Sulfide production and oxidation by heterotrophic bacteria under aerobic conditions

Supplementary methods

H_2S production test in MM1 medium

A mineral medium (MM1) was also used with 0.2% of glucose for *E. coli* and *Pseudomonas aeruginosa* PAO1 or 0.2% of monosodium glutamate for *Cupriavidus pinatubonensis* JMP134. MM1: Na_2HPO_4, 0.15%；K_2HPO_4, 0.05%；$(NH4)_2SO_4$, 0.03%; $CaCl_2$, 0.3 mM; $MgCl_2$, 1 mM; 1% of the trace mineral solution ($FeSO_4.7H_2O$, 0.45%; $ZnSO_4.7H_2O$, 0.144%; $MnSO_4.H_2O$, 0.086%; $CuSO_4$, 0.016%; $CoSO_4.7H_2O$, 0.028%; H_3BO_3, 0.006%; and H_2SO_4, 1%).

Cloning and gene knockout

The genomic DNA of *C. pinatubonensis JMP134*, *P. aeruginosa* PAO1, *B. cereus* ATCC10876 , *S. aureus* ATCC 6538P and *Z. profunda* SM-A87 were extracted by using the Omega Bio-Tek DNA Purification Kit (Omega Bio-Tek, Norcross, USA). All primers used for cloning and gene knockout were listed in supplementary Table 2.

The *pdo* gene fragment (gene ID: 882596) and *sqr2* gene fragment (gene ID: 879884) from *P. aeruginosa* PAO1 were PCR-amplified with primer pairs of PaF1-PaR1 and PaF2-PaR2, respectively. The linearized pBBR1MCS5 plasmid backbone was PCR-amplified with Pm5F1-Pm5R1 primer pair. The three DNA fragments were gel-purified and fused together with an In-fusion kit (Takara, Japan) to generate pBBR1MCS5-Pa*pdo*-Pa*sqr2*. The plasmid pBBR1MCS5-Pa*pdo*-Pa*sqr2* were used as template to get linear

fragment pBBR1MCS5-Pa*sqr2* by PCR amplification with primer pairs of Pm5F2 and Pm5R2. The *pdo* gene fragments from *B. cereus* ATCC10876 (gene ID: BCERE0002_RS16755), *S. aureus* ATCC 6538P (gene ID: 3919304), *Z. profunda* SM-A87 (gene ID: ZPR_RS08580) were PCR-amplified with BcF1 - BcR1, SaF1 - SaR1 and ZpF1 - ZpR1, respectively. The gene fragment and the plasmid fragment were assembled together with the In-fusion kit to produce pBBR1MCS-5::Bcpdo-Pasqr$_2$, pBBR1MCS-5::Sapdo-Pasqr2 and pBBR1MCS-5::Zppdo-Pasqr2, respectively. *E. coli* BL21(DE3) was used as the host. The plasmids pBBR1MCS-5::Pa*sqr*2 and pBBR1MCS-2::Cp*sqr* were constructed with pairs of primers DpF1 and DpR1 and DpF2 and DpR2, respectively by using a site-directed mutagenesis method from pBBR1MCS-5::Bc*pdo*-Pa*sqr*2 or pBBR1MCS-5::Cp*pdo*-Cp*sqr* (Xia et al 2014). Primer pair of DsF1 and DsR1 was used to construct pBBR1MCS-5::Zppdo, pBBR1MCS-5::Sapdo and pBBR1MCS-5::Bcpdo with a similar way. Primer pair of DsF2 and DsR2 was used to construct pBBR1MCS-2::Cp*pdo*. All clones were sequenced to verify that no mutations were introduced during PCR and cloning.

The vector pK18mobsacB$_{tet}$ was used to generate in frame deletions (Harighi 2009). For inactivating Pa*sqr*1, one-kb regions immediately upstream and downstream of Pa*sqr1* were amplified via PCR with primer pairs of Pu11 and Pu12 and primer pairs of Pd11 and Pd12, respectively. The primers were designed to contain 20 bp overlaps for linking the two fragments and for cloning into pK18mobsacB$_{tet}$ at the EcoRI-cut site by using the In-fusion kit to produce pK18mobsacB$_{tet}$-ΔPa*sqr*$_1$ in *E. coli* BL21(DE3). Using the same approach, pK18mobsacB$_{tet}$-ΔPa*sqr*$_2$ for inactivating Pa*sqr*$_2$ and pK18mobsacB$_{tet}$-ΔPa*pdo* for inactivating Pa*pdo* were constructed. The constructs were used to generate in-frame

deletion mutants in *P. aeruginosa* PaO1 via triparental mating essentially the same as reported (Harighi 2009). Briefly, the cells were grown in LB at 37 ^{o}C and harvested when OD_{600nm} reached 0.5 and washed with PBS once. Then the cells were mixed 3:2:1 (*E. coli* donor (6 ml): assistant strain (4 ml): recipient cells (2ml)), centrifuged and re-suspended in 100 μl of LB medium. All the cells were spot on LB agar overnight at 37 ^{o}C. Then the overnight cells were collected, re-suspended in 200 μl of M9 medium, and plated on M9 agar with citrate as the sole carbon source and tetracycline for selection. Since *E. coli* did not grow on citrate, the colonies were *P. aeruginosa* with the plasmid inserted via homologous recombination. Single colony was inoculated into LB without tetracycline at 37^{o}C overnight. The culture was diluted and spread on LB plates supplemented with 12% sucrose to select mutants that have lost the inserted plasmid. Finally, the correct mutants were selected by using colony PCR. The deletion mutants in C. *pinatubonensis* JMP134 were generated in a similar way. The only difference is that the MSM medium was used to instead of M9 medium with 0.5% gluconate as carbon source.

Determine the end-products of sulfide oxidation by recombinant E. coli

E. coli BL21(DE3) and its recombinant cells were cultivated at 25^{O}C. When the A_{600} to 0.4, 0.5 mM IPTG was added and further cultivated for 12h at 25 ^{O}C. The cells were harvested by centrifugation (6 000 × *g*, 10 min) and suspended in 100 mM Tris buffer (pH 8) at an absorbance of 2 at 600 nm. One ml of the cell suspension was transferred to a 15-ml glass tube. Fresh prepared NaHS was added to initiate the reaction. The tube was capped with a rubber stopper and incubated at 25 ^{o}C without shaking to minimize sulfide oxidation by O_2. The pH was selected to prevent H_2S from rapid evaporation, as the

dominant species is HS- at pH 8 (Hughes et al 2009). The sulfide, sulfane sulfur, sulfite and thiosulfate were analyzed at various time intervals.

Analytical procedures

Sulfide was analyzed by a colorimetric method (Fogo and Popowsky 1949), sulfite and thiosulfate were determined by ion chromatography as previously described (Xin et al 2016). Sulfane sulfur, including polysulfides and persulfides, was measured by the cyanolysis method (Kamyshny Jr 2009).

The H_2S spiking test with defined strains

Selected bacteria were culture in LB at 30 ^{o}C with shaking until OD_{600nm} reached between 0.4 and 0.6. *E. coli* was induced with 0.4 mM IPTA and culture for additional 3 hours, and others were not induced. When most bacteria were harvested, the OD_{600nm} was about 1. *Serratia fonticola* DSM4576 and *Corynebacterium vitaeruminis* DSM20294 were also induced with 200 μM NaHS before harvesting. The harvested cells were washed with 50 mM pH 8.0 Tris-HCl buffer and suspended in the Tris buffer with 50 μM DTPA at OD_{600nm} of 1. NaHS was added into the cell suspension to initiate sulfide oxidation, and sulfide was determined at various time points. Three repeats were executed at the same time. Sulfide was analyzed by a colorimetric method (Kamyshny Jr 2009).

Sample collection, treatment, and culturing from various environmental samples

Most samples were collected around Jinan, Shandong, China from June and July of 2015. Soil samples were from a wheat field and forest; freshwater samples were from a lake, and near-shore seawater samples were collected from Qingdao, Shandong, China. To collect soil samples, the top layer of 2 cm was removed and soil from 2-10 centimeter (cm) depth was transferred into a sterile bottle. Water samples were directly collected into sterile bottles. All samples were immediately transported back to lab and processed.

Two grams of soil were added to a 50-ml centrifugation tube containing 20 ml of 10 mM phosphate buffered saline and three grass beads (5 mm in diameter). The samples were vigorously vortexed to disperse bacterial colonies and left on the bench for 2 min. Five ml of the soil leachate or water sample were centrifuged at 8 000 × *g* for 10 min to precipitate bacteria and the pellets were re-suspended in 2 ml of LB medium and transferred into a 15-ml tube, incubated with shaking at 30°C for 24 hours. H_2S was detected with the paper strip containing Lead(II)-acetate. Simultaneously, the soil leachate and water samples were diluted and spread onto LB plates, incubated at 25°C for 48 hours. A circle with twenty colonies on a LB plate was randomly drawn, and the colonies were individually transferred into 2 ml of LB medium in a 15-ml tube to test for H_2S production. The isolated pure bacterial cultures of H_2S-oxidizing and non-oxidizing bacteria were also mixed and tested for H_2S production. The overnight cultures of tested bacteria in LB were mixed to a various ratios according their OD_{600nm} in a fixed volume of 200 microliter (μl) and then transferred into 2 ml of LB medium to test H_2S production. Seawater samples and the bacteria from seawater sample were incubated in LB containing 2% of NaCl and 10 mM $MgSO_4$.

The finding of adjacent sqr and pdo genes

A microbial genomic protein sequence set from NCBI updated until April 15, 2016 was downloaded as a preliminary database for PDO and SQR sorting. This database contains proteins from 4929 completely annotated genomes. The archaeal genomic protein sequence set from NCBI updated until April 6, 2016 was downloaded for PDO and SQR sorting, which are collected from 242 completely annotated genomes. The seeds of PDOs were from previous reports (Liu et al 2014, Shen et al 2015), including 18 PDO sequences and 12 outgroup sequences. The seed sequences of SQRs were collected from two reports (Gregersen et al 2011, Marcia et al 2010). The seeds were used to blast the total GenBank bacterial genomes with a relaxed criterion (e value < $1e^{-5}$). The output proteins were collected, and the adjacent *sqr* and *pdo* genes were selected if their gene identification numbers next to each other with a difference equal or less than 3. The protein pairs were removed if the coding genes were not transcribed in the same direction or the genes were separated by > 500 bp of an intergenic region. Finally, SQRs and PDOs were filtered again by using phylogenetic tree analysis with reported PDOs and SQRs with a pairwise deletion, *p*-distance distribution, and bootstrap analysis of 1,000 repeats.

Some neighboring *pdo* and *sqr* do not have adjacent accession numbers, and they were manually identified from the strains with both *pdo* and *sqr*.

Analysis of PDO and SQR in sequenced bacterial genomes and archaeal genomes

The seed sequences of PDO were collected from reported PDOs and from PDOs whose genes are located next to *sqr*. The seed sequences of SQRs were collected from

similar sources. The seed sequences of PDOs were used to search the downloaded bacterial or archaeal genome database by using Standalone BLASTP algorithm with conventional criteria (e-value$\leqslant 1e^{-10}$, coveryage$\geqslant$60%, identity$\geqslant$35%) to obtain PDO candidates from 4929 bacterial genomes and 242 archaeal genomes. The seed sequences of SQRs were used to search the downloaded bacterial or archaeal genome database by using Standalone BLASTP algorithm with less stringent criteria (e-value$\leqslant 1e^{-10}$, coveryage$\geqslant$50%, identity$\geqslant$30%) to obtain SQR candidates from both bacterial and archaeal genomes. The candidates were stochastically separated into several groups and each group contains lower than 300 candidate sequences. For each group, the seed sequences of PDOs or SQRs were combined with the candidate groups of PDOs or SQRs. Then, the candidate sequences were aligned though MAFFT and further filtered through phylogenetic tree analysis by using a neighbor-joining analysis with the MEGA version 7.0 program, running a pairwise deletion, *p*-distance distribution, and bootstrap analysis of 1,000 repeats. The filtered sequences were picked and grouped into defined subfamily if the sequence was in the same clade with known seed sequences. To further check the accuracy of these filtered sequences, two trees were built for inspection. First, the filtered sequences were grouped together and combined with seed sequences for phylogenetic tree analysis again by using the same parameters as above. Second, the filtered sequences were separated into subfamilies, each subfamily of sequences were combined with seed sequences for phylogenetic tree analysis by using the same parameters. The sequences that could steadily stay in the same clade with seed sequences in these two trees were collected. The collected PDOs and SQRs were further grouped into 193 and 173 unique groups (identity>60% within each group) by using the CD-HIT program, respectively (Li

and Godzik 2006). Representative sequences from each group and the reported sequences were subject to phylogenetic tree analysis to construct the final trees of PDOs and SQRs.

The phylum distribution of *sqr* and *pdo* genes from the 4929 genomes was also analyzed. The strains encoding both *pdo* and *sqr*, only *pdo*, or only *sqr* were normalized with the total strains at the phylum level. If the total strains in a phylum were lower than 50, they were not analyzed.

The 1024 strains with *sqr* were distributed in 283 genera, of which 193 genera contained bacteria with both *sqr* and *pdo*, and their growth modes were manually checked according to the representative strains at the genus level. For details, the methylotrophic bacteria were considering as heterotroph. The facultative aerobic bacteria and the microaerophilic bacteria were all considered as aerobic. The strains have both autotrophic and heterotrophic abilities were considered as heterotrophic. Finally, the statistical results were separated into two groups, aerobic heterotrophs and the others that were either anaerobes or autotrophs.

Analysis of pdo and sqr genes in marine genomes

Additional seed sequences of PDOs and SQRs were obtained from 177 marine microbial genomes from the Gordon and Betty Moore Foundation Marine Microbial Genome Sequencing project (http://www.jcvi.org/cms/research/past-projects/microgenome/overview/). The searches for PDOs and SQRs were done by using the BLASTP program and phylogenetic analysis as described above. Total of 175 PDOs genes and 82 SQRs were found and used as the seed sequences.

Identification of PDOs and SQRs in marine metagenomes of global ocean sampling (GOS) expedition

The seed sequences included the marine sequences, the reported sequences, and the sequences encoded by the neighboring *sqr* and *pdo* genes, and they were used as query sequences for a BLASTP search of the GOS proteins at CAMERA (GOS: all peptides (P) database) with a cutoff E value of $1e^{-10}$. The method used to filter the PDOs and SQRs sequences in GOS data set were as similar as used in the screening for PDO and SQR in sequenced GenBank bacterial genomes described above. The duplicated peptide sequences of PDOs and SQRs due to paired reads were identified by mapping the identified J. Craig Venter Institute (JCVI) PEP numbers to JCVI Read ID and Mate ID; duplicated sequences were combined. The frequency of strains containing PDOs and SQRs were calculated by using a method as reported (Howard et al 2008). The distribution of PDOs and SQRs against each other in the GOS dataset was normalized to six essential single-copy genes (*atpD*, *rpoB*, *gyrB*, *dnaK*, *tufA* and *recA*). Six genes from *E. coli* MG1655 were used as queries to BLASTP with CAMERA ('GOS: all ORF peptides (p)' database) with a limiting E value of 10^{-25}. Comparing this hits returned with preview reports could validated this E-value cut-off. Duplicate sequences due to paired reads were also removed. Each gene's homologues were size-normalized to the length of *recA* gene. The count of each essential gene's homologues were multiplied with the length of *recA* gene and divided with the length of single-copy essential gene. The length for *atpD*, *rpoB*, *gyrB*, *dnaK*, *tufA* and *recA* were 1383 bp, 4029 bp, 2415 bp, 1917 bp, 1185 bp and 1062 bp, respectively. The average length of *pdo* and *sqr* were calculated by using the length of total *pdo* genes and *sqr* genes. The average length of *pdo* and *sqr* were

1014 bp ± 273 bp and 1387 bp ± 213 bp, respectively. The homologues of *pdo* and *sqr* were size-normalized by multiplying the length of *recA* and dividing with the average length of *pdo* and *sqr*, respectively. The abundance of *pdo* and *sqr* were calculated for each sample by using the following formula.

For the *pdo* gene:

(Size-normalized counts of the homologues of *pdo*) / (Average, normalized counts of six single-copy genes) *100

For the *sqr* gene:

(Size-normalized counts of the homologues of *sqr*) / (Average, normalized counts of six single-copy genes) *100

The calculation process was recorded in Table S9.

The expression of pdo and sqr genes in marine metatranscriptomic data sets

Three publicly available metatranscriptomic data sets were from coastal California waters, Monterey Bay and station ALOHA in the North Pacific Gyre (NCBI accession: PRJNA268385, PRJNA183166, and PRJNA244754, respectively). The same query seeds used for marine metagenomic search were used for TBLASTN searches to identify transcripts in the three marine metatranscriptomic data sets. The reads with 60% sequence identity and $\geqslant$46 amino acids of the read length aligning to the query were considered as positive hits. The transcripts were divided by the total transcripts to obtain the transcription ratio for *pdo* and *sqr*, respectively. The detected transcripts of *pdo* and *sqr* were further grouped according to the sampling time of day (6:00 – 18:00) or night (18:00-6:00). For each sampling time, the counts of the transcripts of *pdo* and *sqr* were

divided by the total transcripts at each sampling time to get normalized transcripts per 1,000,000 transcripts.

Supplemental Methods References

Fogo JK, Popowsky M (1949). Spectrophotometric determination of hydrogen sulfide. *Analytical Chemistry* **21:** 732-734.

Gregersen LH, Bryant DA, Frigaard NU (2011). Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. *Front Microbiol* **2:** 116.

Harighi B (2009). Genetic evidence for CheB- and CheR-dependent chemotaxis system in *A. tumefaciens* toward acetosyringone. *Microbiological research* **164:** 634-641.

Howard EC, Sun S, Biers EJ, Moran MA (2008). Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environmental microbiology* **10:** 2397-2410.

Hughes MN, Centelles MN, Moore KP (2009). Making and working with hydrogen sulfide: The chemistry and generation of hydrogen sulfide in vitro and its measurement in vivo: a review. *Free Radic Biol Med* **47:** 1346-1353.

Kamyshny Jr A (2009). Improved cyanolysis protocol for detection of zero-valent sulfur in natural aquatic systems. *Limnology and Oceanography: Methods* **7:** 442-448.

Li W, Godzik A (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22:** 1658-1659.

Liu H, Xin Y, Xun L (2014). Distribution, diversity, and activities of sulfur dioxygenases in heterotrophic bacteria. *Applied and environmental microbiology* **80:** 1799-1806.

Marcia M, Ermler U, Peng G, Michel H (2010). A new structure-based classification of sulfide:quinone oxidoreductases. *Proteins* **78:** 1073-1083.

Shen J, Keithly ME, Armstrong RN, Higgins KA, Edmonds KA, Giedroc DP (2015). Staphylococcus aureus CstB Is a Novel Multidomain Persulfide Dioxygenase-Sulfurtransferase Involved in Hydrogen Sulfide Detoxification. *Biochemistry* **54:** 4542-4554.

Xia Y, Chu W, Qi Q, Xun L (2014). New insights into the QuikChangeTM process guide the use of Phusion DNA polymerase for site-directed mutagenesis. *Nucleic acids research***:** gku1189.

Xin Y, Liu H, Cui F, Liu H, Xun L (2016). Recombinant Escherichia coli with sulfide: Quinone oxidoreductase and persulfide dioxygenase rapidly oxidizes sulfide to sulfite and thiosulfate via a new pathway. *Environmental microbiology*.