Supplemental Figure Legends:

Figure S1. Summer and winter Antarctic SSU rRNA gene sequence distribution. (a) Antarctic winter archaeal (inner most pie; n=718 sequences), winter bacterial (inner ring; n=411 sequences) and summer bacterial (outer ring; n=411 sequences) clone libraries. Sequences were clustered at a distance of 0.03, then sequences from each cluster were classified and nearest relatives identified using BLAST as described in the materials and methods. The legend lists classifications and affiliations of the clusters. For clusters that had the same nearest relative, identical colors are used in the chart; additionally, *Archaea* are in shades of orange, *Alphaproteobacteria* in reds and purples, *Gammaproteobacteria* in blues and *Flavobacteria* in yellows. Clusters representing <1% of the library are shown in gray. (b) Rarefaction curves for summer and winter SSU rRNA gene sequences (411 sequences were sampled from each library for comparative purposes) clustered at a distance of 0.03.

Figure S2. MEGAN comparison of end-sequences from winter and summer environmental genomes. Circles represent the normalized proportion of end sequences in the winter (blue) and summer (red) end-sequences that could be placed at a node in the phylogenetic tree using the least common ancestor algorithm described in Huson et al. (2007).

Figure S3. Phylogenetic relationships of predicted Antarctic RuBisCO proteins. This neighbor joining bootstrap consensus tree (1000 bootstraps) based on 133 aligned amino

acids shows the evolutionary relationships of these proteins to their nearest neighbors. WEG sequences are identified with ANTWFO and AWFS while SEG sequences are identified with ANTSFO. The number of sequences per phylogenetic group is indicated.

Figure S4. Cluster of orthologous group analysis for significant difference between winter and summer environmental genomes using STAMP (Parks and Beiko 2010) . A. COG classes – Fisher exact test, two-sided, testing for the difference between proportions – using Newcombe-Wilson method for calculating CIs - 95% CI and Storey false discovery rate. B. Fisher's exact test for calculating statistical significance – two sided hypothesis test; Testing for the difference between proportions – using Newcombe-Wilson method for calculating CIs - 95% CI; results here used the Storey's false discovery rate method to correct for multiple tests. Q-value indicates the percentage of false positives that should be expected among all features (i.e. if Q=0.047 then expect 4.7 false positives out of 100 features).

Figure S5. Density plot of GC content for eight ocean metagenome samples. Kernel smoothed density plot of the GC content of the end-sequence libraries for the Antarctic summer-winter comparison, four GOS sample sites, and two HOTs sites. Note that all GOS data are from short insert libraries, and the Antarctic and HOTs libraries were large insert fosmid libraries.

Figure S6. End-sequence reads assigned to specific genera of bacterial genomes based on the MEGAN analysis. Companion figure to Figure 3 **A.** Reads assigned to

Gammaproteobacteria, categorized by genera, from winter and summer end sequence libraries. These are the genera based on normalized hits – only those genera that had >20 hits are shown for ease of data presentation. **B.** Reads assigned to the Alphaproteobacteria, categorized by genera and labeled as in A.