

# Lawrence Berkeley National Laboratory

## LBL Publications

### Title

Metagenomes and metatranscriptomes from boreal potential and actual acid sulfate soil materials

### Permalink

<https://escholarship.org/uc/item/5748c8zn>

### Journal

Scientific Data, 6(1)

### ISSN

2052-4463

### Authors

Högfors-Rönholm, Eva  
Lopez-Fernandez, Margarita  
Christel, Stephan  
et al.

### Publication Date

2019

### DOI

10.1038/s41597-019-0222-3

Peer reviewed

# SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

## Metagenomes and metatranscriptomes from boreal potential and actual acid sulfate soil materials

Eva Högfors-Rönholm<sup>1</sup>, Margarita Lopez-Fernandez<sup>2</sup>, Stephan Christel<sup>2</sup>, Diego Brambilla<sup>2</sup>, Marcel Huntemann<sup>3</sup>, Alicia Clum<sup>3</sup>, Brian Foster<sup>3</sup>, Bryce Foster<sup>3</sup>, Simon Roux<sup>3</sup>, Krishnaveni Palaniappan<sup>3</sup>, Neha Varghese<sup>3</sup>, Supratim Mukherjee<sup>3</sup>, T. B. K. Reddy<sup>3</sup>, Chris Daum<sup>3</sup>, Alex Copeland<sup>3</sup>, I-Min A. Chen<sup>3</sup>, Natalia N. Ivanova<sup>3</sup>, Nikos C. Kyrpides<sup>3</sup>, Miranda Harmon-Smith<sup>3</sup>, Emiley A. Eloë-Fadrosh<sup>3</sup>, Daniel Lundin<sup>2</sup>, Sten Engblom<sup>1</sup> & Mark Dopson<sup>2</sup>

Natural sulfide rich deposits are common in coastal areas worldwide, including along the Baltic Sea coast. When artificial drainage exposes these deposits to atmospheric oxygen, iron sulfide minerals in the soils are rapidly oxidized. This process turns the potential acid sulfate soils into actual acid sulfate soils and mobilizes large quantities of acidity and leachable toxic metals that cause severe environmental problems. It is known that acidophilic microorganisms living in acid sulfate soils catalyze iron sulfide mineral oxidation. However, only a few studies regarding these communities have been published. In this study, we sampled the oxidized actual acid sulfate soil, the transition zone where oxidation is actively taking place, and the deepest un-oxidized potential acid sulfate soil. Nucleic acids were extracted and 16S rRNA gene amplicons, metagenomes, and metatranscriptomes generated to gain a detailed insight into the communities and their activities. The project will be of great use to microbiologists, environmental biologists, geochemists, and geologists as there is hydrological and geochemical monitoring from the site stretching back for many years.

### Background & Summary

Naturally occurring deposits containing sulfidic sediments that form acid sulfate soils (ASS) cover over 17 million hectares of coastal areas. These extreme soils occur in North America (e.g. in the South and South West of the U.S) as well as in Europe (e.g. surrounding the Baltic Sea), Asia, and Australia<sup>1</sup>. Exposure of the sulfidic materials to air initializes chemical reactions that produce sulfuric acid<sup>2</sup>, thus creating ASS with a pH < 4, while mobilizing large quantities of acidity and leachable toxic metal(loid)s (e.g. Al, As, Cd, Co, Ni, and Zn). The leached metals and acidity are ultimately transported to the surrounding waters where they cause severe environmental problems<sup>3</sup>, negative economic consequences<sup>4</sup>, and impact human health<sup>5</sup>.

Pyrite oxidation is a complex biogeochemical process that involves a series of chemical reactions aided by microbiological catalysis<sup>1</sup>. Intermediate sulfur species formed during the oxidation process are metabolized by sulfur-oxidizing bacteria<sup>6</sup> while ferrous iron can be further oxidized at low pH by iron-oxidizing bacteria<sup>7</sup>. Ultimately, pyrite oxidation causes acidification and mobilization of trace metals that are leached to recipient waters. These processes have been extensively studied at the Risöfladan experimental field, Vaasa, Finland (Fig. 1) and there is a large amount of available geochemical and microbial 16S rRNA gene sequencing data<sup>8–11</sup>. The previous 16S rRNA gene based investigations from the oxidized zone at the site identified a mixed community of acidophilic bacteria and archaea similar to that found in acid mine drainage environments along with sulfate

<sup>1</sup>Research and Development, Novia University of Applied Sciences, Vaasa, 65200, Finland. <sup>2</sup>Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Linnaeus University, Kalmar, 59231, Sweden. <sup>3</sup>Department of Energy Joint Genome Institute, Walnut Creek, CA, 94598, USA. Correspondence and requests for materials should be addressed to E.H.-R. (email: [eva.hogfors-ronnholm@novia.fi](mailto:eva.hogfors-ronnholm@novia.fi))



**Fig. 1** Vertical soil profile of the transition from oxidized ASS to reduced PASS at the Risöfladan experimental field. Basic geochemical information; pH,  $\text{Fe}^{2+}$ , and total reduced sulfur (TRS); is provided for the oxidized (approximately 75–140 cm depth,  $n = 3$ ), transition (approximately 140–190 cm depth,  $n = 3$ ), and un-oxidized soil layers (approximately > 190 cm depth,  $n = 3$ ).

Sampling zone	Sampling depth (cm below ground)	Replicates	Sample alias	Available data	Ancillary data categories
Oxidized (OX)	75	3	OX-1 to OX-3	16S rRNA gene amplicons, metagenomes, metatranscriptomes	Chemical (pH, EC, redox) and geochemical, Fe- and S- speciation, multi-elements
Transition (TR)	140	3	TR-1 to TR-3	16S rRNA gene amplicons, metagenomes, metatranscriptomes	Chemical (pH, EC, redox) and geochemical, Fe- and S- speciation, multi-elements
Un-oxidized (UN)	190	3	UN-1 to UN-3	16S rRNA gene amplicons, metagenomes, metatranscriptomes	Chemical (pH, EC, redox) and geochemical, Fe- and S- speciation, multi-elements

**Table 1.** Summary of soil samples and available data.

reducing bacteria in the underlying un-oxidized potential acid sulfate soil (PASS) zone<sup>10,11</sup>. Finally, 16S rRNA genes with similarity to low temperature adapted microbes were identified in all soil layers, reflecting the boreal environment in Finland. However, more detailed investigations into the metabolic potential and activities of the contrasting microbial communities in oxidized ASS, the transition zone containing the oxidative front, and the PASS have not been reported.

In this Data Descriptor, we present triplicate biological replicate data for 16S rRNA gene amplicons, metagenomes, and metatranscriptomes from the oxidized, transition, and un-oxidized soils horizons at the Risöfladan experimental field (total nine samples) along with the accompanying geochemical metadata (Table 1). The 16S rRNA gene amplicon sequencing dataset contains on average 369 027 reads (min 83 731, max 634 072) for the nine samples (Table 2). In addition, the nine metagenomes (three biological replicates from the three soil types designated OX-MG, TR-MG, and UN-MG) contained  $3.78 \times 10^9$  paired-end reads ( $2 \times 151$  bp) of raw sequence data (Table 3). Finally, the corresponding nine metatranscriptomes (designated OX-MT, TR-MT, and UN-MT) to the three soil types generated  $1.11 \times 10^{11}$  paired-end reads ( $2 \times 151$  bp) of raw sequence data (Table 3).

The broader goals that motivated the study were to take advantage of multi-omics and geochemical data to gain insights into the metabolic landscape and the molecular mechanisms underlying microbial life in this extreme environment. In addition, based on 16S rRNA gene sequencing many of the microbes in the PASS and ASS are unknown and the data will characterize the structure and function of these microbial populations. Moreover, the microbial metabolic pathways in the actual ASS will be used to answer the question how they catalyze the oxidation process.

Sample	Sequencing Data	Minimum	Mean	Median	Maximum
OX	Raw reads	251 850	318 098	262 356	440 089
TR	Raw reads	83 731	335 070	354 649	566 830
UN	Raw reads	361 455	453 915	366 218	634 072

**Table 2.** Summary of 16S rRNA gene amplicon sequencing of the three soil zones.

Sample	Data	Minimum	Mean	Median	Maximum
OX-MG	Paired-end reads	$3.24 \times 10^8$	$3.70 \times 10^8$	$3.89 \times 10^8$	$3.95 \times 10^8$
	Total bases	$4.90 \times 10^{10}$	$5.58 \times 10^{10}$	$5.88 \times 10^{10}$	$5.97 \times 10^{10}$
TR-MG	Paired-end reads	$3.00 \times 10^8$	$3.82 \times 10^8$	$3.65 \times 10^8$	$4.83 \times 10^8$
	Total bases	$4.53 \times 10^{10}$	$5.78 \times 10^{10}$	$5.51 \times 10^{10}$	$7.29 \times 10^{10}$
UN-MG	Paired-end reads	$4.85 \times 10^8$	$5.06 \times 10^8$	$5.02 \times 10^8$	$5.33 \times 10^8$
	Total bases	$7.32 \times 10^{10}$	$7.65 \times 10^{10}$	$7.57 \times 10^{10}$	$8.04 \times 10^{10}$
OX-MT	Paired-end reads	$1.09 \times 10^8$	$1.43 \times 10^8$	$1.17 \times 10^8$	$2.04 \times 10^8$
	Total bases	$1.65 \times 10^{10}$	$2.16 \times 10^{10}$	$1.76 \times 10^{10}$	$3.08 \times 10^{10}$
TR-MT	Paired-end reads	$9.41 \times 10^7$	$9.73 \times 10^7$	$9.70 \times 10^7$	$1.01 \times 10^8$
	Total bases	$1.42 \times 10^{10}$	$1.47 \times 10^{10}$	$1.46 \times 10^{10}$	$1.52 \times 10^{10}$
UN-MT	Paired-end reads	$1.00 \times 10^8$	$1.31 \times 10^8$	$1.12 \times 10^8$	$1.81 \times 10^8$
	Total bases	$1.51 \times 10^{10}$	$1.98 \times 10^{10}$	$1.69 \times 10^{10}$	$2.73 \times 10^{10}$

**Table 3.** Summary of the metagenome (MG) and metatranscriptomes (MT) sequencing of the three soil zones.

The project will be of great use to the research community, as it will generate biological data for which there is hydrological and geochemical monitoring of the site stretching back for many years. Since ASS are extreme environments, the project will be of direct use to researchers, e.g. microbiologists, environmental biologists, geochemists, and geologists. Scientists interested in the iron and sulfur cycles as well as the contaminants released from the ASS during the oxidation process will additionally have a direct use of the project. As sulfidic sediments are affected during dredging, building of infrastructure, and agriculture; researchers involved in these areas will also have use of this work as it may give an explanation in how to minimize the negative impact of these sediments. Finally, the study is relevant to bioinformaticians that study large data sets.

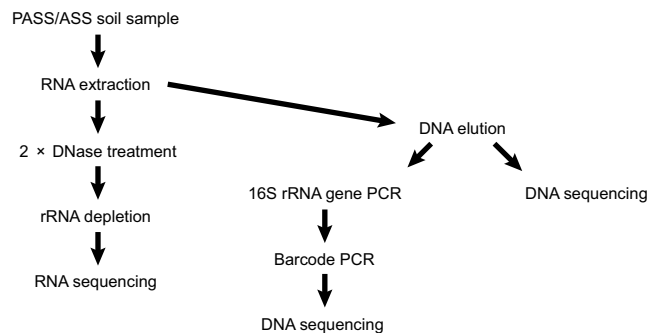
## Methods

**Soil sampling.** Soil was sampled in mid-August 2017 from the Risöfladan experimental field located in Vaasa, Finland ( $63^\circ 02' 50.22''N$ ,  $21^\circ 42' 41.85''E$ ; Figs 1 and 2). Ten replicate soil samples of 5 g were taken from 75 cm below ground (oxidized zone, OX), 140 cm below ground (transition zone, TR), and 190 cm below ground (un-oxidized zone, UN). The samples were placed in separate sterile tubes (30 tubes in total), immediately preserved by freezing in liquid nitrogen, and stored at  $-80^\circ\text{C}$  until RNA and DNA were extracted within ten days of sampling.

**RNA and DNA extraction.** RNA and DNA were extracted simultaneously from 3 g of soil using the RNeasy<sup>®</sup> PowerSoil Total RNA Kit (Qiagen) and the following RNeasy<sup>®</sup> PowerSoil DNA elution Kit (Qiagen). Briefly, 5 g of frozen soil was thawed and carefully mixed before 3 g of soil was transferred to a Bead Tube provided in the RNeasy<sup>®</sup> PowerSoil Total RNA Kit and RNA was extracted according to the manufacturer's instructions. The final RNA pellet was suspended in 25  $\mu\text{L}$  RNase/DNase-free water. After eluting the RNA from the RNA Capture Column provided by the RNeasy<sup>®</sup> PowerSoil Total RNA Kit, the bound DNA was eluted from the RNA Capture Column using the RNeasy<sup>®</sup> PowerSoil DNA Elution Kit according to the manufacturer's instructions. The final DNA pellet was suspended in 50  $\mu\text{L}$  RNase/DNase-Free water.

After extraction, RNA samples were treated twice with RNase-free DNase using the Turbo DNA-free<sup>™</sup> Kit (Invitrogen) according to the manufacturer's instructions and then stored at  $-80^\circ\text{C}$ . In order to obtain sufficient samples for the final three replicates per soil zone, two to three of the extracted RNA and DNA samples were pooled. The same replicate extractions were pooled for RNA and DNA in order for the metatranscriptomes to match the metagenomes and the RNA transcripts to be mapped to the metagenomes. Ribosomal RNA was depleted from the pooled RNA samples with the Ribominus<sup>™</sup> Transcriptome Isolation Kit (Invitrogen) according to the manufacturer's instructions. Nucleic acid concentrations were measured using a Qubit<sup>®</sup> 2.0 Fluorometer (Life Technologies). DNA and RNA samples were stored at  $-80^\circ\text{C}$  before being submitted to the U.S. Department of Energy Joint Genome Institute (JGI) for sequencing.

**16S rRNA gene amplicon sequencing.** Subsamples from the DNA extracts were used for amplifying the V3-V4 region of the microbial 16S rRNA gene by using primers 341F (CCTACGGGNGGCWGCAG) and 805R (GACTACHVGGGTATCTAATCC)<sup>12</sup>, followed by PCR amplification for Illumina sequencing<sup>13</sup> (Fig. 2). Samples were sequenced at the Science for Life Laboratory, Sweden, on the Illumina MiSeq platform<sup>14</sup>. A summary of the 16S rRNA gene amplicon sequencing reads is shown in Table 2.



**Fig. 2** Workflow from potential acid sulfate soil (PASS)/acid sulfate soil (ASS) sampling to deposited sequencing data.

Sample	NCBI BioProject	NCBI SRA accession	
	16S rRNA	MG	MT
OX-1	PRJNA524144	SRP185082	SRP185222
OX-2	PRJNA524144	SRP185084	SRP185094
OX-3	PRJNA524144	SRP185086	SRP185093
TR-1	PRJNA524144	SRP185087	SRP185225
TR-2	PRJNA524144	SRP185088	SRP185224
TR-3	PRJNA524144	SRP185089	SRP185226
UN-1	PRJNA524144	SRP185092	SRP185227
UN-2	PRJNA524144	SRP185091	SRP185228
UN-3	PRJNA524144	SRP185223	SRP185229

**Table 4.** Summary of the 16S rRNA gene, metagenome, and metatranscriptome data records.

**Metagenome and metatranscriptome library construction and sequencing.** For metagenomes, 100 ng of DNA was sheared to 300 bp using the Covaris LE220 and size selected using SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA Biosystems). The prepared libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Sequencing of the flowcell was performed on the Illumina NovaSeq sequencer using NovaSeq XP V1 reagent kits, tbd-sample dependant flowcell, following a tbd-sample dependant indexed run recipe.

For metatranscriptomes, stranded cDNA libraries were generated using the Illumina Truseq Stranded RNA LT kit. Ten ng of total RNA was fragmented using divalent cations and high temperature. The fragmented RNA was reversed transcribed using random hexamers and SSII (Invitrogen) followed by second strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation, and eight cycles of PCR. The prepared libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Sequencing of the flowcell was performed on the Illumina NovaSeq sequencer using NovaSeq XP V1 reagent kits, tbd-sample dependent flowcell, following a tbd-sample dependant indexed run recipe.

The generated reads and bases of all samples can be seen in Table 3.

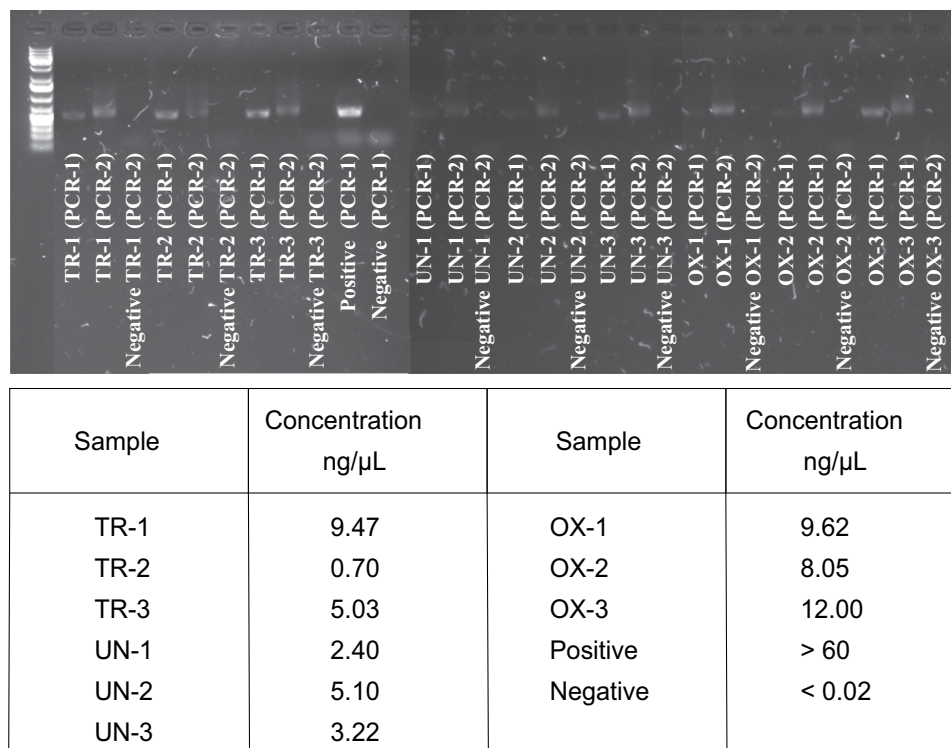
### Data Records

A summary of the data records included in the study is given in Table 4. The raw Illumina MiSeq 16S rRNA gene sequencing reads are available from the NCBI Sequence Read Archive under the NCBI BioProject accession number PRJNA524144<sup>15</sup>. The raw Illumina NovaSeq sequencing reads for all metagenomes are available from the NCBI Sequence Read Archive under accession numbers SRP185082<sup>16</sup>, SRP185084<sup>17</sup>, SRP185086<sup>18</sup>, SRP185087<sup>19</sup>, SRP185088<sup>20</sup>, SRP185089<sup>21</sup>, SRP185092<sup>22</sup>, SRP185091<sup>23</sup> and SRP185223<sup>24</sup>. The raw Illumina NovaSeq sequencing reads for all metatranscriptomes are also available from the NCBI Sequence Read Archive under accession numbers SRP185222<sup>25</sup>, SRP185094<sup>26</sup>, SRP185093<sup>27</sup>, SRP185225<sup>28</sup>, SRP185224<sup>29</sup>, SRP185226<sup>30</sup>, SRP185227<sup>31</sup>, SRP185228<sup>32</sup> and SRP185229<sup>33</sup>.

### Technical Validation

Soil samples were taken aseptically by using sterilized equipment and sterile RNase and DNase-free tubes. RNA and DNA were extracted in an RNase free environment and the quantity of the extracted nucleic acids were measured between every step with a Qubit<sup>®</sup> 2.0 Fluorometer. The quality of the extracted nucleic acids was analyzed using both a NanoDrop<sup>™</sup> 2000 Spectrophotometer and agarose gel electrophoresis using a 1% agarose gel with 0.01% SYBR<sup>™</sup> Safe DNA Gel Stain (Invitrogen) and a Thermo Scientific<sup>™</sup> GeneRuler<sup>™</sup> Ready-to-use 1 kb Plus DNA Ladder.





**Fig. 3** Agarose gel electrophoresis and Qubit® 2.0 Fluorometer DNA concentrations of the V3-V4 region of 16S rRNA gene amplicons from the triplicate samples of the three soil zones. Thermo Scientific™ GeneRuler™ Ready-to-use 1 kb Plus DNA Ladder was used in the gel electrophoresis and can be seen to the left side of the gel. TR-1 to TR-3 are triplicate samples from the transition zone, UN-1 to UN-3 are triplicate samples from the un-oxidized zone, and OX-1 to OX-3 are triplicate samples from the oxidized zone. PCR-1 stands for templates after the first amplification step by using primers 341F and 805R. The positive (DNA from *Pseudoaeromonas citrea*) and negative (RNase/DNase-free water) controls used in the first amplification are named “Positive” and “Negative” in the figure. PCR-2 stands for templates after the second amplification step that attaches the barcodes identifying the sequences in the MiSeq Illumina sequencing. Negative PCR-2 stands for negative controls in the second amplifications step, i.e. mastermix with primer-barcodes without templates.

For 16S rRNA gene amplification, negative and positive controls in the form of RNase/DNase-free water and DNA extracted from *Pseudoaeromonas citrea*, respectively were included in the first amplification step. In addition, negative controls in the form of mastermix with primer-barcodes without templates were included in the second PCR amplification to confirm that both the primers and the barcodes identifying the sequences functioned correctly and that no contamination occurred. The concentration of the 16S rRNA gene amplicons and controls was measured with a Qubit® 2.0 Fluorometer and their quality were analyzed using agarose gel electrophoresis (Fig. 3).

## References

- Karimian, N., Johnston, S. G. & Burton, E. D. Iron and sulfur cycling in acid sulfate soil wetlands under dynamic redox conditions: A review. *Chemosphere* **197**, 803–816 (2018).
- Schippers, A. & Sand, W. Bacterial leaching of metal sulfides proceeds by two indirect mechanisms via thiosulfate or via polysulfides and sulfur. *Appl. Environ. Microbiol.* **65**, 319–321 (1999).
- Sutela, T. & Vehanen, T. The effects of acidity and aluminium leached from acid-sulphate soils on riverine fish assemblages. *Boreal Env. Res.* **22**, 385–391 (2017).
- Macdonald, B. C. T., Denmead, O. T., White, I. & Melville, M. D. Natural sulfur dioxide emissions from sulfuric soils. *Atmos. Environ.* **38**, 1473–1480 (2004).
- Hinwood, A., Horwitz, P. & Rogan, R. Human exposure to metals in groundwater affected by acid sulfate soil disturbance. *Arch. Environ. Contam. Toxicol.* **55**, 538–545 (2008).
- Dopson, M. & Johnson, D. B. Biodiversity, metabolism and applications of acidophilic sulfur- metabolizing microorganisms. *Environ. Microbiol.* **14**, 2620–2631 (2012).
- Bonnefoy, V. & Holmes, D. S. Genomic insights into microbial oxidation and iron homeostasis in extremely acidic environments. *Environ. Microbiol.* **14**, 1597–1611 (2012).
- Åström, M., Österholm, P., Bärlund, I. & Tattari, S. Hydrochemical effects of surface liming, controlled drainage and lime-filter drainage on boreal acid sulfate soils. *Water Air Soil Pollut.* **179**, 107–116 (2007).
- Boman, A., Fröjdö, S., Backlund, K. & Åström, M. E. Impact of isostatic land uplift and artificial drainage on oxidation of brackish-water sediments rich in metastable iron sulfide. *Geochim. Cosmochim. Acta* **74**, 1268–1281 (2010).
- Wu, X. *et al.* Impact of mitigation strategies on microbial community from an Ostrobothnian acid sulfate soil. *Sci. Tot. Environ.* **526**, 215–221 (2015).

11. Wu, X. *et al.* Microbial community potentially responsible for acid and metal release from an Ostrobothnian acid sulfate soil. *FEMS Microbiol. Ecol.* **84**, 555–563 (2013).
12. Herlemann, D. P. *et al.* Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**, 1571–1579 (2011).
13. Hugerth, L. W. *et al.* DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Appl. Environ. Microbiol.* **80**, 5116–5123 (2014).
14. Lindh, M. V. *et al.* Transplant experiments uncover Baltic Sea basin-specific responses in bacterioplankton community composition and metabolic activities. *Front. Microbiol.* **6**, 223 (2015).
15. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP187263> (2019).
16. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185082> (2019).
17. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185084> (2019).
18. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185086> (2019).
19. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185087> (2019).
20. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185088> (2019).
21. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185089> (2019).
22. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185092> (2019).
23. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185091> (2019).
24. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185223> (2019).
25. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185222> (2019).
26. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185094> (2019).
27. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185093> (2019).
28. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185225> (2019).
29. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185224> (2019).
30. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185226> (2019).
31. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185227> (2019).
32. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185228> (2019).
33. *Sequence Read Archive*, <https://identifiers.org/ncbi/insdc.sra:SRP185229> (2019).

## Acknowledgements

(A portion of) this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). E.H.-R. acknowledges the Nessling Foundation (grant number 201700273 and 201800502) and Svensk-Österbottanska Samfundet r.f. (grant number 18/0604) for financial support. M.D. acknowledges The Swedish Research Council Formas (grant number 2018-00760) and the Geological Survey of Sweden (grant number 36-1878/2017) for financial support. M.D. further acknowledges the Science for Life Laboratory (SciLifeLab), the National Genomics Infrastructure (NGI) for complementary computations that were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project b2013127. D.L. and D.B. acknowledges the Swedish Research Council infrastructure project Biodiversity Atlas Sweden (VR, grant number 2017-00688) and the marine strategic research environment EcoChange (Formas) for financial support.

## Author Contributions

E.H.-R., S.E. and M.D. conceived the study and wrote the manuscript with contributions from all authors. E.H.-R. and S.E. performed the sampling, E.H.-R. extracted the biomolecules together with M.L-F. and S.C. and M.L-F. amplified the 16S rRNA gene for amplicon sequencing. S.C., D.B. and D.L. carried out bioinformatic analysis of the data sets. M.H., A.C., Bri.F., Bry.F., S.R., K.P., N.V., S.M., T.B.K.R., C.D., A.C., I.-M.C., N.N.I., N.C.K., M.H.-S. and E.A.E.-F. carried out the sequencing and bioinformatic analysis of the metagenomic and metatranscriptomic data sets.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019