATTTCCATTCATCATGACAGCCCTCCAGAGGTTAGACAAC
Sequence Info	+HWI-EAS367_0010:3:1:2380:6567#0/1
Quality in ASCII	GGG?GEG@GB?<EBE>C8E8ECCEE GDD>CC89AGD8BBA8
Sequence Name	@HWI-EAS367_0010:3:1:2585:6567#0/1
Sequence	ACAGTATTCTGGGGAGGATTAAATTAGATAAACATGCAAGAA
Sequence Info	+HWI-EAS367_0010:3:1:2585:6567#0/1
Quality in ASCII	EGGGGGGGFECG+CDADDGD4BAB9D<G+G?/?;7E,8CAGG



Functional Inference



COGs

Phylogenetic classification of proteins encoded in complete genomes



UniProt

UniProtKB reviewed:yes

BLAST Align Retrieve/ID mapping Peptide search

UniProtKB results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.
- Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

Filter by:

Reviewed (560,459) Swiss-Prot

Popular organisms

- Human (20,431)
- Mouse (17,019)
- A. thaliana (15,868)
- Rat (8,068)

Entry	Entry name	Protein names
☐ Q0ATK2	ACCD_MARMM	Acetyl-coenzyme A carboxylase carbo...
☐ O40976	1A_CMVNT	Replication protein 1a
☐ Q84P24	4CLL6_ARATH	4-coumarate--CoA ligase-like 6
☐ P62260	1433E_RAT	14-3-3 protein epsilon
☐ P0CK20	A25_VACCC	Protein A2.5

BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 14 2014, pages 2068-2069 doi:10.1093/bioinformatics/btu153

Genome analysis Advance Access publication March 18, 2014

Prokka: rapid prokaryotic genome annotation

Torsten Seemann^{1,2}

¹Victorian Bioinformatics Consortium, Monash University, Clayton 3800 and ²Life Sciences Computation Centre, Victorian Life Sciences Computation Initiative, Carlton 3053, Australia

Associate Editor: Alfonso Valencia

ABSTRACT
Summary: The multiplex capability and high yield of current day DNA-sequencing instruments has made bacterial whole genome sequencing a routine affair. The subsequent *de novo* assembly of reads into contigs has been well addressed. The final step of annotating all relevant genomic features on those contigs can be achieved slowly using existing web- and email-based systems, but these are not applicable for sensitive data or integrating into computational pipelines. Here we introduce Prokka, a command line software tool to fully annotate a draft bacterial genome in about 10 min on a typical desktop computer. It produces standards-compliant output files for further analysis or viewing in genome browsers.

Availability and implementation: Prokka is implemented in Perl and is freely available under an open source GPLv2 license from <http://vicbioinformatics.com/>.

Contact: torsten.seemann@monash.edu

Received on November 10, 2013; revised on February 6, 2014; accepted on March 13, 2014

2 DESCRIPTION

2.1 Input

Prokka expects preassembled genomic DNA sequences in FASTA format. Finished sequences without gaps are the ideal input, but it is expected that the typical input will be a set of scaffold sequences produced by *de novo* assembly software. This sequence file is the only mandatory parameter to the software.

2.2 Annotation

Prokka relies on external feature prediction tools to identify the coordinates of genomic features within contigs. These tools are listed in Table 1, and all of them, except for Prodigal, provide coordinates and appropriate labels to describe the feature.

Proteins coding genes are annotated in two stages. Prodigal identifies the coordinates of candidate genes, but does not describe the putative gene product. The traditional way to predict what a gene codes for is to compare it with a large database of known sequences, usually at a protein sequence level, and transfer the annotation of the best significant match.

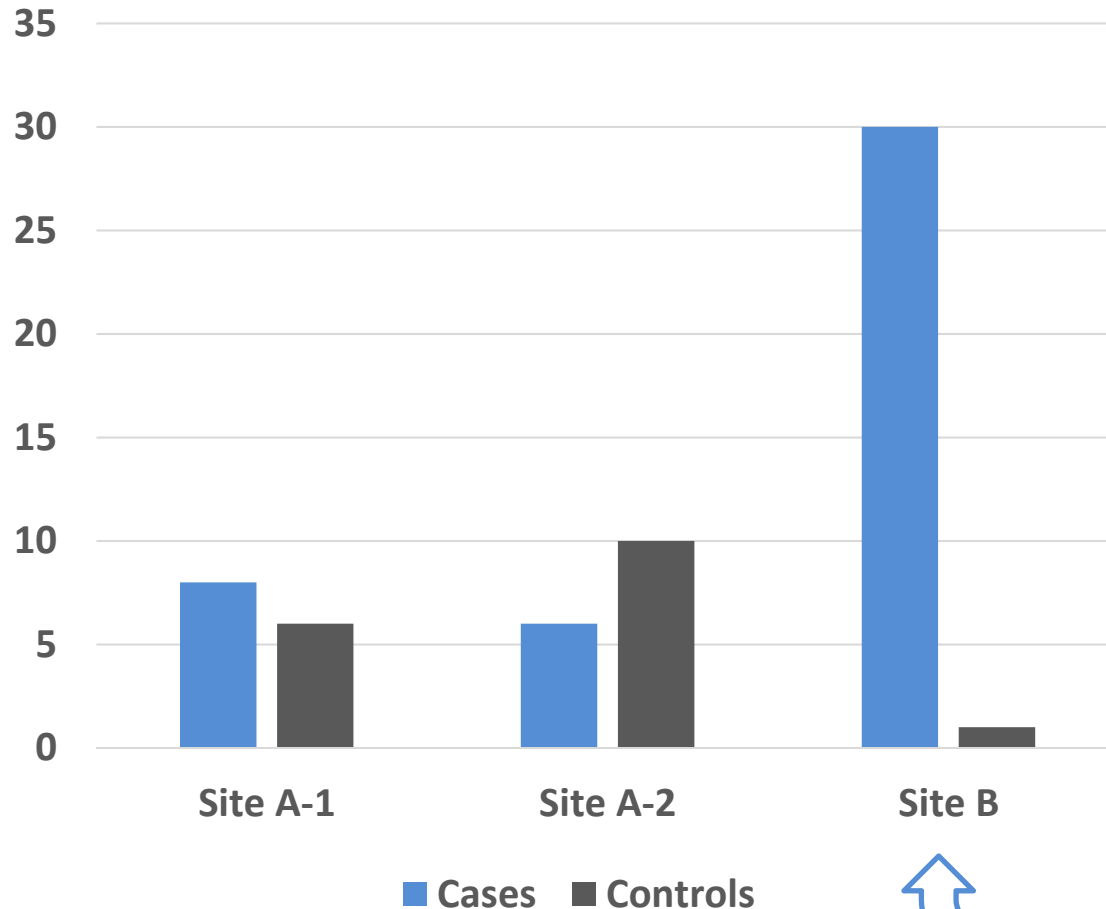
Prokka uses this method, but in a hierarchical manner, starting with a smaller trustworthy database, moving to medium-sized but domain-specific databases, and finally to curated models of protein families. By default, an e-value threshold of 10^{-6} is used with the following series of included databases:

- (1) An optional user-provided set of annotated proteins. These are expected to be trustworthy curated datasets and will be

1 INTRODUCTION

Genome annotation is the process of identifying and labeling all the relevant features on a genome sequence (Richardson and Watson, 2012). At minimum, this should include coordinates of predicted coding regions and their putative products, but it is desirable to go beyond this to non-coding RNAs, signal peptides and so on.

“Arsenic” in Predicted Gene Products

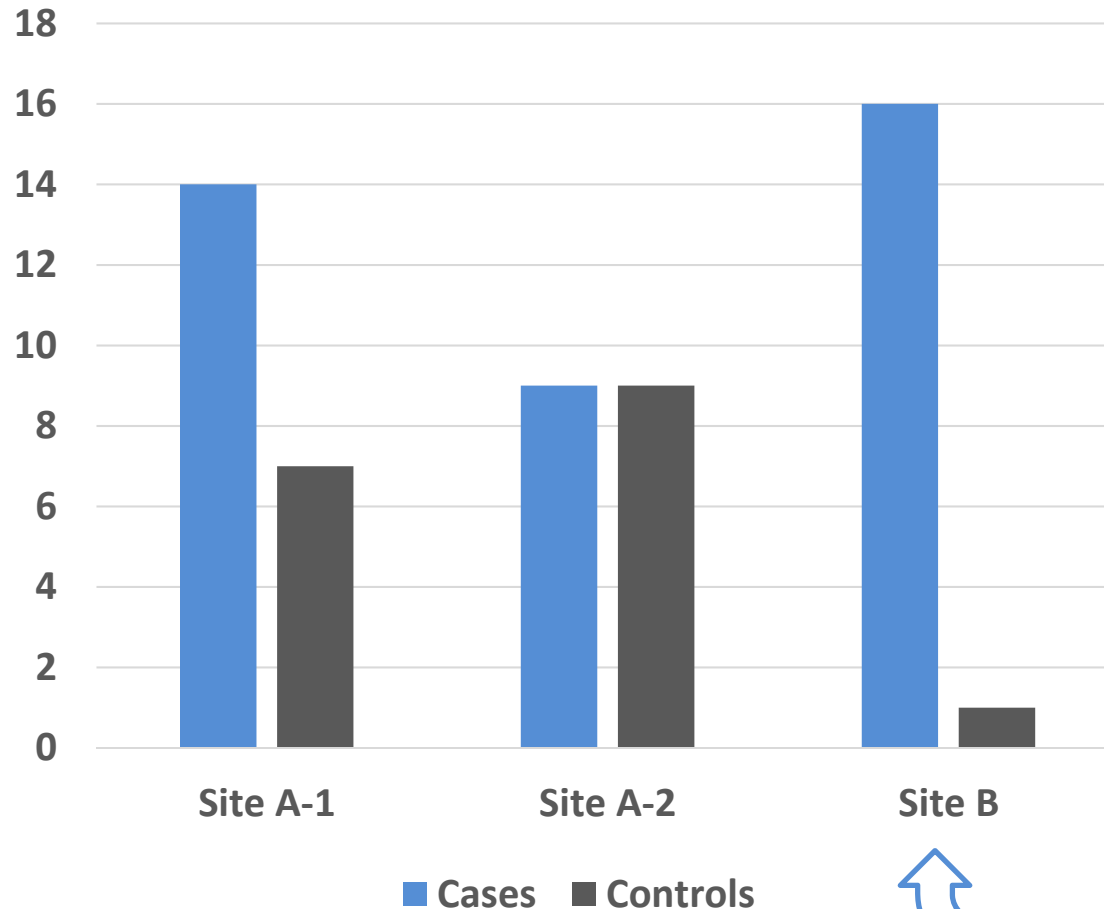


Site B

Arsenate reductase	7	0
Arsenical-resistance protein Acr3	6	0
Protein ArsC	4	0
Arsenical pump-driving ATPase	3	0
Arsenical resistance operon trans-acting repressor ArsD	3	0
Arsenite methyltransferase	3	0
Arsenic resistance transcriptional regulator ArsR1	2	0
Arsenite oxidase subunit AioA	1	0
Arsenite oxidase subunit AioB	1	0
Arsenical pump membrane protein	0	0
Glutaredoxin arsenate reductase	0	0
NADPH-dependent FMN reductase ArsH	0	1



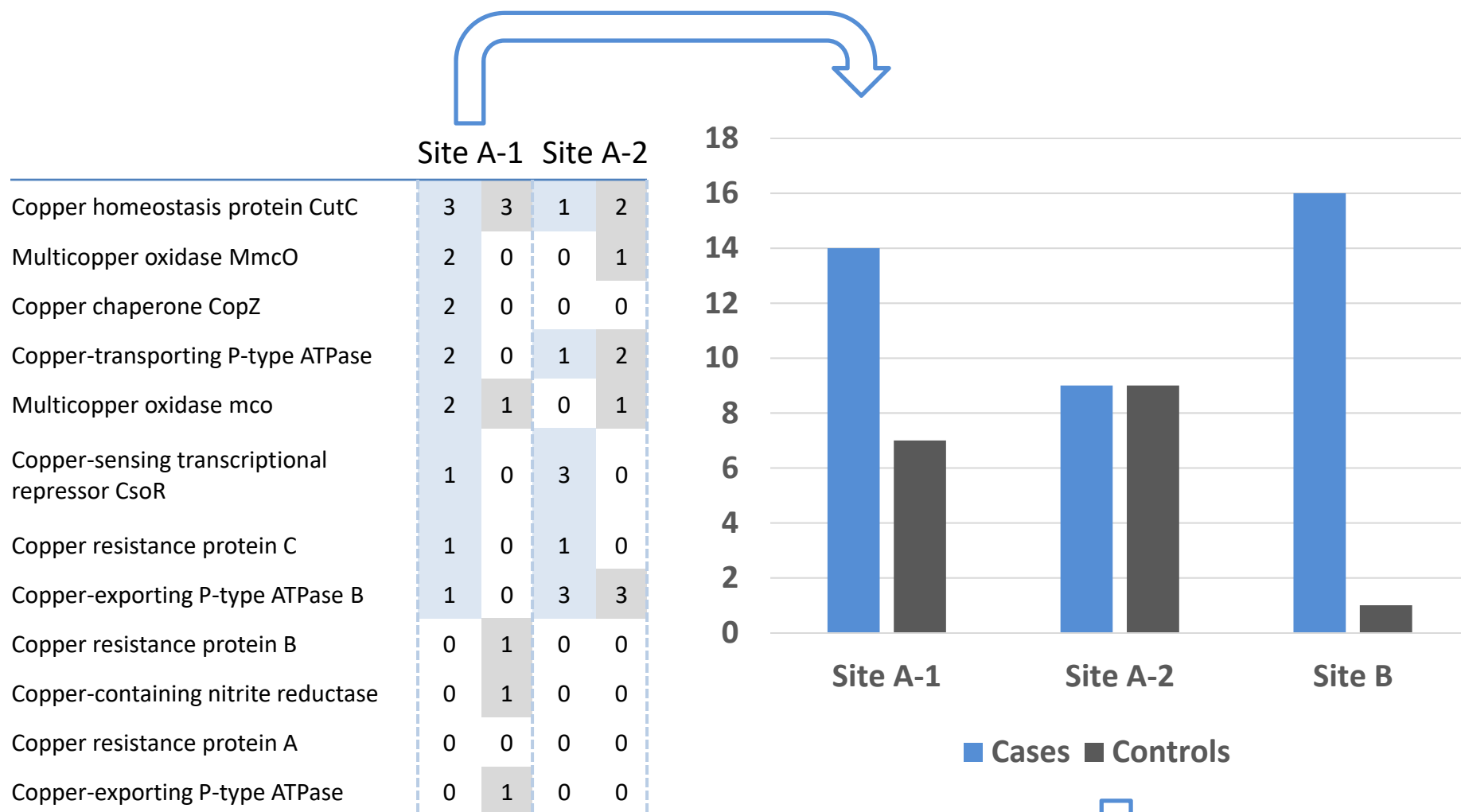
“Copper” in Predicted Gene Products

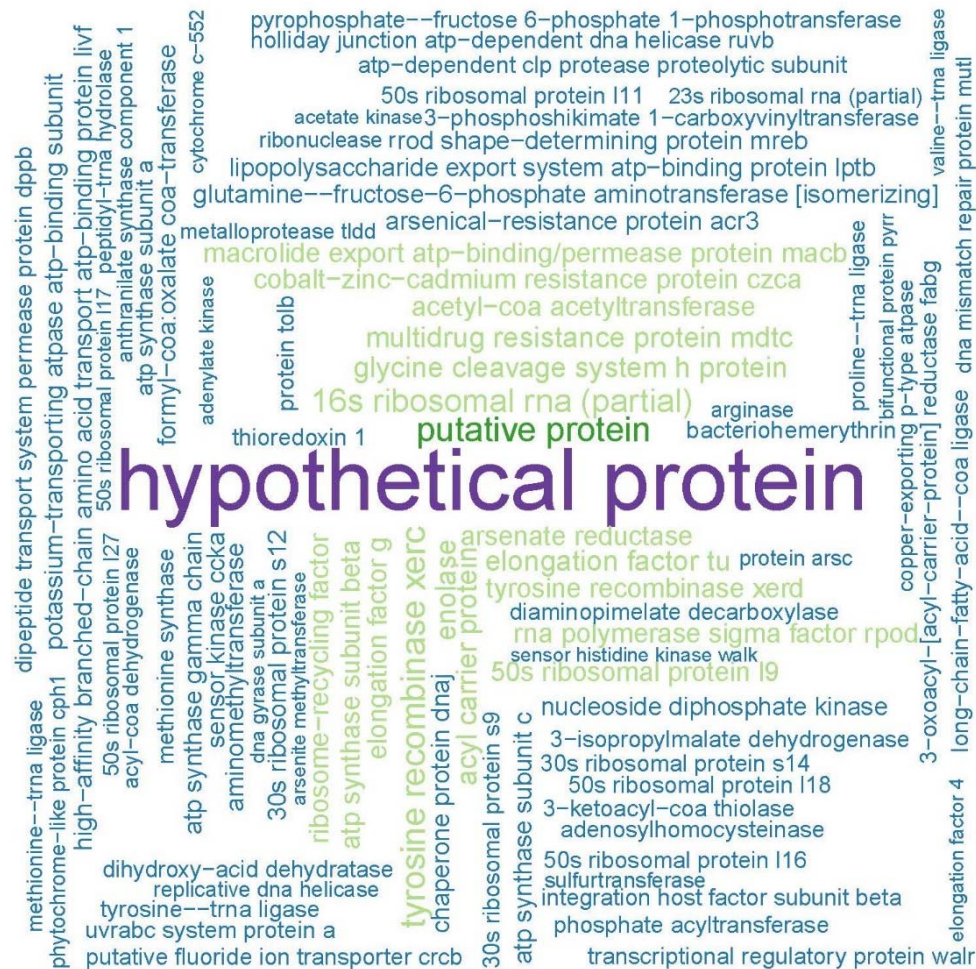


Site B

Copper-exporting P-type ATPase	4	1
Copper resistance protein A	3	0
Copper-exporting P-type ATPase B	3	0
Copper resistance protein B	2	0
Copper-transporting P-type ATPase	2	0
Copper chaperone CopZ	1	0
Multicopper oxidase MmcO	1	0
Multicopper oxidase mco	0	0
Copper-sensing transcriptional repressor CsoR	0	0
Copper-containing nitrite reductase	0	0
Copper homeostasis protein CutC	0	0
Copper resistance protein C	0	0

“Copper” in Predicted Gene Products





Predicting the functions of unknown proteins is an active area of research for my group.

Please contact me if we can help you!

**Please contact me if
we can help you!**

Krista Ternus, PhD

kternus@signaturescience.com

@KristaTernus

Questions?



signature
science[®] LLC