Text S1. Supplementary Methods

Total DNA was extracted from the gut using DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. DNA concentration was quantified with a Qubit© fluorometer (Invitrogen). To obtain appropriate amounts of DNA, three guts were combined to make one sample for polymerase chain reactions (PCR) of 16S and 18S rRNA genes. Depending on available samples two or three replicates per station were used for sequencing of 16S and 18S rRNA genes amplified. Only one water sample was taken at each of the corresponding sampling stations due to availability. A total of 25 *L. h. antarctica* gut samples (from nine sampling stations, 2-3 replicates per station) and six water samples (from six corresponding stations) were analyzed. Gut samples from *L. h. antarctica*, as well as a negative control, were amplified with PCR in singleton (25 µl volume), primers, and GoTaq Green Master Mix (Promega), and Bovine Serum Albumin (BSA). Volumes of each component per 1 µl gut DNA sample were as follows 1 µl 515F-Y primer, 1 µl 926R, 10 µl water, 0.25 µl BSA, and 12.5 µl GoTaq. Water samples were amplified with PCR in singleton (25 µl volume), primers, and GoTaq Green Master Mix only. Volumes of each component per 1 µl water sample DNA were as follows, 1 µl 515F-Y primer, 1 µl 926R, 10 µl water, and 12 µl GoTaq. Primers 515F-Y (5'-GTGYCAGCMGCCGCGGTAA) and 926R (5'-CCGYCAATTYMTTTRAGTTT) targeting both 16S and 18S rRNA genes were used (Quince et al. 2011, Parada et al. 2016). This primer pair encompasses both the V4 and V5 hypervariable regions of 16S and 18S rRNA genes. While the primer only matches 86% of eukaryotic 18S rRNA, it enables simultaneous amplification of both the 16S and 18S rRNA genes (Parada et al. 2016). To optimize amplification, PCR cycling condition for *L. h. antarctica* guts was 94°C for 3 min, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 54°C for 2 min, and extension at 68°C for 1 min. PCR cycling condition for corresponding water samples was 95°C for 3.5 min, followed by 25 cycles of denaturation at 95°C for 30 sec, annealing at 55°C for 1 min, and extension at 75°C for 1 min. The PCR products from each sample were purified with the HighPrep PCR© (MagBio) protocol using magnetic beads. The purified PCR product (5 µl) was used for Stage 2 PCR (Index PCR). The index PCR product was purified according to the Mag-Bind© TotalPure NGS protocol (Omega Bio-tek). The purified index PCR products were quantitated using the Qubit© fluorometer (Invitrogen) and adjusted to equimolar concentrations of 4 nM. These amplicon products were then used as templates for the sequencing run on the MiSeq system (Illumina) producing 2 x 300 bp read lengths for 16S and 2 x 350 bp read lengths for 18S.

Regarding bioinformatic analysis, only forward sequence reads were analyzed with DADA2 to analyze 16S and 18S sequences together (Parada et al. 2016). Negative controls that were positive for *Escherichia* sequences were removed from the samples. To increase 18S sequence coverage, sequences from replicate gut samples at each station were pooled together for analysis in DADA2. Pooled sequences for 16S (prokaryotes) and 18S (eukaryotes) were analyzed separately in Phyloseq (McMurdie & Holmes 2013). Replicate gut samples were visually analyzed with PCoA to ensure adequate similarity before pooling sequences together (Fig. S2). For quality trimming, forward reads were filtered with a 215 bp cutoff to analyze 16S and 18S sequences simultaneously. Quality filtering, dereplication, sample inference, and chimera removal were performed in DADA2. Amplicon Sequence Variants (ASVs) were determined and taxonomic classification of the ASVs were analyzed based on the Silva v132 database for 16S and 18S (Yilmaz et al. 2014). In addition, chloroplast 16S rRNA gene sequences were analyzed separately from other 16S and 18S rRNA genes in gut and water samples with the Ribosomal Database Project (RDP) classifier (Wang et al. 2007).

To identify eukaryotic 18S sequences indicative of *L. h. antarctica* prey, *L. h. antarctica* host 18S sequences were subtracted from sequenced samples for PCoA and relative abundance analyses. Chloroplast 16S sequences were also removed in Phyloseq prior to 18S taxonomic analysis to ensure results for eukaryotes were not based on microalgal chloroplast sequences binning to the 16S database. Gut communities as ASVs were compared amongst each other (β-diversity) and to those in surrounding seawater with PCoA using the vegan package v2.5 with Bray-Curtis dissimilarity distances (Oksanen et al. 2019) and visualized using the ggplot2 package in R (Wickham 2016). Significant differences between water and gut community dissimilarity were tested by multivariate permutational ANOVA (PERMANOVA) using the adonis function of the vegan package (Oksanen et al. 2019). Marti Anderson's PERMDISP2 procedure was used for the analysis of multivariate homogeneity of group dispersions (variances) within the vegan package (Oksanen et al. 2019). The function betadisper was used within vegan as a multivariate analogue of Levene's test for determining homogeneity of variances, a necessary assumption for PERMANOVA. Data were not transformed. A one-way ANOVA was used to analyze statistical differences in diversity between gut and water communities. Rarefaction, Ace, Chao, and Shannon diversity indices were used to determine sample coverage, richness, and α-diversity (richness and evenness), respectively for seawater and pooled gut samples and then analyzed separately for 16S and 18S in Phyloseq. Relative abundances were visualized using the ggplot2 package in R (Wickham 2016). Classes with $> 0.5\%$ of relative abundance were considered in the analysis to identify a broader range of taxa present. 18S ASVs were visualized at the genus level and considered dominant genera with $> 5\%$ of relative abundance. In addition, chloroplast 16S ASVs were visualized at the genus level and considered dominant genera with $> 0.5\%$ of relative abundance to identify a broader range of taxa present. All statistical and bioinformatic analyses were performed in R statistical framework version 3.6.0 (R Statistical Core Team 2019).

References

McMurdie PJ, Holmes S (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8.

Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, Wagner H (2019) Vegan: Community Ecology Package.

Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol 18:1403–1414. https://doi.org/10.1111/1462-2920.13023

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12:38. https://doi.org/10.1186/1471-2105-12-38

R Statistical Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl. Environ. Microbiol. **73**: 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wickham H (2016) Elegant Graphics for Data Analysis. Springer International Publishing, New York, NY.

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C (2014) The SILVA and all-species living tree project (LTP) taxonomic frameworks. Nucleic Acids Res 42:D643–D648. https://doi.org/10.1093/nar/gkt1209

Supplementary Table

Table S1. Summary of 16S and 18S rRNA gene sequencing statistics and diversity indices in water and the *Limacina helicina antarctica* (pteropod) gut. Sampling stations arranged north to south. Sample size (*n*) for each sample with an arrow showing the number of replicate pteropod samples pooled into one sample for analysis (3 → 1). 'QC sequences' indicates the number of sequences available after trimming, filtering and denoising raw sequences. 'Host removed' shows the remaining number of sequences available after subtracting host (*L. h. antarctica*) 18S sequences for pteropod samples (not applicable (NA) for water samples and 16S data). 'Observed' indicates relative sequence coverage (ASVs) for each sample; Chao 1, richness diversity index; ACE, richness diversity index, Shannon, alpha diversity (richness & evenness) index. Grand means (± 1 standard error) across all stations for water (mean of samples) and pteropod guts (mean of pooled samples) are also shown.

Station	Sample	*n*	Target gene	Input sequences	QC sequences	Host removed	Observed ASVs	Chao 1	ACE	Shannon
600.200	Water	1	16S	14393	7408	NA	103	103	103	4.08
300.200	Water	1	16S	7859	5164	NA	89	89	89	4.07
300.040	Water	1	16S	18402	11636	NA	65	65	65	2.80
200.000	Water	1	16S	24469	10510	NA	73	73	73	2.95
100.100	Water	1	16S	2365	788	NA	31	31	31	2.95
100.040	Water	1	16S	18833	11043	NA	88	88	88	3.56
Mean Water			16S	14386 ± 3289	7758 ± 1720	NA	53.83 ± 7.74	57.01 ± 8.13	56.67 ± 8.06	3403 ± 0.24
600.200	Water	1	18S	14393	1900	NA	38	38	38	3.03
300.200	Water	1	18S	7859	395	NA	21	21	21	2.61
300.040	Water	1	18S	18402	526	NA	19	19	19	2.37
200.000	Water	1	18S	24469	789	NA	15	15	15	1.75
100.100	Water	1	18S	2365	662	NA	16	16	16	1.74
100.040	Water	1	18S	18833	1450	NA	20	20	20	1.58
Mean Water			18S	14386 ± 3289	953 ± 241	NA	21.50 ± 2.80	21.50 ± 2.80	21.50 ± 2.80	2.18 ± 0.19
600.200	Pteropod	3 → 1	16S	157650	60033	NA	89	89	89	2.01
400.100	Pteropod	2 → 1	16S	27452	9936	NA	103	103	103	3.10
300.200	Pteropod	3 → 1	16S	45236	5386	NA	84	84	84	3.70
300.040	Pteropod	2 → 1	16S	51213	23825	NA	69	69	69	2.86
200.000	Pteropod	3 → 1	16S	156258	37038	NA	139	139	139	3.25
100.200	Pteropod	3 → 1	16S	83733	30420	NA	73	73	73	2.85
100.100	Pteropod	3 → 1	16S	269043	150808	NA	175	175	175	3.28
100.040	Pteropod	3 → 1	16S	40884	6513	NA	65	65	65	3.47
-100.100	Pteropod	3 → 1	16S	160483	67601	NA	108	108	108	2.18
Mean Pteropod			16S	110216 ± 26921	43506 ± 15328	NA	100 ± 12.08	100 ± 12.08	100 ± 12.08	2.97 ± 0.18
600.200	Pteropod	3 → 1	18S	157650	48171	935	17	17	17	2.07
400.100	Pteropod	2 → 1	18S	27452	9776	4260	26	26	26	1.58
300.200	Pteropod	3 → 1	18S	45236	19816	900	14	14	14	1.90
300.040	Pteropod	2 → 1	18S	51213	15189	58	5	5	5	1.39
200.000	Pteropod	3 → 1	18S	156258	70661	2541	21	21	21	1.75
100.200	Pteropod	3 → 1	18S	83733	25107	935	15	15	15	2.30
100.100	Pteropod	3 → 1	18S	269043	20701	6550	27	27	27	2.18
100.040	Pteropod	3 → 1	18S	40884	26967	96	4	43	4	1.03
-100.100	Pteropod	3 → 1	18S	160483	34181	7424	16	16	16	1.49
Mean Pteropod			18S	110216 ± 26921	30063 ± 6294	2633 ± 933	16.11 ± 2.67	16.11 ± 2.67	16.11 ± 2.67	1.74 ± 0.13

Supplementary Figures

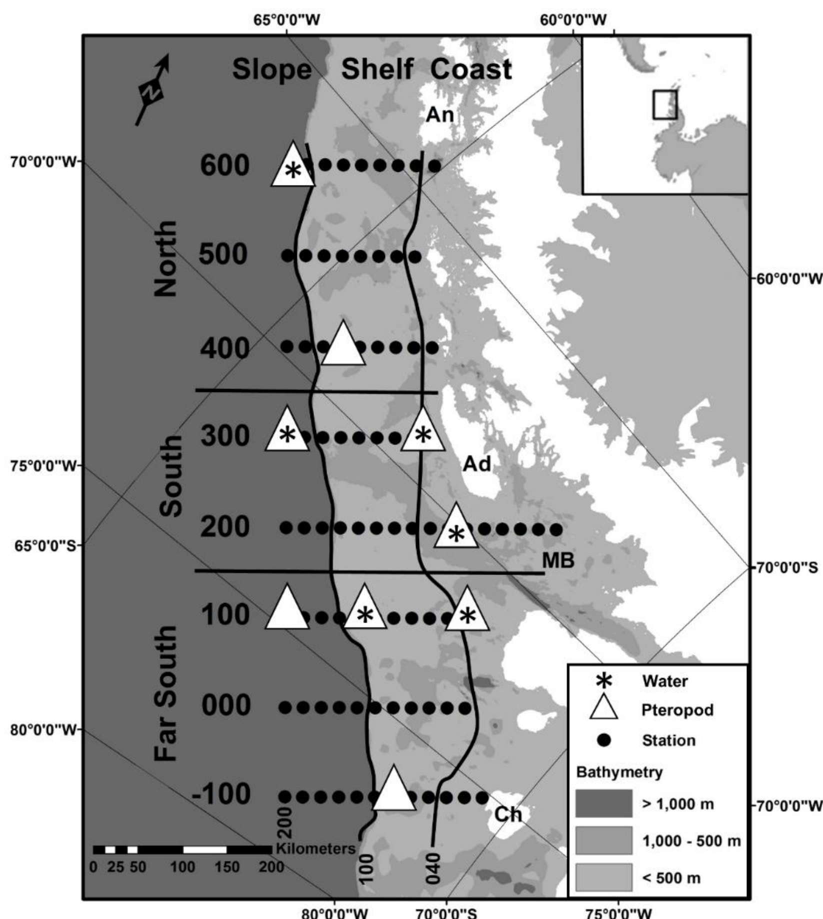

Fig. S1. Map of PAL LTER study region. Study region (highlighted in box) relative to the Antarctic continent. Shades of gray illustrate bathymetry, with light gray indicating the continental shelf and dark gray the continental slope and abyssal plain. Shelf break is represented by light/dark gray interface near 1000 m, extending down to 3000 m. PAL LTER grid lines are numbered from 600 to −100, with the far slope (200), shelf (100), and coast (040) stations indicated for reference (Waters & Smith 1992). Grid lines are distanced 100 km apart and individual stations for a given grid line are 20 km apart. Horizontal lines delineate the "North," "South," and "Far South" subregions. Vertical lines indicate the coastal, shelf, and slope subregions. An, Anvers Island, the location of Palmer Station; MB, Marguerite Bay; Ad, Adelaide Island; Ch, Charcot Island. Triangles indicate *Limacina helicina antarctica* collection stations. Asterisks show stations where surface water samples were collected for comparative sequencing with pteropod guts.

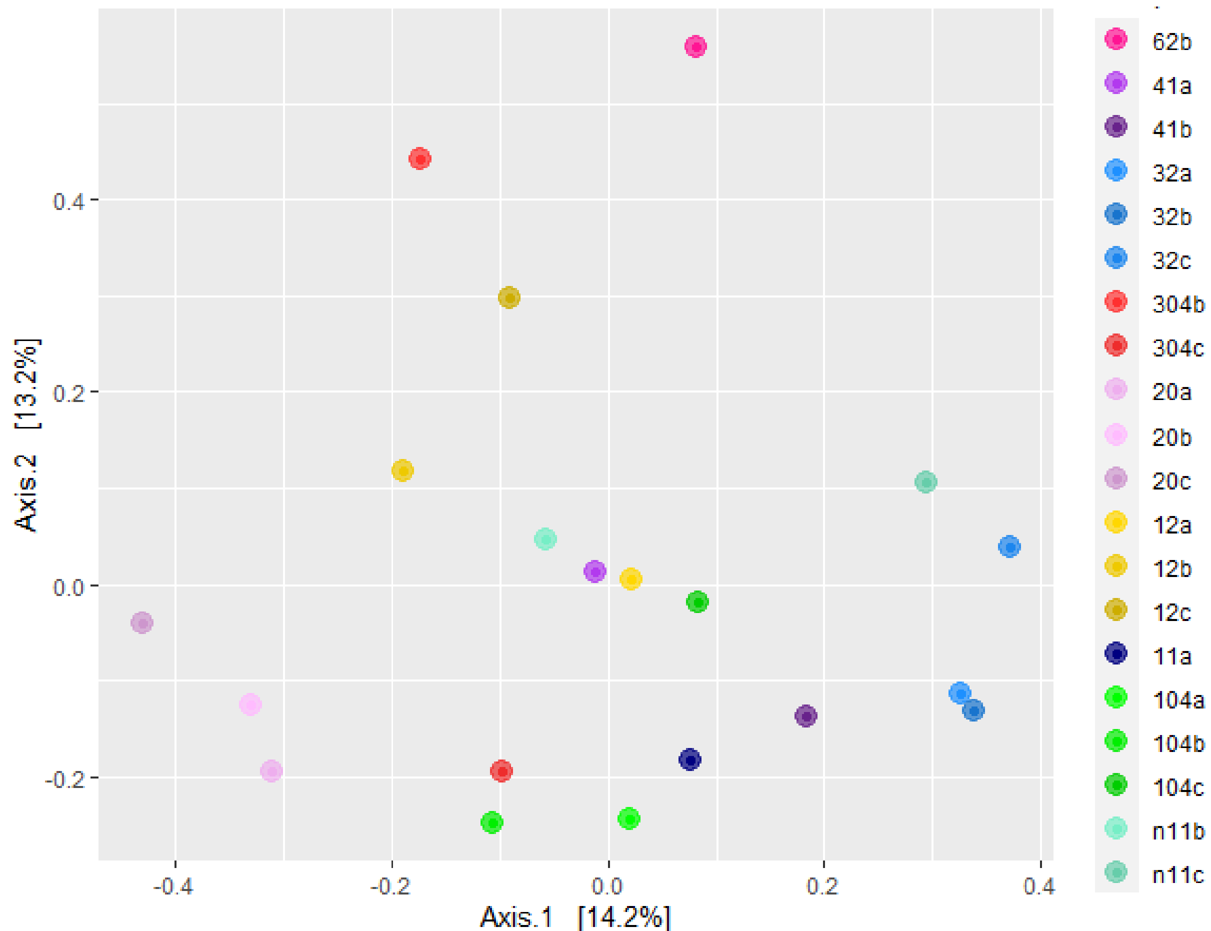

Fig. S2. Principal coordinate analysis of prokaryote communities from *Limacina helicina antarctica* guts based on 16S amplicon sequence variants. Corresponding colors indicate sampling location. Letters next to station numbers represent the number of replicates from triplicate samples. Samples labeled n11 are from station -100.000; Samples labeled 104 are from station 100.040; Samples labeled 11 are from station 100.100; Samples labeled 12 are from station 100.200; Samples labeled 20 are from station 200.000; Samples labeled 304 are from 300.040; Samples labeled 32 are from station 300.200; Samples labeled 41 are from station 400.100; Sample labeled 62 is from 600.200.

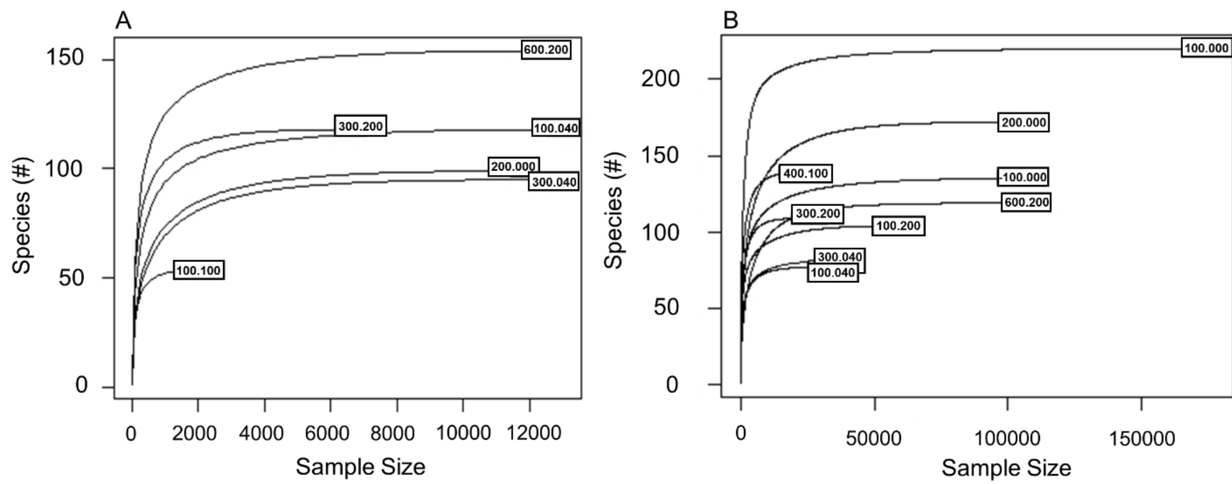

Fig. S3. Rarefaction curves for water samples (A) and *Limacina helicina antarctica* gut contents (B) at each sampling station based on combined 16S and 18S sequences.

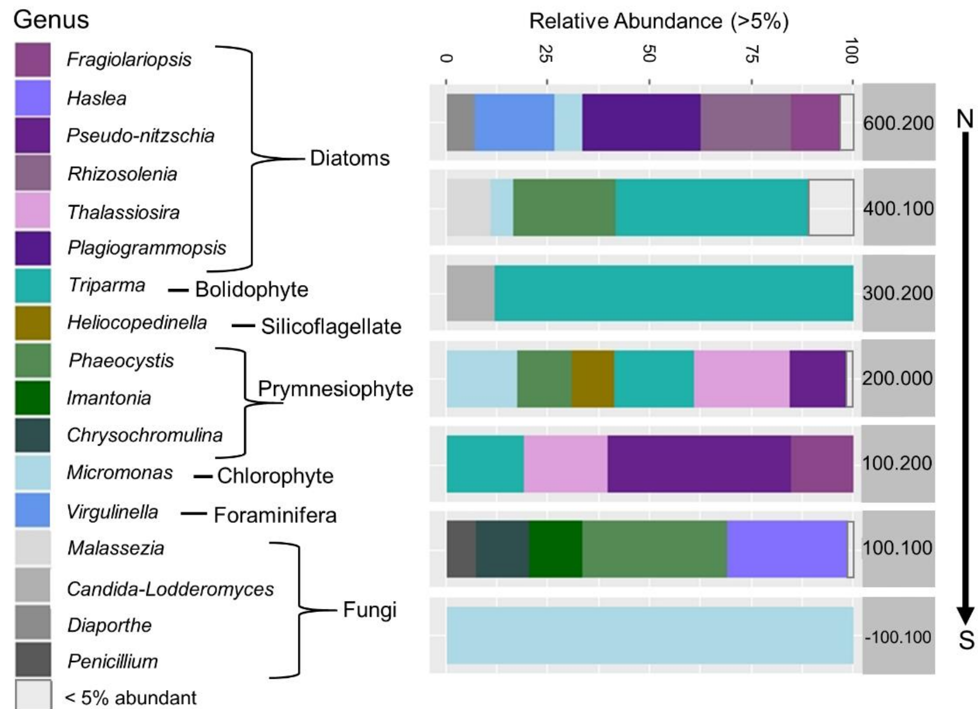

Figure S4. Eukaryotic community composition by genus present within *Limacina helicina antarctica* pteropod guts at sampling stations as determined by 18S amplicon sequence variants. *L. h. antarctica* host sequences were removed for this analysis. Genera shown represent > 5% relative abundance. Genus levels are listed in key to left. Key to right identifies each sampling station number, arranged from north (N, top) to south (S, bottom). Note: samples from station 300.040 and 100.040 were removed from this analysis due to low number of sequences. Dinoflagellates, ciliates, and salps are not included here because their taxonomies could not be resolved to genus level.

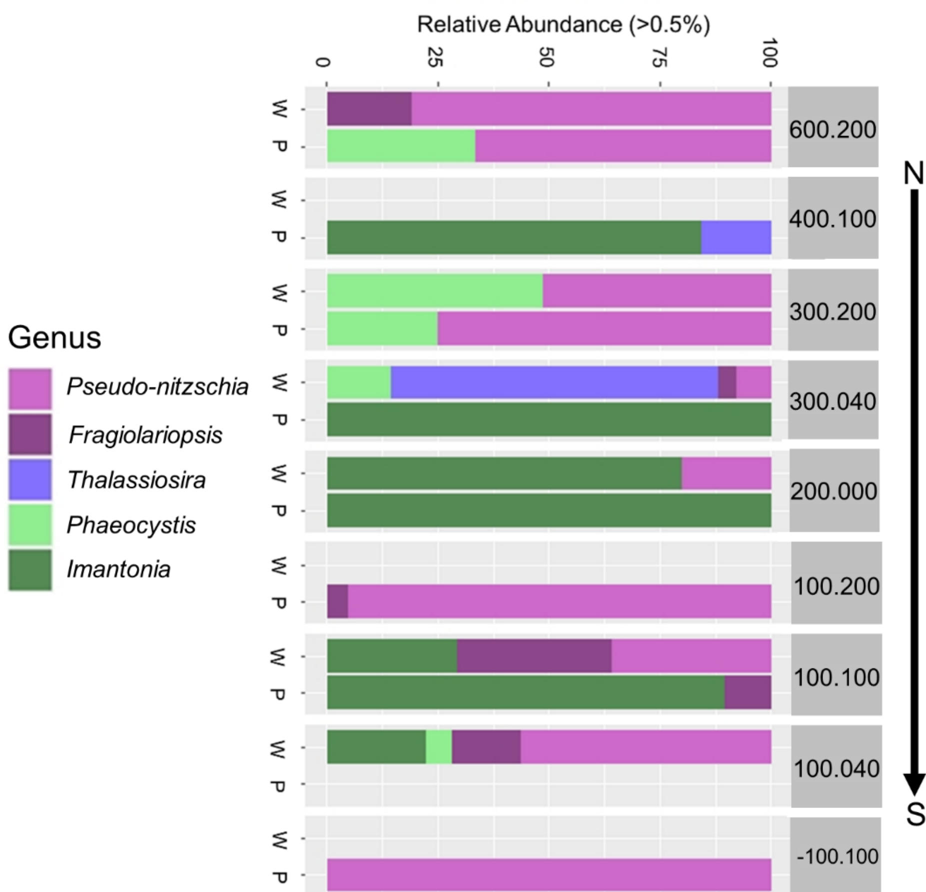

Fig. S5. Phytoplankton community composition by genus present in seawater (W) and within *Limacina helicina antarctica* pteropod guts (P) at sampling stations as determined by chloroplast 16S amplicon sequence variants (ASVs). Genera shown represent > 0.5% relative abundance. Genera are listed in key to left. Key to right identifies each sampling station number, arranged from north (top) to south (bottom). Note: there were no detectable chloroplast 16S ASVs in gut sample 100.040.