# Integrating biogeochemistry with multi-omic sequence information in a model oxygen minimum zone
# - SI Appendix -

Stilianos Louca, Alyse K. Hawley, Sergei Katsev, Monica Torres-Beltran, Maya P. Bhatia, Sam Kheirandish, Céline C. Michiels, David Capelle, Gaute Lavik, Michael Doebeli, Sean A. Crowe, Steven J. Hallam

## S1 Data acquisition

### S1.1 Sampling site and time

Saanich Inlet (SI) a seasonally anoxic fjord on the coast of Vancouver Island, British Columbia Canada has been the site of intensive study for many decades (1, 2). The presence of a shallow glacial entrance sill at $75\ \mathrm{m}$ depth limits mixing and ventilation of basin waters below approximately $100\ \mathrm{m}$, resulting in stratification and oxygen depletion during spring and summer (Fig. 1a). Shifts in coastal currents in late summer and fall lead to an influx of denser, oxygenated and nutrient-rich water into the Inlet shoaling anoxic basin waters upward in a process known as deep water renewal (2, 3). Consistent partitioning of the microbial community along the redox cline and similarity to other OMZ microbial communities make Saanich Inlet a model ecosystem for studying the intersection between environmental sequence information and biogeochemical activity along defined redox gradients (3–5).

The fjord has a maximal depth of $232\ \mathrm{m}$ at the sampling site SI03 ($123^\circ\ 30.300'$ W, $48^\circ\ 35.500'$ N).

Sampling is conducted monthly during daylight hours using a combination of 5 and $8\ \mathrm{L}$ Niskin bottles and $12\ \mathrm{L}$ Go-Flo bottles attached to a nonconducting wire. A Sea-Bird CTD (conductivity, temperature and depth) sensor attached to the bottom of the wire provides depth profiles for temperature, salinity, PAR/Irradiance, conductivity, density, and dissolved oxygen (Sea-Bird Electronics$^{\mathrm{TM}}$). Water sampling for multiple chemical and microbial parameters proceeds directly from the bottles in the following order: First, samples are taken for dissolved $\mathrm{O_2}$ measurements via Winkler titration, followed by sampling of dissolved gases. Next, samples are taken for RNA, then protein followed by ammonium, hydrogen sulfide and nitrite. Finally, salinity is measured for a subset of depths for CTD calibration, and samples are taken for DNA.

Chemical data were acquired on January 13, 2010 (cruise SI041_01/13/10), February 10 (SI042_02/10/10), March 10 (SI043_03/10/10), April 7 (SI044_04/07/10), July 7 (SI047_07/07/10)

and August 11 (SI048_08/11/10). All molecular sequencing was performed using samples collected on February 10, 2010 (SI040_02/08/10) at depths 100 m, 120 m, 135 m, 150 m, 200 m for metagenomes and metatranscriptomes and at depths 97 m, 100 m, 120 m, 150 m, 165 m, 200 m for metaproteomes.

## S1.2 Chemical and physical depth profiles

Temperature, salinity and depth were measured using the CTD sensor described above. The Winkler titration method was used to measure dissolved oxygen ($O_2$) concentrations (6) and calibrate CTD measurements. Samples were collected into Winkler glass Erlenmeyer flasks using latex tubing, overflowing three times to ensure no air contamination, manganese (III) sulphate and potassium iodide were added in succession, inverted to mix and stored at room temperature. Samples were titrated using an automatic titrator. CTD data were processed and manually curated using the Sea-Bird Seasoft™ software.

Samples for dissolved nutrient (nitrate, nitrite, sulphate and silicate) analyses were collected into $60\ \mathrm{mL}$ syringes and filtered through a $0.22\ \mu\mathrm{m}$ Millipore Acrodisc™ into $15\ \mathrm{mL}$ falcon tubes. Prior to analysis all samples were stored on ice. Nitrate ($NO_3^-$) samples were stored at $-20°\mathrm{C}$ in the laboratory, and later analyses carried out using a Bran Luebbe autoanalyser using standard colorimetric methods. For nitrite ($NO_2^-$) analysis, $2\ \mathrm{mL}$ of sample water were supplemented with $100\ \mu\mathrm{L}$ sulfanilamide and $100\ \mu\mathrm{L}$ nicotinamide adenine dinucleotide in $4\ \mathrm{mL}$ plastic cuvettes. Prepared standards were supplemented with reagents at the same time. Cuvettes were inverted for mixing, and temporarily stored on ice for not more than 4 hrs. Concentration was measured using a Cary60® spectrometer, based on absorbance at $452\ \mathrm{nm}$.

Samples for ammonium ($NH_4^+$) and hydrogen sulphide ($H_2S$) were collected directly from Niskin and GoFlo bottles into $15\ \mathrm{mL}$ amber scintillation vials and $15\ \mathrm{mL}$ falcon tubes aliquoted with $200\ \mu\mathrm{L}$ 20% zinc acetate respectively. Samples were stored on ice prior to analysis. For $NH_4^+$ analysis, amber vials for standard curve and samples were pre-aliquoted with $7.5\ \mathrm{mL}$ O-phthaldialdehyde (OPA) reagent respectively. $5\ \mathrm{mL}$ of sample water in triplicate and standard solutions were transferred into OPA pre-aliquoted amber vials. Vials were inverted and stored up to 4 hours. From each standard solution and sample water vial, $300\ \mu\mathrm{L}$ were transferred into a 96 well round bottom plate. Fluorescence at $380_{\mathrm{ex}}/420_{\mathrm{emm}}$ was read using a Varioskan plate reader. For $H_2S$ analysis, $300\ \mu\mathrm{L}$ samples were transferred in triplicate to a 96 well plate, and finally Hach Reagent 1 and 2 ($6\ \mu\mathrm{L}$ per well) were added. Absorbance at $670\ \mathrm{nm}$ was read after 5 min incubation using a Varioskan™ plate reader.

Water for dissolved nitrous oxide ($N_2O$) analysis was collected using Go-flo or Niskin bottles, and was transferred via a Teflon tube into $30\ \mathrm{mL}$ or $60\ \mathrm{mL}$ borosilicate glass serum vials. Vials were overflown three times their volume in order to remove any bubbles from the vial or tubing. Vials were subsequently spiked with $50\ \mu\mathrm{L}$ saturated mercuric chloride using a pipette. Vials were then crimp-sealed with a butyl-rubber stopper and aluminum cap, and stored in the dark at $4°\mathrm{C}$ until they were analyzed. Dissolved nitrous-oxide concentrations were measured using a purge-and-trap

auto-sampler coupled with a gas-chromatography mass-spectrometer (7).

## S1.3 Metagenomics, metatranscriptomics and metaproteomics

Metagenome and metaproteome datasets were generated using the same methods as described in Hawley *et al.* (8). Metaproteome sequence coverage was quantified using normalized spectral abundance factors (NSAF) (9). Metatranscriptome samples were filtered in the field onto $0.2\ \mu\mathrm{m}$ sterivex filter with inline pre-filter of $2.7\ \mu\mathrm{m}$ pre-filter, adding $1.8\ \mathrm{mL}$ of RNAlater® (Qiagen) and freezing on dry ice before transferring to $-80°\mathrm{C}$. RNA later was removed by washing Sterivex filter with Ringer's solution before proceeding with cell lysis in the filter cartridge. Total RNA was extracted using the mirVana™ miRNA extraction kit (Ambion), DNA was removed using the TURBO DNA-free™ kit (Ambion) and total RNA was purified using RNeasy™ MiniElute Cleanup Kit (Qiagen). RNA concentration and quality was determined using a Bioanalyzer. Production of cDNA libraries and sequencing was carried out at the Joint Genome Institute using the TruSeq® Stranded Total RNA Sample preparation Guide, including depletion of ribosomal RNA using Ribo-Zero. Assembled metagenomic sequences (contigs) were run through Metapathways (10) for annotation using a combination of RefSeq (11), KEGG (12), COG (13) and MetaCyc (14) databases. KEGG annotations of metagenomic contigs are provided as Dataset S1. Mass-spectrometry (metaproteomics) run information are provided as Dataset S2. Contig coverage by metagenome or metatranscriptome reads was quantified using RPKM values (Appendix S1.4). KEGG-annotated contigs were assigned to the selected process proxy genes of the model (Table 1.3 in the Appendix); gene coverage at each depth was then quantified by summing all assigned contig RPKM values. Because metaproteomes were missing at depth 100 m (the upper bound of our simulation domain), and in order to increase statistical power when evaluating our protein models, we used linear interpolation between depths 97 m and 120 m to estimate protein NSAF values at 100 m depth ("unit imputation").

Nitrate reductase (narGHIJ) assigned to planctomycetes showed a decline with depth, suggesting that it may be acting in reverse as a nitrite oxidase (15). In fact, narGHIJ counts affiliated with planctomycetes (narGHIJ-P) dominated all other nxr-associated counts in the metagenomes, metatranscriptomes and metaproteomes. We thus associated nxr with narGHIJ-P. However, because planctomycetes perform anammox in deeper depths (16), we observed a secondary peak in the *narGHIJ-P* DNA closer to the SNTZ that did not dissipate completely in bottom waters. Given this ambiguity in the interpretation of detected narGHIJ-P genes, we omitted the narGHIJ-P metagenomes and only used the narGHIJ-P metatranscriptomes and metaproteomes. For more details see Appendix S3.4.

All nosZ-related protein sequences mapped to a nosZ homolog found in the strictly aerobic *Roseobacter Maritimibacter alkaliphilus* HTCC2654 (17, 18) and showed strong inconsistencies with nosZ metagenomic and metatranscriptomic profiles. nosZ genes have been found to be enriched on particles, likely because they constitute a more anaerobic niche (19). Our metaproteomes were pre-filtered to remove eukaryotes and particles and are expected to be impoverished in nosZ proteins, facilitating a potential masking by related but functionally different proteins. We thus omitted the nosZ metaproteomic data from our analysis.

Table S1: KEGG orthologous groups (KOG) identified with each gene in the metagenomes or metatranscriptomes. The abundance of each gene was the sum of RPKM values (Appendix S1.4) assigned to all included KOGs.

| **gene or pathway** | **KOGs** | **restrictions** |
|---|---|---|
| ROM | K12536, K05648 | ABC transporters in *Pelagibacter* and *Roseobacter* |
| amo | K10945, K10946 | |
| nxr | K00370, K00371 | narGHIJ in *Planctomycetacea* |
| | K00374, K00373 | |
| hzo | K10535 | hao in *Planctomycetacea* |
| PDNO (norBC) | K04561, K02305 | |
| nosZ | K00376 | |
| sat | K00958 | |
| aprAB | K00394, K00395 | |
| dsrAB | K11180, K11181 | |
| nirKS | K00368, K15864 | |
| napAB | K02567, K02568 | |
| narGHIJ | K00370, K00371 | |
| | K00374, K00373 | |

## S1.4 Quantifying metagenomic and metatranscriptomic data using RPKM

Relative open reading frame (ORF) abundance in the metagenomic and metatranscriptomic datasets was determined for quantitative assessment of pathway coverage. This was achieved by adapting the reads per kilobase per million mapped (RPKM) coverage measure as described by Konwar *et al.* (20). Briefly, unassembled Illumina reads were mapped to assembled contigs using the short-read aligner BWA-MEM. The resulting SAM file is then inputed into the MetaPathways v2.5 software (20), which generates an RPKM value per ORF that is extended to an RPKM per pathway via summation. For the case of determining the abundance of pathways expressed in the metatranscriptome relative to those present in the metagenome, the unassembled metatranscriptome reads were mapped back to the assembled metagenome contigs. The RPKM calculation is a simple proportion of the number of reads mapped to a particular section of sequence normalized for ORF length and sequencing depth.

## S1.5 Process rate measurements

Rate measurements for anammox and denitrification were carried out as follows: Sample water from each depth was collected anaerobically with sterile nitrile tubing directly into $200\ \mathrm{mL}$ glass serum bottles, six per depth, and capped with butyl-rubber stopper and aluminum cap and stored at $10^{\circ}\mathrm{C}$ for approximately 1 hr while collection was completed. The protocol described by Holtappels *et al.* (21),and briefly outlined here, was then followed. One sample from each depth was bubbled with

He for 30 min to decrease concentration of $N_2$. The following substrates were then added: $^{15}NH_4^+$ alone, $^{15}NH_4^+$ and $^{14}NO_2^-$ combined, $^{15}NO_2^-$ alone, $^{15}NO_2^-$ and $^{14}NH_4^+$ combined or $^{15}NO_3^-$ alone. A blank for each depth was also bubbled with $He$. Sample water was then transferred from the serum bottle into a $12\ mL$ exetainer, capped and stored upside down. Samples in exetainers were then killed with $50\ \mu L$ saturated $HgCl$ at time intervals of 0 min, 6 hr, 12 hr, 24 hr, 48 hr and 72 hrs. Partial pressures of $^{29}N_2$ and $^{30}N_2$ evolved during the incubations were measured by gas chromatography coupled to isotope ratio mass spectroscopy. Rates of anammox and denitrification were calculated as described by Holtappels *et al.* (21).

Rate measurements using N isotope methods require a compromise between ensuring detection of labeled tracer elements and avoiding excessive perturbation of ambient substrate concentrations (22, §2.1). Due to the extremely low in-situ substrate levels in some of our samples (Fig. 2 in the main article), tracer substrate concentrations in the ex-situ incubator ($25\ \mu M\ NH_4^+$, $2\ \mu M\ NO_2^-$ and $5\ \mu M\ NO_3^-$) significantly exceeded in-situ concentrations. On the other hand, denitrification and anammox-related genes were found throughout the OMZ water column (Fig. 3a in the main article). Hence, rates measured in the incubator are only potential rates that likely overestimate actual in-situ rates, especially in substrate-depleted regions far from the SNTZ. For example, Dalsgaard *et al.* (23) reports a 2–4 fold increase of anammox rates following the addition of $10\ \mu M\ NH_4^+$ in anoxic water column experiments. Similarly, Wenk *et al.* (24) found high potential denitrification rates in nitrate-depleted regions of a meromictic lake. We thus corrected our rate measurements for differences between in-situ and incubator substrate concentrations, as described below.

The simplest approach would be to multiply measured rates with the ratios of in-situ over ex-situ substrate concentrations, as has been done in previous ex-situ incubation experiments (25). However, such a linear rescaling implicitly assumes that substrate half-saturation constants are much higher than both the in-situ as well as ex-situ concentrations, an assumption that may not be justifiable in regularly substrate-depleted natural environments. For example, members of the Scalindua candidate clade, which is well represented in Saanich Inlet (16), exhibit nitrite half-saturation constants as low as $0.45\ \mu M$ (26). To avoid an implicit assumption of 1st order kinetics, and for consistency with the assumptions of our model, we corrected our rates using Michaelis-Menten kinetic curves (Appendix S2.4) with the same half-saturation constants as used in our model (Appendix S2.7). Specifically, if $R^*_{\mathrm{hzo}}(z)$ is the measured ex-situ (i.e. potential) anammox rate at some particular depth, then the corrected in-situ rate was assumed to be

$$R_{\mathrm{hzo}} = R^*_{\mathrm{hzo}}(z) \cdot \frac{\frac{\left[NH_4^+\right]}{K_{NH_4^+}+\left[NH_4^+\right]} \frac{\left[NO_2^-\right]}{K_{NO_2^-}+\left[NO_2^-\right]}}{\frac{\left[NH_4^+\right]^*}{K_{NH_4^+}+\left[NH_4^+\right]^*} \frac{\left[NO_2^-\right]^*}{K_{NO_2^-}+\left[NO_2^-\right]^*}}. \tag{1}$$

Here, $K_{NH_4^+}$ and $K_{NO_2^-}$ are anammox half-saturation constants for $NH_4^+$ and $NO_2^-$, respectively (Appendix S2.7), $[NH_4^+]$ and $[NO_2^-]$ are the corresponding measured in-situ concentrations and $[NH_4^+]^*$ and $[NO_2^-]^*$ are the concentrations in the incubator at the beginning of the experiment, i.e. $[NH_4^+]^* = [NH_4^+] + 25\ \mu M$ and $[NO_2^-]^* = [NO_2^-] + 2\ \mu M$. Measured denitrification rates were corrected in a similar way to account for differences in $NO_3^-$ concentrations.

## S1.6 qPCR quantification of SUP05 abundances

All metagenomic, metatranscriptomic and metaproteomic profiles presented here only provide relative — rather than absolute — biomolecule abundances. This remains the de facto standard for multi-omic data sets, owing largely due to methodological challenges involved in absolute DNA, mRNA and protein quantification (but see Smets *et al.* (27) for recent advancements). As we explain below (section 2.9), multi-omic depth profiles were linearly rescaled to facilitate comparison with our model predictions — expressed in absolute gene counts, however this comes at the cost of additional rescaling parameters.

In order to perform an independent validation of modeled gene concentrations, we compared the predicted PDNO gene concentrations to independent cell-count estimates for SUP05 (the dominant nitrate reducer in Saanich Inlet; 16), obtained through quantitative polymerase chain reaction (qPCR). qPCR quantification of SUP05 abundances was performed for water samples collected at 8 distinct depths from the same location and time as for multi-omic sequencing (Fig. 3a in the main article). Water samples (volume $\sim 1$ L) were filtered in the field onto $0.2$ $\mu$m sterivex filters. Samples were not pre-filtered in order to obtain an accurate estimate of total in-situ SUP05 abundances. We used a custom SUP05-specific primer set (Ba519F–1048R) to amplify the 519–1048 region of the SUP05 16S rRNA gene, and followed the protocol described by Hawley *et al.* (8) to estimate the starting template concentration. qPCR was performed in triplicate for each sample. We multiplied the average template concentration for each sample by the volume of extracted fluid ($\sim 200 - 400$ $\mu$L), divided by the volume of filtered seawater, to obtain an estimate for the concentration of SUP05 16S gene copies in seawater. To correct for multiple 16S gene copies in single cells, we divided this concentration by the 16S gene copy number (3.767), estimated for members of the SUP05 clade based on closely related fully sequenced reference genomes. Specifically, we used the 16S gene copy number assigned by the Tax4Fun pipeline (28) to the clade "*Oceanospirillales;SUP05 cluster;uncultured gamma proteobacterium*" in the SILVA 123 database (29). Note that Tax4Fun (28) uses a probabilistic model to assign multiple reference genomes with varying weights to each clade in the SILVA database. Hence, the effective 16S gene copy number assigned by Tax4Fun to each clade is the weighted harmonic mean of the 16S gene copy numbers in each reference genome assigned to that clade.

# S2 Mathematical model

## S2.1 Overview

The gene-centric model describes the spatiotemporal dynamics of $8$ metabolite concentrations and $6$ gene (DNA) concentrations along the Saanich Inlet water column between depths 100–200 m. Each gene is a proxy for a particular redox pathway that couples the oxidation of an external electron donor to the reduction of an external electron acceptor (Appendix S2.2). The model assumes that each cell occupies a single metabolic niche, associated with one of the modeled pathways and thus

one of the considered proxy genes. Reaction rates (per gene) depend on the concentrations of all used metabolites according to 1st order or 2nd order (Michaelis-Menten) kinetics (30, 31) (Appendix S2.4). In turn, the production or depletion of metabolites at any depth is determined by the reaction rates at that depth, taking into account reaction stoichiometry (Appendix S2.3). The production of genes (or more precisely, their host cells) at any depth is driven by the release of energy from their catalyzed reactions, and is proportional to the Gibbs free energy multiplied by the reaction rate (Appendix S2.5) (32). In addition, genes are subject to exponential decay as well as eddy-diffusion and sinking. Metabolites are also subject to eddy-diffusion.

Mathematically, the model is defined as a set of partial differential equations (PDE) for the gene and metabolite concentrations across time and depth. More precisely, the DNA concentration of the $r$-th gene ($\Gamma_r$, copies per volume) at any a given depth $z$ changes according to

$$\frac{\partial \Gamma_r}{\partial t} = - q_r \Gamma_r + \frac{1}{c} Z_r H_r \Gamma_r - \mathrm{v} \frac{\partial \Gamma_r}{\partial z} + \frac{\partial}{\partial z} \left( K(z) \frac{\partial}{\partial z} \Gamma_r \right), \tag{2}$$

and the concentration of the $m$-th metabolite ($C_m$, mole per volume) changes according to

$$\frac{\partial C_m}{\partial t} = \sum_r {}_{mr} H_r \Gamma_r + \frac{\partial}{\partial z} \left( K(z) \frac{\partial C_m}{\partial z} \right). \tag{3}$$

Both the DNA concentrations $\Gamma_r$ and metabolite concentrations $C_m$ depend on time $t$ and depth $z$. The first term in equation (2) corresponds to cell death, with $q_r$ being the exponential death rate for cells hosting gene $r$ in the absence of any metabolites. The 2nd term in (2) corresponds to gene production proportional to the per-gene reaction rate $H_r$ (which in turn depends on metabolite concentrations, see Appendix S2.4). The biomass production coefficient $Z_r$ is a linear function of the Gibbs free energy of the reaction catalyzed by gene $r$ and depends on the reaction quotient (Appendix S2.5). In particular, $Z_r$ increases when product concentrations are low and decreases when substrate concentrations are low. $c$ is the average dry cell mass, which is used to convert biomass production into cell production. The 3rd term in equation (2) corresponds to cell sinking at a constant speed $\mathrm{v}$. The last term in equation (2) and equation (3) corresponds to diffusive transport (33), with $K$ being the vertical eddy-diffusion coefficient. The 1st term in equation (3) corresponds to production or depletion of metabolites due to microbial metabolism. Reaction rates are transformed into metabolite fluxes via the stoichiometric matrix $S$, with entry $S_{mr}$ corresponding to the stoichiometric coefficient of metabolite $m$ in reaction $r$ (Appendix S2.3).

The differential equations (2) and (3) give the rate of change of each metabolite and gene profile, if the profiles are known at a given moment in time. Once all boundary conditions (Appendix S2.6), model parameters (Appendix S2.7) and initial profiles are specified, the model predicts the profiles at any future time point. Steady state profiles were obtained by running simulations of the model until convergence to equilibrium. Because the predicted profiles depend on model parameters, parameters can be calibrated such that the predicted steady state profiles best reproduce the measured data: We used chemical depth profiles to fit poorly known model parameters, thus obtaining a model calibrated to Saanich Inlet's OMZ (Appendix S2.8). This calibrated model was then used to make predictions about steady state DNA profiles, which were compared to measured metagenomic profiles (sections S1.3 and S2.9). This comparison, described in the main article, serves as a test of the model's ability to explain metagenomic profiles in Saanich Inlet's OMZ. Reaction rates at each depth are automatically calculated using the kinetics described in Appendix S2.4.

## S2.2 Considered pathways

The model considers key dissimilatory redox pathways involved in nitrogen and sulfur cycling. When comparing model predictions to molecular data, each pathway was represented by a single gene. For example, nitrous oxide reduction (nosZ gene) coupled to hydrogen sulfide oxidation (dsr, apr and sat genes) is formally represented by nosZ. Other pathways considered by the model were aerobic ammonium oxidation (amo), aerobic nitrite oxidation (nxr), partial denitrification of nitrate to nitrous oxide (PDNO) coupled to sulfide oxidation, anammox (hzo) and remineralization of organic matter via aerobic respiration (ROM). PDNO comprises 3 denitrification steps: Reduction of nitrate to nitrite (narGHIJ or napAB genes), reduction of nitrite to nitric oxide (nirKS genes) and reduction of nitric oxide to nitrous oxide (norBC genes), all three of which are suspected to be predominantly performed by SUP05 $\gamma$-proteobacteria (16, 34). The first denitrification step was assumed to be leaky, so that a small fraction of nitrite was released into the extracellular environment (35). PDNO was represented by norBC genes when comparing the model to molecular data (Fig. 3a in the main article, but see Figures S4d,e,f for narGHIJ, napAB and nirKS multimolecular data). Aerobic ammonium oxidation included a weak production of nitrous oxide (nitrifier denitrification (36)), although the inclusion of this process did not noticeably affect model predictions because most of the nitrous oxide was produced by PDNO. Aerobic respiration of organic matter included the release of ammonium and sulfate at ratios adjusted to measured C:N:S ratios for marine bacterial biomass (37). The choice of pathways follows the hypotheses made by Hawley *et al.* (16) based on metagenomic and metaproteomic depth profiles, as well as reports of nitrous oxide reduction in Saanich Inlet's OMZ (38). Hydrogen sulfide is assumed to originate from the sediments via diffusion, where high rates of sulfate reduction have been observed (39, 40) (Appendix S3.1 for a discussion of this assumption). Figure 1a in the main article gives an overview of the described reaction network. The detailed reaction stoichiometry is given in section S2.3.

## S2.3 Pathway stoichiometry

We list the stoichiometry of the dissimilatory redox pathways considered by the model:

- Remineralization of organic matter through aerobic respiration:

$$\frac{1}{6}\mathrm{POM} + \mathrm{O_2} \xrightarrow{\mathrm{ROM}} \mathrm{CO_2} + \mathrm{H_2O} + \nu\mathrm{NH_4^+} + \sigma\mathrm{SO_4^{2-}} \tag{4}$$

  where POM corresponds to

$$(\mathrm{C_6H_{12}O_6})(\mathrm{NH_4^+})_{6\nu}(\mathrm{SO_4^{2-}})_{6\sigma} \tag{5}$$

  and

$$1 : \nu : \sigma = 1 : 0.184 : 0.0113 \tag{6}$$

  correspond to typical molar $\mathrm{C} : \mathrm{N} : \mathrm{S}$ ratios in marine bacterial biomass (37).

- Aerobic ammonium oxidation:

$$\mathrm{NH_4^+} + \frac{1}{2}(3 - L_{\mathrm{amo}}) \times \mathrm{O_2} \tag{7}$$

$$\xrightarrow{\mathrm{amo}} (1 + L_{\mathrm{amo}}/2) \times \mathrm{H_2O} + (1 - L_{\mathrm{amo}}) \times \mathrm{NO_2^-} + (L_{\mathrm{amo}}/2) \times \mathrm{N_2O} + (2 - L_{\mathrm{amo}}) \times \mathrm{H^+}. \tag{8}$$

  Here $L_{\mathrm{amo}}$ is a parameter representing the fraction of N released as $\mathrm{N_2O}$ via nitrifier denitrification, compared to the total $\mathrm{NH_4^+}$ consumed (36). For example, if $L_{\mathrm{amo}} = 0$, then ammonium is completely oxidized and released as nitrite.

- Aerobic nitrite oxidation:

$$2\mathrm{NO_2^-} + \mathrm{O_2} \xrightarrow{\mathrm{nxr}} 2\mathrm{NO_3^-}. \tag{9}$$

- Anaerobic ammonium oxidation (anammox):

$$\mathrm{NH_4^+} + \mathrm{NO_2^-} \xrightarrow{\mathrm{hzo}} \mathrm{N_2} + 2\mathrm{H_2O}. \tag{10}$$

- Partial denitrification to nitrous oxide (PDNO) coupled to hydrogen sulfide oxidation:

$$(1 - L_{\mathrm{PDNO}}/2) \times \mathrm{H_2S} + 2\mathrm{NO_3^-} \xrightarrow{\mathrm{PDNO}} \tag{11}$$

$$(1 - L_{\mathrm{PDNO}}) \times \mathrm{N_2O} + 2L_{\mathrm{PDNO}} \times \mathrm{NO_2^-} \tag{12}$$

$$+ (1 - L_{\mathrm{PDNO}}/2) \times \mathrm{SO_4^{2-}} + (1 - L_{\mathrm{PDNO}}) \times \mathrm{H_2O} + L_{\mathrm{PDNO}} \times \mathrm{H^+}. \tag{13}$$

  Here, $L_{\mathrm{PDNO}}$ is a parameter representing the fraction of $\mathrm{NO_2^-}$ leaked to the extracellular medium during PDNO, compared to the total $\mathrm{NO_3^-}$ consumed (35).

- Nitrous oxide reduction coupled to hydrogen sulfide oxidation:

$$\mathrm{H_2S} + 4\mathrm{N_2O} \xrightarrow{\mathrm{nosZ}} 4\mathrm{N_2} + \mathrm{SO_4^{2-}} + 2\mathrm{H^+}. \tag{14}$$

- Nitrate reduction to ammonium (DNRA, identified with the nirBD gene):

$$\mathrm{H_2S} + \mathrm{NO_3^-} + \mathrm{H_2O} \xrightarrow{\mathrm{DNRA}} \mathrm{NH_4^+} + \mathrm{SO_4^{2-}}. \tag{15}$$

  DNRA was eventually omitted from the model for reasons described in Appendix S3.2.

## S2.4 Reaction kinetics

Respiration of organic matter involves the hydrolysis of particulate organic mater (POM) to dissolved organic matter (DOM), which is subsequently broken down to simpler organic molecules by fermenters that provide non-fermenting organotrophs with a reactive DOM pool. However, reactive DOM rarely accumulates and most of the DOM pool is expected to be refractory (41). Furthermore, POM degradation has been found to be strongly correlated to bacterial growth in subeuphotic zones,

likely due to limiting POM hydrolysis rates (42). We thus modeled organic matter respiration rates as a first-order function of POM concentrations (43). More precisely, the gene-specific ROM reaction rate, $H_{\mathrm{ROM}}$, is a function of metabolite concentrations $\mathbf{C}$ given by

$$H_{\mathrm{ROM}}(\mathbf{C}) = A_{\mathrm{ROM}} F_T \times \frac{C_{\mathrm{POM}} C_{\mathrm{O_2}}}{C_{\mathrm{O_2}} + K_{\mathrm{ROM,O_2}}}, \tag{16}$$

where $K_{\mathrm{ROM,O_2}}$ is the oxygen half-saturation constant, $A_{\mathrm{ROM}}$ is a first-order rate constant ("affinity") and $F_T$ is the unitless thermodynamic potential factor given by Reed et al. (31) (equation S1)

Half-saturation constants reported for nitrous oxide oxidation are typically in the order of $0.37 - 2.5$ $\mu$M $\mathrm{N_2O}$ (44, 45) and $40$ $\mu$M $\mathrm{H_2S}$ (46), which are well above the typical $\mathrm{N_2O}$ and $\mathrm{H_2S}$ concentrations in the Saanich Inlet OMZ (Fig. 2 in the main article). Sulfide-driven nitrous oxide reduction in Saanich Inlet is therefore likely limited both by electron donor as well as electron acceptor availability. We thus modeled nitrous oxide reduction using first order substrate kinetics with oxygen inhibition:

$$H_{\mathrm{nosZ}}(\mathbf{C}) = A_{\mathrm{nosZ}} F_T \times \frac{C_{\mathrm{N_2O}} C_{\mathrm{H_2S}} K_{\mathrm{nosZ,O_2}}}{C_{\mathrm{O_2}} + K_{\mathrm{nosZ,O_2}}}, \tag{17}$$

where $K_{\mathrm{nosZ,O_2}}$ is the oxygen half-inhibition constant and $A_{\mathrm{nosZ}}$ is a first-order rate constant.

All other gene-specific reaction rates ($H_r$) are modeled using Michaelis-Menten kinetics with possible inhibition (30, 31):

$$H_r(\mathbf{C}) = V_r F_T \times \prod_{\substack{m \text{ reactant} \\ \text{of reaction } r}} \frac{C_m}{K_{rm} + C_m} \tag{18}$$

$$\times \prod_{\substack{n \text{ inhibitor} \\ \text{of reaction } r}} \frac{K^\star_{rn}}{K^\star_{rn} + C_n}. \tag{19}$$

Here, $V_r$ is the maximum gene-specific reaction rate and $K_{rm}$ and $K^\star_{rn}$ are half-saturation and half-inhibition constants, respectively, given in Appendix S2.7. The only explicitly modeled inhibition was oxygen inhibition for anammox (hzo), PDNO and nitrous oxide reduction (nosZ).

## S2.5 Gibbs free energy and gene growth

Following Roden *et al.* (32) and Reed *et al.* (31), we set

$$Z_r = 2.08\gamma^{\mathrm{e}}_r - 0.0211\Delta G_r, \tag{20}$$

(in g biomass per mole reaction flux) where $\gamma^{\mathrm{e}}_r$ is the negative stoichiometric coefficient of the electron donor in the reaction,

$$\Delta G_r = \Delta G^o_r + R_{\mathrm{g}} T \ln Q_r \tag{21}$$

is the Gibbs free energy of the reaction (in kJ per mol), $\Delta G_r^o$ is the standard Gibbs free energy of the reaction and

$$Q_r = \prod_m C_m^{S_{mr}} \tag{22}$$

is the reaction quotient (47). Each $\Delta G_r^o$ depends on the local temperature and pressure and was calculated using the `CHNOSZ` R package (48).

## S2.6 Boundary conditions

Uniquely solving the partial differential equations (2) and (3) requires appropriate boundary conditions (BC) for all genes and metabolites at the top and bottom boundaries ($100\ \mathrm{m}$ and $200\ \mathrm{m}$, respectively). For all metabolites except $N_2$, $N_2O$, $SO_4^{2-}$ and $O_2$, BCs were fixed values set to the average measurements from cruises 41 (SI041_01/13/10), 42 (SI042_02/10/10) and 43 (SI043_03/10/10). For $N_2$ and $N_2O$, lower BCs were set to Neumann (zero flux). For $O_2$, we used Dirichlet BCs (fixed value) with values equal to the average measurements from cruise 42 and 44 (SI044_04/07/10), because $O_2$ data were unavailable for cruises 41 and 43. For $SO_4^{2-}$ we used Dirichlet BCs set to $28\ \mathrm{mM}$ on both sides (43). Metabolite boundary conditions are summarized in Table S2. These boundary conditions result in a net oxygen and nitrate influx from the top as well as an ammonium and sulfide influx from the sediments (40, 49, 50).

Gene boundary conditions were either set to fixed zero (*hzo* and *norBC* top BCs, ROM, *amo* and *nxr* bottom BCs) or to fixed relative gradients (ROM, *amo*, *nxr*, *nirBD* and *nosZ* top BCs, *hzo*, *nirBD*, *norBC* and *nosZ* bottom BCs), with the relative gradient inferred from the metagenomic profiles.

Table S2: Top ($100\ \mathrm{m}$) and bottom ($200\ \mathrm{m}$) boundary conditions for metabolites in the gene-centric partial differential equation model. Numerical values denote Dirichlet boundary conditions. 'N' denotes zero-flux Neumann conditions.

| **Metabolite** | **Top** ($\mu$M) | **Bottom** ($\mu$M) |
|---|---|---|
| $NH_4^+$ | 0 | 8.67 |
| $O_2$ | 77.23 | 0 |
| $NO_3^-$ | 27.59 | 0 |
| $NO_2^-$ | 0.045 | 0 |
| $N_2$ | $4.8 \times 10^{-4}$ | **N** |
| $SO_4^{2-}$ | $28 \times 10^3$ | $28 \times 10^3$ |
| $H_2S$ | 0 | 14.07 |
| $N_2O$ | $24.49 \times 10^{-3}$ | **N** |

## S2.7 Model parameterization

Half-saturation and half-inhibition constants for all involved pathways are listed in Table S3. Maximum cell-specific reaction rates were set to $V_{\mathrm{amo}} = 1.23 \times 10^{-13}\ \mathrm{mol/(cell \cdot d)}$ (51), $V_{\mathrm{nxr}} = 3.26 \times 10^{-13}\ \mathrm{mol/(cell \cdot d)}$ (51) and $V_{\mathrm{hzo}} = 2 \times 10^{-14}\ \mathrm{mol/(cell \cdot d)}$ (52, 53). The nitrifier denitrification fraction $L_{\mathrm{amo}}$ was set to $10^{-4}$, according to nitrifier denitrification fractions of marine ammonium oxidizing archaea measured by Santoro *et al.* (36, Fig 2) over varying $\mathrm{NO_2^-}$ concentrations, and the fact that in Saanich Inlet $\mathrm{NO_2^-}$ concentrations are typically below $2\ \mu\mathrm{M}$ (Fig. 2 in the main article). Because of a lack of reliable information, the rate constants $A_{\mathrm{ROM}}$, $V_{\mathrm{PDNO}}$ and $A_{\mathrm{nosZ}}$, as well as the PDNO leakage fraction $L_{\mathrm{PDNO}}$, were calibrated to chemical profiles as described in Appendix S2.8 and in the main article. Calibration yielded $A_{\mathrm{ROM}} = 5.11 \times 10^{-9}\ \mathrm{L/(cell \cdot d)}$, $V_{\mathrm{PDNO}} = 2.18 \times 10^{-14}\ \mathrm{mol/(cell \cdot d)}$, $A_{\mathrm{nosZ}} = 0.098\ \mathrm{L/(cell \cdot d)}$ and $L_{\mathrm{PDNO}} = 0.352$. An overview of fixed and calibrated reaction-kinetic parameters is provided in Table S3. The sensitivity of the model to parameter variation is illustrated in Appendix S2.12.

The dry cell mass was assumed to be $c = 5 \times 10^{-13}\ \mathrm{g}$, for consistency with the mass used by Roden *et al.* (32) to obtain the regression formula (20). Cell death rates were set to $q_{\mathrm{ROM}} = 0.063\ \mathrm{d^{-1}}$ in accordance with turnover times estimated by Whitman *et al.* (54) for marine prokaryotic heterotrophs above $200\ \mathrm{m}$; to $q_{\mathrm{amo}} = 0.024\ \mathrm{d^{-1}}$ in accordance with average values reported for ammonium oxidizing bacteria (55); to $q_{\mathrm{nxr}} = 0.054\ \mathrm{d^{-1}}$ corresponding to values estimated for nitrite oxidizers (56) and to $0.0033\ \mathrm{d^{-1}}$ for all other genes, in accordance with turnover times estimated by Whitman *et al.* (54) for marine prokaryotes below $200\ \mathrm{m}$.

The concentration of $\mathrm{H^+}$ was fixed to $8.5\ \mathrm{nM}$, corresponding to pH$= 8.07$ (57). The total dissolved inorganic carbon (DIC) was fixed to $2141\ \mu\mathrm{M}$, corresponding to a surface DIC of $2180\ \mu\mathrm{mol/kg}$ (58) and a surface water density of $1018\ \mathrm{kg/m^3}$. Accordingly, the dissolved $\mathrm{CO_2}$ concentration was fixed at $28\ \mu\mathrm{M}$ according to aquatic carbonate equilibrium at the given pH and DIC (59). The particulate organic carbon (POC) profile was calculated from data reported for February 2011 by Luo *et al.* (60) and POM was set to $(1/6)\times$POC (Fig. S1c in the Appendix). Fixing the POM profile circumvents poorly understood physical processes contributing to organic matter fluxes in Saanich Inlet (60). $\mathrm{CO_2}$, $\mathrm{H^+}$ and POM concentrations, while fixed, were still included in the reaction quotients (Appendix S2.5) as well as the reaction-kinetics (Appendix S2.4).

The diapycnal eddy diffusion coefficient $K$ was set to $N^{-2} \cdot 3.7 \times 10^{-10}\ \mathrm{W \cdot kg^{-1}}$, where $N$ is the buoyancy frequency (61, 62). The latter was calculated using temperature and salinity profiles from January 13, 2010, using the `oce` R package (63) (Fig. S1 in the supplement) after loess-smoothing temperature at degree $2$ and salinity at degree $1$, with a span of $75\%$. We chose this time point because the two subsequent temperature and salinity measurements (February 10th and March 10th) were unreliable due to technical problems with our CTD. The cell sinking speed $\mathrm{v}$ was set to $0.1\ \mathrm{m/day}$, in accordance with previous marine microbial ecological models (64, 65).

Table S3: Reaction-kinetic parameters used in the gene-centric model, either calibrated or taken from the literature: Half-saturation substrate concentrations ($K$), half-inhibition concentrations ($K^\star$), cell-specific maximum rates for 2nd order kinetics ($V$), 1st order kinetic constants ($A$, "affinities"), nitrifier denitrification fraction ($L_{\mathrm{amo}}$) and PDNO leakage fraction ($L_{\mathrm{PDNO}}$). The exact role of each parameter is explained in Appendix S2.4. Additional (non-kinetic) fixed model parameters are provided in Appendix 2.7. Clades with members that have been found active in the Saanich Inlet OMZ (16) are marked with a "†".

| **reaction** | **parameter** | **value** | **units** | **organism/region** | **Source** |
|---|---|---|---|---|---|
| ROM | $K_{\mathrm{O_2}}$ | 0.121 | $\mu$M | *Escherichia coli* | (66) |
| | $A$ | 5.11 | $\mathrm{nL}/(\mathrm{cell}\cdot\mathrm{d})$ | | calibr. |
| amo | $K_{\mathrm{NH_4^+}}$ | 0.133 | $\mu$M | *Ca.* Nitrosopumilus maritimus† | (67) |
| | $K_{\mathrm{O_2}}$ | 3.91 | $\mu$M | *Ca.* Nitrosopumilus maritimus† | (68) |
| | $V$ | 123 | $\mathrm{fmol}/(\mathrm{cell}\cdot\mathrm{d})$ | *Nitrosomonas* spp.† | (51) |
| | $L_{\mathrm{amo}}$ | $10^{-4}$ | – | marine ammonia oxidizing archaea† | (36) |
| nxr | $K_{\mathrm{NO_2^-}}$ | 11.7 | $\mu$M | *Nitrospira* spp.† | (69) |
| | $K_{\mathrm{O_2}}$ | 0.78 | $\mu$M | Chilean OMZ | (70) |
| | $V$ | 326 | $\mathrm{fmol}/(\mathrm{cell}\cdot\mathrm{d})$ | *Nitrobacter* sp. | (51) |
| hzo | $K_{\mathrm{NH_4^+}}$ | 3 | $\mu$M | *Ca.* Scalindua sp.† | (26) |
| | $K_{\mathrm{NO_2^-}}$ | 0.45 | $\mu$M | *Ca.* Scalindua sp.† | (26) |
| | $K^\star_{\mathrm{O_2}}$ | 0.2 | $\mu$M | Peruvian OMZ | (31, 71) |
| | $V$ | 20 | $\mathrm{fmol}/(\mathrm{cell}\cdot\mathrm{d})$ | *Planctomycetales*† | (52, 53) |
| PDNO | $K_{\mathrm{NO_3^-}}$ | 2.9 | $\mu$M | marine anoxic basin | (72) |
| | $K_{\mathrm{H_2S}}$ | 2 | $\mu$M | Saanich Inlet OMZ | (73) |
| | $K^\star_{\mathrm{O_2}}$ | 0.1 | $\mu$M | Eastern South Pacific OMZ | (65, 74) |
| | $V$ | 21.8 | $\mathrm{fmol}/(\mathrm{cell}\cdot\mathrm{d})$ | | calibr[1]. |
| | $L_{\mathrm{PDNO}}$ | 35.2 | % | | calibr[2]. |
| nosZ | $K^\star_{\mathrm{O_2}}$ | 0.971 | $\mu$M | low-oxygen activated sludge | (75) |
| | $A$ | 0.098 | $\mathrm{L}/(\mathrm{cell}\cdot\mathrm{d})$ | | calibr. |

[1] Frey *et al.* (76) reports cell-specific thiosulphate-driven denitrification rates for *Sulfurimonas gotlandica* in the range $24.2 - 74.3\ \mathrm{fmol}/(\mathrm{cell}\cdot\mathrm{d})$.

[2] Reported fractions of nitrite leakage during incomplete denitrification ($L_{\mathrm{PDNO}}$) range from $0\%$ to $87\%$ (77–79).

## S2.8 Calibrating reaction-kinetic parameters to data

As described in the previous section, most model parameters were obtained from the literature, however a subset of reaction-kinetic parameters ($A_{\mathrm{ROM}}$, $V_{\mathrm{PDNO}}$, $L_{\mathrm{PDNO}}$ and $A_{\mathrm{nosZ}}$; overview in Table S3) had to be calibrated due to the lack of available information. Here we describe the statistical methods used to calibrate unknown reaction-kinetic model parameters to available chemical depth profile data. The steady state solution of the model defines a mapping from a given choice of parameter values (collectively written as a vector $\mathbf{p}$) to predicted depth profiles for metabolite concentrations, $C_1, C_2, ...$ We assumed that measured concentrations ($\tilde{C}_1, \tilde{C}_2, ..$) are normally distributed:

$$\tilde{C}_i = C_i + \sigma_i \cdot \varepsilon_i. \tag{23}$$

Here, $\varepsilon_i$ is a standard-normally distributed error and $\sigma_i$ is the (unknown) standard deviation of measurement errors (henceforth referred to as *error scale*). We allowed for a different $\sigma_i$ for each metabolite to account for variations in the magnitude of measurement errors.

In the context of our spatial model, the concentrations $C_i$ are predicted as functions of depth, $z$, i.e. $C_i = C_i(z; \mathbf{p})$. Calibration data is given as tuples $(z_{ij}, \tilde{C}_{ij})$, where each $\tilde{C}_{ij}$ is a measurement of the $i$-th concentration at some depth $z_{ij}$ and $j$ enumerates all measurements of the $i$-th concentration. The overall log-likelihood function for such a data set is given by

$$l(\boldsymbol{\sigma}, \mathbf{p}) = -\sum_{i,j} \ln\left(\sigma_i \sqrt{2\pi}\right) \tag{24}$$

$$-\sum_{i,j} \frac{1}{2\sigma_i^2} \left[\tilde{C}_{ij} - C_i(z_{ij}; \mathbf{p})\right]^2. \tag{25}$$

The model was calibrated by maximizing the log-likelihood $l(\boldsymbol{\sigma}, \mathbf{p})$ by choice of the error scales $\sigma_i$ and the parameter values $\mathbf{p}$. This calibration method is known as *maximum-likelihood* (ML) *estimation*, and is widespread in statistical regression and physics (80). Maximization of the log-likelihood was performed using the MATLAB® function `fmincon`, which uses repeated simulations and gradual exploration of parameter space (81). The following chemical concentration data were used for calibration: $NH_4^+$, $NO_3^-$, $NO_2^-$, $N_2O$ and $H_2S$ from cruises 41–43, and $O_2$ from cruises 42 and 44.

## S2.9 Calibrating multi-omic data units

Metagenomic, metatranscriptomic and metaproteomic data are given only in relative units. For example, the correspondence between metagenomic RPKM values and actual DNA concentrations in the water column is, a priori, unknown. In fact, RPKM values for different genes may correspond to different gene concentrations due to detection biases (82–84). Furthermore, model predictions regarding RNA and protein abundances are in arbitrary units because the transcriptional, translational and enzymatic efficiency of proteins is unknown and differs between proteins.

In order to compare model predictions to multi-omic sequence data, we assumed that each measured DNA, mRNA and protein abundance profile is related to the corresponding model prediction by

a constant linear conversion factor. Conversion factors were estimated via maximum-likelihood estimation, separately for each molecule to account for detection biases. More precisely, for each data set we assumed a normal error distribution as already described in Appendix S2.8. Hence, measured environmental biomolecule concentrations, for example *amo* DNA concentrations, are distributed as

$$\tilde{\Gamma}_i = \Gamma_i/\beta_i + \sigma_i \cdot \varepsilon_i, \tag{26}$$

where $\varepsilon_i$ are uncorrelated standard-normally distributed errors, scaled by an unknown factor $\sigma_i$, and $\beta_i$ is the unknown proportionality factor between *amo* metagenomic RPKM values $\tilde{\Gamma}_i$ and actual DNA concentrations. The log-likelihood of a measured depth profile comprising $N_i$ data points, $(z_{i1}, \tilde{\Gamma}_{i1}), .., (z_{iN_i}, \tilde{\Gamma}_{iN_i})$, is thus

$$l_i(\sigma_i, \mathbf{p}) = -\sum_{j=1}^{N_i} \ln\left(\sigma_i\sqrt{2\pi}\right) - \sum_{j=1}^{N_i} \frac{1}{2\sigma_i^2}\left[\tilde{\Gamma}_{ij} - \Gamma_i(z_{ij};\mathbf{p})/\beta_i\right]^2. \tag{27}$$

For any fixed model parameter choice $\mathbf{p}$ (and therefore fixed predictions $\Gamma_i$), the log-likelihood $l_i(\sigma_i; \mathbf{p})$ is maximized by choosing

$$\beta_i = \sqrt[N_i]{\prod_{j=1}^{N_i} \frac{\Gamma_i(z_{ij};\mathbf{p})}{\tilde{\Gamma}_{ij}}}, \tag{28}$$

(i.e. the geometric mean of model predictions over measurements) and

$$\sigma_i^2 = \frac{1}{N_i}\sum_{j=1}^{N_i} \left|\tilde{\Gamma}_{ij} - \Gamma_j(z_{ij};\mathbf{p})/\beta_i\right|^2. \tag{29}$$

Choosing $\beta_i$ as in equation (28) yields maximum-likelihood estimates for the appropriate conversion factors between metagenomic units (RPKM) and actual DNA concentrations (genes/L) (see table S4 in the supplement). Inserting the estimated $\beta_i$ and $\sigma_i$ back into equation (27) yields the log-likelihood of the particular metagenomics data set and for a particular choice of model parameters $\mathbf{p}$. A similar approach was used to compare metatranscriptomic and metaproteomic data sets to model predictions (Appendix S2.10).

The estimated proportionality factors $\beta_i$ are listed in table S4 of the supplement, and range from $3.9 \times 10^4$ genes $\cdot\, \mathrm{L}^{-1} \cdot \mathrm{RPKM}^{-1}$ for *norBC* up to $3.3 \times 10^7$ genes $\cdot\, \mathrm{L}^{-1} \cdot \mathrm{RPKM}^{-1}$ for ROM. These differences may be due to variable DNA extraction efficiencies across cells, uneven community sampling due to filter-size partitioning (19) or differences in gene copy numbers per cell. Additionally, the assumption of a common cell mass for all modeled genes may have resulted in an inaccurate conversion of predicted biomass production to gene production. However, the good overall agreement between predicted functional gene concentrations and SUP05 abundances (Fig. 3 in the main article) suggests that this may only be a minor problem.

Table S4: Proportionality factors ($\beta$) between environmental gene abundances and metagenomic RPKM values (in $\mathrm{genes}\cdot\mathrm{L}^{-1}\cdot\mathrm{RPKM}^{-1}$), as defined in Appendix S2.9. Estimated by comparing the predictions of the calibrated model with metagenomic data from February 10, 2010. Unambiguous metagenomic data was not available for *nxr* (see Appendix S3.4).

| gene | $\beta$ |
|---|---|
| ROM | $4.1 \times 10^7$ |
| *amo* | $1.0 \times 10^6$ |
| *nxr* | NA |
| *hzo* | $3.2 \times 10^5$ |
| *norBC* | $3.4 \times 10^4$ |
| *nosZ* | $4.5 \times 10^4$ |

## S2.10 Predicting metatranscriptomic and metaproteomic profiles

A priori, the gene-centric model makes no predictions regarding mRNA or protein dynamics; in fact transcription and translation are circumvented by assuming that the release of energy manifests directly as DNA replication. To explore the possibility of explaining mRNA and protein distributions in Saanich Inlet's OMZ, we extended the model to a set of hypothetical mechanisms driving the production, decay and dispersal of these molecules. More precisely, we assumed that mRNA and protein production rate at a particular depth is proportional to the total reaction rate at that depth ($H_r\Gamma_r$), and that mRNA and proteins disperse similarly to genes (Appendix S2.10). The assumption that mRNA and protein production rates are proportional to reaction rates is motivated by observations of a positive relation between transcription and translation rates and metabolic activity or growth (85–87). A linear relation, in particular, may be justified by the fact that increased enzyme dilution rates at elevated cell growth must be balanced (at the population level) by correspondingly increased translation (and hence transcription) rates (88).

This simple description introduces two unknown parameters per mRNA or protein: The proportionality factor that converts reaction rates to molecule production rates, and the decay time of molecules following production. We calibrated both parameters using metatranscriptomic and metaproteomic depth profiles and then checked how well the latter could be reproduced. Our methodology is described for mRNA in detail below. Protein dynamics were modeled and compared to metaproteomic data in a similar way.

As mentioned, our first assumption was that the mRNA production rate (transcripts produced per time and per volume of seawater) at a particular depth is proportional to the total reaction rate (mol per time and per volume of seawater) at that depth. We also assumed that mRNA molecules disperse via diffusion and sinking similarly to genes, as they are hosted by the same cells. Thus, environmental mRNA concentrations satisfy the partial differential equation

$$\partial_t T_r = -T_r/\tau_r + R_r/\alpha_r - \mathrm{v}\partial_z T_r + \partial_z K \partial_z T_r, \tag{30}$$

where $T_r(t, z)$ is the mRNA concentration corresponding to the $r$-th reaction, $\tau_r$ is the decay time

of the mRNA molecule, $R_r(t,z) = H_r(t,z)\Gamma_r(t,z)$ is the total reaction rate at depth $z$ and $\alpha_r$ is an unknown proportionality constant. We considered $T_r$ in the same units as the multi-omic data (i.e. RPKM for metatranscriptomes and NSAF for metaproteomes). Consequently, $\alpha_r$ is the ratio between the $r$-th reaction rate (mol per time per vole of seawater) and the corresponding RPKM (or NSAF) "production rate" (RPKM per time), and thus not only depends on the particular reaction, but also on our sampling protocol and sequencing pipeline.

For each gene $r$, the transcript profile $T_r$ will satisfy the same boundary conditions as the DNA profile $\Gamma_r$, provided that the latter are either zero value (Dirichlet), zero flux (Neumann) or fixed relative gradient boundary conditions (Appendix S2.6). We calculated the steady state solution of equation (30), $T_r^\star$, by solving the time-invariant equation

$$0 = -T_r^*/\tau_r + R_r/\alpha_r - \mathrm{v}\partial_z T_r^* + \partial_z K \partial_z T_r^* \tag{31}$$

using the MATLAB function `bvp4c` (81). This was done after the gene-centric model had already reached steady state, at which point the reaction rates $R_r$ are time-independent functions of depth.

Note that the steady state profile $T_r^\star(z)$ is proportional to $1/\alpha_r$, all else being equal. Hence, by comparing $T_r^\star$ to metatranscriptomic data (for some given $\tau_r$), the constant $\alpha_r$ can be calibrated via maximum-likelihood estimation as described in Appendix S2.9. On the other hand, maximizing the log-likelihood in equation (27) (separately for each gene) by choice of $\alpha_r, \tau_r$ and the corresponding error scale, yields an estimate of the decay time $\tau_r$. This was done through repeated solutions of equation (31) with varying $\tau_r$ and using the interior-point optimization algorithm implemented by the MATLAB function `fmincon` (81). We confined the fitted $\tau_r$ to between $10^{-4}$ and $10^5$ days.

After calibration of the decay time $\tau_r$ and proportionality factor $\alpha_r$, we calculated the coefficients of determination,

$$R_r^2 = 1 - \frac{\sum_j \left[\tilde{T}_{rj} - T_r(z_{rj})\right]^2}{\sum_j \left[\tilde{T}_{rj} - \overline{T}_r\right]^2}, \tag{32}$$

to evaluate how well the mRNA model explained the metatranscriptomic data. Here, $\tilde{T}_{r1}, \tilde{T}_{r2}, ..$ are measured mRNA abundances at depths $z_{r1}, z_{r2}, ..$ and $\overline{T}_r$ is their average. For any given gene $r$, $R_r^2$ is a measure for the *goodness of fit* of the above model to the multi-omic data. Table S5 in the supplement lists the results for all genes for which $R_r^2 \geq 0.5$.

The statistical significance (*P-value*) of the obtained $R^2$ was defined as the probability of obtaining the same or greater $R^2$ by applying the same procedure to a random data set, with independent normally distributed values with mean and standard deviation set to the original sample mean and standard deviation. We estimated the P-values for cases where $R_r^2 \geq 0.9$ using Monte Carlo simulations of $1000$ random data sets: all of them were estimated below $0.005$.

## S2.11 Calculating metabolic fluxes between pathways

Dissimilatory metabolic reactions can be interpreted as sources and sinks of metabolites distributed along the water column, producing and consuming metabolites at rates given by the first term in

Table S5: Proportionality factors ($\alpha$) between mRNA or protein production rates and reaction rates (in $\mathrm{mol}/(\mathrm{L}\cdot\mathrm{RPKM})$ or $\mathrm{mol}/(\mathrm{L}\cdot\mathrm{NSAF})$, respectively), exponential mRNA or protein decay times ($\tau$) and coefficients of determination ($R^2$), estimated as described in Appendix S2.10. Only cases with $R^2 \geq 0.4$ are shown.

| molecule | | $\alpha$ | $\tau$ (days) | $R^2$ |
|---|---|---|---|---|
| nxr | mRNA | $9.7 \times 10^{-10}$ | 52.2 | 0.93 |
| nosZ | mRNA | $2.5 \times 10^{-8}$ | 222 | 0.95 |
| ROM | protein | $3.9 \times 10^{-3}$ | 89.2 | 0.59 |
| amo | protein | $9.6 \times 10^{-6}$ | 16.6 | 0.88 |
| nxr | protein | $7.5 \times 10^{-5}$ | 123 | 0.42 |
| norBC | protein | $4.9 \times 10^{-4}$ | 329 | 0.69 |

equation (3). Due to diffusive transport (2nd term in equation (3)), metabolite fluxes from sources to sinks need not be localized and can span across different depths. Furthermore, some metabolites are partly transported across the OMZ boundaries, towards or from the top layers or the sediments. In the following we describe our approach for calculating steady-state metabolite fluxes across individual reactions.

Let us focus on a particular metabolite and consider a single hypothetical particle created at time 0 at depth $x$. Let $G(t, x, y)$ be the Green's function of the dispersal-destruction model, so that $G(t, x, \cdot)$ is the distribution density of a particle (created at depth $x$) at depth $y$ and after time $t$. Note that $G(t, x, \cdot)$ may integrate to less than unity if the particle has a positive probability of being consumed anywhere in the water column. The probability rate at which that particle is consumed by any sink $j$ at time $t$ is then

$$\int dy\, G(t, x, y) \frac{\lambda_j(y)}{C(y)}, \tag{33}$$

where $\lambda_j(y)$ gives the rate at which sink $j$ consumes particles at depth $y$ and $C(y)$ is the steady state metabolite concentration at that depth. Since each sink corresponds to a pathway consuming the metabolite, $\lambda_j(y)$ is given by the community-wide reaction rate at $y$ multiplied by the appropriate stoichiometric coefficient. The probability that the particle will eventually be destroyed by sink $j$ is given by

$$\int_0^\infty dt \int dy\, G(t, x, y) \frac{\lambda_j(y)}{C(y)}. \tag{34}$$

The total rate at which particles created by source $i$ are destroyed by sink $j$ across the entire OMZ, denoted $F_{ij}$, is

$$F_{ij} = \int dx\, b_i(x) \int_0^\infty dt \int dy\, G(t, x, y) \frac{\lambda_j(y)}{C(y)}, \tag{35}$$

where $b_i(x)$ is the rate at which the metabolite is produced by source $i$ at depth $x$. Switching integrals in (35) gives

$$F_{ij} = \int dy \, \frac{\lambda_j(y)}{C(y)} \int_0^\infty dt \, \vartheta_i(t, y), \tag{36}$$

where

$$\vartheta_i(t, y) = \int dx \, b_i(x) G(t, x, y) \tag{37}$$

is the solution to the dispersal-destruction model with initial distribution $b_i(x)$:

$$\partial_t \vartheta_i(t, y) = -\frac{\vartheta_i(t, y)}{C(y)} \sum_j \lambda_j(y) + \partial_y \left[ K(y) \partial_y \vartheta_i(t, y) \right], \tag{38}$$

$$\vartheta_i(0, y) = b_i(y). \tag{39}$$

Particles crossing the domain boundary are considered to be lost. Hence, Dirichlet (Neumann) boundary conditions in the original model correspond to zero-value (zero-flux) boundary conditions for $\vartheta_i$. The total boundary loss rate of particles created by source $i$ is the remainder

$$F_{i,o} = \int dx \, b_i(x) - \sum_j F_{ij}. \tag{40}$$

Similarly, the rate at which particles flow in at the boundary and are destroyed by sink $j$ is given by

$$F_{o,j} = \int dx \, \lambda_j(x) - \sum_i F_{ij}. \tag{41}$$

We solved equation (38) using the MATLAB® function `pdepe` and evaluated all integrals in equation (36), (40) and (41) using the trapezoid integration scheme (81).

## S2.12 Local sensitivity analysis

We evaluated the sensitivity of the model predictions to small changes in model parameters using normalized local sensitivity coefficients (NLSC) (89). NLSCs compare the relative changes in model output variables ($V_j$, integrated over all depths) to the relative changes of model parameters ($p_i$) by means of partial derivatives, evaluated at the default (e.g. fitted) parameter values:

$$\mathrm{NLSC}_{ij} = \left| \frac{p_i}{V_j} \frac{\partial V_j}{\partial p_i} \right|. \tag{42}$$

Hence, $\mathrm{NLSC}_{ij}$ is a measure for the relative effects that parameter $i$ has on the output variable $j$. The partial derivative in equation (42) was approximated numerically by changing $p_i$ by $1\%$ from its default value. The results are summarized in figure S3 in the supplement.

The sensitivity of the model varied strongly among parameters. For example, the kinetic constants for ROM (aerobic remineralization of organic matter) had a relatively strong effect on chemical as well as gene concentration profiles by modulating the availability of oxygen and ammonium near and above the SNTZ. On the other hand, the kinetic constants for PDNO and nosZ (which constitute the denitrification pathway) had relatively little effects on the predicted chemical profiles, as long as both were increased or decreased in unison. Similar observations were made for amo and nxr, which constitute the nitrification pathway. Moreover, the total predicted gene concentrations (Fig. 3 in the main article) were robust against parameter changes and only varied within an order of magnitude as long as the calibrated geochemical profiles matched the data moderately well. This suggests that geochemical fluxes are good predictors for microbial growth, but less suited for estimating reaction-kinetic parameters, especially when these are correlated (90).

# S3 Caveats and special notes

## S3.1 The role of sulfate reduction

The choice of pathways included in the model was based on metaproteomics data by Hawley *et al.* (16). None of the proteins associated with sulfur-metabolism were mapped to known sulfate reducers, suggesting that these proteins may act in sulfur oxidation and that sulfate reduction only played a minor role in Saanich Inlet's OMZ at the time of sampling. In particular, an NCBI BLASTP search mapped all detected dsrA and aprAB proteins to SUP05 (91). All other taxonomically resolved sulfite reductase proteins were mapped to *Candidatus* Ruthia magnifica, a sulfur-oxidizing endosymbiont (92). The mRNA depth profiles of sat, aprAB and dsrAB (Figs S4a,b,c in the Appendix), which comprise the dissimilatory sulfide oxidation pathway (or sulfate reduction pathway when reversed), show a clear peak at the SNTZ, consistent with the metatranscriptomic profiles of norBC and nosZ (Fig. 3 in the main article). These multimolecular data suggest that the sat, aprAB and dsrAB enzymes act predominantly in sulfur oxidation. The high sat, aprAB and dsrAB gene concentrations at the bottom might be due to sediment resuspension, cell sinking from the more productive SNTZ or cell diffusion from the sulfate reducing sediments (60, 93).

Due to the much higher organic matter concentrations in the sediments, heterotrophic sulfate reduction and anaerobic remineralization is correspondingly higher in the sediments than in the water column (39, 40). Hence, most of the $H_2S$ and $NH_4^+$ in the sulfidic part of the OMZ is expected to originate from the adjacent sediments via diffusion. An influx of $H_2S$ and $NH_4^+$ predominantly from the sediments is compatible with the measured steep $H_2S$ and $NH_4^+$ gradients (Figs 2 b,f in the main article), as well as the gradual upward progression of the $H_2S$ and $NH_4^+$ fronts following annual renewal (Figs 1b and 2 b,f in the main article). Sediments have previously been indicated as the main sulfide sources in other OMZs, such as the the Eastern Boundary upwelling system (94) or the central Namibian coastal upwelling zone (95).

Due to the lack of rate measurements heterotrophic sulfate reduction and cryptic sulfur cycling cannot be completely ruled out. However, calibrating the above model to the chemical data (Fig.

2 in the main article), while including sulfate reduction as an additional pathway, dramatically decreases the *goodness of fit*. This is because an additional $\mathrm{H_2S}$ source in the OMZ shifts the SNTZ further up, thereby increasing the main discrepancy between the model and the data. Hence, on grounds of parsimony, we eventually omitted sulfate reduction from the model and assumed that $\mathrm{H_2S}$ originates from the sediments via diffusion.

We note that similar theoretical work by Reed *et al.* (31) did suggest the existence of a cryptic sulfur cycle in the Arabian Sea OMZ. However, the latter is located more than $1\ \mathrm{km}$ above the sediments and hydrogen sulfide influx from the sediments into the OMZ is not possible due to elevated oxygen levels below the OMZ (96).

## S3.2 The role of DNRA

It has been previously hypothesized that dissimilatory nitrate reduction to ammonium (DNRA) might be active in Saanich Inlet's OMZ, possibly providing ammonium to anammox bacteria (16, 35, 97). So far DNRA was not detected in any of our incubation experiments, although we cannot rule out cryptic DNRA due to rapid ammonium consumption by anammox (97). Measured ammonium profiles in Spring 2010 did not indicate a significant ammonium source at or below the SNTZ (Fig 2 b in the main article). Similarly, Schunck *et al.* (94) reports negligible DNRA for a sulfidic OMZ off the coast of Peru.

Nevertheless, we tested an extension of our model with DNRA as an additional pathway. Calibrating the model to the same data (January–March 2010) consistently predicted negligible DNRA rates, and the goodness of fit (in terms of the log-likelihood) did not significantly improve with the inclusion of DNRA. On grounds of parsimony we thus eventually omitted DNRA from the model. We mention that calibrating the model to chemical data from September 2009 (16) indicated significant DNRA as well as anammox rates (both in the order of $1\ \ \mathrm{mmol\,N/(m^2 \cdot d)}$), suggesting that DNRA-fed anammox activity fluctuates strongly throughout the year. High spatiotemporal variability of $N$-loss activities are known for other OMZs and may be associated with fluctuations in surface primary production, as well as fluctuations in electron acceptor availability driven by annual deep water renewal (98–100).

## S3.3 The role of aerobic sulfide oxidation

Extensive previous work points towards $\mathrm{NO_3^-}$ and other nitrogen compounds as dominant electron acceptors for $\mathrm{H_2S}$ oxidation in Saanich Inlet during periods of strong stratification (3, 5, 16, 101, 102). For example, as shown in Fig. 1b in the main article, the upper boundary of $\mathrm{H_2S}$ concentrations closely follows the lower boundary of $\mathrm{NO_3^-}$ — rather than $\mathrm{O_2}$ — over time, especially during the period considered in this study (early 2010). The strong similarity between sulfur cycling gene profiles and denitrification gene profiles (February 10, 2010; Fig. S4) provides further evidence for the tight coupling between denitrification and sulfide oxidation at that time. Similarly, nitrogen

compounds have been shown to be the dominant electron acceptors for sulfide oxidation in the Peruvian OMZ (94), and Canfield *et al.* (103) established a strong link between sulfide oxidation and nitrate reduction in the Chilean OMZ. Note that during renewal events in Fall, $O_2$ can indeed become an important electron acceptor for $H_2S$ oxidation in Saanich Inlet (3). This does not, however, affect this study, which focuses on periods of intense stratification near steady state conditions.

We note that we had initially considered aerobic sulfide oxidation as an additional reaction in our model. Preliminary calibrations to geochemical data showed that the model's explanatory power was significantly compromised by this reaction, because diffusive $O_2$ fluxes into the sulfidic zone could not account for the $O_2$ needed for sulfide oxidation (in addition to $O_2$ needed for nitrification). In fact, in our simulations ammonium ended up competing with $H_2S$ for $O_2$, which in turn negatively affected the accuracy of the predicted $NO_3^-$ profile. While lateral intrusions of oxygenated water could in principle account for the additional $O_2$ needed for sulfide oxidation, spatiotemporal $O_2$ profiles do not provide any indication of such intrusions during this period of intense stagnation (Fig. 1b in the main article). We thus omitted aerobic sulfide oxidation from our final model.

### S3.4 Planctomycetes and nxr

Our molecular data suggest that the anammox bacteria planctomycetes (23) are also aerobically oxidizing nitrite to nitrate in the oxycline (16) using the nitrate oxidoreductase narGHIJ (15). Metatranscriptomic and metaproteomic profiles of planctomycete-associated narGHIJ sequences peak at about $120\ \mathrm{m}$ depth and decrease rapidly below that (Fig. 3 in the main article), while planctomycete-associated HAO (anammox-associated hydroxylamine-oxidoreductase (15)) sequences are most abundant at $150\ \mathrm{m}$ depth and at appreciable levels all the way down to $200\ \mathrm{m}$. As a consequence, narGHIJ is expected to also proliferate in regions where it is not actually being transcribed. Indeed, metagenomic data show a bimodal profile of Planctomycete-associated narGHIJ sequences, with local maxima at $120\ \mathrm{m}$ and $150\ \mathrm{m}$ depths, corresponding to the putative maxima of nitrite oxidation and anammox activity. Due to this bimodality we did not include *narGHIJ* nor *nxr* metagenomic profiles in our analysis.

## S4 Simulation code

All simulations, model calibration and sensitivity analysis were performed with MATLAB® (81). The complete code is available upon request from Stilianos Louca. In the code, the biochemical model is defined as a list of genes, a list of metabolites and a stoichiometric matrix for all involved pathways. In addition, the user can specify optional depth profile data sets for chemical concentrations as well as metagenomics, metatranscriptomics and metaproteomics. These are then automatically compared to the model predictions, or used for model calibration. The diffusion coefficient can be provided as an external data set (e.g. calculated from standard CTD data), or internally as a mathematical function. Boundary conditions for the partial differential equations can be specified as Dirichlet (i.e. fixed value) or Neumann (i.e. fixed derivative), independently for

each metabolite or gene. The set of parameters to be calibrated or perturbed for sensitivity analysis can be customized in the code.

# S5 Inverse linear transport modeling (ILTM)

Chemical concentration profiles were used to estimate denitrification and anammox rates across the water column, independently of the gene-centric model and the rate measurements described in Appendix S1.5. In short, a steady state diffusion model was used to estimate the net metabolite production (or consumption) rates that "best" explained the observed depth profiles. This so called inverse linear transport modeling (ILTM) approach is widespread in oceanography and atmospheric sciences, were known global distributions of compounds such as trace gases are used to estimate unknown sources and sinks (104, 105).

In the following, we explain our procedure for estimating the net production profile, $\rho(z)$, for a particular metabolite with a given concentration profile, $\hat{C}(z)$. All calculations were performed in Mathworks MATLAB®. Each profile $\hat{C}(z)$ was obtained through Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation of the actual measured concentrations. ILTM was applied separately to concentration profiles from cruises 47 and 48, as well as to the chemical profiles used for model calibration (cruises 41–44, Appendix S2.8) after averaging across replicates at each depth.

Our starting point is the diffusive transport model

$$0 = \rho + \frac{\partial}{\partial z}\left[K(z) \cdot \frac{\partial C}{\partial z}\right], \tag{43}$$

which describes the steady-state distribution $C(z)$ across depth $z$, given a particular net production profile $\rho(z)$ and eddy diffusion coefficient $K(z)$. The eddy diffusion coefficient was calculated as described in Appendix S2.7. Our goal is to determine the appropriate $\rho(z)$ that "best" explains the observed steady state profile $\hat{C}(z)$, through the following steps:

1. Calculate the discretized Green's function (106) of the above partial differential equation (PDE) with zero Dirichlet boundary conditions: Let $G_{nm}$ be an approximation for $G(z_n, z_m)$, where $G$ solves the time-independent PDE

$$0 = \frac{\partial}{\partial x}\left[K(x)\frac{\partial}{\partial x}G(x,y)\right] + \delta(x-y) \tag{44}$$

on the domain $\Omega := [\text{top}, \text{bottom}]$, with boundary conditions

$$G(x,y)\big|_{x\in\partial\Omega} = 0. \tag{45}$$

In practice, $G_{nm}$ can be set to $dz_m \cdot G(z_n, z_m)$, where $G$ is the solution to the PDE system

$$0 = \frac{\partial}{\partial x}\left[K(x)\frac{\partial}{\partial x}G(x,z_m)\right] + H(x - z_m + dz_m/2)H(z_m + dz_m/2 - x)/dz_m, \tag{46}$$

$$G(x,z_m)\big|_{x\in\partial\Omega} = 0. \tag{47}$$

Here, $H$ is the Heaviside step function and $dz_m$ is the grid's step at $z_m$, assumed to be chosen small enough ($dz = 2$ m in our case).

2. Note that for any candidate net production profile $\rho(x)$, the sum

$$\sum_m G_{nm} \cdot \rho(z_m) \tag{48}$$

becomes an approximation for $C^o(z_n)$, where $C^o$ is a solution to the following steady-state transport problem with zero Dirichlet boundary conditions:

$$0 = \frac{\partial}{\partial x}\left[D(x)\frac{\partial C^o}{\partial x}\right] + \rho(x), \quad C^o(x)\big|_{x\in\partial\Omega} = 0. \tag{49}$$

3. For the given measured concentrations $\hat{C}(x)$ at the domain boundary $x \in \{\text{top}, \text{bottom}\}$, calculate the particular solution $C^{\mathrm{p}}$ to the transport problem with given boundary values but no sources:

$$0 = \frac{\partial}{\partial x}\left[K(x)\frac{\partial C^{\mathrm{p}}}{\partial x}\right], \quad C^{\mathrm{p}}(x)\big|_{x\in\partial\Omega} = \hat{C}(x). \tag{50}$$

After solving for $C^{\mathrm{p}}$, evaluate $C^{\mathrm{p}}$ on the grid, i.e. set $C^{\mathrm{p}}_n = C^{\mathrm{p}}(z_n)$.

4. Note that for any candidate net production profile $\rho(x)$, the sum $C := C^o + C^{\mathrm{p}}$ is a solution to the full PDE problem

$$0 = \frac{\partial}{\partial x}\left[K(x)\frac{\partial}{\partial x}C(x)\right] + \rho(x), \quad C(x)\big|_{x\in\partial\Omega} = \hat{C}(x). \tag{51}$$

Similarly, the sum

$$C^{\mathrm{p}}_n + \sum_m G_{nm} \cdot \rho(z_m) \tag{52}$$

is an approximation for $C(z_n)$.

5. Note that $C^{\mathrm{p}}$ corresponds to the hypothetical steady-state profile that would result purely from transport across the domain boundary, in the absence of any sources or sinks in its interior. Similarly, the difference $B = \hat{C} - C^{\mathrm{p}}$ is the part that cannot be explained by transport across boundaries, but must rather be attributed to production and consumption inside $\Omega$. Hence, using the particular discretized solution $C^{\mathrm{p}}_n$, the discretized profile $\hat{C}_n = \hat{C}(z_n)$ and the discretized steady-state transport kernel $G_{nm}$, one could in principle estimate $\rho_m = \rho(z_m)$ by minimizing the sum of squared residuals (SSR)

$$\mathrm{SSR} = \sum_n \Big| \sum_m G_{nm} \cdot \rho_m - B_n \Big|^2, \tag{53}$$

where $B_n = \hat{C}_n - C^{\mathrm{p}}_n$. The above problem is a classical linear least-squares problem if one considers $G_{nm}$ as a matrix ($\mathbb{G}$) and $\rho_m$, $\hat{C}_n$, $C^{\mathrm{p}}_n$ as vectors ($\boldsymbol{\rho} \in \mathbb{R}^M$, $\hat{\mathbf{C}} \in \mathbb{R}^N$ and $\mathbf{C}^{\mathrm{p}} \in \mathbb{R}^N$):

$$\mathrm{SSR} = \left\| \mathbb{G} \cdot \boldsymbol{\rho} - \mathbf{B} \right\|^2. \tag{54}$$

The minimum SSR is then obtained for

$$\boldsymbol{\rho} = \tilde{\mathbb{G}} \cdot (\hat{\mathbf{C}} - \mathbf{C}^{\mathrm{p}}), \tag{55}$$

where $\tilde{\mathbb{G}}$ is the Moore-Penrose pseudoinverse of $\mathbb{G}$. Put simply, the so estimated $\boldsymbol{\rho}$ is the net production profile that "best" explains the observed steady-state concentration profile $\hat{\mathbf{C}}$, after subtracting the part $\mathbf{C}^{\mathrm{p}}$ explained by transport across the domain boundaries.

6. The least-squares estimator in Eq. (55) becomes unstable if the reference profile $\hat{C}$ stretches linearly (or almost linearly) across large depth intervals, leading to spurious oscillations in the estimated profile $\rho$. To address this problem, we "penalized" strong oscillations in the estimated net production profile by instead minimizing the modified SSR

$$\mathrm{SSR}^* = \|\mathbb{G} \cdot \boldsymbol{\rho} - \mathbf{B}\|^2 + M^{-2} \|\xi \boldsymbol{\rho}\|^2, \tag{56}$$

where $\xi$ is an appropriately chosen regularization parameter (107) that quantifies the penalty imposed on large $|\boldsymbol{\rho}|$. The above regularization method is known as Tikhonov regularization. A larger Tikhonov factor $\xi$ will typically result in a smoother $\boldsymbol{\rho}$ but also a poorer overall fit, since goodness of fit is sacrificed in favor of small $\boldsymbol{\rho}$. We manually chose $\xi$ as large as possible but still small enough such that the residual $\|\mathbb{G} \cdot \boldsymbol{\rho} - \mathbf{B}\|$ remained much smaller than $\|\mathbf{B}\|$.

7. Assuming that $H_2S$ is mostly consumed by denitrification (PDNO and nosZ) according to the stoichiometry given in Appendix S2.3, one mol of consumed $H_2S$ corresponds to $8 \cdot (1 - L_{\mathrm{PDNO}})/(5 - 3L_{\mathrm{PDNO}})$ mol N released as $N_2$. Similarly, one mol of consumed $NH_4^+$ by anammox corresponds to 2 mol N released as $N_2$, however nitrification likely also contributes to $NH_4^+$ consumption in the more oxygenated layers. Hence, whenever the net $NO_3^-$ production was positive, the net $NO_3^-$ production rate was subtracted from the net $NH_4^+$ consumption rate, yielding an estimate for $NH_4^+$ consumption purely by anammox.

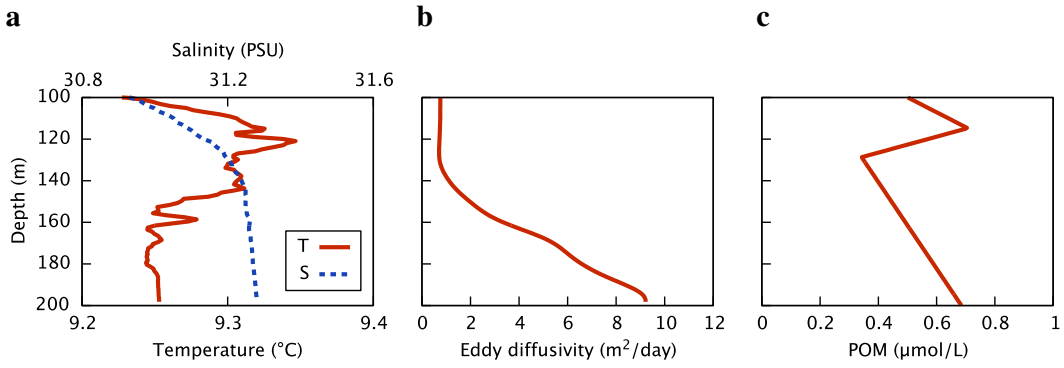


Figure S1: (a) Temperature and salinity profiles at Saanich Inlet main station, January 13, 2010. (b) Corresponding smoothened eddy diffusivity profile, as used in the simulations. (c) Fixed POM profile used in the simulations.

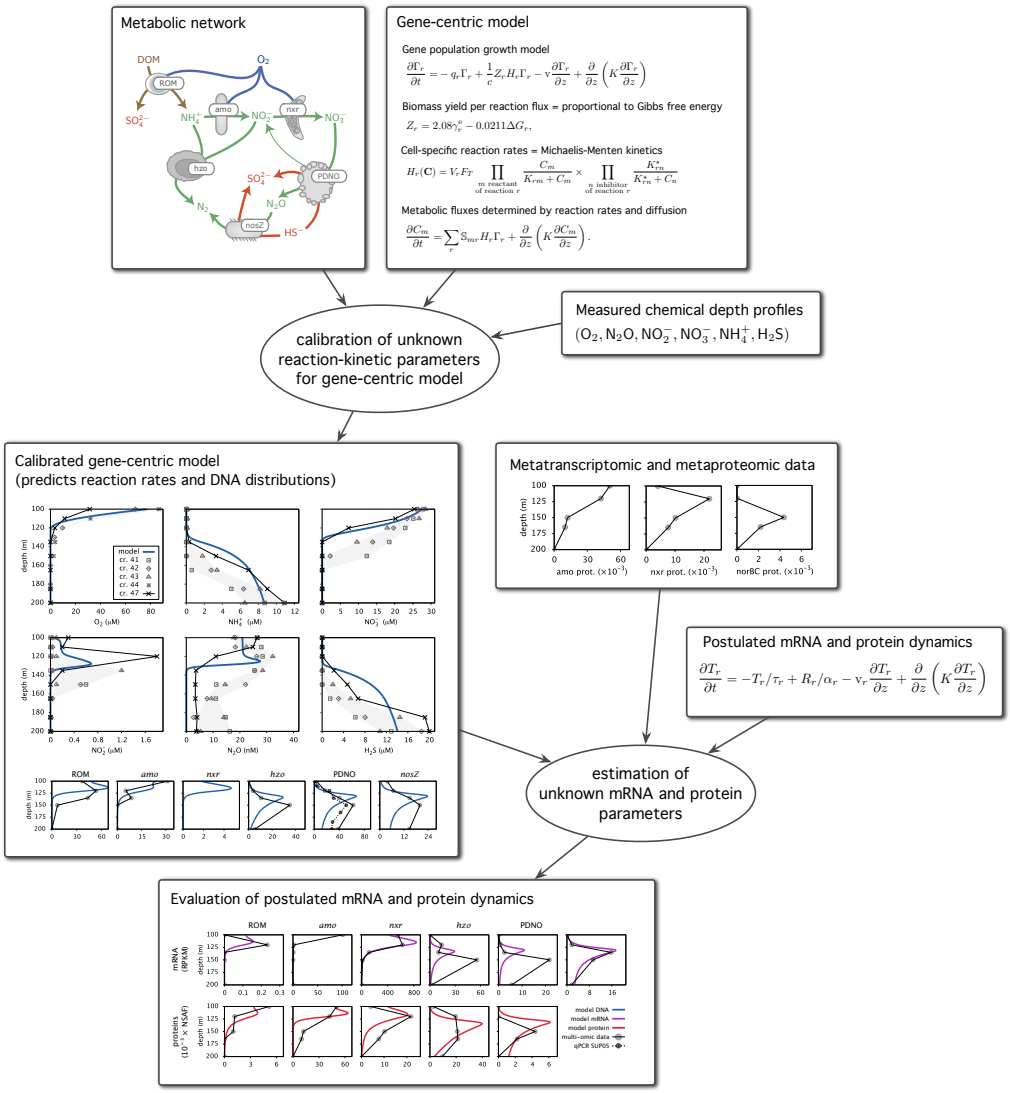


Figure S2: **Overview of this study.** Previous geochemical and multi-omic investigations provide conceptual information on the metabolic network in the Saanich Inlet OMZ (3, 5, 16, 34, 101, 102). This information was used to construct a gene-centric biogeochemical mathematical model, which describes the population dynamics of individual genes and metabolic process rates. Unknown reaction-kinetic parameters of the model were calibrated using geochemical depth profiles. The predictions of the calibrated gene-centric model were then validated using independent metagenomic sequence data, qPCR-based abundance estimates for SUP05 as well as process rate measurements. A subsequent extension of the model describes the production, dispersal and decay of mRNA and protein molecules based on the reaction rates predicted by the calibrated gene-centric model. Unknown parameters for the mRNA and protein dynamics are estimated using metatranscriptomic and metaproteomic data. The "goodness of fit" to these multi-omic data is used to further evaluate the gene-centric model, to assess the adequacy of the postulated mRNA and protein dynamics and to gain insight into potentially important but omitted mechanisms of mRNA and protein regulation at ecosystem scales.

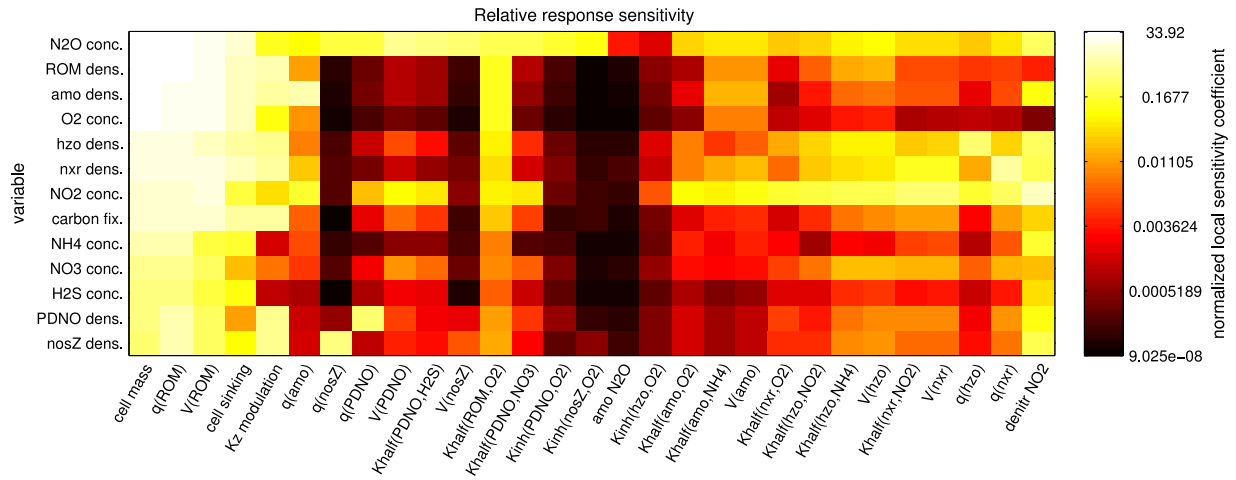


Figure S3: Local sensitivity heatmap of the calibrated model by means of normalized local sensitivity coefficients. A brighter color corresponds to a higher sensitivity. "Khalf" stands for half-saturation constants, "Kinh" for half-inhibition constants, "V" for maximum cell-specific reaction rates and "q" for cell death rates. The heatmap is hierarchically clustered using UPGMA linkage and Euclidean metric. Methodological details are given in Appendix S2.12.

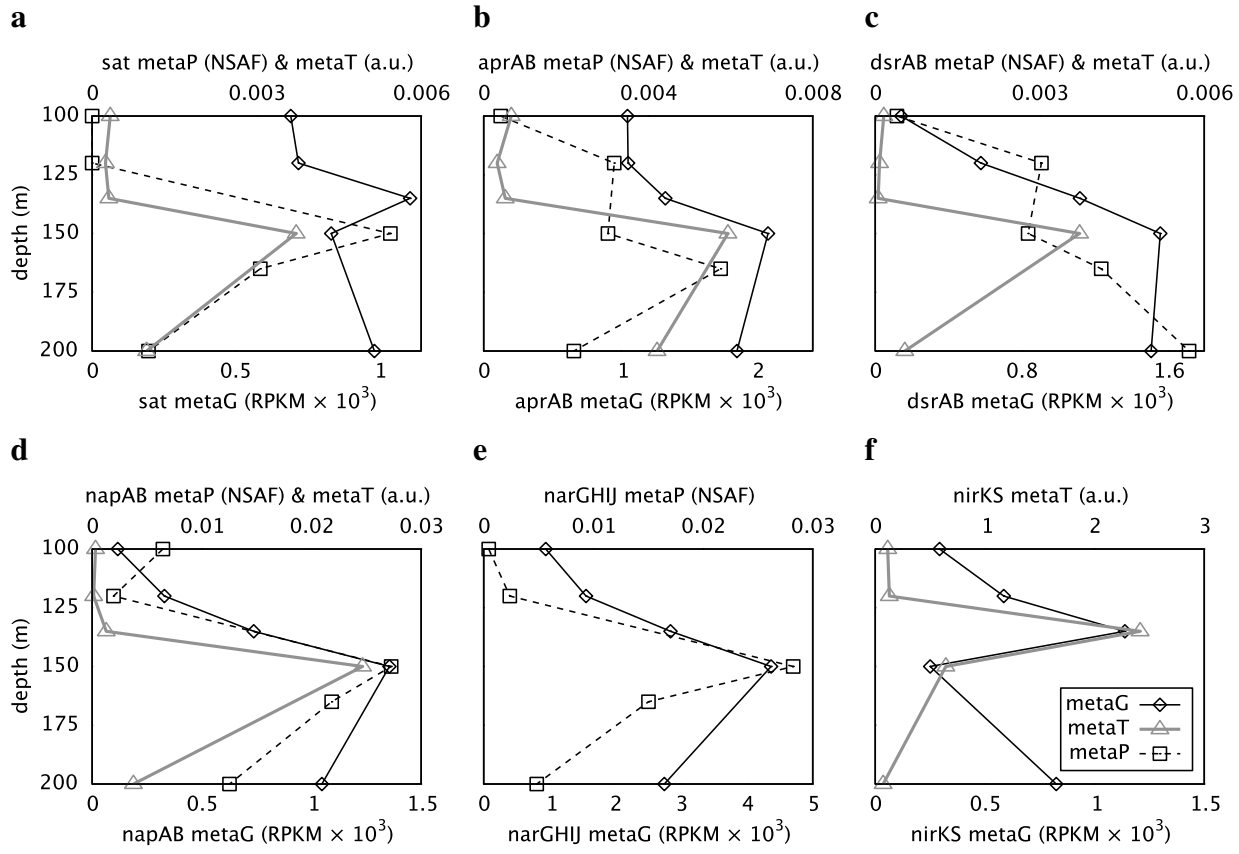


Figure S4: Metagenomic, metatranscriptomic and metaproteomic depth profiles of (a) sulfate adenylyltransferase (sat), (b) adenylylsulfate reductase (aprAB) and (c) sulfite reductase (dsrAB) genes, which together comprise the sulfide oxidation pathway (or sulfate reduction pathway, if reversed), as well as (d) periplasmic nitrate reductase napAB, (e) nitrate reductase narGHIJ and (f) NO-forming nitrite reductase nirKS. Data taken on February 10, 2014. All of the dsrAB, aprAB and most of the napAB protein sequences were mapped to the $\gamma$-proteobacterial SUP05 clade (34). All detected narGHIJ protein sequences were either mapped to SUP05 or to the anammox planctomycete bacteria *Candidatus* Scalindua profunda and KSU-1 (108) (only SUP05 proteins are shown). Similarly, only non-planctomycete-annotated narGHIJ and nirKS DNA abundances are shown.

# References


[1] Herlinveaux RH (1962) Oceanography of Saanich Inlet in Vancouver Island, British Columbia. *J Fish Res Board Can* 19(1):1–37.

[2] Cross SF, Chandler PC (1996) Saanich inlet study - surface circulation patterns, (Province of British Columbia, Ministry of Environment, Lands and Parks), Technical report.

[3] Zaikova E et al. (2010) Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* 12(1):172–191.

[4] Wright JJ, Konwar KM, Hallam SJ (2012) Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* 10(6):381–394.

[5] Walsh DA, Hallam SJ (2011) in *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*, ed. de Bruijn FJ. (John Wiley & Sons, Hoboken, NJ, USA), pp. 253–267.

[6] Montgomery H, Thom N, Cockburn A (1964) Determination of dissolved oxygen by the Winkler method and the solubility of oxygen in pure water and sea water. *J Appl Chem* 14(7):280–296.

[7] Capelle D, Dacey J, Tortell P (in review) An automated, high through-put method for accurate and precise measurements of dissolved nitrous-oxide and methane concentrations in natural waters. *Limnol Oceanogr Meth*.

[8] Hawley AK et al. (2013) in *Microbial Metagenomics, Metatranscriptomics, and Metaproteomics*, Methods in Enzymology, ed. DeLong EF. (Academic Press) Vol. 531, pp. 305–329.

[9] Paoletti AC et al. (2006) Quantitative proteomic analysis of distinct mammalian mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci USA* 103(50):18928–18933.

[10] Konwar KM, Hanson NW, Pagé AP, Hallam SJ (2013) Metapathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. *BMC Bioinformatics* 14(1):202.

[11] Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I (2014) RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res* 42(1):D553–9.

[12] Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27–30.

[13] Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28(1):33–36.

[14] Caspi R et al. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 42(D1):D459–D471.

[15] Strous M et al. (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* 440(7085):790–794.

[16] Hawley AK, Brewer HM, Norbeck AD, Paša-Tolić L, Hallam SJ (2014) Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc Natl Acad Sci USA* 111(31):11395–11400.

[17] Lee K, Choo YJ, Giovannoni SJ, Cho JC (2007) *Maritimibacter alkaliphilus* gen. nov., sp. nov., a genome-sequenced marine bacterium of the *Roseobacter* clade in the order *Rhodobacterales*. *Int J Syst Evol Microbiol* 57(7):1653–1658.

[18] Thrash JC et al. (2010) Genome sequences of *Pelagibaca bermudensis* HTCC2601T and *Maritimibacter alkaliphilus* HTCC2654T, the type strains of two marine *Roseobacter* genera. *J Bacteriol* 192(20):5552–5553.

[19] Ganesh S, Parris DJ, DeLong EF, Stewart FJ (2014) Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J* 8(1):187–211.

[20] Konwar KM et al. (2015) MetaPathways v2.5: quantitative functional, taxonomic and usability improvements. *Bioinformatics* 31(20):3345–3347.

[21] Holtappels M, Lavik G, Jensen MM, Kuypers MM (2011) in *Research on Nitrification and Related Processes, Part A*, Methods in Enzymology, ed. Klotz MG. (Academic Press) Vol. 486, pp. 223–251.

[22] Klotz MG (2011) *Research on Nitrification and Related Processes*, Methods in Enzymology. (Elsevier) Vol. 486.

[23] Dalsgaard T, Thamdrup B, Canfield DE (2005) Anaerobic ammonium oxidation (anammox) in the marine environment. *Res Microbiol* 156(4):457–464.

[24] Wenk CB et al. (2013) Anaerobic ammonium oxidation (anammox) bacteria and sulfide-dependent denitrifiers coexist in the water column of a meromictic south-alpine lake. *Limnol Oceanogr* 58(1):1–12.

[25] Ward B (2005) Temporal variability in nitrification rates and related biogeochemical factors in Monterey Bay, California, USA. *Mar Ecol Prog Ser* 292:97–109.

[26] Awata T et al. (2013) Physiological characterization of an anaerobic ammonium-oxidizing bacterium belonging to the "*Candidatus* Scalindua" group. *Appl Environ Microbiol* 79(13):4145–4148.

[27] Smets W et al. (2016) A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biol Biochem* 96:145–151.

[28] Aßhauer KP, Wemheuer B, Daniel R, Meinicke P (2015) Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31(17):2882–2884.

[29] Pruesse E et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35(21):7188–7196.

[30] Jin Q, Roden EE, Giska JR (2013) Geomicrobial kinetics: Extrapolating laboratory studies to natural environments. *Geomicrobiol J* 30(2):173–185.

[31] Reed DC, Algar CK, Huber JA, Dick GJ (2014) Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proc Natl Acad Sci USA* 111(5):1879–1884.

[32] Roden EE, Jin Q (2011) Thermodynamics of microbial growth coupled to metabolism of glucose, ethanol, short-chain organic acids, and hydrogen. *Appl Environ Microbiol* 77(5):1907–1909.

[33] Lebowitz J, Spohn H (1982) Microscopic basis for Fick's law for self-diffusion. *J Stat Phys* 28(3):539–556.

[34] Walsh DA et al. (2009) Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 326(5952):578–582.

[35] Lam P et al. (2009) Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA* 106(12):4752–4757.

[36] Santoro AE, Buchwald C, McIlvin MR, Casciotti KL (2011) Isotopic signature of $\mathrm{N_2O}$ produced by marine ammonia-oxidizing archaea. *Science* 333(6047):1282–1285.

[37] Fagerbakke K, Heldal M, Norland S (1996) Content of carbon, nitrogen, oxygen, sulfur and phosphorus in native aquatic and cultured bacteria. *Aquat Microb Ecol* 10(1):15–27.

[38] Cohen Y (1978) Consumption of dissolved nitrous oxide in an anoxic basin, Saanich Inlet, British Columbia. *Nature* 272(5650):235–237.

[39] Devol AH, Anderson JJ, Kuivila K, Murray JW (1984) A model for coupled sulfate reduction and methane oxidation in the sediments of Saanich Inlet. *Geochim Cosmochim Acta* 48(5):993–1004.

[40] Ahmed S, King S, Clayton Jr J (1984) Organic matter diagenesis in the anoxic sediments of Saanich Inlet, British Columbia, Canada: a case for highly evolved community interactions. *Mar Chem* 14(3):233–252.

[41] Ogawa H, Amagai Y, Koike I, Kaiser K, Benner R (2001) Production of refractory dissolved organic matter by bacteria. *Science* 292(5518):917–920.

[42] Hoppe H, Ducklow H, Karrasch B (1993) Evidence for dependency of bacterial-growth on enzymatic-hydrolysis of particulate organic-matter in the mesopelagic ocean. *Mar Ecol Prog Ser* 93(3):277–283.

[43] Middelburg JJ, Vlug T, Jaco F, Van Der Nat W (1993) Organic matter mineralization in marine systems. *Glob Plan Change* 8(1):47–58.

[44] Schulthess RV, Gujer W (1996) Release of nitrous oxide ($N_2O$) from denitrifying activated sludge: Verification and application of a mathematical model. *Water Res* 30(3):521–530.

[45] Wicht H (1996) A model for predicting nitrous oxide production during denitrification in activated sludge. *Water Sci Technol* 34(5–6):99–106.

[46] Jing C, Ping Z, Mahmood Q (2010) Influence of various nitrogenous electron acceptors on the anaerobic sulfide oxidation. *Bioresour Technol* 101(9):2931–2937.

[47] Dean J (1998) *Lange's Handbook of Chemistry*. (McGraw-Hill Professional), 15 edition.

[48] Dick JM (2008) Calculation of the relative metastabilities of proteins using the CHNOSZ software package. *Geochem Trans* 9(10).

[49] Murray JW, Grundmanis V, Smethie Jr. WM (1978) Interstitial water chemistry in the sediments of Saanich Inlet. *Geochim Cosmochim Acta* 42(7):1011–1026.

[50] Devol AH, Ahmed SI (1981) Are high rates of sulphate reduction associated with anaerobic oxidation of methane? *Nature* 291(5814):407–408.

[51] Remacle J, De Leval J (1978) in *Microbiology*, ed. Schlessinger D. (American Society for Microbiology, Washington), pp. 352–356.

[52] Strous M et al. (1999) Missing lithotroph identified as new planctomycete. *Nature* 400(6743):446–449.

[53] Kuypers MMM et al. (2005) Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc Natl Acad Sci USA* 102(18):6478–6483.

[54] Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* 95(12):6578–6583.

[55] Geets J, Boon N, Verstraete W (2006) Strategies of aerobic ammonia-oxidizing bacteria for coping with nutrient and oxygen fluctuations. *FEMS Microbiol Ecol* 58(1):1–13.

[56] Hao X, Wang Q, Zhang X, Cao Y, van Mark Loosdrecht C (2009) Experimental evaluation of decrease in bacterial activity due to cell death and activity decay in activated sludge. *Water Res* 43(14):3604–3612.

[57] Raven J et al. (2005) *Ocean acidification due to increasing atmospheric carbon dioxide*. (The Royal Society).

[58] Wong C et al. (2002) Seasonal and interannual variability in the distribution of surface nutrients and dissolved inorganic carbon in the northern North Pacific: influence of El Niño. *J Oceanogr* 58(2):227–243.

[59] Zeebe R, Wolf-Gladrow D (2001) $CO_2$ *in Seawater: Equilibrium, Kinetics, Isotopes*, Elsevier oceanography series. (Elsevier).

[60] Luo Y, Miller LA, De Baere B, Soon M, Francois R (2014) POC fluxes measured by sediment traps and $^{234}Th$:$^{238}U$ disequilibrium in Saanich Inlet, British Columbia. *Mar Chem* 162:19–29.

[61] Gargett A (1984) Vertical eddy diffusivity in the ocean interior. *J Mar Res* 42(2):359–393.

[62] Fennel K, Boss E (2003) Subsurface maxima of phytoplankton and chlorophyll: Steady-state solutions from a simple model. *Limnol Oceanogr* 48(4):1521–1534.

[63] Kelley D (2014) *oce: Analysis of oceanographic data*. R package version 0.9-13.

[64] Fennel K et al. (2006) Nitrogen cycling in the Middle Atlantic Bight: Results from a three-dimensional model and implications for the North Atlantic nitrogen budget. *Global Biogeochem Cycles* 20(3).

[65] Azhar MA, Canfield DE, Fennel K, Thamdrup B, Bjerrum CJ (2014) A model-based insight into the coupling of nitrogen and sulfur cycles in a coastal upwelling system. *J Geophys Res Biogeosci* 119(3):264–285.

[66] Stolper DA, Revsbech NP, Canfield DE (2010) Aerobic growth at nanomolar oxygen concentrations. *Proc Natl Acad Sci USA* 107(44):18755–18760.

[67] Martens-Habbena W, Berube PM, Urakawa H, de la Torre JR, Stahl DA (2009) Ammonia oxidation kinetics determine niche separation of nitrifying archaea and bacteria. *Nature* 461(7266):976–979.

[68] Limpiyakorn T et al. (2013) amoA-encoding archaea in wastewater treatment plants: a review. *Appl Microbiol Biotechnol* 97(4):1425–1439.

[69] Blackburne R, Vadivelu VM, Yuan Z, Keller J (2007) Kinetic characterisation of an enriched *Nitrospira* culture with comparison to *Nitrobacter*. *Water Res* 41(14):3033–3042.

[70] Bristow LA et al. (2013) High sensitivity of ammonia and nitrite oxidation rates to nanomolar oxygen concentrations. *Mineral Mag* 77(5):636–804.

[71] Kalvelage T et al. (2011) Oxygen sensitivity of anammox and coupled n-cycle processes in oxygen minimum zones. *PLoS ONE* 6(12):e29299.

[72] Jensen MM, Petersen J, Dalsgaard T, Thamdrup B (2009) Pathways, rates, and regulation of $N_2$ production in the chemocline of an anoxic basin, Mariager Fjord, Denmark. *Mar Chem* 113(1):102–113.

[73] Blumberg K, Michiels C, Jones C, Crowe SA, Hallam S (2014) *Saanich Inlet oxygen mimimum zone: SUP05-driven Michaelis-Menten sulfide oxidation kinetics*, Harnessing "omics" to advance the geosciences: New paradigms and platforms for observing Earth systems. (Geological Society of America, Vancouver, BC, Canada), No. 191-9.

[74] Thamdrup B, Dalsgaard T, Revsbech NP (2012) Widespread functional anoxia in the oxygen minimum zone of the eastern south pacific. *Deep Sea Res Part I* 65(0):36–45.

[75] Von Schulthess R, Wild D, Gujer W (1994) Nitric and nitrous oxides from denitrifying activated sludge at low oxygen concentration. *Water Sci Technol* 30(6):123–132.

[76] Frey C, Hietanen S, Jürgens K, Labrenz M, Voss M (2014) N and o isotope fractionation in nitrate during chemolithoautotrophic denitrification by *Sulfurimonas gotlandica*. *Environ Sci Technol* 48(22):13229–13237.

[77] Betlach MR, Tiedje JM (1981) Kinetic explanation for accumulation of nitrite, nitric oxide, and nitrous oxide during bacterial denitrification. *Appl Environ Microbiol* 42(6):1074–1084.

[78] Almeida JS, Reis MAM, Carrondo MJT (1995) Competition between nitrate and nitrite reduction in denitrification by *Pseudomonas fluorescens*. *Biotechnol Bioeng* 46(5):476–484.

[79] Oh J, Silverstein J (1999) Acetate limitation and nitrite accumulation during denitrification. *J Environ Eng* 125(3):234–242.

[80] Eliason SR (1993) *Maximum Likelihood Estimation: Logic and Practice*. (SAGE Publications, Newbury Park, CA).

[81] MATLAB (2010) *version 7.10.0 (R2010a)*. (The MathWorks Inc., Natick, Massachusetts).

[82] Poretsky RS et al. (2005) Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol* 71(7):4121–4126.

[83] Morgan JL, Darling AE, Eisen JA (2010) Metagenomic sequencing of an *in vitro*-simulated microbial community. *PLoS ONE* 5(4):e10209.

[84] Beszteri B, Temperton B, Frickenhaus S, Giovannoni SJ (2010) Average genome size: a potential source of bias in comparative metagenomics. *ISME J* 4(8):1075–1077.

[85] Amy PS, Pauling C, Morita RY (1983) Recovery from nutrient starvation by a marine *Vibrio* sp. *Appl Environ Microbiol* 45(5):1685–1690.

[86] Kemp PF, Lee S, LaRoche J (1993) Estimating the growth rate of slowly growing marine bacteria from RNA content. *Appl Environ Microbiol* 59(8):2594–2601.

[87] Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* 123(6):1025–1036.

[88] Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* 323(5):785–793.

[89] Christopher Frey H, Patil SR (2002) Identification and review of sensitivity analysis methods. *Risk Anal* 22(3):553–578.

[90] Knightes CD, Peters CA (2000) Statistical analysis of nonlinear parameter estimation for Monod biodegradation kinetics using bivariate data. *Biotechnol Bioeng* 69(2):160–170.

[91] Marshall KT, Morris RM (2013) Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J* 7(2):452–455.

[92] Roeselers G et al. (2010) Complete genome sequence of *Candidatus* Ruthia magnifica. *Stand Genomic Sci* 3(2):163.

[93] Nissenbaum A, Presley BJ, Kaplan IR (1972) Early diagenesis in a reducing fjord, saanich inlet, british columbia — i. chemical and isotopic changes in major components of interstitial water. *Geochim Cosmochim Acta* 36(9):1007–1027.

[94] Schunck H et al. (2013) Giant hydrogen sulfide plume in the oxygen minimum zone off Peru supports chemolithoautotrophy. *PLoS ONE* 8(8):e68661.

[95] Brüchert V et al. (2003) Regulation of bacterial sulfate reduction and hydrogen sulfide fluxes in the central Namibian coastal upwelling zone. *Geochim Cosmochim Acta* 67(23):4505–4518.

[96] Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ (2012) Microbial oceanography of anoxic oxygen minimum zones. *Proc Natl Acad Sci USA* 109(40):15996–16003.

[97] Prokopenko MG et al. (2013) Nitrogen losses in anoxic marine sediments driven by *Thioploca*-anammox bacterial consortia. *Nature* 500(7461):194–198.

[98] Anderson JJ, Devol AH (1973) Deep water renewal in Saanich Inlet, an intermittently anoxic basin. *Estuar Coast Mar Sci* 1(1):1–10.

[99] Lam P, Kuypers MM (2011) Microbial nitrogen cycling processes in oxygen minimum zones. *Annu Rev Mar Sci* 3(1):317–345.

[100] Jensen MM et al. (2011) Intensive nitrogen loss over the Omani Shelf due to anammox coupled with dissimilatory nitrite reduction to ammonium. *ISME J* 5(10):1660–1670.

[101] Anderson J (1984) *The oxic/anoxic interface in Saanich Inlet*, Canadian technical report of hydrography and ocean sciences eds. Juniper S, Brinkhurst R. (Department of Fisheries and Oceans Canada), No. 83, pp. 17–23.

[102] Juniper S, Brinkhurst R (1986) Water-column dark $\mathrm{CO}_2$ fixation and bacterial-mat growth in intermittently anoxic Saanich Inlet, British-Columbia. *Mar Ecol Prog Ser* 33(1):41–50.

[103] Canfield DE et al. (2010) A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* 330(6009):1375–1378.

[104] Steinkamp K (2011) Ph.D. thesis (ETH Zurich).

[105] Martinez-Camara M, Béjar Haro B, Stohl A, Vetterli M (2014) A robust method for inverse transport modeling of atmospheric emissions using blind outlier detection. *Geosci Model Dev* 7(5):2303–2311.

[106] Roach G (1982) *Green's Functions*. (Cambridge University Press).

[107] Björck A (1996) *Numerical Methods for Least Squares Problems*. (Society for Industrial and Applied Mathematics).

[108] van de Vossenberg J et al. (2013) The metagenome of the marine anammox bacterium '*Candidatus* Scalindua profunda' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ Microbiol* 15(5):1275–1289.