Assembly of a reference transcriptome for the gymnosome pteropod *Clione limacina* and profiling responses to short-term CO_2 exposure

Ali A. Thabet[ab*], Amy E. Maas[ac*], Samy A. Saber[d] and Ann M. Tarrant[a#]

a. *Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543*
b. *Zoology Dept., Faculty of Science, Al-Azhar University in Assiut, Assiut, Egypt.*
c. *Bermuda Institute of Ocean Sciences, St. George's GE01, Bermuda*
d. *Zoology Dept., Faculty of Science, Al-Azhar University, Cairo, Egypt.*

* denotes equal authorship contribution

Corresponding author

Abstract:

The gymnosome (unshelled) pteropod *Clione limacina* is a pelagic predatory mollusc found in polar and sub-polar regions. It has been studied for its distinctive swimming behavior and as an obligate predator on the closely related thecosome (shelled) pteropods. As concern about ocean acidification increases, it becomes useful to compare the physiological responses of closely-related calcifying and non-calcifying species to acidification. The goals of this study were thus to generate a reference transcriptome for *Clione limacina*, to expose individuals to CO_2 for a period of 3 days, and to explore differential patterns of gene expression. Our Trinity assembly contained 300,994 transcripts of which ~26% could be annotated. In total, only 41 transcripts were differentially expressed following the CO_2 treatment, consistent with a limited physiological response of this species to short-term CO_2 exposure. The differentially expressed genes identified in our study were largely distinct from those identified in previous studies of thecosome pteropods, although some similar transcripts were identified, suggesting that comparison of these transcriptomes and responses may provide insight into differences in OA responses among phylogenetically and functionally distinct molluscan lineages.

Key Words: ocean acidification, zooplankton, gene expression, next generation sequencing, mollusc, invertebrate

1. Introduction:

Recently, shelled pteropods (Gastropoda: Thecosomata, "thecosomes") have become the focus of research due to the sensitivity of their aragonitic shells to ocean acidification (Manno et al. in review). The genus that has been most extensively studied, *Limacina*, is abundant in the epipelagic zone, and globally *Limacina spp.* are ecologically important in food webs and carbon fluxes (Lalli and Gilmer 1989; Hunt et al. 2008). Numerous studies have demonstrated that in polar or upwelling regions the shells of these pelagic snails are impacted by undersaturation of waters with respect to calcium carbonate (i.e. Bednaršek et al. 2012; Bednarsek and Ohman 2015). As a consequence, a number of transcriptomic resources have recently become publically available for thecosomes, including studies of changes in gene expression in response to short-term CO_2 exposures (Koh et al. 2015; Maas et al. 2015; Johnson and Hofmann 2016; Moya et al. 2016). Responsive genes have varied among species, but have broadly included genes with roles in biomineralization, neural function, and energetic metabolism

In polar and sub-polar regions, the unshelled pteropods (Gymnosomata, "gymnosomes") *Clione limacina limacina* (Phipps 1774) and *Clione limacina antarctica* (Smith, 1902) are major predators of *Limacina* spp. Gymnosomes, unlike the species in the sister thecosome order (Klussmann-Kolb and Dinapoli 2006), only have calcium carbonate shells during their larval veliger stage. As adults, they lose their cup-shaped veliger shell and transition to a streamlined body shape that allows them to be efficient hunters of *Limacina* spp., which are their exclusive prey (Lalli and Gilmer 1989). They are active swimmers with a locomotory system that is easily observed and quantified (Satterlie et al. 1985; Gilmer and Lalli 1990; Borrell et al. 2005). The coevolution of the gymnosome–thecosome clade has led to the development of highly specialized mechanisms of prey-capture and predator-evasion in what appears to be a predator–

prey arms race (a reciprocal relationship by virtue of its specificity; Brodie and Brodie 1999). Predation involves tactile recognition of the prey species, specific prey-capture swimming behavior, rapid capture using highly specialized buccal cones, and complete extraction of the prey from its shell, using numerous hooks and a toothed radula. Highly efficient digestion and assimilation follow extraction (Conover and Lalli 1972; Conover and Lalli 1974; Lalli and Gilmer 1989). Thus, the predator avoidance and bafflement properties of the thecosome shell are intimately associated with the success of the gymnosome feeding mechanism.

Due to their obligate trophic linkage, *Clione* spp. are found in the same water masses as *Limacina* spp., experience the same changes in ocean chemistry, and serve as a closely related non-calcifying contrast to the thecosomes. The aims of this study were to conduct a short-term CO_2 exposure on wild-caught juveniles of *Clione limacina limacina* (hereafter *C. limacina*), assemble the transcriptome *de novo,* and perform differential gene expression (DE) analyses to better understand how these non-calcifying pteropods may respond to ocean acidification.

2. Materials & Methods:

2.1 Larval collection:

Juvenile *Clione limacina* (post-metamorphosis, but retaining ciliary bands and lacking full parapod development) were collected from multiple tows near ~42 2.0° N and ~70 14.0° N in the Gulf of Maine on April 27th 2015 (Table 1). Tows were conducted with a specialized Reeve net with a 333 µm mesh net and large ~20 L cod end from the R/V *Tioga*. About 190 juveniles were maintained in 1 L jars of seawater collected *in situ*. To minimize thermal stress, jars were initially held in an 8 ± 1°C refrigerator and then transported in coolers to the laboratory within 12 hours of collection.

84 **2.2 CO_2 Exposures:**

85 Short-term CO_2 exposure was performed in an 8 ± 1°C walk-in environmental chamber at
86 Woods Hole Oceanographic Institution. Seawater, collected via a submersible pump from ~30 m
87 in the same region as pteropod collection, was filtered (1 μm pore size) and then pre-equilibrated
88 with CO_2 (400 ppm and 1200 ppm) for approximately 12 hours prior to the start of the experiment.
89 This water was then distributed into 12 L carboys (3 per treatment), each containing 6 L of CO_2-
90 equilibrated seawater. Thirty juveniles of *C. limacina* were randomly assigned to each carboy,
91 where they were maintained without feeding. After three days, they were gently siphoned onto a
92 200 μm mesh net, then transferred to crystallization dishes. Surviving swimming individuals were
93 counted under a stereo microscope and then preserved in pools of 7-10 individuals in RNAlater.

94 A water sample was taken for dissolved inorganic carbon (DIC) and total alkalinity (TA)
95 when the experiment was started to validate the carbonate chemistry associated with each
96 treatment. In addition, temperature, pH and salinity were measured on days 1 and 3. TA was
97 measured using an Apollo SciTech alkalinity auto-titrator, an Orion 3 Star pH meter, and a Ross
98 combination pH electrode based on a modified Gran titration method (Wang and Cai 2004). DIC
99 was analysed with a DIC auto-analyser (AS-C3, Apollo SciTech, Bogart, GE, USA) via
100 acidification and non-dispersive infrared CO_2 detection (LiCOR 7000: Wang and Cai 2004). The
101 saturation state of aragonite (Ω_A), pCO_2, and pH were calculated from DIC and TA with the
102 CO2SYS software (Pierrot et al. 2006), using constants K_1 and K_2 by Dickson and Millero
103 (1987), and the $KHSO_4$ dissociation constant from Dickson (1990). Temperature was measured
104 using a mercury thermometer, and pH was determined using a USB 4000 spectrometer with an
105 Ls-1 light source and a FIA-Z-SMA-PEEK 100 mm flow cell (Ocean Optics, Dunedin, FL,

106 USA) following the protocol of White et al. (2013). Salinity was measured using a seawater
107 refractometer (Hanna Instruments, model 96822).
108
109 **2.3 RNA Extraction and Illumina Sequencing:**
110 Total RNA was extracted from pools of *C. limacina* juveniles using the Aurum Total
111 RNA Fatty and Fibrous Tissue Kit (Bio-Rad). RNA purity and integrity were quantified using a
112 Nanodrop ND-1000 spectrophotometer, and quality was assessed using a Bioanalyzer 2100.
113 Purified RNA samples were submitted to the Genomic Services Laboratory at HudsonAlpha
114 (Huntsville, AL, USA) in four replicates of each treatment for library construction and
115 sequencing. Libraries were synthesized using NEBNext® Ultra Directionality Kit (New England
116 BioLabs) from 1 μg total RNA per sample. Samples were multiplexed, and sequenced in one
117 lane of the Illumina HiSeq2000 platform as 100 base pair (bp) paired-end reads.
118 ***2.4 De novo* Transcriptome Assembly and Differential Expression Analysis:**
119 Adapter sequences, low quality bases (phred score <20 bp), and the first 15bp at the 5'
120 end of the reads were removed from raw HiSeq data using the Trimmomatic program (v 0.33) in
121 paired-end mode (Bolger et al. 2014) (ILLUMINACLIP:TruSeq2-PE.fa:2:40:13 LEADING:20
122 TRAILING:20 SLIDINGWINDOW:4:15 HEADCROP:15 MINLEN:30). All reads from both
123 treatments were then concatenated, and the transcriptome was assembled de novo using the
124 Trinity software package (v.2.1.1) (Haas et al. 2013). The default parameters were slightly
125 modified with a minimum kmer coverage of 2, a maximum internal gap in the same path of 15
126 and a maximum difference within the same path of 4. Any transcript that was not present in at
127 least three samples was excluded from the assembly. This reduced dataset was then further
128 clustered to a 99% similarity level with the program cd-hit est (Li and Godzik 2006), and the

129 longest contig within each cluster was retained. The resulting assembly was annotated and used
130 for downstream DE analysis. Transcripts were annotated using BLASTX searches of the NCBI
131 non-redundant (nr) and Swiss-Prot databases with an e-value threshold < 1.0 e^{-5}. Annotated
132 transcripts were then searched against the InterPro database using the Blast2Go program with
133 default parameters. Results from nr, Swiss-Prot and InterPro searches were integrated within
134 Blast2Go, and gene ontology (GO) terms were assigned using the default parameters with an e-
135 value threshold of < 1.0 e^{-6}. GO terms were compared at the second level and categorized to
136 biological process, molecular function and cellular components. Finally, Kyoto Encyclopaedia of
137 Genes and Genomes (KEGG) pathways were populated with this set of annotated transcripts.
138
139 Using scripts bundled with Trinity (Haas et al. 2013), reads from each sample were
140 mapped back to the reference transcriptome assembly using Bowtie2 (v 2.2.6) (Langmead and
141 Salzberg 2012) in paired-read mapping mode with default parameters. Abundances of the
142 assembled transcripts was estimated using RSEM (Li and Dewey 2011). Subsequently, counts of
143 transcripts were TMM- (trimmed-mean of M-values) and then FPKM-normalized (fragments per
144 feature kilobase per million reads mapped) to account for differences in RNA production across
145 samples (Robinson and Oshlack 2010) and transcript length, respectively. DE analysis of
146 transcripts and isoforms between the two CO_2 treatments was performed using the edgeR
147 package (Robinson et al. 2010) with a p-value cutoff and false discovery rate (FDR) of 0.005 and
148 a minimum 4-fold change in expression. We then reciprocally searched for similar DE sequences
149 from other published pteropod studies using TBLASTX with an e-value cutoff of 1.0 e^{-5}.
150 **3. Results:**

151 Temperature, salinity and carbonate chemistry parameters were near targeted values during the
152 exposure period (Table 2). The ambient treatment was consistent among replicates and over time
153 (~440 µatm CO_2, ΩAr =1.6); the high treatment was slightly more variable, but consistently
154 distinct from ambient (~1000-1080 µatm CO_2, ΩAr = 0.80-0.76). Percent survival of *C. limacina*
155 juveniles in the high CO_2 was slightly but significantly lower than the ambient (79% and 97%
156 respectively; $t(4) = 5.060$, $p = 0.007$).

157 **3.1 Gene Compliment and Annotation**

158 High-throughput sequencing produced a total of 289 million paired-end (100 bp) raw
159 reads. After quality trimming, 255 million paired-end reads (88%) were initially assembled into
160 293,756 trinity components ('genes') that contained 477,401 trinity transcripts ('isoforms').
161 transcriptome (Table 3). After filtering and clustering, the final transcriptome was composed of
162 300,994 contigs and 181,861 unigenes. The size of transcripts of the final assembly ranged from
163 201 bp to 30,094 bp with the average length of 604 bp, median length 359 bp and an N50 of 816
164 bp. 95.5 % of trimmed reads were successfully mapped back to the reference

165 BLASTX searches of the nr database resulted in 77,265 transcripts (25.7%) with at least
166 one hit. For most of these (51,386), the top BLAST hit was to *Aplysia californica,* another
167 gastropod (Figure 1). Functional annotation with at least one GO term was accomplished for
168 41,252 transcripts (73% and 17.8% of annotated and total transcripts, respectively; Figure 2).
169 Transcripts were associated with a total of 65 different GO terms. Within the biological
170 processes group, the most common terms were cellular process (20%), metabolic process (19%),
171 single-organism process (17%) and biological regulation (9%). Within the cellular component
172 group, the most common assignments were to cell, cell part (20%), organelle (16%) and
173 membrane (15%). In the molecular function category, most of the annotated genes were related

174 to binding (60%) and catalytic activity (40%). Of the transcripts with at least one BLAST hit,
175 InterProScan analysis showed 76,543 (~99%) of annotated transcripts had at least one InterPro
176 protein domain. The most frequently identified domains were P-loop containing nucleoside
177 triphosphate hydrolase (IPR027417), Zinc finger related domains (IPR007087, IPR013087,
178 IPR015880) and immunoglobulin-like fold (IPR013783) (Supplementary File 1). Also, BLASTX
179 queries against Swiss-Prot database resulted in annotation of 44,949 (~15%) of *Clione* transcripts
180 (Supplementary File 1). Annotated gene lists are provided in Supplementary File 2.

181 Analysis of KEGG (Kyoto Encyclopedia of Genes and Genomes) metabolic pathways
182 revealed 16,273 (~21%) of the transcripts annotated via BLASTX versus nr had at least one
183 significant match in the KEGG database and were involved in 135 pathways, which included
184 1240 enzymes (Supplementary File 2). Analyses of these pathways suggests that there is good
185 coverage of a number of essential metabolic pathways such as glycolysis/gluconeogenesis and
186 the TCA cycle.

187 **3.2 Differential expression analysis:**

188 Short-term exposure (3-days) to high CO_2 resulted in a small change in the gene
189 expression profile of juvenile *C. limacina* compared with individuals reared under ambient
190 conditions. In total, 41 transcripts corresponding to 39 genes were DE between treatments (Table
191 4; Supplementary File 3). Of these, 28 transcripts were up-regulated and 13 down-regulated.
192 Annotation of these transcripts was limited, with 71% remaining unidentified. Those that were
193 down-regulated in the high CO_2 treatment included a dynactin subunit and an adhesion g protein-
194 coupled receptor e3. Those transcripts that were up-regulated included myosin, androglobin, a
195 nuclear receptor and chondroitin proteoglycan.

196 In comparing these results with lists of differentially expressed genes from published CO_2
197 exposure studies conducted with other pteropod species, no DE genes were shared with *Clio*
198 *pyramidata* (Genbank accesssion PRJNA210933; Maas et al. 2015). However, four homologous
199 sequences (based on reciprocal BLAST) responded similarly to CO_2 exposure in both *Limacina*
200 *retroversa* (Genbank accession PRJNA260534; Maas et al. in prep.) and *C. limacina* (Table 4);
201 these included a mucin, androglobin/calponin-7, and methuselah-like 3. Although the
202 transcriptome and raw reads are available for *Helicinoides inflatus* (Genbank accession
203 PRJNA312154; Moya et al. 2016), the full list of DE genes is not; of those DE genes that were
204 reported in this dataset (221/573 transcripts), there were no DE transcripts that were shared with
205 our analysis. Sequences of DE genes from *Limacina helicina* (Koh et al. 2015) are not currently
206 available online, but a comparison of the BLAST-based annotation of DE genes from their study
207 did not reveal any similarities (putative homologs) with the DE genes we identified in *C.*
208 *limacina*.

209

210 **4. Discussion:**

211 In the present study, we generated a transcriptome for the unshelled pteropod *Clione*
212 *limacina* using juveniles exposed to ambient and elevated CO_2 for a period of three days. Only
213 26% of the transcripts could be annotated through BLAST-based searches. Using similar methods
214 and the same cutoff value, researchers were recently able to annotate 37% percent of the transcripts
215 from the thecosome *Limacina helicina antarctica* (Johnson and Hofmann 2016). Not surprisingly,
216 both studies have found that most annotated pteropod transcripts (66% present study of *C.*
217 *limacina,* 62% *L. helicina antarctica*) are most similar to sequences from the closest species with
218 a sequenced genome, the California sea hare, *Aplysia californica*. The overall lack of annotation

219 is an ongoing problem with non-model transcriptomes and represents one of the greatest hurdles
220 to our ability to interpret changes in gene expression. Despite this limitation, the addition of the
221 *Clione* transcriptome allows for directed searches of genes of interest that may help to improve
222 our understanding of the phylogenetic diversity and complexity of the molluscan lineages.

223 We found that *C. limacina* juveniles have limited transcriptional response to short-term
224 elevated CO_2, which may reflect the short duration of exposure and/or be a consequence of
225 physiological lack of sensitivity to this level of CO_2. (Maas et al. 2015) found a similar limited
226 number of DE genes (29) in the shelled pteropod *Clio pyramidata* following shorter-term (12 h)
227 exposure to elevated CO_2. In contrast, comparable CO_2 exposures (3 d) of two other thecosome
228 species resulted in substantial changes in gene expression including hundreds of DE genes (Koh
229 et al. 2015; Moya et al. 2016). Although all studies have been limited by low annotation success,
230 and the lack of published sequence data makes direct comparison difficult in many cases, there
231 do appear to some similarities in the sorts of transcripts that are DE between calcifying and non-
232 calcifying pteropods. The transcripts DE in *C. limacina* that were found to be similar with *L.*
233 *retroversa* included a mucin, which has gel-like properties and has been previously implicated in
234 aragonite biomineralization in molluscs (Marin et al. 2000), androglobin/calponin-7, which has
235 calcium-binding characteristics (Hoogewijs et al. 2011), and methuselah-like 3, which is thought
236 to play a role in aging and reproduction (Li et al. 2014).

237 Moya et al. (2016) reported that exposure of the thecosome *Heliconoides inflatus* to elevated
238 CO_2 for three days resulted in up-regulation of many genes involved in neuron function. These
239 included a $GABA_A$ receptor subunit as well as other glutamatergic and cholinergic synapse
240 associated transcripts. In contrast, we did not find any of the genes associated with nervous
241 function or GABA signaling to be differentially expressed in *C. limacina*. GABA receptors and

242 the GABA peptide play a number of roles in molluscs, including chemoreception (Murphy and
243 Hadfield 1997), settlement and metamorphosis (Morse et al. 1980; García-Lavandeira et al.
244 2005), stress avoidance behavior (Kavaliers et al. 1999) and as one of the neurotransmitters
245 associated with the mechanisms of feeding (Richmond et al. 1994; Díaz-Ríos et al. 1999).
246 Interestingly, GABA receptor activity is one of the physiological functions that has been shown
247 to be most sensitive to acidification in marine vertebrates. In fish experiments, OA has been
248 shown to directly influence chemosensory and anxiety related behavior; these traits have been
249 traced to the GABA receptor by the use of antagonist and agonist chemicals such as gabazine
250 and muscimol (Nilsson et al. 2012; Chung et al. 2014; Hamilton et al. 2014). Importantly, a
251 recent study using conch snails has identified that the escape response of this invertebrate is
252 negatively influenced by ocean acidification (Watson et al. 2014). Individuals kept in high CO_2
253 for 5-7 days were less likely to avoid predators. Application of the GABA antagonist chemical
254 gabazine restored anti-predator behavior, indicating that the same pathway which has been
255 identified in fish is sensitive in molluscs.

256 Specifically within gymnosomes, neurological regulation of feeding behavior has been well-
257 studied, and the eversion and coordination of the buccal cones is reliant upon the excitation of
258 GABA receptors (Arshavsky et al. 1993; Norekian and Satterlie 1993; Norekian and Malyshev
259 2005). Lack of modulation of GABA signaling in *C. limacina* may indicate either that these
260 regulatory processes are robust to CO_2 exposure, that the animals are unable to raise a
261 transcriptional response, or that since individuals were not fed in our experiment there was no
262 GABA-related feeding function to measure. A further study where individuals were offered prey
263 during exposure would help to further explore this lack of gene expression response. This
264 difference in transcriptional response between the two species may, however, indicate a

265 difference in physiological sensitivity or acclimatization potential and points to an important area
266 for future functional studies.

267

268 Overall, future studies with a focus on those genes that appear to respond similarly to
269 CO_2 exposure in multiple pteropod species, and including the earliest life stages of gymnosomes
270 (which do calcify as veligers) will be valuable to compare and contrast the physiological
271 response across the pteropod lineage. Integrative studies of respiration, calcification,
272 transcriptomics and proteomics, leveraging the newly available gymnosome transcriptome, will
273 be useful as we continue to seek to understand how both the calcifying and non-calcifying
274 molluscs will respond to climate change in the coming decades.

275

276 **Acknowledgements:**

277 A. Thabet is grateful for a fellowship from the Egyptian Culture and Education Bureau and for
278 mentoring from Drs. M.M. Sarhan and M.M. Fouda. Funding for this research was provided by a
279 National Science Foundation grant to Lawson, Maas, and Tarrant (OCE-1316040). Additional
280 support for field sampling was provided by the Tom Haas Fund at the New Hampshire
281 Charitable Foundation.

282

References:

Arshavsky YI, Deliagina TG, Gamkrelidze GN, Orlovsky GN, Panchin YV, Popova LB, Shupliakov OV (1993) Pharmacologically induced elements of the hunting and feeding behavior in the pteropod mollusk *Clione limacina.* I. Effects of GABA. Journal of Neurophysiology 69: 512-521

Bednarsek N, Ohman M (2015) Changes in pteropod distributions and shell dissolution across a frontal system in the California Current System. Marine Ecology Progress Series 523: 93-103

Bednaršek N, Tarling G, Bakker D, Fielding S, Jones E, Venables H, Ward P, Kuzirian A, Lézé B, Feely R (2012) Extensive dissolution of live pteropods in the Southern Ocean. Nature Geoscience 5: 881-885

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120

Borrell BJ, Goldbogen JA, Dudley R (2005) Aquatic wing flapping at low Reynolds numbers: swimming kinematics of the Antarctic pteropod, *Clione antarctica*. Journal of Experimental Biology 208: 2939

Brodie ED, Brodie ED (1999) Predator-prey arms races. Bioscience 49: 557-568

Chung W-S, Marshall NJ, Watson S-A, Munday PL, Nilsson GE (2014) Ocean acidification slows retinal function in a damselfish through interference with $GABA_A$ receptors. The Journal of Experimental Biology 217: 323-326

Conover RJ, Lalli CM (1972) Feeding and growth in *Clione limacina* (Phipps), a pteropod mollusc. Journal of Experimental Marine Biology and Ecology 9: 279-302

Conover RJ, Lalli CM (1974) Feeding and growth in Clione limacina (Phipps), a pteropod mollusc. II. Assimilation, metabolism, and growth efficiency. Journal of Experimental Marine Biology and Ecology 16: 131-154

Díaz-Ríos M, Suess E, Miller MW (1999) Localization of GABA-like immunoreactivity in the central nervous system of *Aplysia californica*. Journal of Comparative Neurology 413: 255-270

Dickson AG (1990) Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K

Deep Sea Research Part A= 37: 755-766

Dickson AG, Millero FJ (1987) A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. . Deep Sea Research Part A 34: 1733-1743

García-Lavandeira M, Silva A, Abad M, Pazos AJ, Sánchez JL, Luz Pérez-Parallé M (2005) Effects of GABA and epinephrine on the settlement and metamorphosis of the larvae of four species of bivalve molluscs. Journal of Experimental Marine Biology and Ecology 316: 149-156

Gilmer RW, Lalli CM (1990) Bipolar variation in Clione, a gymnosomatous pteropod. Am. Malacol. Bull 81: 67-75

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 8: 1494-1512

Hamilton TJ, Holcombe A, Tresguerres M (2014) CO_2-induced ocean acidification increases anxiety in Rockfish via alteration of $GABA_A$ receptor functioning. Proceedings of the Royal Society B: Biological Sciences 281: 20132509

Hoogewijs D, Ebner B, Germani F, Hoffmann FG, Fabrizius A, Moens L, Burmester T, Dewilde S, Storz JF, Vinogradov SN (2011) Androglobin: a chimeric globin in metazoans that is preferentially expressed in mammalian testes. Molecular biology and evolution: msr246

Hunt BPV, Pakhomov EA, Hosie GW, Siegel V, Ward P, Bernard K (2008) Pteropods in Southern Ocean ecosystems. Progress in Oceanography 78: 193-221

Johnson KM, Hofmann GE (2016) A transcriptome resource for the Antarctic pteropod *Limacina helicina antarctica*. Marine Genomics 28: 25-28
Kavaliers M, Perrot-Sinal T, Desjardins D, Cross-Mellor S, Wiebe J (1999) Antinociceptive effects of the neuroactive steroid, 3a-hydroxy-5a-pregnan-20-one and progesterone in the land snail, *Cepaea nemoralis*. Neuroscience 95: 807-812
Klussmann-Kolb A, Dinapoli A (2006) Systematic position of the pelagic Thecosomata and Gymnosomata within Opisthobranchia (Mollusca, Gastropoda)–revival of the Pteropoda. Journal of Zoological Systematics & Evolutionary Research 44: 118-129
Koh H, Lee J, Han S, Park H, Shin S, Lee S (2015) A transcriptomic analysis of the response of the arctic pteropod *Limacina helicina* to carbon dioxide-driven seawater acidification. Polar Biology 38: 1727-1740
Lalli CM, Gilmer RW (1989) Pelagic Snails: The Biology of Holoplanktonic Gastropod Mollusks. Stanford University Press, Stanford, CA
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357-359
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 1-16
Li C, Zhang Y, Yun X, Wang Y, Sang M, Liu X, Hu X, Li B (2014) Methuselah-like genes affect development, stress resistance, lifespan and reproduction in Tribolium castaneum. Insect molecular biology 23: 587-597
Li W, Godzik A (2006) CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659
Maas AE, Lawson GL, Tarrant AM (2015) Transcriptome-wide analysis of the response of the thecosome pteropod *Clio pyramidata* to short-term CO_2 exposure. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics: 1-9
Manno C, Bednaršek N, Tarling GA, Peck VL, Comeau S, Adhikari D, Bakker DCE, Bauerfeind E, Bergan AJ, Berning MI, Buitenhuis E, Burridge A, Chierici M, Flöter S, Fransson A, Gardner J, Howes E, Keul N, Kimoto K, Kohnert P, Lawson GL, Lischka S, Maas AE, Mekkes L, Oakes RL, Pebody C, Peijnenburg K, Seifert M, Skinner J, Thibodeau PS, Wall-Palmer D, Ziveri P (in review) Shelled pteropods in peril: assessing vulnerability in a high CO_2 ocean. Biological Reviews
Marin F, Corstjens P, De Gaulejac B, de Vrind-De Jong E, Westbroek P (2000) Mucins and molluscan calcification Molecular characterization of mucoperlin, a novel mucin-like protein from the nacreous shell layer of the fan mussel Pinna nobilis (Bivalvia, Pteriomorphia). Journal of Biological Chemistry 275: 20667-20675
Morse D, Duncan H, Hooker N, Baloun A, Young G (1980) GABA induces behavioral and developmental metamorphosis in planktonic molluscan larvae Federation proceedings, pp 3237-3241
Moya A, Howes EL, Lacoue-Labarthe T, Forêt S, Hanna B, Medina M, Munday PL, Ong JS, Teyssié JL, Torda G (2016) Near future pH conditions severely impact calcification, metabolism and the nervous system in the pteropod *Heliconoides inflatus*. Global Change Biology
Murphy BF, Hadfield MG (1997) Chemoreception in the nudibranch gastropod *Phestilla sibogae*. Comparative Biochemistry and Physiology Part A: Physiology 118: 727-735
Nilsson GE, Dixson DL, Domenici P, McCormick MI, Sørensen C, Watson S-A, Munday PL (2012) Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function. Nature Climate Change 2: 201-204
Norekian TP, Malyshev AY (2005) Coordinated excitatory effect of GABAergic interneurons on three feeding motor programs in the mollusk Clione limacina. Journal of Neurophysiology 93: 305-315
Norekian TP, Satterlie RA (1993) FMRFamide and GABA produce functionally opposite effects on prey-capture reactions in the pteropod mollusk *Clione limacina*. The Biological Bulletin 185: 248-262

376 Pierrot D, Lewis E, Wallace D (2006) CO2SYS DOS Program developed for CO_2 system calculations.
377 Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of
378 Energy ORNL/CDIAC-105
379 Richmond J, Murphy A, Lukowiak K, Bulloch A (1994) GABA regulates the buccal motor output of
380 Helisoma by two pharmacologically distinct actions. Journal of Comparative Physiology A 174:
381 593-600
382 Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression
383 analysis of digital gene expression data. Bioinformatics 26: 139-140
384 Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of
385 RNA-seq data. Genome Biol 11: R25
386 Satterlie RA, Labarbera M, Spencer AN (1985) Swimming in the Pteropod Mollusc, *Clione Limacina*: I.
387 Behaviour and Morphology. Journal of Experimental Biology 116: 189-204
388 Wang ZA, Cai W-J (2004) Carbon dioxide degassing and inorganic carbon export from a marsh-dominated
389 estuary (the Duplin River): A marsh CO_2 pump. Limnology and Oceanography 49: 341-354
390 Watson S-A, Lefevre S, McCormick MI, Domenici P, Nilsson GE, Munday PL (2014) Marine mollusc
391 predator-escape behaviour altered by near-future carbon dioxide levels. Proceedings of the
392 Royal Society B: Biological Sciences 281: 20132377
393 White MM, McCorkle DC, Mullineaux LS, Cohen AL (2013) Early exposure of bay scallops (*Argopecten*
394 *irradians*) to high CO_2 causes a decrease in larval shell growth. PLoS ONE 8: e61065
395
396

397 **Data Accessibility:**

398 Raw sequences and assembled transcriptome are archived as NCBI BioProject PRJNA314884.
399 This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank
400 under the accession GESV00000000. The version described in this paper is the first version,
401 GESV01000000.

402

403 **Author Contributions:**

404 AAT, AEM and AMT conceived of the study, analyzed the data and wrote the article. AT
405 conducted the experiments, took carbonate chemistry samples, performed the extractions, and
406 assembled the transcriptome. AAT and AEM annotated the transcriptome and performed DE
407 analysis.

408

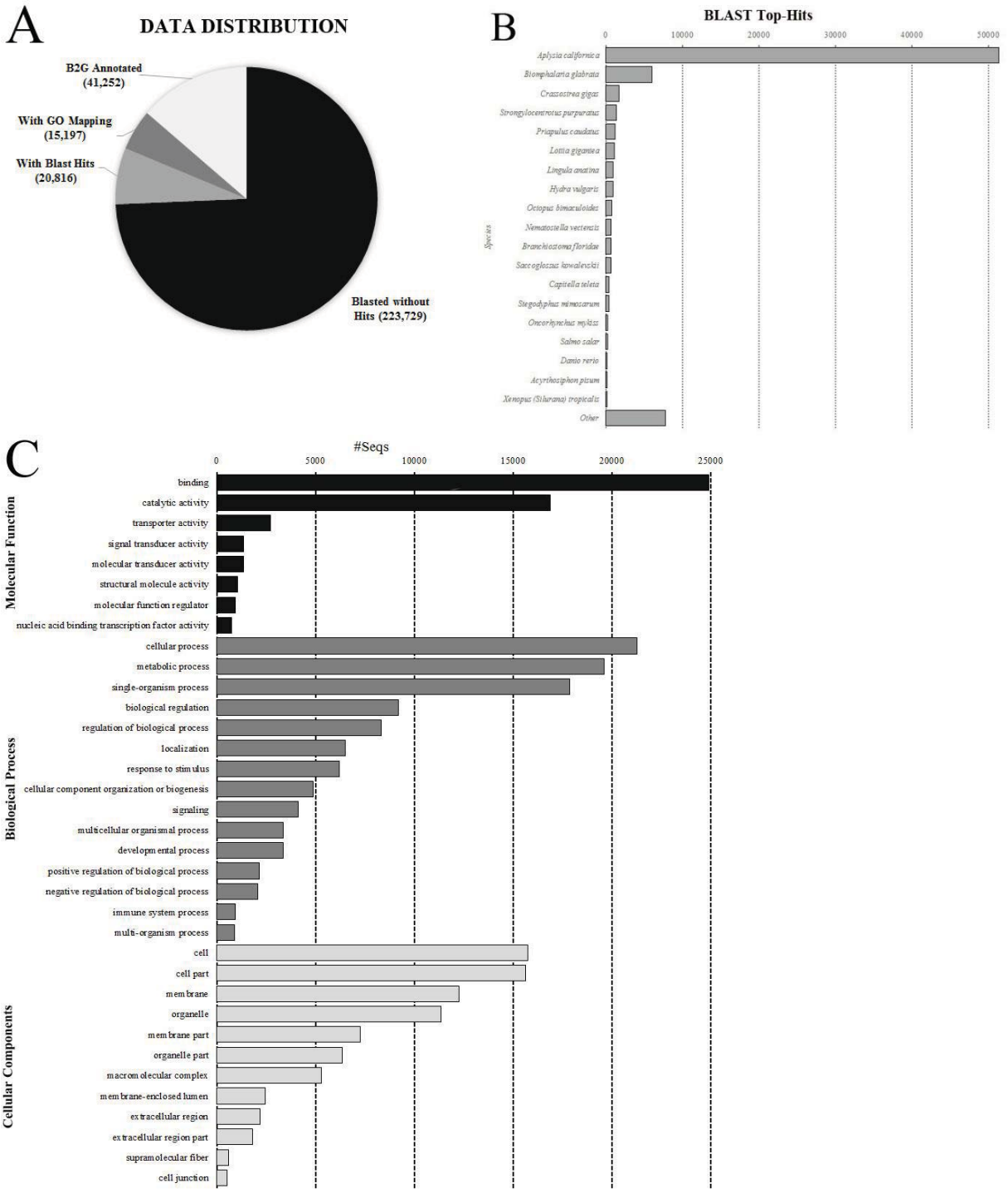

Figure 1: BLAST results from searches against the NCBI nr database; A) Data distribution of annotated transcripts. Categories include fully annotated sequences (with a Blast2Go annotation score > 55), sequences with only GO mapping and BLAST annotation, sequences with BLAST annotation only, and sequences with no annotation; B) distribution of top BLAST hit species and C) Functional gene ontology (GO) terms (level 2) for the main three categories.

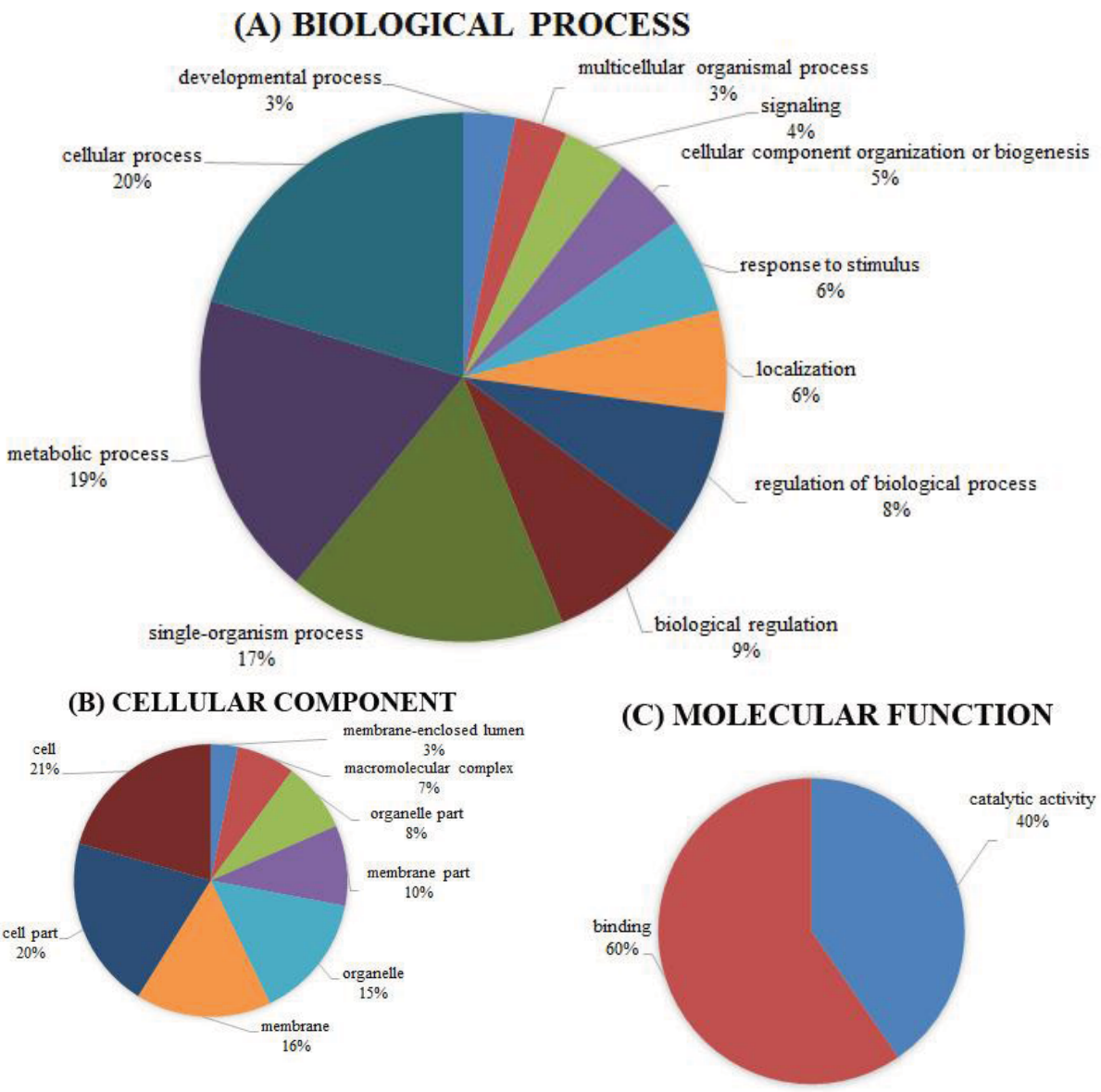

Figure 2: Level 2 gene ontology (GO) terms most commonly represented in the *Clione limacine* transcriptome, using annotation derived from a BLASTX search versus nr combined with IPS annotation as implemented within Blast2GO. Results sorted according to Biological Process (A), Cellular Component (B) and Molecular Function (C).

Table 1: Characteristics of the *Clione limacina* transcriptome sequencing project, compliant with the MIGS standards.

Item	Description
Investigation_type	Eukaryote
Species	*Clione limacina*
Project_name	Transcriptome sequencing and differential expression of *Clione limacina*
Geographic location	Gulf of Maine, USA
Collected_by	A. Thabet
Collection_date	April 27th 2015
Lat_lon	~42 2.0° N ~70 14.0° N
Environment (biome)	Marine – pelagic
Rel_to_oxygen	Aerobe
Motility	Yes

Transcriptome assembly data	
Sequencing_meth	Illumina
Sequencing technology	HiSeq2000
Assembly method	*De novo*
Assembly name	GESV00000000 (Genbank)
Assembly	Trinity v.2.1.1
Finishing_strategy	Draft
Annot_source	Blast2GO

Table 2: Average pH, salinity, temperature, and carbonate chemistry calculations. Total Alkalinity (TA), Dissolved Inorganic Carbon (DIC) and pH were all measured on day 1 and were used to calculate the pCO_2 and aragonite saturation state (ΩAr) on day 1. On day 3 the measured pH was paired with the previous TA reading to make the calculations.

Treat.	Day	TA ($\mu mol\ kg^{-1}$)	DIC ($\mu mol\ kg^{-1}$)	pH ± S.error	calc. pCO_2 (μatm)	calc. ΩAr	Salinity (psu)	Temp. (ºC) ± S.error
Ambient	1	2218	2081	7.99 ± 0.001	439	1.60	33	8
	3	*		7.99 ± 0.003	438	1.62	33	8.23± 0.09
High	1	2223	2191	7.58 ± 0.018	999	0.80	33	8
	3	*		7.63 ± 0.06	1077	0.76	33	8.23± 0.09

Table 3: Statistical information of reads generated by next-generation sequencing and *de novo* transcriptome composition and annotation success.

Statistic	Count (Total)	Count (Reduced[1])	% Annotated
Total raw reads	288,944,446		
Reads after trimming	247,594,216		
Total assembled bases	258,267,445	181,834,683	
Total transcripts	477,401	300,994	
Total "genes"	293,756	181,879	
N50	710	816	
Average length	540.99	604.06	
Median length	310	359	
Min length	200	200	
Max length	30,190	30,190	
GC content	35.57	35.72	
Annotation Success (1.0 e-5 threshold)			
nr			
BLASTX		77,265	25.6%
GO		53581	17.8%
Swis-Prot			
BLASTX		45,510	15.1%
GO		43339	14.4%

[1] After exclusion of transcripts found in <3 samples and cd-hit clustering (see Methods).

Table 4. Annotation of DE transcripts. In response to CO_2 exposure 28 transcripts were up-regulated and 13 were down-regulated (full list Supplementary File 3). Of these 6 up-regulated and 3 down-regulated transcripts could be annotated based on similarity to the nr database. Four of these DE genes were homologs (reciprocal BLAST hits) of DE genes from *L. retroversa* (Genbank accession PRJNA260534) and the DE pattern (up- or down-regulation) was the same in both studies.

Accession #	logFC	PValue	FDR	Length	nr BLAST results	nr e-Value	*L. retroversa* DE homology	homology e-Value
DN165269_c0_g1_i9	-7.73	2.87E-06	0.0265	1044	Uncharacterized protein C6orf203	1.02E-06		
DN166076_c0_g1_i3	-7.49	8.74E-12	2.63E-06	993	---NA--- (mucin via *L. retroversa*)		c30099_g1_i1, c30099_g1_i2	2.90E-08, 1.50E-08
DN167508_c3_g1_i4	-4.17	9.28E-07	0.0175	3330	Probable nuclear hormone receptor HR3; Nuclear receptor subfamily 1 group F member 4	5.18E-21		
DN168496_c0_g1_i11	-5.82	2.38E-09	0.0001	6419	Calpain-7/Androglobin	3.55E-149		
DN168496_c0_g1_i26	-4.85	2.99E-06	0.0265	6395	Calpain-7/Androglobin	1.28E-152	c10333_g1_i1, c10929_g1_i1	4.1E-47, 1.3E-126
DN172691_c0_g2_i1	-6.30	4.52E-07	0.0124	1312	27 kDa inositol polyphosphate phosphatase-interacting A	9.64E-39		
DN175037_c0_g2_i4	-6.03	1.40E-06	0.0221	1568	G- coupled receptor GRL101	3.54E-162		
DN160157_c0_g1_i3	6.51	2.60E-07	0.0078	1388	RNA-directed DNA polymerase from mobile element jockey; Reverse transcriptase	4.01E-24		
DN162313_c1_g2_i1	4.89	1.56E-06	0.0224	1762	Probable G- coupled receptor Mth-like 3; methuselah-like 3	2.62E-14	c49632_g1_i1	1.60E-07
DN171745_c1_g2_i3	7.95	2.75E-06	0.0265	1766	Dynactin subunit 6	1.49E-56		

Supplementary File 1: Summary figures depicting supplementary transcriptome annotation statistics and distribution of functional categories.

Table S1: Top 20 InterPro domains in *Clione limacina* juvenile transcriptome.

	IPS Domain	#Seqs
IPR027417	P-loop containing nucleoside triphosphate hydrolase	1836
IPR007087	Zinc finger, C2H2	1704
IPR013087	Zinc finger C2H2-type/integrase DNA-binding domain	1542
IPR015880	Zinc finger, C2H2-like	1467
IPR013783	Immunoglobulin-like fold	1047
IPR000477	Reverse transcriptase domain	1026
IPR020846	Major facilitator superfamily domain	824
IPR016187	C-type lectin fold	811
IPR016186	C-type lectin-like	798
IPR007110	Immunoglobulin-like domain	701
IPR011009	Protein kinase-like domain	691
IPR015943	WD40/YVTN repeat-like-containing domain	657
IPR013083	Zinc finger, RING/FYVE/PHD-type	655
IPR001304	C-type lectin	648
IPR011992	EF-hand domain pair	631
IPR017986	WD40-repeat-containing domain	630
IPR020683	Ankyrin repeat-containing domain	621
IPR016040	NAD(P)-binding domain	582
IPR000719	Protein kinase domain	550
IPR002048	EF-hand domain	529

Table S2: Top 20 KEGG pathways in *Clione limacina* with the highest number of enzymes.

Pathway	#Seqs	#Enzs	Pathway ID
Biosynthesis of antibiotics	861	123	01130
Purine metabolism	2913	53	00230
Pyrimidine metabolism	269	33	00240
Cysteine and methionine metabolism	315	31	00270
Amino sugar and nucleotide sugar metabolism	169	27	00520
Glycine, serine and threonine metabolism	275	27	00260
Glycolysis / Gluconeogenesis	224	25	00010
Alanine, aspartate and glutamate metabolism	210	24	00250
Pyruvate metabolism	237	23	00620
Valine, leucine and isoleucine degradation	244	23	00280
Aminoacyl-tRNA biosynthesis	219	22	00970
Arginine and proline metabolism	186	22	00330
Glycerophospholipid metabolism	137	22	00564
Fructose and mannose metabolism	117	20	00051
Carbon fixation pathways in prokaryotes	183	19	00720
Tryptophan metabolism	243	19	00380
Inositol phosphate metabolism	126	18	00562
Phosphatidylinositol signaling system	158	18	04070
Drug metabolism - other enzymes	224	17	00983
Citrate cycle (TCA cycle)	142	17	00020

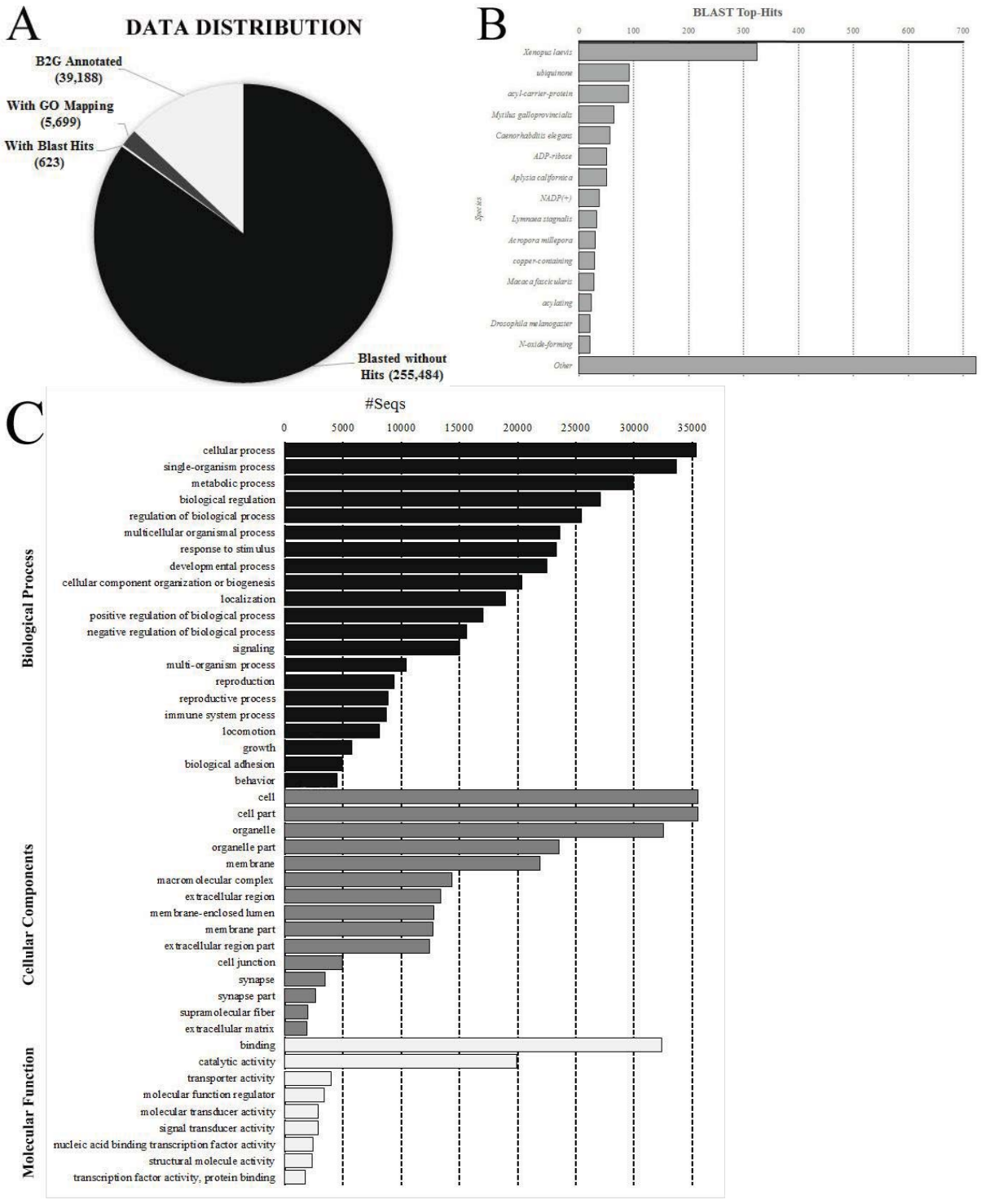

Figure S1: BLAST results from searches against the Swiss-Prot database; A) Data distribution of annotated transcripts Categories include fully annotated sequences (with a Blast2Go annotation score > 55), sequences with only GO mapping and BLAST annotation, sequences with BLAST annotation only, and sequences with no annotation; B) distribution of top BLAST hit species and C) Functional gene ontology (GO) terms (level 2) for the main three categories.

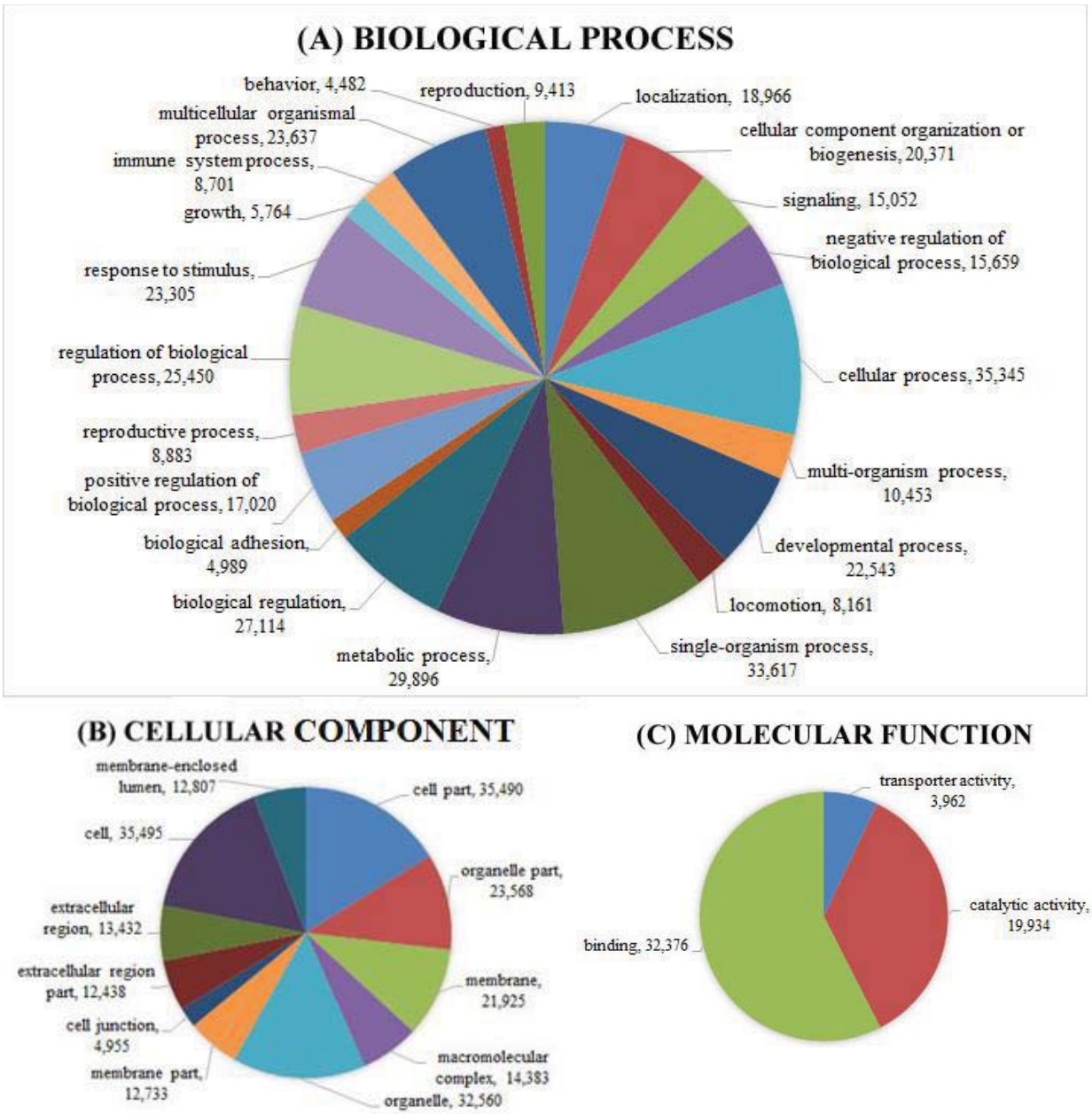

Figure S2 : Distribution of most common gene ontology (GO) terms for the three categories (level 2); Biological Process (A), Cellular Component (B) and Molecular Function (C) using the results from the BLAST search versus Swiss-Prot combined with IPS annotation in Blast2Go.