Supplementary information

Distribution of additional *nifH* clusters. Compared to the dominant clusters P1, P4 and P8, clusters P2, P3, P6 and P7 showed lower *nifH* abundance. Nevertheless, they show a specific distribution along vertical and horizontal gradients with generally higher abundances at 16°S, where N/P ratios were close to zero in surface and the OMZ had high P* values reaching 2 µM (Fig. S1).

Effects of O_2 and NO_3^{2-} on the distribution of *nifH* clusters.

To investigate the effects of parameters O_2 and NO_3^-, which had been dropped from the RDA model during stepwise selection but are generally assumed important factors for N_2-fixation, we additionally evaluated a model comprising those two variables (Fig. S4). While the overall model and above-stated marginal effects remained significant, marginal effects attributed to O_2 and NO_3^- were insignificant ($P_{O2} = 0.162$, $P_{NO3-} = 0.553$). (Variance inflation factors were generally higher than in the parsimonious model and >5 for PO_4^{3-} and O_2. Explained variance improved by 0.2% compared to the parsimonious model ($R^2_{adj.\,full} = 9.0\%$). The full model produced five constrained axes, the first three of which bundled about 84%, 12%, and 3% of the explained variance, respectively. The first and second RDA axis is shown in Fig. S3.) , although an RDA with O_2 and NO_3^- as the only explanatory variables would also be significant (not shown), our results suggest that both their effects can be substituted by the effect of PO_4^{3-} in the parsimonious model with negligible loss of explanatory power.

Tab. S1: Overlap in sequences of novel detected clusters and previous phylogenetic studies of the *nifH* gene in oceanic waters

Cluster	Fernandez et al.[1]		Farnelid et al.[2]		Halm et al.[3]	
	accession number	location	accession number	location	accession number	location
P1						
P2						
P3			GU196924.1 (99,4% identity)	Indian Ocean		
P4	HM801228.1 (100% identity)	Galathea Expedition III				
P5						
P6						
P7			GU196873.1; GU196871.1 (97,9% identity)	Indian Ocean		
P8	HM801335.1 (100% identity)	Galathea Expedition III	GU196971.1; GU196969.1; GU196945.1 (100% identity)	Indian Ocean	HM210363.1 (100% identity)	South Pacific Gyre during the cruise KNOX02RR on the Roger Revelle, CTD 9 100m

Analysis of glucose and O$_2$ addition effects. Effects of glucose and O$_2$ addition on N fixation and gene abundances of selected novel clusters were analyzed in R. The systematic part of all models consisted of both factors and their interaction term ("Response variable ~ Glucose + O$_2$ + Glucose:O$_2$"). Standard analysis of variance based on ordinary least squares was not applicable due to strong heteroscedasticity of residuals in all fitted models. Therefore, generalized least squares were used for the modelling of N fixation and cluster P4 gene abundances with function gls of the R package nlme [4]. This method allowed for different standard deviations per Glucose and O$_2$ level, respectively, chosen using the Akaike Information Criterion (AIC). Likewise, to model P7 gene abundances, a generalized linear model (function glm of the stats package) was employed. This was based on a gamma-distributed response variable with a log link between the stochastic (response variable) and systematic model parts. Since such a model requires a positive response variable, an arbitrary small constant (10^{-10}) was added to all P7 gene abundances to avoid zero values. Visual analysis of residuals revealed no adverse patterns in any final model.

Table S2: Summary of statistical analysis of the response of N fixation and cluster P4/P7 gene abundances on glucose and O$_2$ addition. GLS, generalized least squares; GLM, generalized linear model. Degrees of freedom were 8 for all t-value test statistics. t-values and corresponding p-values are marked in bold-type for effects significant at the 5% level.

	Response variable					
	N fixation		P4		P7	
Model type	GLS[a]		GLS[b]		GLM[c]	
Effect	**t-value**	**p-value**	**t-value**	**p-value**	**t-value**	**p-value**
Glucose	**7.06**	**<0.001**	-2.04	0.076	**206.96**	**<0.001**
O$_2$	0.29	0.780	**-3.71**	**0.006**	0.00	1.000
Glucose:O$_2$	1.71	0.126	**6.76**	**<0.001**	**-6.99**	**<0.001**

[a] different standard deviations per Glucose level
[b] different standard deviations per O$_2$ level
[c] gamma distribution with log link; constant 10^{-10} added to avoid zero values

Fig. S1: Distribution of O$_2$, P*, NO$_2^-$ and novel *nifH* clusters: P*, NO$_2^-$ (µM), *nifH* clusters P2, P3, P6 and P7 and total nifH (log$_{10}$ copies L^{-1}), the oxycline (concentrations in µM O$_2$) is indicated with white contour lines along (A) 10°S and (B) 16°S.

Fig. S2: Depth integrated horizontal distribution of functional key genes of N-loss processes and N$_2$ fixation in the upper 350 m in the Peruvian OMZ: archaeal *amoA* (functional marker for archaeal ammonia oxidation), *nirS* (functional marker for denitrification), *hzo* (functional marker for anammox), *nifH* for diazotrophs clusters P1, P8 and total *nifH* (log$_{10}$ copies per L). The section shows the vertical distribution of the dominant *nifH* cluster P8 (log$_{10}$ copies L^{-1}, measured by qPCR) along a

North-South transect at 85.83°W as indicated by the black box on the map, the oxycline (μM O_2) is indicated as contour lines.

Fig. S3: Absolute abundances of individual *nifH* clusters mapped onto the sample points in the parsimonious redundancy analysis (RDA) model (see Fig. 4). The area of the circles is proportional to ln+1-transformed *nifH* cluster abundances (see scale insert in the upper-left panel). Fill color corresponds to the respective *nifH* cluster color in Fig. 4. Samples with zero-abundance of a respective *nifH* cluster are represented as grey dots.

Fig. S4: Redundancy analysis (RDA) of Hellinger-transformed *nifH* cluster abundances. (A) The distance biplot (scaling) shows relations between samples (weighted sums of species scores) and environmental variables in the full model. (B) Correlation biplot (scaling 2) shows relations between *nifH* cluster vectors and environmental variables in the full model.

Fig. S5: **Vertical profiles of N_2- fixation**: (A) Water column N_2- fixation determined by 24h $^{15}N_2$-incubation experiments along the initial vertical distribution of *nifH* clusters P1 and P4 at station #3 (Fig.1, map, 10°S/ 81.3°W) and (B) along O_2 and nutrient gradients.
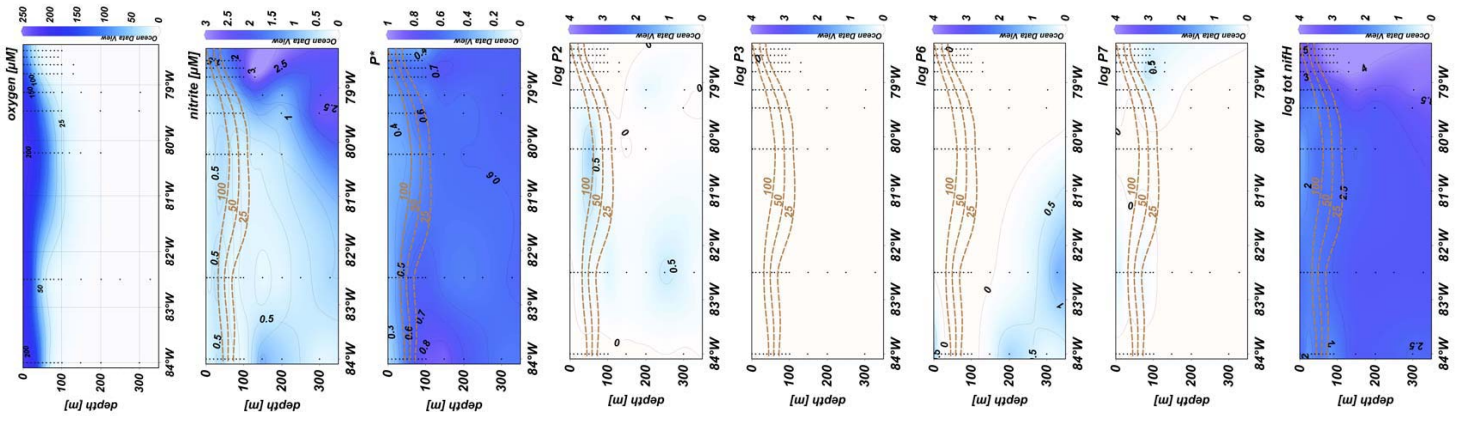
Fig. S6: **Vertical profiles of N_2- fixation**: 24h $^{15}N_2$- incubation experiments at two stations at 10°S (# 805 at 79.134 °W is marked with black triangles, and # 811 at 81.361 °W with open circles).

Fig. S7: N_2- fixation was triggered in incubation experiments (#807, 10.001°S 78.38 °W, 20 m depth) by glucose and glucose/ oxygen addition
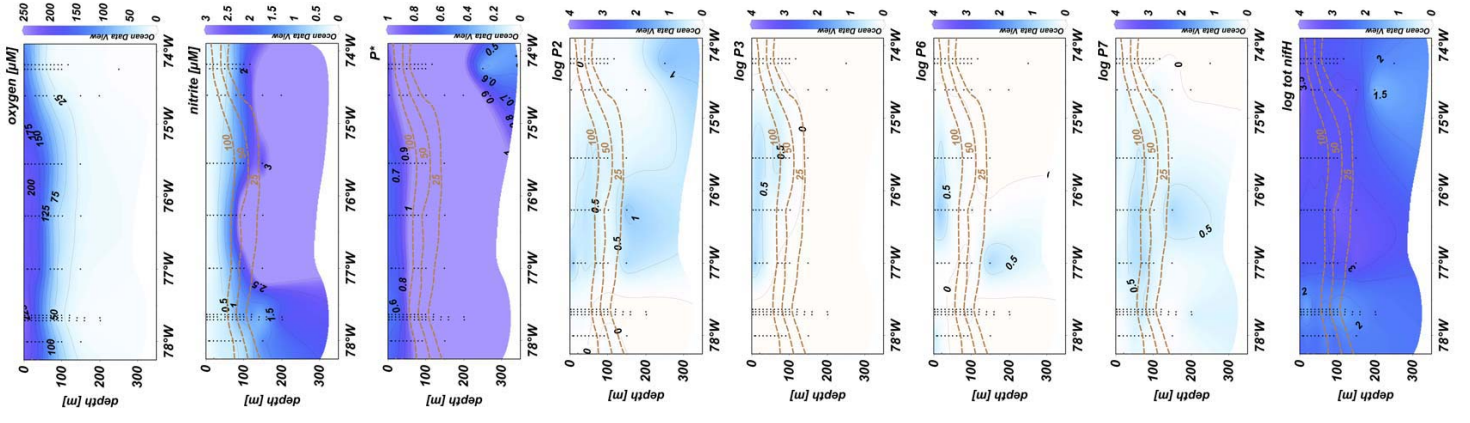
Fig. S8: Correlation of *nifH* P1 and NO_2^- : NO_2^- [μM], *nirS, hzo* and *amoA* sections section overlaid by P1 [$\log_{10}$ copies L^{-1}] indicated by dashed white contour lines.

References

[1]   C. Fernandez, L. Farias, and O. Ulloa, *Plos One* **6** (6), 9 (2011).
[2]   H. Farnelid, A.F. Andersson, S. Bertilsson et al., *PLoS One* **6** (4) (2011).
[3]   H. Halm, P. Lam, T.G. Ferdelman et al., *submitted* (2011).
[4]   José C. Pinheiro, Douglas M. Bates, Saikat DebRoy et al., nlme: Linear and nonlinear mixed effects models (2013).

A

B

log total *nifH* [copies m^{-2}] log *amoA* [copies m^{-2}] log *nirS* [copies m^{-2}] log *hzo* [copies m^{-2}]

log P1 *nifH* [copies m^{-2}] log P8 *nifH* [copies m^{-2}] log P8 *nifH* [copies L^{-2}]

Panel A

Nitrogen fixation [nmol N L^{-1} d^{-1}]

Depth [m]

n.d.

Nitrogen fixation
P1
P4

A

nifH gene abundance [copies L^{-1}]

Panel B

NO$_3^-$ x 10, NO$_2^-$, NH$_4^+$ [µmol kg^{-1}]

B

Oxygen [µmol kg^{-1}]