Supplemental figure and table legends

Supplemental figure 2.1: Functional analysis of the host and symbiont light organ proteomes. (A). COG category counts for all symbiont proteins present in the light organ (including putative identifications). **(B)**. KOG category counts for all host proteins present in the light organ (including putative identifications) using representative light organ ESTs. (COG/ KOG key: J- translation, ribosomal structure, and biogenesis, A- RNA processing and modification, K- transcription, L- replication, recombination and repair, B- chromatin structure and dynamics, D- cell cycle control, cell division and chromosome partitioning, Y- nuclear structure, V- defense mechanisms, T- signal transduction mechanisms, M- cell wall, membrane and envelope biogenesis, N- cell motility, Z- cytoskeleton, W- extracellular structures, U- intracellular trafficking, secretion and vesicular transport, O- posttranslational modification, protein turnover and chaperones, C- energy production and conversion, G- carbohydrate transport and metabolism, E- amino acid transport and metabolism, F- nucleotide transport and metabolism, H- coenzyme transport and metabolism, I- lipid transport and metabolism, P- inorganic ion transport and metabolism, Q- secondary metabolites biosynthesis, transport and catabolism, R- general function prediction only, S- function unknown)

Supplemental file 2.1: Symbiont proteins detected in light organ exudates and central cores by MudPIT and LC MS/MS.

Supplemental file 2.2: BUDAPEST analysis of host proteins detected in light organ exudates and central cores by MudPIT and LC MS/MS.

Supplemental file 2.3: Additional symbiont proteins detected in light organ exudates and central cores by MudPIT and LC MS/MS categorized by functions relevant to survival in the light organ crypts.

Supplemental file 3.1: Comparison of WT and *vig*$^-$ CFUs per light organ from 48 hr colonized squid.

Supplemental file 4.1: ProteinPilot false discovery analysis.

Supplemental file 4.2: Pairwise comparisons of iTRAQ reporters before and after normalization.

Supplemental file 4.3: ProteinPilot summary of hemocyte peptides identified and quantified by iTRAQ after Cyclic Loess normalization. (**Terminology-N:** rank of a particular protein in respect to the other proteins identified. **Unused**: The "protscore" for a particular protein. Originating from the confidence of all of the unique peptides for that protein. **Total**: The "protscore" for a particular protein using all of the available peptides identified for that specific protein. **%Cov(95)**: The number of amino acids matching the identified protein sequence with confidence greater than 95%, divided by the total number of amino acids in the protein. **Accessions**: Identification number for the protein from the squid protein sequence database. **Names**: Protein sequence name and taxonomic homology. **Used**: Displays if a peptide has been selected for quantitation (1) or not (0). **Annotation**: Shows how the "used" function is determined. Set to "auto" and determined by ProteinPilot. Auto selects peptides that are not shared with other proteins and contain at least 2 iTRAQ reporter ions. **Contribution**:

The contribution of the specific peptide to the unused protscore of a particular protein. **Confidence**: Highest confidence for the peptide identification. **Sequence**: Identified peptide sequence. **Modifications**: Indicates presence and position of iTRAQ reagents on the peptide sequence. **Cleavages**: Indicates any missed peptide sequence cleavage sites. Peptides with missed cleavages were not included for this analysis. **Delta Mass**: The difference in mass between the precursor molecular weight and the theoretical molecular weight of the peptide sequence. **Precursor MW**: Monoisotopic mass of the ion fragmented in this analysis. **Precursor *m/z***: Mass to charge ratio for the ion fragmented in this analysis. **Theoretical MW**: Theoretical mass of the ion fragmented in this analysis calculated from the peptide sequence and modifications. **Theoretical *m/z***: Theoretical mass to charge ratio. **Sc**: Score of the peptide determined by the matching ions from the associated spectrum. **Spectrum**: Identifier for the peptide MS/MS spectrum in this analysis. **Time**: The retention time for the particular spectrum. **Label Error**: The error associated with each iTRAQ ion area measured for a peptide. **Label Area**: The normalized area of each iTRAQ ion for a peptide (Log2 transformed). **Background**: Background calculated for each peptide by ProteinPilot.)

Supplemental file 4.4: Fold change summary of all hemocyte proteins identified by iTRAQ with 2 or more peptides (sym relative to cured hemocytes). Proteins highlighted in gray are mentioned in the text as housekeeping proteins maintaining a Log2 fold change near zero. Proteins highlighted in red are more abundant in sym and have a Log2 fold change greater than 0.80. Bold red proteins are significantly more abundant (p-value <0.05). Proteins highlighted in blue are less abundant in sym and have a Log2 fold change less than -0.80. Bold blue proteins are significantly less abundant (p-value <0.05). (Terminology-**Accessions**: Identification number for the protein from the squid protein sequence database. **Names**: Protein sequence name and taxonomic homology. **Unused**: The "Protscore" for a particular protein. Originating from the confidence of all of the unique peptides for that protein. **Fold Change**: The fold change for the average of the areas measured for the sym iTRAQ ions (114 and 115) relative to the average of the areas measured for the cured iTRAQ ions (116 and 117). **Average Log2 Fold Change**: Fold change of sym relative to cured Log2 transformed. **Median**: Log2 of the median fold change. **Standard Deviation:** Standard deviation of the Log2 fold change. **Coefficient of Variance**: Coefficient of variance of the Log2 fold change. **P-Value**: P-value from a paired t-test on the average of the areas measured for the sym iTRAQ ions (114 and 115) and the average of the areas measured for the cured iTRAQ ions (116 and 117) for proteins with 2 or more peptides. **Peptides**: Number of peptides contributing to the iTRAQ ratio.)

Supplemental file 4.5: Label-free spectral counting summary. Sheet 1: Sym and cured hemocyte normalized spectral abundance factors (NSAFs) for each protein and each technical replicate. Sheet 2: Significantly different proteins (p-value <0.05, unpaired t-test). Proteins highlighted in red are more abundant in sym and have p-value <0.05. Bold red proteins are significantly more abundant and meet the Log2 fold change cutoff of greater than 0.59 (1.5 fold). Proteins highlighted in blue are less abundant in sym and have a p-value <0.05. Bold blue proteins are significantly less abundant and meet the Log2 fold change cutoff of less than -0.59. Sheets 3-6: Peptide summaries for both sym and cured technical runs. The Locus identifies the accession number of the matched protein. All numbers in this row correspond to the first row of column headers. The second row of column headers represents the numbers immediately below the accession number row. (Terminology- **Sequence Count:** Number of peptides identified for a respective protein. **Spectrum Count:** Number of spectra identified for a respective protein. **Sequence Coverage:** % of amino acids matched to the protein sequence by the identified peptides. **Length:** Number of amino acids in the matched protein sequence. **MolWt:** Molecular weight determined by the protein sequence (Daltons). **pI:** Isoelectric point of the protein based on the protein sequence. **Validation Status:** "U" means unvalidated. **Descriptive Name:** Protein

name. **Asterisks (*):** Designates a unique peptide. **Xcorr:** cross correlation value determined by the SEQUEST search. **deltCN:** The difference between the top Xcorr and the 2nd Xcorr divided by the top Xcorr. **ObsM+H+:** Observed protonated mass of the parent ion. **CalcM+H+:** Calculated protonated mass of the parent ion. **Total Intensity:** The intensity of the parent ion identified. **SpR:** Spectral rank during SEQUEST preliminary scoring. **SpScore:** SEQUEST preliminary scoring based on number of matching fragment ions. **IonProportion:** % of matching fragment ions identified in the spectra. **Redundancy:** The number of times a respective peptide was identified. **Sequence:** peptide sequence identified by SEQUEST.) **Sheets 7-10: Protein summaries for both sym and cured technical runs.**

Supplemental file 4.6: Secondary controls for cathepsin L2 immunocytochemistry. Representative control maximum projection confocal images of sym hemocytes. The hemocytes were only stained with an Alexa Fluor 488 goat anti-rabbit secondary antibody. All hemocytes were counterstained with the nuclear stain DRAQ5 (n= 14 hemocytes, DIC- differential interference contrast image, scale bar measurements- 10 μ M).

Appendix II Supplemental file A1.1: Complete protein and peptide summary of the hemocytes.

Appendix II Supplemental file A1.2: Complete list of hemocyte proteins not identified by the transcriptome.