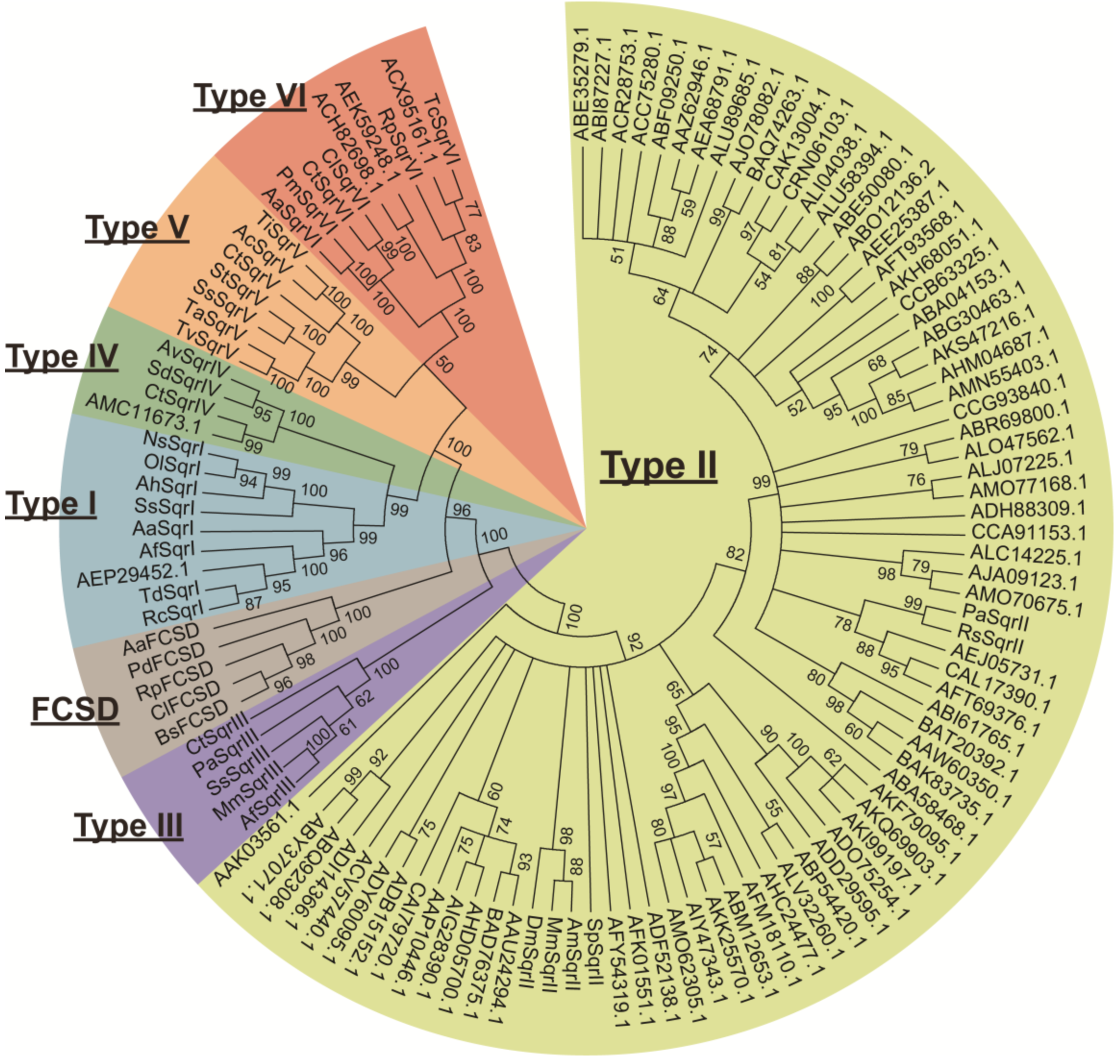
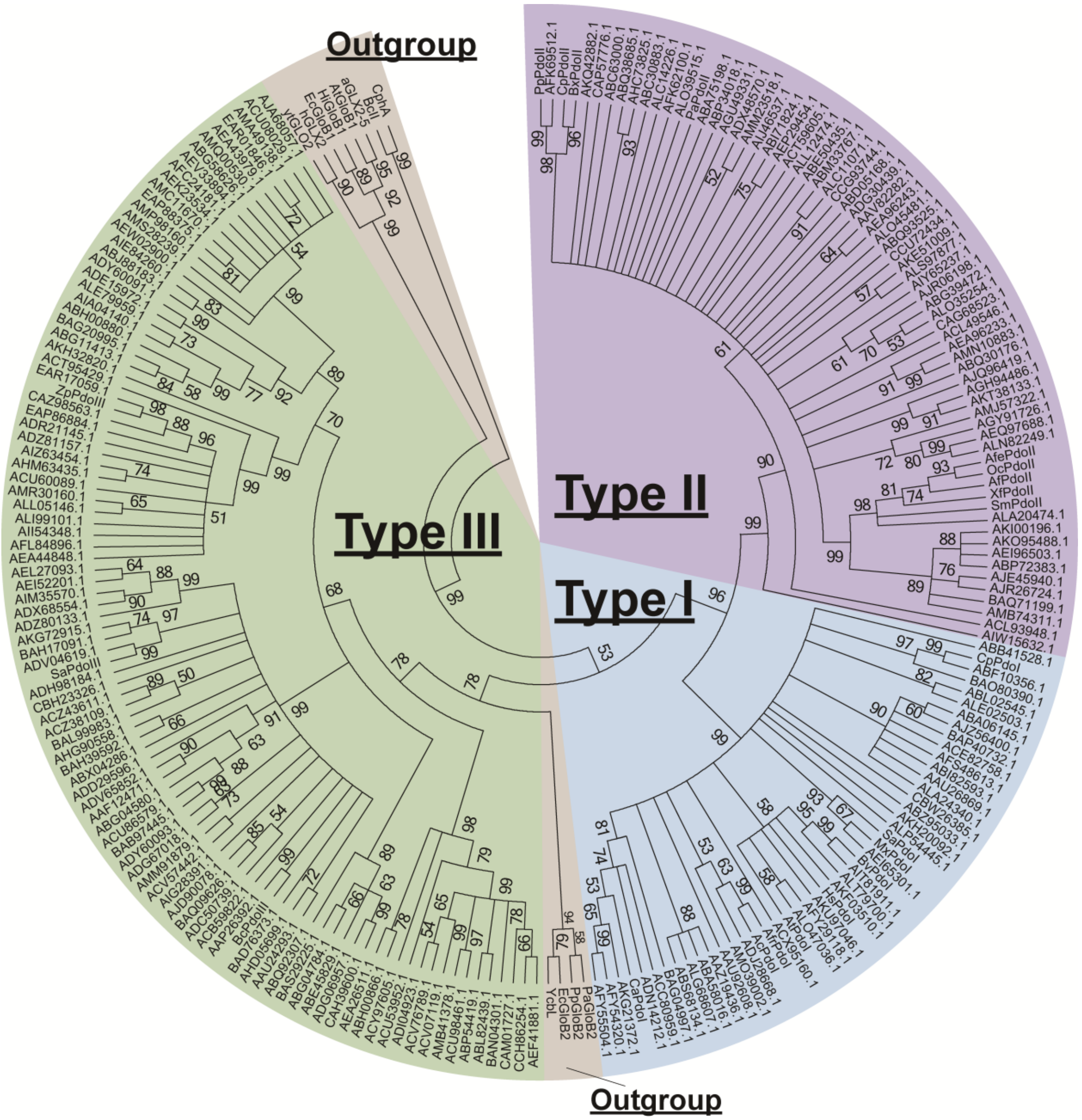
Sulfide production and oxidation by heterotrophic bacteria under aerobic conditions

Supplementary figures



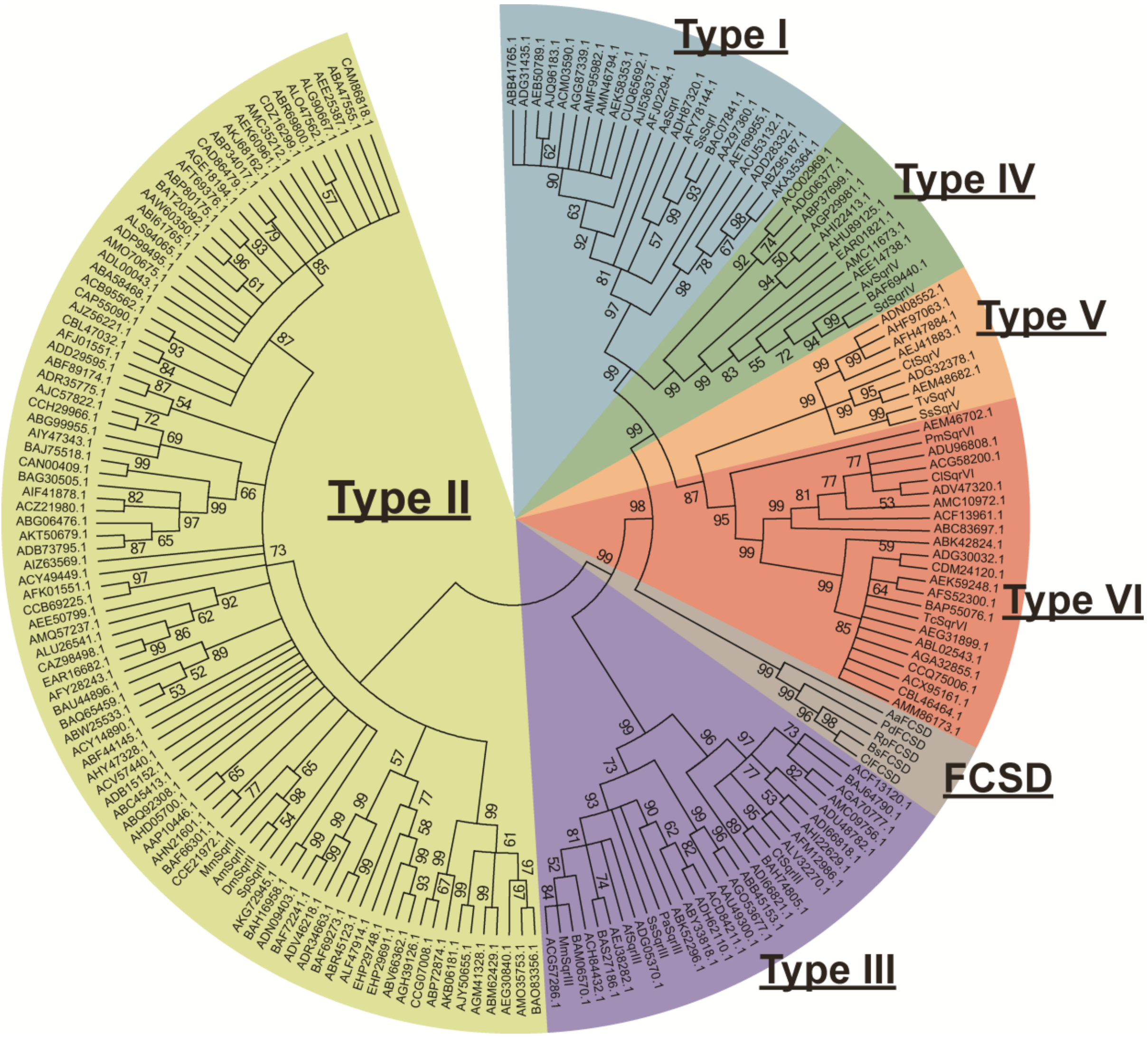
Supplementary Figure S1. The phylogenetic tree of representative SQRs whose genes are physically linked to *pdo* in bacterial genomes. The representative 77 SQR sequences were used for phylogenetic tree construction with reported seed sequences. Five flavocytochrome c sulfide dehydrogenase (FCSD) proteins were included as reference. These sequences were aligned by using clustalW, and the tree was built by using MEGA6. Reference proteins are listed below, with the organism origin and accession number: RcSqrI, *R. capsulatus* (CAA66112.1); TdSqrI, *T. denitrificans* (AAM52227.1); AfSqrI, *A. ferrooxidans* (ABF71988.1); AaSqrI, *A. aeolicus* (NP_214500.1); SsSqrI, *S. sp.* (NP_942192.1); AhSqrI, *A. halophytica* (AAF72963.1); OlSqrI, *O. limnetica* (AAF72962.1); NsSqrI, *N. sp.* (NP_488552.1);

PaSqrII, *P. aeruginosa* (NP_251035.1); RsSqrII, *R. solanacearum* (NP_519663.1); SpSqrII, *S. pombe* (NP_596067.1); DmSqrII, *D. melanogaster* (NP_647877.1); AmSqrII, *A. marina* (ABV22505.1); MmSqrII, *M. musculus* (EDL28115.1); AfSqrIII, *A. fulgidus* (NP_069393.1); MmSqrIII, *M. magnetotacticum* (ZP_00055086.1); CtSqrIII, *C. tepidum* (NP_661917.1); SsSqrIII, *S. solfataricus* (NP_343961.1); PaSqrIII, *P. aerophilum* (NP_560139.1); SdSqrIV, *S. denitrificans* (YP_393133.1); AvSqrIV, *A. vinosum* (ZP_04773162.1); CtSqrIV, *C. tepidum* (NP_661023.1); CtSqrV, *C. tepidum* (NP_661769.1); AcSqrV, *A. caldus* (ZP_05293376.1); TiSqrV, *T. intermedia* (ZP_05500015.1); TvSqrV, *T. volcanium* (NP_111725.1); TaSqrV, *T. acidophilum* (NP_394588.1); SsSqrV, *S. solfataricus* (NP_343636.1); StSqrV, *S. tokodaii* (NP_378484.1); TcSqrVI, *T. crunogena* (ABB41976.1); RpSqrVI, *R. palustris* (YP_484673.1); AaSqrVI, *A. aeolicus* (NP_213539.1); PmSqrVI, *P. marina* (YP_002730289.1); CtSqrVI, *C. tepidum* (NP_661978.1); ClSqrVI, *C. luteolum* (YP_375032.1); BsFCSD, *B. sp.* (ZP_02001369.1); ClFCSD, *C. limicola* (AAL68892.1); PdFCSD, *P. denitrificans* (CAA55826.2); RpFCSD, *R. palustris* (YP_001990584.1); AaFCSD, *A. aeolicus* (NP_213158.1).



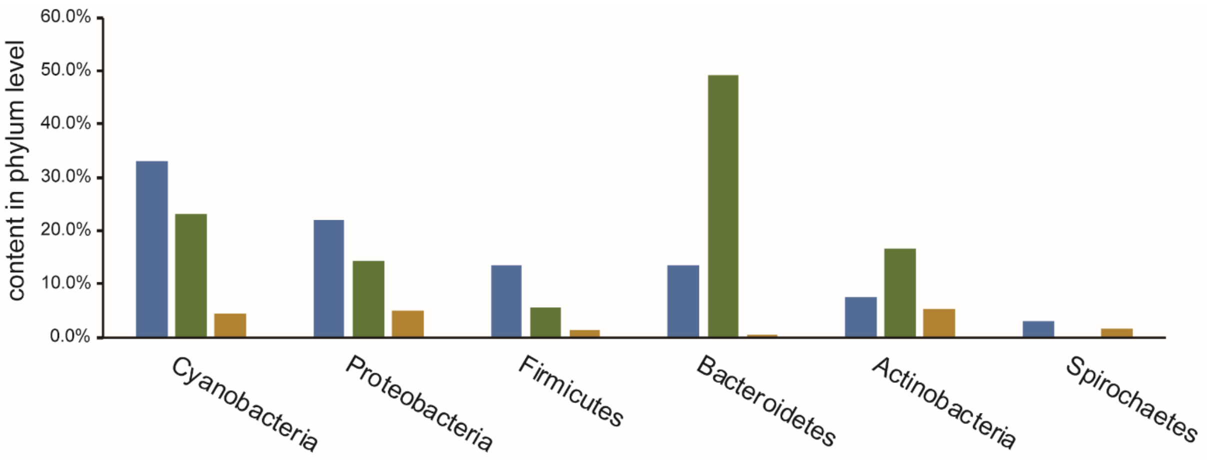
Supplementary Figure S2. The phylogenetic tree of selected PDOs in sequenced bacterial genomes from GenBank. The representative 193 PDO sequences were used for phylogenetic tree construction with reported seed sequences. These sequences were aligned by using clustalW, and the tree was built by using MEGA6. Reference proteins are listed below, with the organism origin and accession number: Type I PDOs: SaPdoI (YP_003957083.1), CpPdoI (YP_297536.1), MxPdoI (YP_633997.1), HsPdoI (NP_055112.2), CaPdoI (YP_007162862.1), BvPdoI (ZP_00420127.1), AtPdoI

(NP_974018.3), AcPdoI (AEK59246.1), AfrPdoI (YP_002424776.1). Type II PDOs: OcPdoII, *O. carboxidovorans* (YP_002287251.1); XfPdoII (NP_298058.1), SmPdoII (NP_435818.1), CpPdoII (YP_297791.1), PaPdoII (NP_251605.1), BxPdoII (YP_554628.1), AfaPdoII (AAK89929.1), AfePdoII (ZP_11421028.1); PpPdoII, *P. putida* (ABQ76243.1). SaPdoIII, *S. aureus* (WP_000465474.1); BcPdoIII, *B. cereus* (EEK49737.1); ZpPdoIII, *Z. profunda* (ADF52140.1); PaGloB2, *P. aeruginosa* (NP_249523.1); PpGloB2, *P. putida* (ABQ76961.1); EcGloB2, *E. coli* (NP_415447.1); YcbL, *S. enterica* (CAD05397.1); EcGloB1, *E. coli* (NP_414748.1); HiGloB1, *H. influenzae* (ADO96205.1); AtGloB1, *A. fabrum* (NP_356997.2); aGLX2-5, *A. thaliana* (NP_850166.1); ytGLO2, *S. cerevisiae* (CAA71335.1); hGLX2, *H. sapiens* (CAA62483.1); BcII, *B. cereus* (AAA22276.1); CphA, *A. hydrophila* (CAA40386.1).

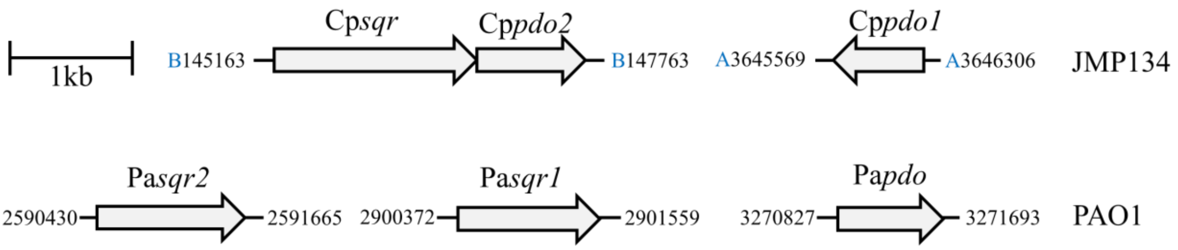


Supplementary Figure S3. The phylogenetic tree of selected SQRs in sequenced bacterial genomes from GenBank. The representative 173 SQR sequences were used for phylogenetic tree construction with reported seed sequences. These sequences were aligned by using clustalW, and the tree was built by using MEGA6. Reference proteins are listed below, with the organism origin and accession number: AaSqrI, *A. aeolicus* (NP_214500.1); SsSqrI, *S. sp.* (NP_942192.1); SpSqrII, *S. pombe* (NP_596067.1); DmSqrII, *D. melanogaster* (NP_647877.1); AmSqrII, *A. marina* (ABV22505.1); MmSqrII, *M. musculus* (EDL28115.1); AfSqrIII, *A. fulgidus* (NP_069393.1); MmSqrIII, *M. magnetotacticum* (ZP_00055086.1); CtSqrIII, *C. tepidum* (NP_661917.1); SsSqrIII, *S. solfataricus* (NP_343961.1); PaSqrIII, *P. aerophilum* (NP_560139.1); SdSqrIV, *S. denitrificans* (YP_393133.1); AvSqrIV, *A. vinosum* (ZP_04773162.1); CtSqrV, *C. tepidum*

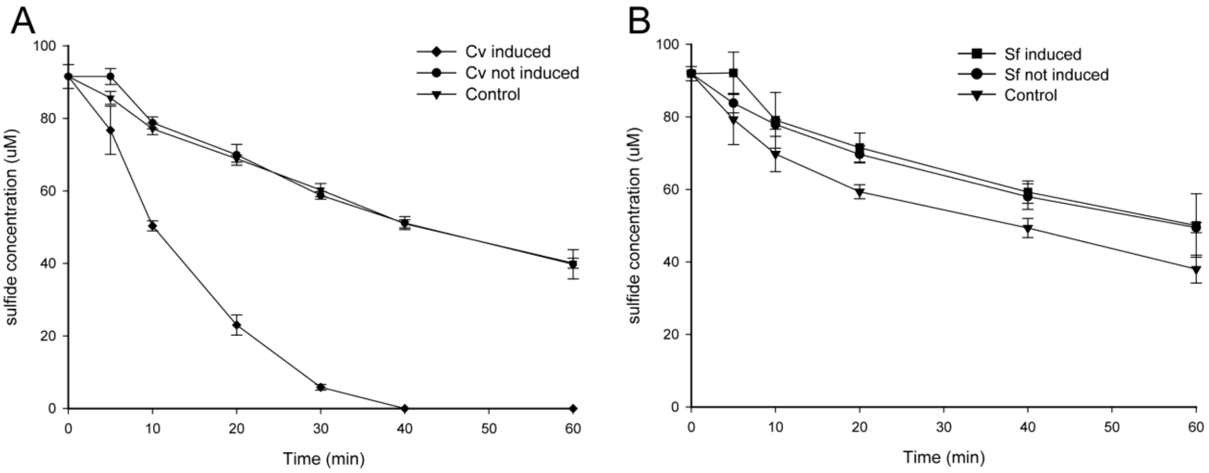
(NP_661769.1); TvSqrV, *T. volcanium* (NP_111725.1); SsSqrV, *S. solfataricus* (NP_343636.1); TcSqrVI, *T. crunogena* (ABB41976.1); PmSqrVI, *P. marina* (YP_002730289.1); ClSqrVI, *C. luteolum* (YP_375032.1).



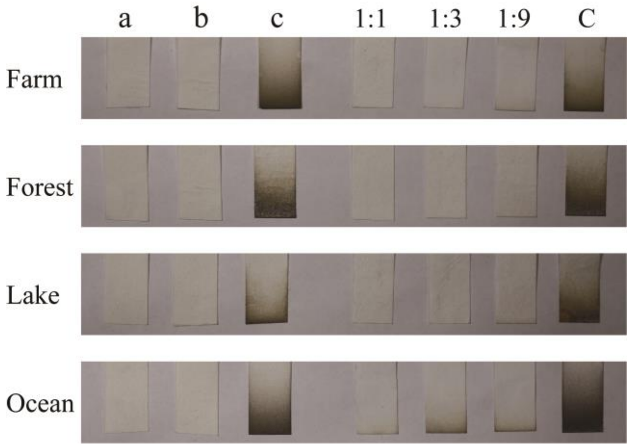
Supplementary Figure S4. The relative abundance of strains encoding *pdo* and *sqr* at the phylum level. The strains encoding *pdo* and *sqr,* only *pdo*, or only *sqr* were separately counted at the phylum level. The relative abundances of them were calculated by dividing the total number of strains in the same phylum. The blue, green and yellow columns represent the relative abundance of strains encoding *pdo* and *sqr*, only *pdo*, or only *sqr*, respectively.



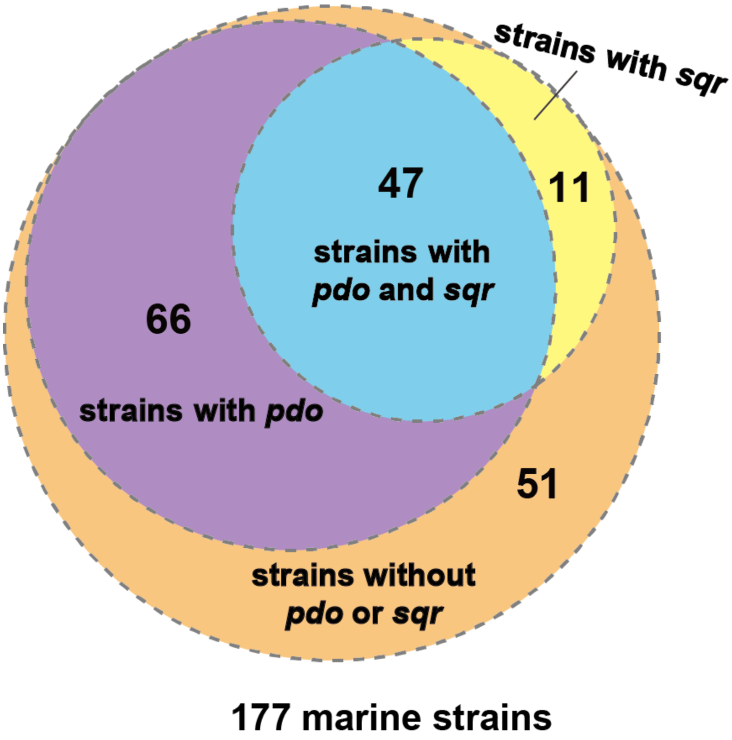
Supplementary Figure S5. Organization of the *sqr* and *pdo* genes in *C. pinatubonensis* JMP134 and *P. aeruginosa* PAO1. The gene ID of Cp*sqr*, Cp*pdo2* (also known as Cn*sdoA*), Cp*pdo1* (also known as Cn*ethe1*), Pa*sqr2*, Pa*sqr1* and Pa*pdo* (also known as Pa*sdoA*) from NCBI were 3613170, 3613171, 3609713, 879884, 882794 and 882596, respectively. The numbers indicate the positions of putative genes in corresponding genome, and blue “A” and “B” denoted chromosome A and chromosome B in *C. pinatubonensis* JMP134.



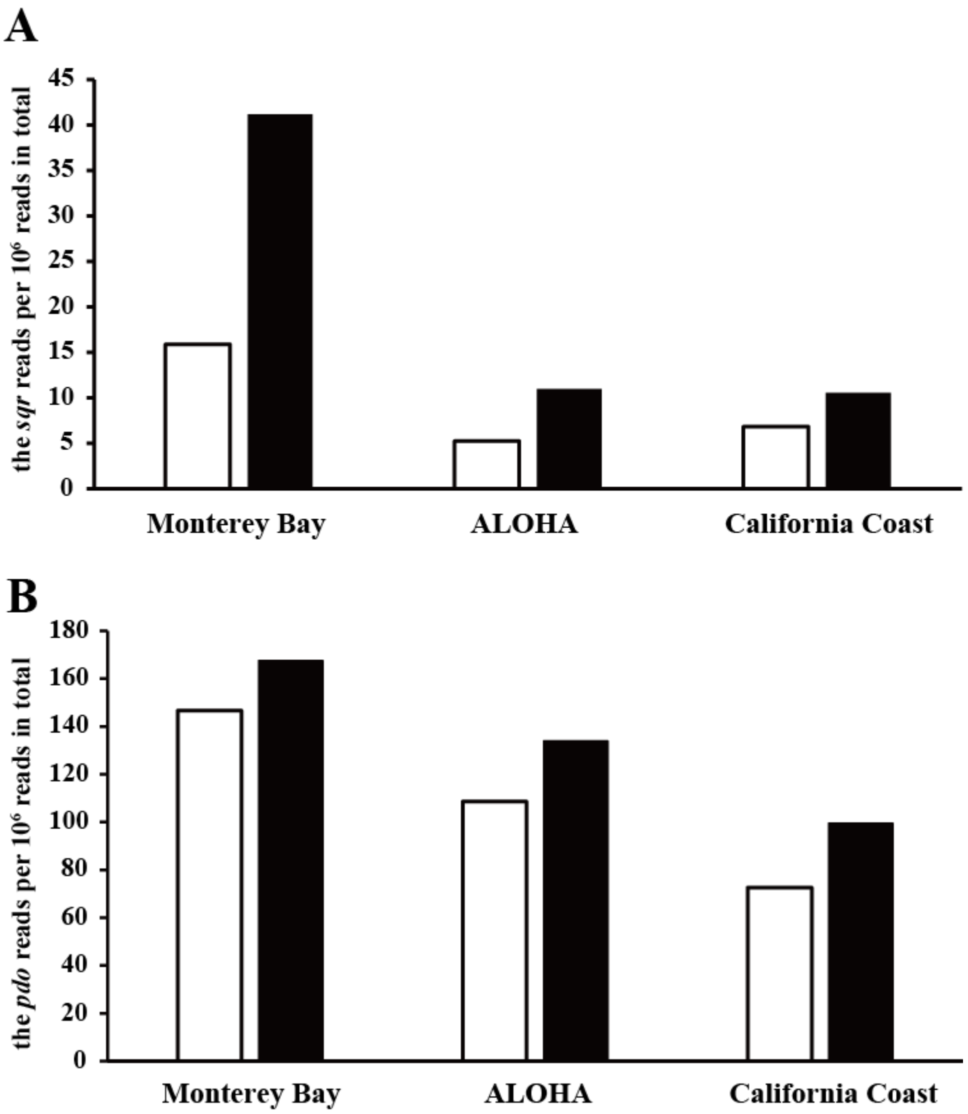
Supplementary Figure S6. H$_2$S oxidation by *C. vitaeruminis* DSM20294 and *S. fonticola* DSM4576. Cells were suspended in 50 mM Tris buffer (pH 8.0) with 50 µM DTPA at OD$_{600nm}$ of 1, and NaHS was added to initiate the reaction. Samples were taken at indicated time points for sulfide detection. Controls were done in the same buffer without bacterial cells. **A**) *C. vitaeruminis* DSM20294 (Cv); **B**) *S. fonticola* DSM4576 (Sf). Both induced and non-induced cells were tested. For the induction, the cells were spiked with 200 µM NaHS and cultured for 2 hours before harvesting. All data are average of three samples with standard deviation (error bar).



Supplementary Figure S7. **Testing H_2S production with environmental samples and bacterial isolates.** Columns are labeled as follows: **a**, cultures inoculated with environmental samples; **b**, H_2S-oxidizing isolates; **c**, H_2S-non-oxidizing isolates; **1:1, 1:3, 1:9**, ratios of oxidizer and non-oxidizers in mixed cultures; and **C**, the non-oxidizer only as controls. The cultures in LB was incubated at 30°C for 24 h. According to 16S rRNA gene sequence, the selected H_2S oxidizers were most similar to *Acinetobacter calcoaceticus* H3， *Pseudomonas moraviensis* IARI-HHS1-33， *Pseudomonas putida* S6, and *Marinomonas* species BSi20567; the non-oxidizers were most similar to *Arthrobacter oryzae* OS115, *Streptomyces gardneri* NBRC 12865 *Aeromonas sobria* JY081016-1, and *Pseudoalteromonas phenolica* AN40 from the farm, forest, lake, and seawater samples.



Supplementary Figure S8. The distribution of PDOs and SQRs in sequenced marine bacterial genomes. The selected PDOs or SQRs from the marine bacterial genomes downloaded from from the Gordon and Betty Moore Foundation Marine Microbial Genome Sequencing project were listed with its related genome IDs to form a table. The genomes containing PDO or SQR were separately summarized. Then the genomes containing both PDO and SQR were combined with the genome IDs. The figure is drawn by using the Vennerable package in R language and modified with Adobe Illustrator CS6.



Supplementary Figure S9. The normalized transcripts of *sqr* and *pdo* per 1 000 000 total reads at different sampling sites. A) The transcripts for *sqr* at three different sampling sites; **B**) The transcripts for *pdo* at three different sampling sites. The white column indicates the normalized transcripts occurring at day time, and the black column indicates the normalized transcripts occurring at night time.