Title: Deep, Staged Transcriptomic Resources for the Novel Coleopteran Models *Atrachya menetriesi* and *Callosobruchus maculatus*

Short title: Staged developmental transcriptomes for two beetle species

Matthew A. Benton[1 ¶], Nathan J. Kenny[2 ¶], Kai H. Conrads[1], Siegfried Roth[1*] and Jeremy A. Lynch[3*]

[1] Institute for Developmental Biology, University of Cologne, Zülpicherstrasse 47b, Cologne 50674, Germany

[2] Simon F.S. Li Marine Science Laboratory of School of Life Sciences and Center for Soybean Research of the State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong

[3] Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois, United States of America

[¶] Authors contributed equally

* Co-corresponding authors

Matthew A. Benton: matthewabenton@gmail.com

Nathan J. Kenny: nathanjameskenny@gmail.com

Kai H. Conrads: kai.conrads@gmx.de

Siegfried Roth: siegfried.roth@uni-koeln.de

Jeremy A. Lynch: jlynch42@uic.edu

Abstract:

Despite recent efforts to sample broadly across metazoan and insect diversity, current sequence resources in the Coleoptera do not adequately describe the diversity of the clade. Here we present deep, staged transcriptomic data for two coleopteran species, *Atrachya menetriesi* (Faldermann 1835) and *Callosobruchus maculatus* (Fabricius 1775). Our sampling covered key stages in ovary and early embryonic development in each species. We utilized this data to build combined assemblies for each species which were then analysed in detail. The combined *A. menetriesi* assembly consists of 228,096 contigs with an N50 of 1,598 bp, while the combined *C. maculatus* assembly consists of 128,837 contigs with an N50 of 2,263 bp. For these assemblies, 34.6% and 32.4% of contigs were identified using Blast2GO, and 97% and 98.3% of the BUSCO set of metazoan orthologs were present, respectively. We also carried out manual annotation of developmental signalling pathways and found that nearly all expected genes were present in each transcriptome. Our analyses show that both transcriptomes are of high quality. Lastly, we performed read mapping utilising our timed, stage specific RNA samples to identify differentially expressed contigs. The resources presented here will provide a firm basis for a variety of experimentation, both in developmental biology and in comparative genomic studies.

Key Words: Beetle; Coleoptera; Transcriptome; Ovary; Embryo; Staged; Atrachya menetriesi; Callosobruchus maculatus; Chrysomelidae

Introduction:

The order Coleoptera is the most speciose clade of animals currently known. Despite the best efforts of generations of biologists, its species are only sparsely sampled and are yet to be comprehensively described, with approximately 90% of coleopteran diversity as yet

uncategorized (e.g. [1,2]). A similar discrepancy exists at the molecular level; while several genomic resources are available in this clade, their number and phylogenetic distribution is only just beginning to accurately sample that of the Coleoptera as a whole. A wide range of transcriptomic data is available in whole organisms (for example [3–5], among others), specific body parts (such as [6,7]), and in several cases for staged embryos following RNA interference-mediated gene knock-down [8–10]. The i5K project [11] will also greatly advance our knowledge of the genomic complement of Coleopterans, with 69 species of this Order listed on that database as nominated for genomic sequencing as of the 07/04/16 (url: http://arthropodgenomes.org/wiki/i5K_nominations) and several genomes publically available [12–14]. However, the majority of this information is largely still outside of the public domain, the Coleoptera are still relatively undersampled compared to the Diptera and Hymenoptera, and, in particular, timed embryonic resources (e.g. [15]) are rare in the literature.

The true phylogeny of the Coleoptera is still under investigation but, in general, four suborders, 17 superfamilies, and 168 families are recognised [16]. The structure of the coleopteran clade can be seen summarised in Fig 1A. Coleopterans have long been used for research into embryology, and in the pre-molecular era, the Chrysomelidae (summarised phylogeny shown in Fig 1B) was one of the best studied superfamilies. For example, the first functional embryonic experiments in any insect were carried out in the Colorado potato beetle, *Leptinotarsa decemlineata* (sub-family Chrysomelinae, see Fig 1B), leading to the discovery of the function of pole cells and the existence of germ plasm in insects [17,18]. Further, the larval cuticle preparation technique, so vital for arthropod developmental biology, was first perfected in the bean beetle, *Callosobruchus maculatus* [19]. Chrysomelid beetles are also interesting from an ecological and economic viewpoint, as members of the group are usually pest species, perhaps most famously the aforementioned Colorado potato beetle, which is a major pest of potato crops in America, Asia and Europe [20].

78 *Figure 1:* A) Cladogram of Coleoptera simplified from that determined by [16]. Black asterisk
79 indicates the superfamily in which the Chrysomelidae are located. B) Cladogram of
80 Chrysomelidae simplified from that determined by [21]. Black and grey asterisks indicate the
81 sub-families in which *A. menetriesi* and *C. maculatus* are located (respectively). C) *A.*
82 *menetriesi* and D) *C. maculatus* adults.

83

84 The Chrysomelidae are represented in public databases by a number of ongoing
85 transcriptomic and genomic projects. In particular, both *L. decemlineata* (Bioproject
86 PRJNA171749) and *C. maculatus* [22] are the subject of ongoing genomic sequencing.
87 However, while a number of transcriptomes are planned or published in this clade (e.g. [5,6,23–
88 27]) none yet sample across embryonic time points in a fashion which would allow insight into
89 the genetic mechanisms behind key developmental stages or deep sampling of developmentally
90 important genes. Here we present transcriptomic sequences for two species of chrysomelid
91 beetle, the false melon beetle, *Atrachya menetriesi,* and the aforementioned *C. maculatus* (each
92 of which are pictured along with their relative phylogenetic positions in Fig 1).
93 Compared to *C. maculatus,* the false melon beetle *A. menetriesi* (Faldermann), is a
94 relatively unknown and understudied species. It is native to Japan, where it is an agricultural
95 pest, and feeds on a variety of plants such as clover and lettuce. Although there has only been
96 a small amount of research carried out on this beetle, the work that has been done has
97 highlighted several interesting developmental traits, including the possibility of generating twin
98 embryos after egg bisection [28], or up to four, seemingly complete, embryos following
99 treatment with low temperatures [28,29]. Another interesting trait exhibited by this beetle is that
100 almost all eggs enter diapause at a certain stage, and this is only broken in the wild by winter
101 conditions. However, a small proportion of eggs skip this diapause and continue to adulthood.
102 The ratio of diapause to non-diapause eggs varies in different parts of Japan [30], and is a
103 heritable trait [31]. Further, *A. menetriesi* embryos undergo a very short germ band mode of

development [32], contrasting strongly with beetles such as *C. maculatus* (see below). As the last common ancestor of these two species is estimated to have existed only 80 million years ago [21], how their developmental mechanisms have diverged so greatly is a potentially fascinating area for future study. Research on these topics would greatly benefit from modern molecular studies.

C. maculatus (Fabricius) is native to West Africa [33] but is now found worldwide, and is a common pest of stored legumes. It is also known as the southern cowpea weevil, however, it is not a true weevil. As noted above, this beetle was the focus of active developmental research in the pre-molecular era, with special focus on segmentation [34,35]. Segmentation in *C. maculatus* has also been studied more recently via immunohistochemistry for the even-skipped protein [36]. This recent work confirmed previous reports that the embryos undergo the long germ mode of development, similar to dipterans like *Drosophila melanogaster* and hymenopterans like *Nasonia vitripennis* [37]. It is commonly believed that the long germ mode of development evolved independently in dipterans and hymenopterans, and given the phylogenetic distribution of short and intermediate germ development within the Coleoptera [38,39], it seems likely that the long germ mode of development also evolved independently in the clade to which *C. maculatus* belongs. A comparison of the molecular basis of development in *C. maculatus* and other long germ insects, plus with more closely related species that feature short germ development, such as well studied flour beetle *Tribolium castaneum* (super-family Tenebrionoidea, see Fig 1A) and *A. menetriesi*, would yield crucial information on how developmental pathways have evolved to generate the long germ mode of development. *C. maculatus* has also been studied in other fields and is a useful system for undergraduate lab teaching [22] which could be aided further with deeper sequence resources.

In order to facilitate research on *A. menetriesi* and *C. maculatus*, as well as wider investigations in the Coleoptera and beyond, we present here deep, multi-stage transcriptomic resources from a range of key time points in the development of these two beetle species.

Using RSEM-based methods we have compared transcript abundance across these life stages, which will allow the investigation of genes that play key roles in developmental changes in these species, particularly at the maternal-zygotic transition. We have carried out extensive searches for key genes in developmental patterning and cell signalling pathways, and from our analyses we conclude that the transcriptomes for both species are of very high coverage, with almost all expected genes being present with long (likely full) average open reading frames. We have already made extensive use of these resources for our own studies on the embryonic development of these two beetles, and are confident that they will be of broad utility to a range of fields in genomics and developmental biology.

Materials and Methods:

Animal Husbandry

A. menetriesi eggs were collected from the wild and kindly provided by Dr Yoshikazu Ando, and were reared at 25°C on wet sand or soil and fed fresh lettuce. *C. maculatus* beetles were kindly provided by Dr Joel Savard, and were reared at 30°C on dry black-eyed peas.

RNA Extraction and Sequencing

RNA was extracted from dissected ovaries and timed embryonic samples using a TRIzol RNA extraction kit according to the manufacturer's protocols. RNA quantity and quality was checked using a Thermo Scientific Nanodrop 2000C Spectrophotometer and 1 µg was sent for sequencing by the Cologne Centre for Genomics on a Illumina HiSeq 2000 sequencer after sample preparation using a TruSeq RNA Library Preparation Kit (Illumina). Adaptor trimming and initial quality control was performed by the provider according to their proprietary standards, with no orphan reads kept. This cleaned data was then made available to us for download from

an external server. Paired end read quality after sequencing was assessed using the FastQC program [40] and no residual adaptor was observed, as detailed in the Results.

Transcriptome Assembly and Comparative Expression Analyses

Assemblies used in our final analysis were made using Trinity version 2013_08_14 [43], with the default settings (--min_contig_length 200). Trimmomatic [41] was assayed (Illuminaclip Leading:3 Trailing:3 Slidingwindow:4:15 Minlen:36) but not utilised for assemblies presented here, as described in results. Full assemblies were made using reads from all time points, and individual assemblies were then constructed using reads from each sampled time point individually. All assemblies are available from Figshare online (*A. menetriesi* DOI: 10.6084/m9.figshare.2056464.v2, *C. maculatus* DOI: 10.6084/m9.figshare.2056467.v2). DeconSeq standalone version 0.4.3 [44] was run on full assemblies with settings -i 95 -c 95, using the bact, fungi, hsref, and prot databases. Comparative expression analyses were performed using RSEM [61] as packaged in the Trinity module (cross sample normalisation: Trimmed Mean of M-values), to compare staged RNA samples with the combined assembly. Results shown here are the ‘as-isoform’ data, although ‘as-gene’ data is also provided in Additional Files. BUSCO v1.1b1 [48] was used to assess gene complement completeness.

Functional Annotation

Our combined assemblies were automatically assigned homologs and annotated according to gene ontology (GO) terms using Blast2GO [49,50]. Initially, BLASTx was run using BLAST 2.2.29+ against the NCBI nr database as downloaded to a local server on the 17/01/2015, with settings -evalue 0.001 -max_target_seqs 5 -outfmt 5. GO term distribution within the *D. melanogaster* and *H. sapiens* genomes were downloaded from B2GO-FAR [62] and calculated using the Combined Graph function of Blast2GO. KEGG KAAS mapping was automatically

179 performed using the KEGG KAAS tool (http://www.genome.jp/tools/kaas/), single-directional
180 best hit with default BLAST settings, and with the eukaryote dataset as a basis for annotation.
181
182 *Gene Identification*
183 Gene sequences were manually identified and their homology confirmed by independently using
184 tBLASTn [63] searches using gene sequences of known homology downloaded from the NCBI
185 nr database as queries against standalone databases created on a local server using BLAST
186 2.2.29+ or the CLC Main workbench. Genes putatively identified using this method were
187 reciprocally BLASTed against the online NCBI nr database using BLASTx to confirm their
188 identity. Where identity was uncertain, phylogenetic analysis was used to confirm identity.
189
190 *Phylogenetic Tree Construction*
191 Sequences were aligned using MAFFT 7 [64] unless otherwise stated under the L-INS-i
192 strategy. Alignments were then saved and exported to MEGA 6, where regions of poor
193 alignment were manually excluded and maximum likelihood phylogenetic trees were
194 constructed using the LG model, 1000 bootstrap replicates as indicated, 4 gamma categories
195 and invariant sites, and all other default prior settings [65].
196

197 ## Results and Discussion

198 *RNA Extraction and Sample Selection*
199 Ovaries and embryos were collected as described in Materials and Methods, and as seen in Fig
200 2. The chosen time-windows cover a variety of important stages in the development of these
201 species, and when combined can reasonably be expected to contain the majority of embryonic
202 transcripts in their expressed complement. Briefly, RNA was collected from dissected ovaries
203 from each species and from four embryonic time windows for *A. menetriesi* and two embryonic

time windows for *C. maculatus*. Samples were sequenced on the Illumina HiSeq 2000 platform (one lane per species), adaptor trimming was performed, along with preliminary assessment of read quality, and data was made available for download from an external server.

Figure 2: Summary of RNA sources, life stages sampled and sequencing results, with *A.menetriesi* data presented at left and *C. maculatus* data at right.

Read Quality and Assembly Metrics:

Fig 2 summarises initial read information, and these reads are available from the NCBI SRA, Bioproject Accession numbers: PRJNA293391 and PRJNA293393. To confirm quality of read data, FastQC [40] was run on all read data. This showed Phred quality scores were high, with median scores always exceeding 28 through to the 101st base, and generally in the mid-30s. A slight bias was found in initial nucleotide sequence. To ensure that Illumina adaptors were removed in their entirety, Trimmomatic v.0.32 [41] was used to check for residual adaptor sequence, but none were observed. We therefore posit that this bias is due to known biases in Illumina hexamer binding [42] rather than sequencing-based artifacts.

Using Trinity [43] as an assembler, Trimmomatic-corrected read assemblies were then compared with assemblies based on uncorrected reads. Initial assays of Trimmomatic-treated read assemblies empirically found them to be less well assembled than those using un-trimmed reads, with a shorter N50 and less total sequence recovery (total number of bp, of which some contigs could result from read errors). Given the advantages of a longer N50 and a preference to preserve as much information as possible, trimmed assemblies were therefore discarded in favour of untrimmed assemblies given the reasonable coverage requirement, and these 'untrimmed' assemblies were used for all further analyses.

228 Assays of initial assemblies noted a small amount of fungal contamination in the data for
229 both species. This contamination likely comes from the environments in which the beetles are
230 cultured. To correct for this, DeconSeq [44] was run on the Trinity output for both species, with
231 high stringency as noted in the methods. A total of 336 and 206 contigs with some homology to
232 fungal sequence was removed from the *A. menetriesi* and *C. maculatus* total assemblies as a
233 result, before read mapping was performed. In the *A. menetriesi* assembly, we observed
234 particularly high contamination with the protists, notably *Dictyostelium discoideum* and
235 *Naegleria gruberi,* and some of this almost certainly remains in the transcriptome, as is normal
236 with most 'omics' experiments. Metrics for final assemblies after the removal of contamination
237 can be seen in Table 1, alongside the results of assemblies for individual time points.

238

239 *Table 1:* Assembly metrics for individual time point and combined transcriptomic resources

	A. menetriesi						*C. maculatus*			
	Ovary	0-24 hpf	24-48 hpf	48-72 hpf	72-100 hpf	Combined	Ovary	0-7 hpf	7-24 hpf	Combined
Number of contigs	100,084	106,548	120,497	140,868	126,400	228,096	98,082	57,879	80,665	128,837
Max contig length (bp)	20,759	17,765	18,398	17,591	17,180	20,757	30,742	22,941	22,956	29,933
Mean contig length (bp)	983.95	1027.36	924.27	879.03	914.2	844.66	1074.86	1216.55	1106.56	991.04
Median contig length	471	490	453	426	448	401	416	526	474	391

(bp)										
N50 contig length (bp)	1,894	1,993	1,721	1,643	1,686	1,598	2,450	2,570	2,308	2,263
# contigs in N50	14,196	15,200	17,511	19,814	18,105	31,196	11,695	7,874	10,978	15,155
# contigs > 1kb	28,255	31,944	32,307	34,520	33,002	52,217	27,763	19,919	25,529	33,250
# bases, total	98,477,173	109,463,362	111,371,655	123,827,226	115,555,146	192,664,213	105,424,220	70,412,726	89,260,499	127,682,507
# bases in contigs > 1kb	68,737,074	78,526,348	75,208,106	80,796,007	77,199,414	122,986,547	78,519,052	55,032,546	67,296,566	91,466,099
GC Content % (2dp)	33.54	33.14	33.14	33	33.17	32.82	39.26	39.21	38.59	38.7

The final, combined timepoint, contamination-removed assemblies, comprising 228,096 and 128,837 contigs for *A. menetriesi* and *C. maculatus* respectively, contain a large number of well-assembled transcripts, with the number of contigs greater than 1 kb in length (52,217 in *A. menetriesi,* 33,250 in *C. maculatus*) and a high N50 (1,598 bp *A. menetriesi,* 2,263 bp *C. maculatus*) indicating a very well assembled dataset. This size is sufficient to span most protein coding domains, allowing easy inference of homology, and will be the full length of many transcripts. It is important to note that many of the contigs in our assembly will represent splice variants of single genes, and some genes will have multiple splice variants, which will affect these statistics. However, excellent recovery of splicing variation itself will be useful for a range of later analysis. GC content of the final assembled transcriptomes closely mirrors that of reads

252 (32.82% in *A. menetriesi,* 38.7% in *C. maculatus*; reads 36-38%/40-43%, respectively,
253 depending on library), and is therefore similar to that expected.

254 To gain an understanding of whether full-length transcripts were present in our data, we
255 ran TransDecoder v 2.01 [45] to identify open reading frames (ORFs) and filtered for the results
256 that were at least 100 amino acids long. For the *A. menetriesi* assembly, this analysis yielded
257 71,961 raw and 51,912 filtered contigs, while for the *C. maculatus* assembly, the analysis
258 yielded 65,433 raw and 36,535 filtered contigs. The mean average length of the predicted
259 polypeptides, after the filtering step, was 351 (*A. menetriesi*) and 448 (*C. maculatus*) amino
260 acids. This is long enough for us to be confident that our assembly adequately spans coding
261 regions, as the average eukaryotic protein length is 361 amino acids [46]. Together, these
262 analyses suggest that we have recovered the vast majority of coding sequence in our combined
263 assemblies, with sufficient length to adequately span ORFs, a conclusion further supported by
264 gene annotation data as discussed further below. The numbers of ORFs presented here are
265 considerably more than the 16,404 gene models observed in the *T. castaneum* genome
266 (*Tribolium* Genome Sequencing Consortium, 2008), and this is likely due to both spurious ORFs
267 in our dataset and to multiple splice variants.

268 Further information was gained using the Ortholog Hit Ratio method detailed in [47],
269 which describes the proportion of best-hit ortholog sequence represented by a dataset. When
270 compared to the *T. castaneum* Tcas3.31.pep.all.fa peptide set with a BlastX cut off of E^{-5}, the
271 average ratio of the full length ortholog present in all of our blast hits was 0.4369 (*A. menetriesi*)
272 and 0.5079 (*C. maculatus*). These statistics compare very well with those found in other
273 organisms and previous studies, such as that in [47], and further indicate that the *C. maculatus*
274 transcriptome may be slightly better assembled than the *A. menetriesi* dataset. This difference
275 is likely due to a major difference in the level of genetic heterogeneity between our samples
276 from the two species, as *C. maculatus* have been cultured in our lab for many years, while *A.*
277 *menetriesi* was recently sourced from the wild.

Timed RNA Expression - Differential Expression Analysis

Key developmental stages for both *A. menetriesi* and *C. maculatus* can be easily observed following fixation and nuclei staining (data not shown). Briefly, egg lay to uniform blastoderm stage takes approximately 24 hours in *A. menetriesi*, and 7 hours in *C. maculatus*. Germband formation, gastrulation and elongation occur from 24-100 hours in *A. menetriesi*, and from 7-24 hours in *C. maculatus*. The latter period was subdivided in *A. menetriesi* according to characteristic stages of short germ development pertinent to our research interests. These stages, along with assayed sample periods, are shown diagrammatically in Fig 2. As well as being used for combined assemblies as described earlier, RNA extracted from mature ovaries and from embryos collected during the aforementioned time periods was also individually assembled using Trinity, allowing this information to be used to find time/stage-specific transcripts within our dataset.

The timed transcriptome assemblies for each species often possess better assembly quality when compared to the combined assemblies by metrics such as N50 and mean contig length (Fig 2, Table 1). As a result of being made up of a subset of the total reads, the individual assemblies do not possess the breadth of the combined assemblies, with fewer contigs, especially at long contig length (e.g. greater than 1kb). As such, the staged transcriptomes were used solely for comparison of expression levels across time, while the combined assemblies were used for gene family analyses. We emphasise that no technical replicate was performed for these comparisons, and any conclusions drawn from them should hold this consideration in regard. With this limitation in mind, we carried out differential expression analyses and observed broad trends in expression, as can be seen in Fig 3.

Figure 3: Overview of results of differential expression analysis performed by RSEM within the Trinity framework. A) and B) show the sample correlation matrix for *A. menetriesi* and *C. maculatus* respectively. C) and D) show relative expression of each differentially expressed contig, considered as isoforms, across time.

First, we generated matrices from general comparison between time points in order to find the most similar samples (Fig 3 A,B). Generally, these results are congruent with steady changes in gene expression across the course of development, with most time points being most similar to those immediately preceding and following them. However, a split can be seen in *A. menetriesi* (Fig 3 A,B) between the ovary and 0-24 hour samples and all others, with the three later samples resembling each other more closely than the 0-24 hour dataset resembles the 24-48 hour sample. This could be due to the maternal-zygotic transition, which will occur at some point in this time frame. This cannot be seen for *C. maculatus*, and could be due to more admixture of RNA within the last 7-24 hour sample. Focused analysis on when the maternal-zygotic transition occurs in each species is required to resolve this question.

Next, we clustered the results of our differential expression calculations (Fig 3 C,D). The results shown are those with RSEM considering each isoform separately (rather than taking into account clustering into genes performed by Trinity). Numerous up- and down-regulated contigs can be seen at each time point, with some time points more obviously possessing or lacking a subset of genes found in the combined transcriptomes.

While replicates have not been performed and we have not analysed up and down regulated transcripts in detail, these data are available to download from Additional Files S6 and S7 attached to this document online. These data will act as a good initial guide for those interested in tracking differential expression of specific genes across development and will be useful for hypothesis building.

Basic Gene Annotation

To gain an understanding of the depth of coverage of our datasets we used the BUSCO library of well-annotated genes [48], which are known to be highly conserved in single copy across the Metazoa, as a basis for comparison with our combined, assembled transcriptomes. Of the BUSCO set of 843 metazoan orthologs, the *A. menetriesi* assembly possesses 801 (95%) complete (of which 225, 26%, appear to be duplicated), 16 fragmented (1.8%) and 26 missing genes (3.0%), for a total recovery of 97% of the BUSCO dataset. The *C. maculatus* assembly contains 815 (96%) complete (with 202, 23%, appear to be duplicated), 13 (1.5%) fragmented and 15 (1.7%) missing genes (98.3% recovery). This extremely high level of recovery gives us confidence that at least all housekeeping genes expected to be present in these species are found in our datasets, which strongly suggests that these transcriptomes contain the vast majority of the gene cassette of these species.

The number of potentially duplicated genes in our BUSCO analyses likely reflects the construction of our transcriptomes from mixed RNA samples, with the allelic variation that this implies. Discerning true duplicates from allelic and splice variant data is largely contingent on the availability of well-assembled genomes. However, the recovery of these putative duplicates in our assemblies underlines that our RNA sequencing and assembly was of good depth and quality (respectively). With this information in-hand, a range of investigations will be made possible, particularly into the regulation and expression of developmentally important genes.

A further understanding of the content of our assemblies was gained from Blast2GO analysis [49,50]. Genes were annotated using b2gpipe, on the basis of BLASTx (*E* value cutoff, 10^{-3}) comparisons made against the nr database as downloaded on the 26 January 2015. Of 228,096 (*A. menetriesi)* and 128,837 (*C. maculatus)* contigs in each assembly, 78,879 (34.6%) and 41,744 (32.4%) possessed a hit in the nr databases above the threshold. After further annotation with ANNEX and Interproscan, a total of 36,315 (15.9%) and 13,096 (10.2%) contigs

were assigned to one or more GO categories. These numbers, while only a fraction of the total number of contigs within our transcriptomes, more closely reflect the expected eukaryotic protein complement in number. Blast2GO annotations are given in Additional Files S2 and S3

Fig 4A shows the distribution of species best hit by BLASTx comparison of contigs from *A. mentriesi and C. maculatus* with the nr database. For both species, *T. castaneum* is the best represented species - a reflection of both the phylogenetic position of this species and its well annotated genome. The fact that contigs in our transcriptomes match *T. castaneum* and other coleopteran and insect species more closely than those of other species suggests that gene orthology will be easy to assign in many cases, and that an abnormally high rate of molecular evolution is not observed in our species.

Figure 4: Blast2GO results A) Distribution of BLASTx best hits by species, showing metazoans only, for *A. menetriesi* in orange, *C. maculatus* in blue. B) Distribution of GO terms expressed as a percentage of annotated contigs which were assigned a term within each of the three (Molecular Function, Cellular Component, Biological Process) categories of GO ID.

The distribution of GO terms within our datasets are shown in Fig 4B, alongside those of the well-annotated *D. melanogaster* and *Mus musculus* proteomes. In general, our datasets resemble one another more than they mirror that of the two sequenced genomes noted. Our transcriptomic resources empirically seem under-represented relative to *D. melanogaster* and *M. musculus* in developmentally interesting categories such as ‘Protein Binding’ (Molecular Function), ‘Multicellular Organism Development’ and ‘Cell Differentiation’ (Biological Process). Given our results for more targeted investigations, found below, we feel this is likely a result of poor annotation of these by Blast2GO, rather than true absence from the transcriptome.

At a gross level, in the intracellular 'Cellular Component' GO categories, our data appears more similar to that seen in the two 'model' species than in Molecular Function or Biological Process categories, although this has not been tested statistically. This would suggest these well-conserved structural components were more readily assigned GO categories than the developmentally interesting categories listed above. While not all GO categories are as well-annotated by this process as may be desired, the broad classification of our data into a wide range of GO categories of all levels of GO distribution demonstrates that Blast2GO annotations of our data are a useful starting point for more focussed investigations and identification of specific genes and pathways of interest.

Gene Family Recovery

To extend the semi-automated analyses presented above, we performed more targeted analysis of individual, developmentally important gene families. Both the Hox family of transcription factors and the TGF-β cassette were exceptionally well recovered in our dataset. The Hox genes, and in particular the ANTP HOXL class, which pattern the anterior-posterior body axis, are recovered almost in their entirety in both species when compared to well-catalogued databases (e.g. [51]). This can be seen in Fig 5, which shows the phylogenetic distribution of recovered ANTP HOXL class sequences from our transcriptomes alongside previously annotated members of this class.

Figure 5: Phylogenetic inter-relationships of ANTP HOXL class genes, as reconstructed by MEGA 6 using the LG+Freqs model with 4 gamma categories and invariant sites, based on a 59 amino acid alignment spanning the homeodomain. Numbers at base of nodes represent bootstrap percentages of 1000 replicates. Scale bar at base of phylogeny gives substitutions

per site at given unit distance. Red underline indicates *A. menetriesi* and *C. maculatus* sequences, coloured boxes are used to delineate known gene families (and in the case of Hox 6/7/8, a superfamily).

In *A. menetriesi,* sequences for all the ANTP HOXL class genes are recovered in our transcriptome, as can be seen in Fig 5, with sequences and alignments in Additional File S1. We note, however, that the *A. menetriesi Hox 2 / maxillopedia (mxp)/ proboscipedia (pb)* sequence bears some BLAST similarity with *Hox 4 / Dfd* sequences, while the putative *Hox 4 / Deformed (Dfd)* recovered does not include the homeodomain sequence - whether it has lost this crucial domain, or if this portion of the sequence is simply not recovered in our assembly is at present unknown. The putative *A. menetriesi Hox 4 / Dfd* sequence is given in Additional File S1, although it is not shown in the phylogeny in Fig 5 due to its truncated length.

While 2 (*A. menetriesi)* and 4 (*C. maculatus*) *Hox 3 / zerknüllt (zen)* variants are seen in our species, these are identical at the coding level, and therefore seem to be splice or allelic variants, rather than the two paralogous genes seen in *T. castaneum* [52]. Similarly, no evidence of the *caudal* (*cad*) duplication seen in *T. castaneum* [53] can be found in our transcriptomic assemblies, suggesting this is perhaps specific to the flour beetle lineage. Both *zen* and *caudal* are important embryonic patterning genes, and comparison of these genes in our two species and *Tribolium* would be an excellent situation in which to study how sub- and neo-functionalisation occurs. Likely allelic or splice variants are also observed for other HOXL genes in both *A. menetriesi* and *C. maculatus*. It should be noted that these could represent very recent duplications, or the effect of gene conversion, although this can only be tested fully with the advent of a complete genomic resource. No *Pdx/Xlox* gene was seen, adding further circumstantial evidence for the broad scale loss of this gene across the Insecta [54] and possibly the wider Arthropoda [55].

Our *C. maculatus* transcriptome also contains the full complement of ANTP HOXL genes, although, similar to the case of *Hox 4 / Dfd* in *A. menetriesi*, the 4 recovered *abdominal-A (abd-A)* homologs lack the whole homeodomain sequence, with several residues missing. These truncations excluded them from the phylogeny shown in Fig 5, but the sequences for these putative homologs are given in Additional File S1.

Even more so than in *A. menetriesi,* a remarkable diversity of potential splice/allelic variants are noted in *C. maculatus,* particularly for the *Hox 6/8* superfamily and *Hox 9-13* / *Abdominal-B* (*Abd-B*) gene family. Of the *Hox 6/8* gene superfamily, normally represented by four genes in *T. castaneum* (*prothoraxless* (*ptl)*, *fushi tarazu* (*ftz)*, *Ultrathorax* (*Utx)* and *abd-A*), 12 *ptl*, 1 *ftz*, 4 *Utx* and 4 *abd-A* representatives were found in our analysis. Furthermore, up to 26 different potential allelic or splice variants of abdB are recorded. As our transcriptome is made of mixed embryonic samples, it is perhaps not surprising that a diversity of putative splice/allelic variants are observed, but the excellent recovery of this data confirms the deep coverage provided by our sequencing and assembly given the coverage possessed by all isoforms.

The TGF-β cassettes of the insects have been very well described previously (e.g. [56,57]). Our datasets recover almost the full expected complement of the Coleoptera. A slight exception to this is *Activin (Act)*, a partial sequence of which is recovered for both species: a portion of the propeptide which does not span the mature signal peptide sequence. Whether this is a consequence of loss of the mature domain in these species or low levels of expression at the sampled timepoints remains to be established. A *BMPx* ortholog of clear homology to genes of that family can be found in *C. maculatus,* but has been excluded from the tree seen in Fig 6 as it is incomplete in length. Its sequence can be found in Additional File S1, and we have no doubt as to its identity due to high levels of sequence conservation between it and the *T. castaneum* and *A. menetriesi* orthologs of this gene.

Figure 6: Maximum Likelihood Phylogeny of TGF-β ligands, as determined using MEGA under the LG+Freqs model with 4 gamma categories and invariant sites, on the basis of a 72 amino acid alignment of mature peptide sequences. The given scale depicts the number of substitutions per site per unit length. Bootstrap percentages (of 1000 replicates) are given at base of nodes. *A. menetriesi* and *C. maculatus* sequences are underlined in red. Coloured boxes represent known gene families with representatives in our transcriptomic resources, while all gene families, including those not found in our datasets, are indicated at right.

The *glass bottom boat (gbb)* duplication observed in *T. castaneum* cannot be found in our data, but we can recover a range of splice or allelic variants for other genes, especially in *A. menetriesi*. These do not differ in the protein coding regions, which leads us to suspect that these are not from gene duplications (unless the duplication(s) occurred very recently). The phylogeny shown in Fig 6 confirms the homology of all genes, and splice/allelic variant numbers observed are given there in brackets, with all sequences available in Additional File S1.

We also note the discovery of an additional putative TGF-β ligand in *C. maculatus*. This gene has been previously automatically annotated as *derriere* in *T. castaneum*, (XP_008191586.1) and if it is truly of this family, which is also known as *GDF1/3/Univin/Vg1*, this would be a surprise, as its presence outside the Deuterostomia is controversial [58]. If proof could be found for this being a *bona fide GDF1/3/Univin/Vg1*, the presence of this gene in more than one coleopteran could suggest that this might in fact be ancestrally present in all bilaterian species, but further investigation is warranted before strong conclusions can be drawn in this regard. We could gain no phylogenetic support for placing either of these beetle sequences in the GDF1/3 clade, and it may well be that these sequences instead represent a coleopteran novelty.

Our recovery of not only the full expected complements of these vital developmental genes, but also a remarkable diversity of alternative variants, demonstrates the depth of our assemblies as a resource, given the high coverage for each of these variants. Whether used as the basis for simple cloning or more sophisticated analysis of patterns of gene variation and diversification, these transcriptomes will be of wide utility to the field of coleopteran and insect developmental biology.

Pathway Recovery

As well as examining specific gene families, we investigated a number of broader pathways commonly studied in insects [59]. This allows us to both note how well-recovered such pathways are in our species as a measure of transcriptome utility, as well as note interesting differences between these pathways in our species when compared to others. We did this using both automated (KEGG KAAS mapping) and manual (BLAST based) methods. Some representative results of KEGG KAAS mapping can be seen in Fig 7, and all KEGG annotations can be downloaded from Additional Files S4 and S5.

Figure 7: KEGG style pathway maps showing recovery in our transcriptome resources of A) the Wnt signalling pathway in canonical and non-canonical contexts, B) the Hedgehog signalling pathway and C) the Notch signalling pathway. Coloration of genes indicates presence, absence or ancestral absence from the Coleoptera as detailed in the key, which also gives other information as noted. All genes automatically annotated by KEGG KAAS server, with the exception of *PAR-1,* which was manually annotated.

KEGG KAAS mapping uses BLAST results to annotate known pathways, and gives a rapid overview of the recovery of these. Here we have shown the well-known Wnt, Notch and Hedgehog pathways to indicate the depth of our transcriptomes, and show how they may be useful for future research. However, these maps often use terminology based on vertebrate nomenclature, and contain genes known to be absent from particular clades. We have therefore indicated in Fig 7 (using unshaded boxes as shown in the Key) genes that may be absent ancestrally in the Coleoptera, based on their absence from the *T. castaneum* pathway. Of genes expected to be present in the Coleoptera we find almost total recovery in our transcriptomes. In the three pathways examined, only three genes noted to be present in *T. castaneum* were noted as absent from both of our transcriptome datasets, all in the Wnt cascades (Fig 7A). The expected Hedgehog cassette was recovered *in toto* (Fig 7B) and in the Notch signalling cascade (Fig 7C), only APH-1 was noted to be absent, and only from the *C. maculatus* transcriptome. We must note that these may not be true absences - KEGG KAAS mapping is based on automatic BLAST assignation, and if these sequences are divergent in our transcriptomes they may have been missed by this analysis.

We also examined pathways manually, using reciprocal BLAST hits and closer manual investigation to confirm the identities of individual genes, the results of which can be seen in Table 2. The anterior-posterior patterning genes *cad* (mentioned earlier) and *hunchback (hb)* are present in both species. Of the germline establishment and localization genes examined, *nanos* was surprisingly absent in both species, while *bruno (bru;* also known as *arrest)*, *exuperantia (exu)*, *tudor* (*tud;* 2 copies in *A. menetriesi)*, *oskar (osk)*, *vasa (vas)* and *valois (vls)* were present. Interestingly, *pumilio (pum)* is present in *C. maculatus* in single copy (although it is divided across two contigs), while *A. menetriesi* possesses a total of seven copies. The different *A. menetriesi pum* copies varied both at the nucleotide level and in their amino acid sequences, strongly suggesting that they are in fact paralogs. In depth analysis of these genes is required to uncover why they have undergone several rounds of duplication. Orthologs of the

529 *Drosophila* gene *swallow (swa)* could not be found in either of our transcriptome resources, nor
530 is it present in several other insects (data not shown) and we suggest it may therefore be a
531 schizophoran novelty.

532

533 *Table 2*: Genes identified by manual annotation

Pathway/Gene	*A. menetriesi*	*C.maculatus*	*Pathway/Gene*	*A. menetriesi*	*C.maculatus*
Maternal Effect:			Pair rule:		
caudal	present	present	*even skipped*	present	present
hunchback	present	present	*fushi tarazu*	present (see Hox)	present (see Hox)
			hairy	present	present
Germline:			*odd paired*	present	present
bruno/arrest	present	present	*odd skipped*	present	present
exuperantia	present	present	*paired*	present	present
nanos	absent	absent	*runt*	present	present
oskar	present	present	*sloppy paired 1*	present	present
pumilio	present - 7 copies	present			
swallow	absent	absent	Segment Polarity:		
tudor	present - 2 copies	present	*armadillo*	present	present
valois	present	present	*cubitus interruptus*	present	present
vasa	present	present	*engrailed*	present	present
			fused	present	present
Gap genes:			*gooseberry*	present	present
buttonhead	present	present	*gooseberry-neuro*	absent	absent

empty spiracles	present	present	*hedgehog*	present	present (but on 3 contigs)
giant	present	present	*pangolin*	present	present
huckebein	present	present	*patched*	present	present
knirps	present	present	*wingless*	present	present
Krüppel	present	present			
orthodenticle 1	present	present			
orthodenticle 2	present	present			
tailless	present	present - 2 copies			

Canonical gap genes *Krüppel (Kr)*, *knirps (kni)*, *giant (gt)*, *huckebein (hkb)*, *tailless* (*tll;* 2 paralogs in *C. maculatus*), *buttonhead (btd)*, *empty spiracles (ems)* and both *orthodenticle* orthologs (*Otd* and *Otd2*) were recovered in both species examined here. The *C. maculatus* paralogs of *tll* exhibited differences at both the nucleotide and amino acid level along their entire lengths (data not shown) confirming that they are paralogs. Given the important embryonic role of *tailless* in other insects (for example [60]), this duplication would be excellent opportunity to study gene duplication and evolution. The pair rule genes e*ven skipped (eve), hairy (h), fushi tarazu (ftz), odd paired (opa), odd skipped (odd), paired (prd), runt (run)* and *sloppy paired 1 (slp1)* were present in single copy. The segment polarity genes were also present in both species, with the notable absence of *gooseberry-neuro (gsb-n)* from our datasets. The genes *armadillo (arm), cubitus interruptus (ci), engrailed (en), fused (fu), gooseberry (gsb), hedgehog (hh), pangolin (pan), patched (ptc)* and *wingless (wg)* were all present, and their sequences can be found in Additional File S1.

549 All of these pathways are commonly studied in insects, and the annotations provided
550 here, along with preliminary timed expression data, will provide a basis for a wide range of
551 targeted investigations into the embryonic development of these two species, and how these
552 pathways have changed over the course of evolution. Furthermore, the excellent recovery of
553 these pathways by both automated (KEGG-KAAS) and manual annotation gives us high
554 confidence in the completeness of our transcriptomic resources. This confirms the results of our
555 BUSCO analysis, and our datasets are therefore likely to contain the vast majority of transcribed
556 genes in these two species, with only lowly expressed and temporally restricted genes absent
557 from these transcriptome resources.
558

559 ## Conclusions

560 Our production of deep transcriptomic sequence data for *A. menetriesi* and *C. maculatus* will
561 assist in the inference of character gain and loss across the Coleoptera, aid in future
562 phylogenetic efforts, and allow a range of investigations into the embryonic development of
563 these species at the molecular level. The status of these organisms as common agricultural
564 pests also suggests that such resources may allow targeted control mechanisms to be
565 developed for these species. This data will be another key building block in our understanding of
566 the transcriptomic basis to embryological development, and provide a window into the basic
567 biology of the most successful clade of animals.
568

569 ## Acknowledgements:

570 The authors would like to thank the members of their laboratories for their support and
571 discussions. In addition, we thank Dr Y Ando for providing *A. menetriesi* eggs and advice on
572 establishing cultures, Dr J Savard for providing *C. maculatus*, and Dr F Marletaz for help in

running BUSCO analyses. The efforts of editors and reviewers in considering this manuscript are gratefully acknowledged.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the NCBI SRA repository [Bioproject Accession numbers: PRJNA293391 and PRJNA293393] and in the Figshare repository [DOIs: 10.6084/m9.figshare.2056464.v2, 10.6084/m9.figshare.2056467.v2].

References:

1. Stork NE. Insect diversity - facts, fiction and speculation. Biol J Linn Soc. 24-28 OVAL RD, LONDON, ENGLAND NW1 7DX: ACADEMIC PRESS LTD; 1988;35: 321–337. doi:10.1111/j.1095-8312.1988.tb00474.x
2. Grimaldi D, Engel M. Evolution of the Insects. Cambridge University Press; 2005.
3. 1KITE: 1000 Insect Transcriptome Evolution [Internet]. 2015. Available: http://www.1kite.org/
4. Keeling CI, Henderson H, Li M, Yuen M, Clark EL, Fraser JD, et al. Transcriptome and full-length cDNA resources for the mountain pine beetle, Dendroctonus ponderosae Hopkins, a major insect pest of pine forests. Insect Biochem Mol Biol. England; 2012;42: 525–536. doi:10.1016/j.ibmb.2012.03.010
5. Kumar A, Congiu L, Lindstrom L, Piiroinen S, Vidotto M, Grapputo A. Sequencing, De Novo assembly and annotation of the Colorado Potato Beetle, Leptinotarsa decemlineata, Transcriptome. PLoS One. United States; 2014;9: e86012. doi:10.1371/journal.pone.0086012
6. Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, ffrench-Constant RH.

598 Pyrosequencing of the midgut transcriptome of the poplar leaf beetle Chrysomela
599 tremulae reveals new gene families in Coleoptera. Insect Biochem Mol Biol. England;
600 2009;39: 403–413. doi:10.1016/j.ibmb.2009.04.001
601 7. Chen H, Lin L, Xie M, Zhang G, Su W. De novo sequencing, assembly and
602 characterization of antennal transcriptome of Anomala corpulenta Motschulsky
603 (Coleoptera: Rutelidae). PLoS One. United States; 2014;9: e114238.
604 doi:10.1371/journal.pone.0114238
605 8. Oberhofer G, Grossmann D, Siemanowski JL, Beissbarth T, Bucher G. Wnt/ -catenin
606 signaling integrates patterning and metabolism of the insect growth zone. Development.
607 2014;141: 4740–4750. doi:10.1242/dev.112797
608 9. Jacobs CGC, Braak N, Lamers GEM, van der Zee M. Elucidation of the serosal cuticle
609 machinery in the beetle Tribolium by RNA sequencing and functional analysis of
610 Knickkopf1, Retroactive and Laccase2. Insect Biochem Mol Biol. Elsevier Ltd; 2015;60:
611 7–12. doi:10.1016/j.ibmb.2015.02.014
612 10. Stappert D, Frey N, von Levetzow C, Siegfried R. Genome wide identification of Tribolium
613 dorsoventral patterning genes. (Under review). Development.
614 11. i5k-Consortium. The i5K Initiative: advancing arthropod genomics for knowledge, human
615 health, agriculture, and the environment. J Hered. United States; 2013;104: 595–600.
616 doi:10.1093/jhered/est050
617 12. Richards S, Gibbs R a, Weinstock GM, Brown SJ, Denell R, Beeman RW, et al. The
618 genome of the model beetle and pest Tribolium castaneum. Nature. 2008;452: 949–55.
619 doi:10.1038/nature06784
620 13. Keeling CI, Yuen MM, Liao NY, Roderick Docking T, Chan SK, Taylor G a, et al. Draft
621 genome of the mountain pine beetle, Dendroctonus ponderosae Hopkins, a major forest
622 pest. Genome Biol. 2013;14: R27. doi:10.1186/gb-2013-14-3-r27
623 14. Cunningham CB, Ji L, Wiberg RAW, Shelton J, McKinney EC, Parker DJ, et al. The

624 Genome and Methylome of a Beetle with Complex Social Behavior, Nicrophorus
625 vespilloides (Coleoptera: Silphidae). Genome Biol Evol. 2015;7: 3383–96.
626 doi:10.1093/gbe/evv194
627 15. Yin A, Pan L, Zhang X, Wang L, Yin Y, Jia S, et al. Transcriptomic study of the red palm
628 weevil Rhynchophorus ferrugineus embryogenesis. Insect Sci. Australia; 2015;22: 65–82.
629 doi:10.1111/1744-7917.12092
630 16. Hunt T, Bergsten J, Levkanicova Z. A comprehensive phylogeny of beetles reveals the
631 evolutionary origins of a superradiation. Science (80-). 2007;438: 1–4. Available:
632 http://www.sciencemag.org/content/318/5858/1913.short
633 17. Hegner RW. Effects of Removing the Germ-Cell Determinants from the Eggs of Some
634 Chrysomelid Beetles. Preliminary Report. 1908;16: 19–26.
635 18. Hegner RW. The origin and early history of the germ-cells in some chrysomelid beetles. J
636 Morphol. 1909;20: 231–296.
637 19. Meer JM Van Der. Optical clean and permanent whole mount preparation for phase-
638 contrast microscopy of cuticular structures of insect larvae. Dros Inf Serv. 1977;52.
639 20. Jolivet PH, Cox ML, Petitpierre E, editors. Novel aspects of the biology of Chrysomelidae.
640 Kluwer Academic Publishers; 1994.
641 21. Gómez-Zurita J, Hunt T, Kopliku F, Vogler AP. Recalibrated tree of leaf beetles
642 (Chrysomelidae) indicates independent diversification of angiosperms and their insect
643 herbivores. PLoS One. 2007;2: e360. doi:10.1371/journal.pone.0000360
644 22. Blumer LS, Beck CW. Bean Beetles: A Model Organism for Inquiry-based Undergraduate
645 Laboratories [Internet]. 2015. Available: http://www.beanbeetles.org/
646 23. Pauchet Y, Wilkinson P, Chauhan R, Ffrench-Constant RH. Diversity of beetle genes
647 encoding novel plant cell wall degrading enzymes. PLoS One. United States; 2010;5:
648 e15635. doi:10.1371/journal.pone.0015635
649 24. Kirsch R, Wielsch N, Vogel H, Svatos A, Heckel DG, Pauchet Y. Combining proteomics

and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. BMC Genomics. England; 2012;13: 587. doi:10.1186/1471-2164-13-587

25. Flagel LE, Bansal R, Kerstetter RA, Chen M, Carroll M, Flannagan R, et al. Western corn rootworm (Diabrotica virgifera virgifera) transcriptome assembly and genomic analysis of population structure. BMC Genomics. England; 2014;15: 195. doi:10.1186/1471-2164-15-195

26. Strauss AS, Wang D, Stock M, Gretscher RR, Groth M, Boland W, et al. Tissue-specific transcript profiling for ABC transporters in the sequestering larvae of the phytophagous leaf beetle Chrysomela populi. PLoS One. United States; 2014;9: e98637. doi:10.1371/journal.pone.0098637

27. Chi YH, Salzman R a, Balfe S, Ahn J-E, Sun W, Moon J, et al. Cowpea bruchid midgut transcriptome response to a soybean cystatin--costs and benefits of counter-defence. Insect Mol Biol. 2009;18: 97–110. doi:10.1111/j.1365-2583.2008.00854.x

28. Miya K, Kobayashi K. The embryonic development of Atrachya menetriesi. Faldermann (Coleoptera, Chrysomelidae). II. Analyses of early development by ligation and low temperature treatment. J Fac Agric Iwate Univ. 1974;12: 39–55.

29. Miya K, Ando Y, Kurihara M. Formation of duplicated embryos by treatment of low temperature in Atrachya menetriesi Faldermann (Chrysomelidae, Coleoptera). Proc 26th Ann Meet Ent Soc Japan. 1966;9.

30. Ando Y. Geographic Variation In The Incidence Of Non-Diapause Eggs Of The False Melon Beetle, Atrachya-Menetriesi Faldermann (Coleoptera, Chrysomelidae). Appl Entomol Zool. 1979;14: 193–202. Available: <Go to ISI>://A1979GZ01300008

31. Ando Y, Miya K. Diapause character in the false melon beetle, Atrachya menetriesi Faldermann, produced by crossing between diapause and non diapause strains. Bull Fac Agri Iwate Univ. 1968;9: 87–96.

32. Miya K. The embryonic development of a Chrysomelid Beetle, Atrachya menetriesi.

676 Faldermann (Coleoptera) I. The stages of development and changes of external form. J
677 Fac Agric Iwate Univ. 1965;7: 155–166.
678 33. Tran BMD, Credland PF. Consequences of inbreeding for the cowpea seed beetle,
679 Callosobruchus maculatus (F)(Coleoptera: Bruchidae). Biol J Linn Soc. 1995;56: 483–
680 503. doi:10.1111/j.1095-8312.1995.tb01106.x
681 34. Meer J van der. The specification of metameric order in the insect Callosobruchus
682 maculatus Fabr. (Coleoptera) I. Incomplete segment patterns can result from constriction-
683 induced cytological damage. J Embryol Exp …. 1979;51: 1–26. Available:
684 http://dev.biologists.org/content/51/1/1.short
685 35. Meer JM Van Der. Parameters influencing reversal of segment sequence in posterior egg
686 fragments of Callosobruchus (Coleoptera). Roux's Arch Dev Biol. 1984; 339–356.
687 36. Patel NH, Condron BG, Zinn K. Pair-rule expression patterns of even-skipped are found
688 in both short- and long-germ beetles. Nature. 1994;367: 429–434. doi:10.1038/367429a0
689 37. Lynch JA, Brent AE, Leaf DS, Pultz MA, Desplan C. Localized maternal orthodenticle
690 patterns anterior and posterior in the long germ wasp Nasonia. Nature. 2006;439: 728–
691 32. doi:10.1038/nature04445
692 38. Anderson DT. The Development of Holometabolous Insects. In: Counce SJ, Waddinton
693 CH, editors. Developmental systems Insects, Vol 1. New York: Academic Press; 1972.
694 pp. 165–242.
695 39. Davis GK, Patel NH. SHORT, LONG, AND BEYOND: Molecular and Embryological
696 Approaches to. Annu Rev Entomol. 2002; 669–99.
697 40. Andrews S. FastQC: A quality control tool for high throughput sequence data [Internet].
698 2010. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
699 41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
700 data. Bioinformatics. England; 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
701 42. Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused

by random hexamer priming. Nucleic Acids Res. England; 2010;38: e131. doi:10.1093/nar/gkq224

43. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. United States; 2011;29: 644–652. doi:10.1038/nbt.1883

44. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One. United States; 2011;6: e17288. doi:10.1371/journal.pone.0017288

45. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. England; 2013;8: 1494–1512. doi:10.1038/nprot.2013.084

46. Brocchieri L, Karlin S. Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Res. England; 2005;33: 3390–3400. doi:10.1093/nar/gki615

47. O'Neil ST, Dzurisin JD, Carmichael RD, Lobo NF, Emrich SJ, Hellmann JJ. Population-level transcriptome sequencing of nonmodel organisms Erynnis propertius and Papilio zelicaon. BMC Genomics. 2010;11: 310. doi:10.1186/1471-2164-11-310

48. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. England; 2015;31: 3210–3212. doi:10.1093/bioinformatics/btv351

49. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. England; 2005;21: 3674–3676. doi:10.1093/bioinformatics/bti610

50. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. England; 2008;36: 3420–3435. doi:10.1093/nar/gkn176

728 51. Zhong Y-F, Butts T, Holland PWH. HomeoDB: a database of homeobox gene diversity.
729 Evol Dev. United States; 2008;10: 516–518. doi:10.1111/j.1525-142X.2008.00266.x
730 52. van der Zee M, Berns N, Roth S. Distinct Functions of the Tribolium zerknullt Genes in
731 Serosa Specification and Dorsal Closure. Curr Biol. 2005;15: 624–636.
732 doi:10.1016/j.cub.2005.02.057
733 53. Schulz C, Schröder R, Hausdorf B, Wolff C, Tautz D. A caudal homologue in the short
734 germ band beetle Tribolium shows similarities to both, the Drosophila and the vertebrate
735 caudal expression patterns. Dev Genes Evol. Springer; 1998;208: 283–9. Available:
736 http://www.ncbi.nlm.nih.gov/pubmed/9683744
737 54. Hui JH, Raible F, Korchagina N, Dray N, Samain S, Magdelenat G, et al. Features of the
738 ancestral bilaterian inferred from Platynereis dumerilii ParaHox genes. BMC Biol. 2009;7:
739 43. doi:10.1186/1741-7007-7-43
740 55. Kenny NJ, Shen X, Chan TFTH, Wong NWY, Chan TFTH, Chu KH, et al. Genome of the
741 Rusty Millipede, Trigoniulus corallinus, Illuminates Diplopod, Myriapod, and Arthropod
742 Evolution. Genome Biol Evol. 2015;7: 1280–1295. doi:10.1093/gbe/evv070
743 56. Van der Zee M, da Fonseca RN, Roth S. TGFbeta signaling in Tribolium: vertebrate-like
744 components in a beetle. Dev Genes Evol. Germany; 2008;218: 203–213.
745 doi:10.1007/s00427-007-0179-7
746 57. Ozuak O, Buchta T, Roth S, Lynch JA. Ancient and diverged TGF-beta signaling
747 components in Nasonia vitripennis. Dev Genes Evol. Germany; 2014;224: 223–233.
748 doi:10.1007/s00427-014-0481-0
749 58. Kenny NJ, Namigai EKO, Dearden PK, Hui JHL, Grande C, Shimeld SM. The
750 Lophotrochozoan TGF-beta signalling cassette - diversification and conservation in a key
751 signalling pathway. Int J Dev Biol. Spain; 2014;58: 533–549. doi:10.1387/ijdb.140080nk
752 59. Gilbert SF, editor. Developmental Biology. 10th ed. Sunderland, MA: Sinauer Associates,
753 Inc.; 2013.

754 60. Wilson MJ, Dearden PK. Tailless patterning functions are conserved in the honeybee
755 even in the absence of Torso signaling. Dev Biol. Elsevier Inc.; 2009;335: 276–287.
756 doi:10.1016/j.ydbio.2009.09.002
757 61. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or
758 without a reference genome. BMC Bioinformatics. England; 2011;12: 323.
759 doi:10.1186/1471-2105-12-323
760 62. Gotz S, Arnold R, Sebastian-Leon P, Martin-Rodriguez S, Tischler P, Jehl M-A, et al.
761 B2G-FAR, a species-centered GO annotation repository. Bioinformatics. England;
762 2011;27: 919–924. doi:10.1093/bioinformatics/btr059
763 63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool.
764 J Mol Biol. ENGLAND; 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2
765 64. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
766 improvements in performance and usability. Mol Biol Evol. United States; 2013;30: 772–
767 780. doi:10.1093/molbev/mst010
768 65. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary
769 Genetics Analysis version 6.0. Mol Biol Evol. United States; 2013;30: 2725–2729.
770 doi:10.1093/molbev/mst197

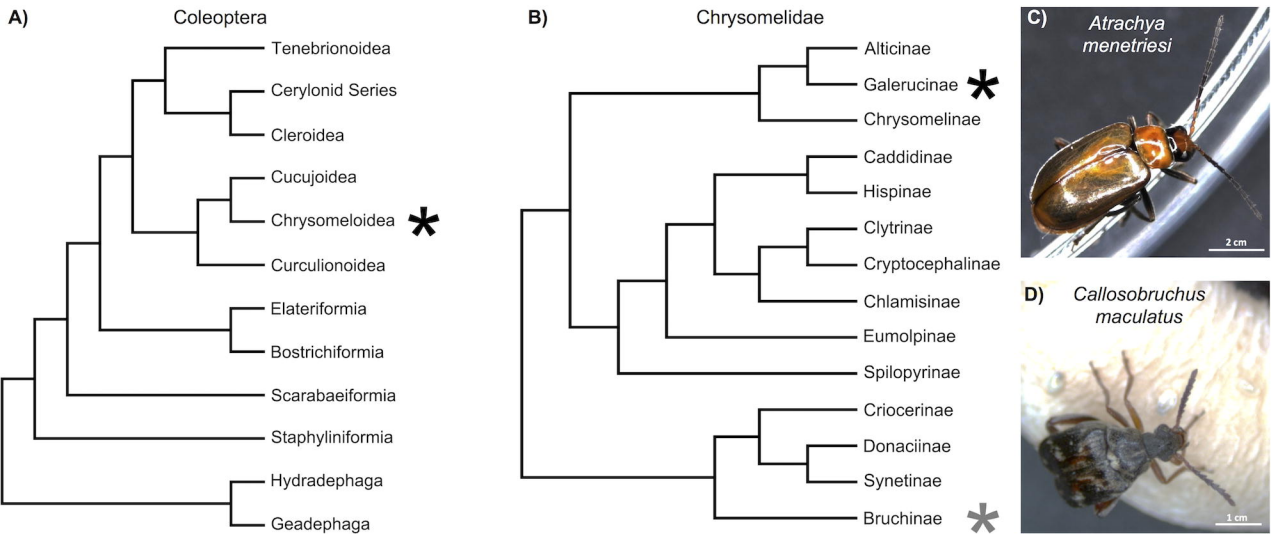
771

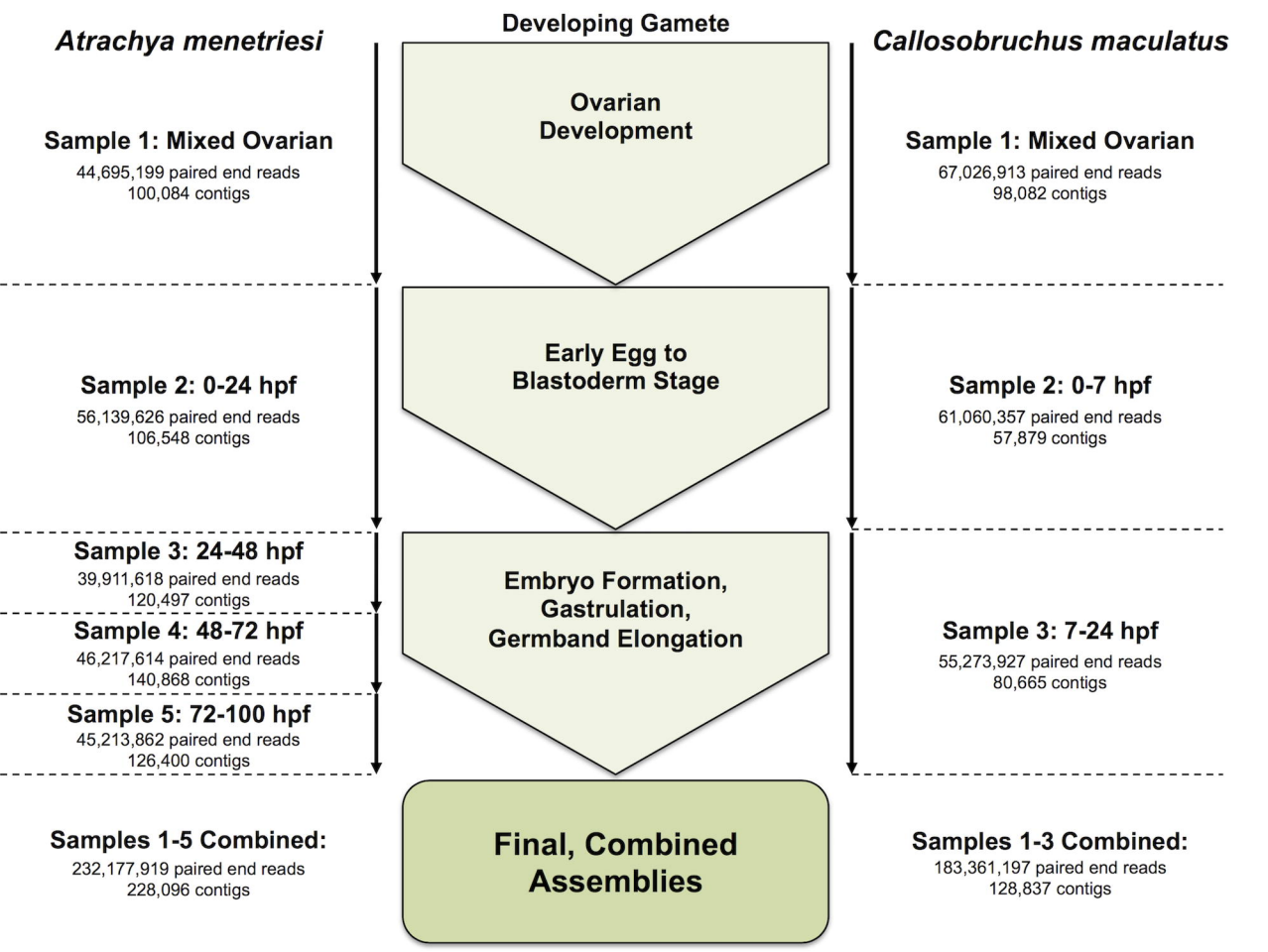
772 Supporting Information Captions:

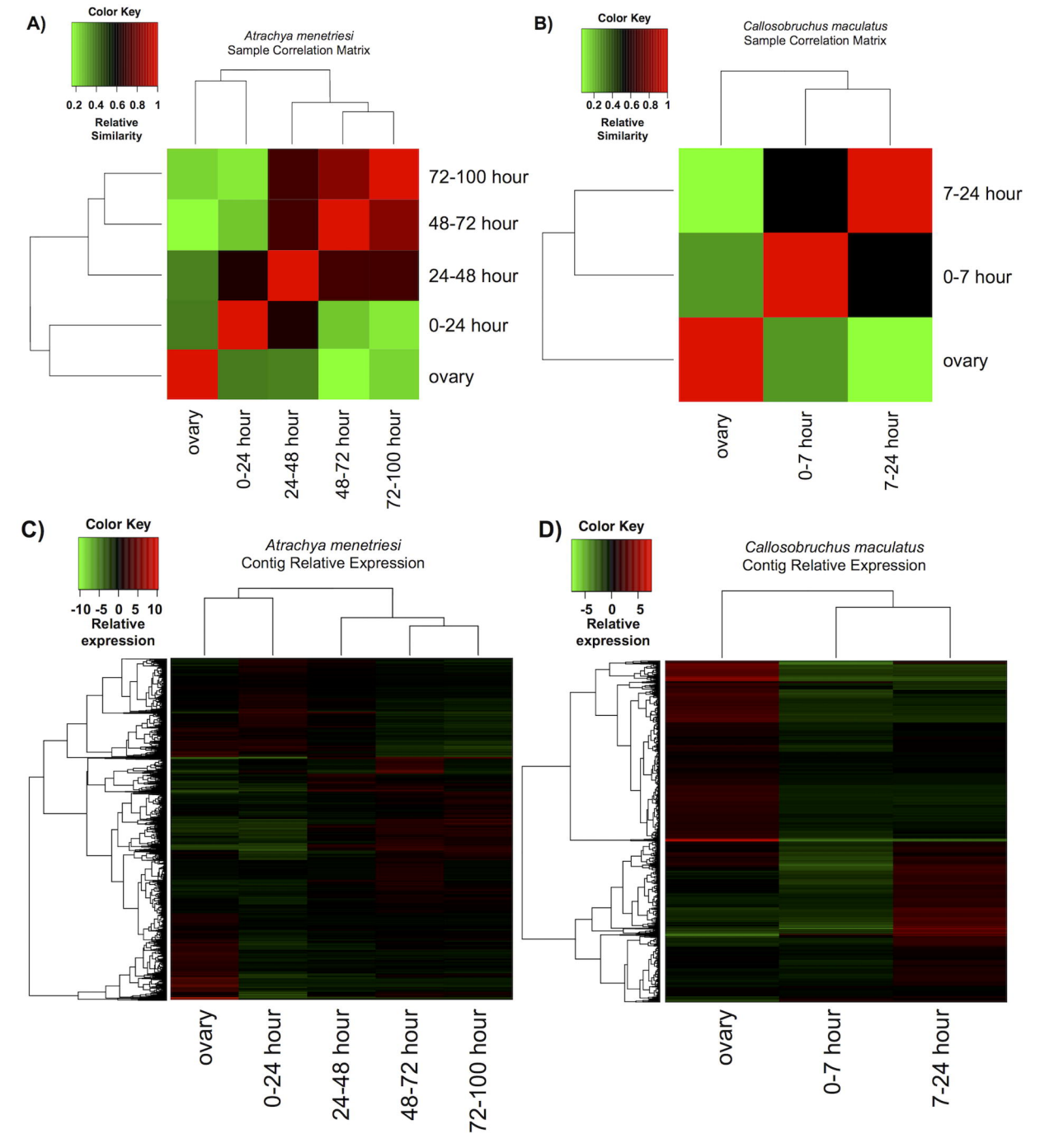
773 Additional File S1: Sequences of all genes referred to in text, and alignments used in
774 phylogenetic analyses (.xls)
775 Additional File S2: Annotations of transcriptome (Blast2GO .annot file) *A. menetriesi*
776 Additional File S3: Annotations of transcriptome (Blast2GO .annot file) *C. maculatus*
777 Additional File S4: KEGG data, *A. menetriesi*
778 Additional File S5: KEGG data, *C. maculatus*

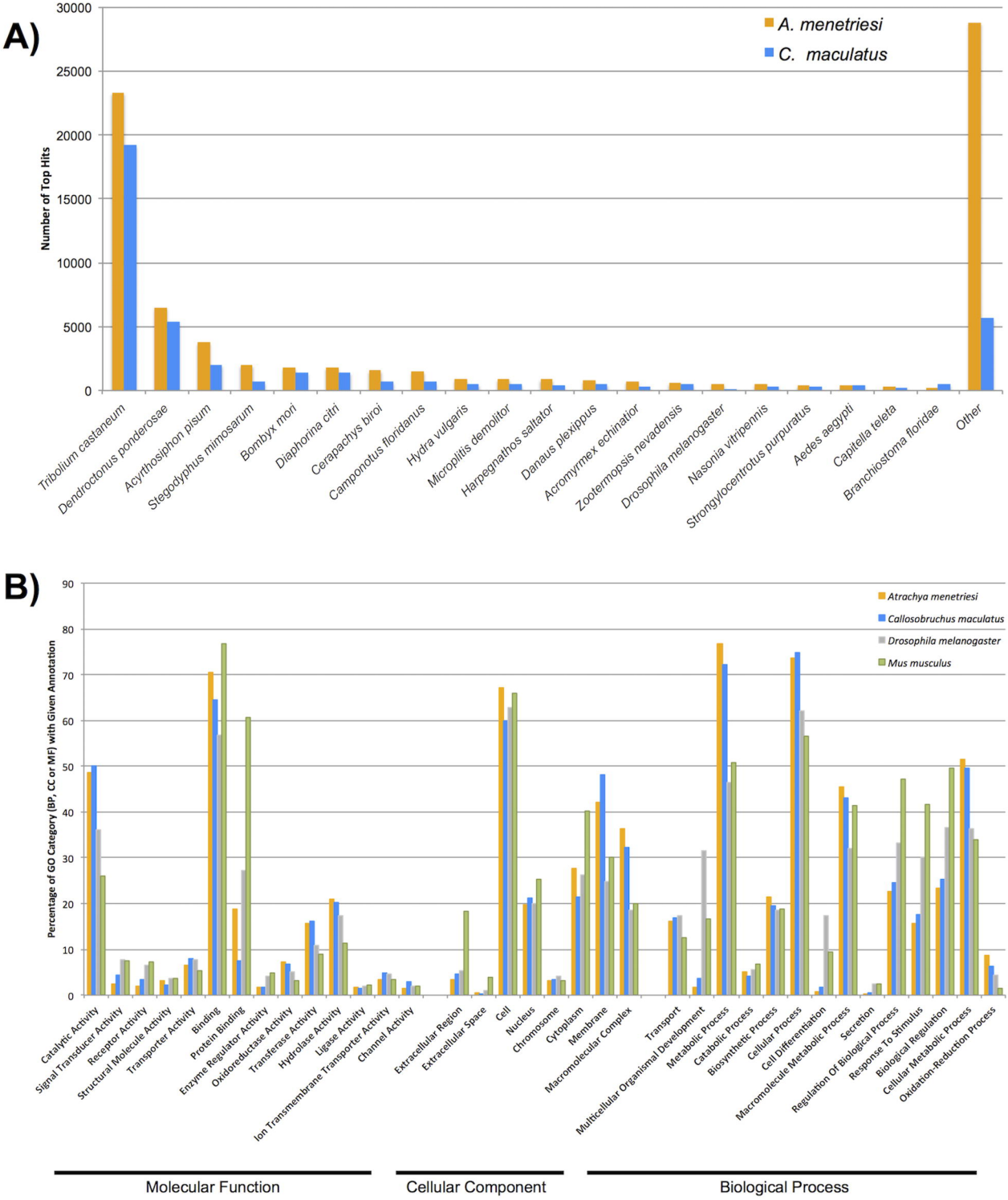
779 Additional File S6: Comparative expression data, *A. menetriesi*

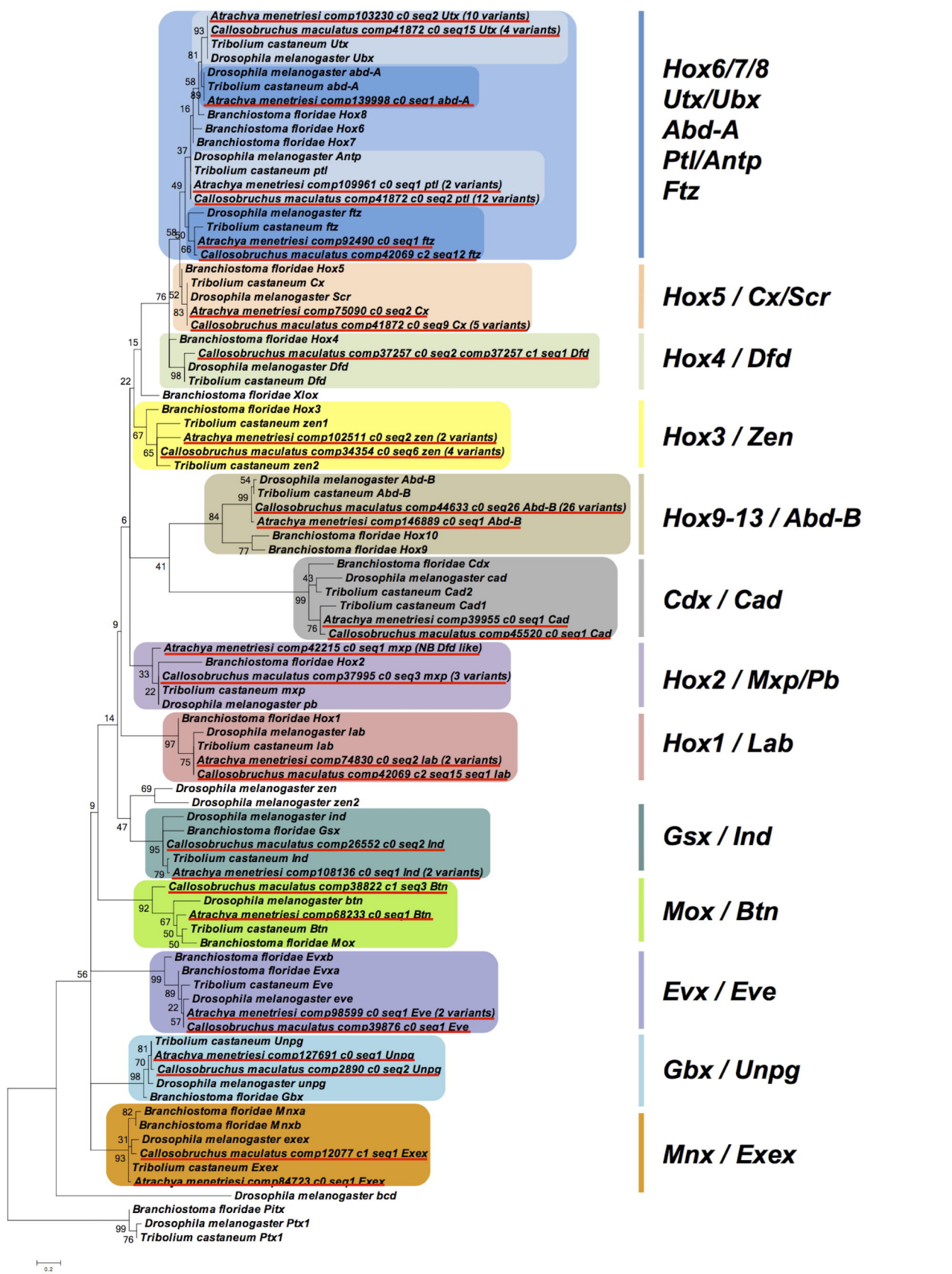
780 Additional File S7: Comparative expression data, *C. maculatus*

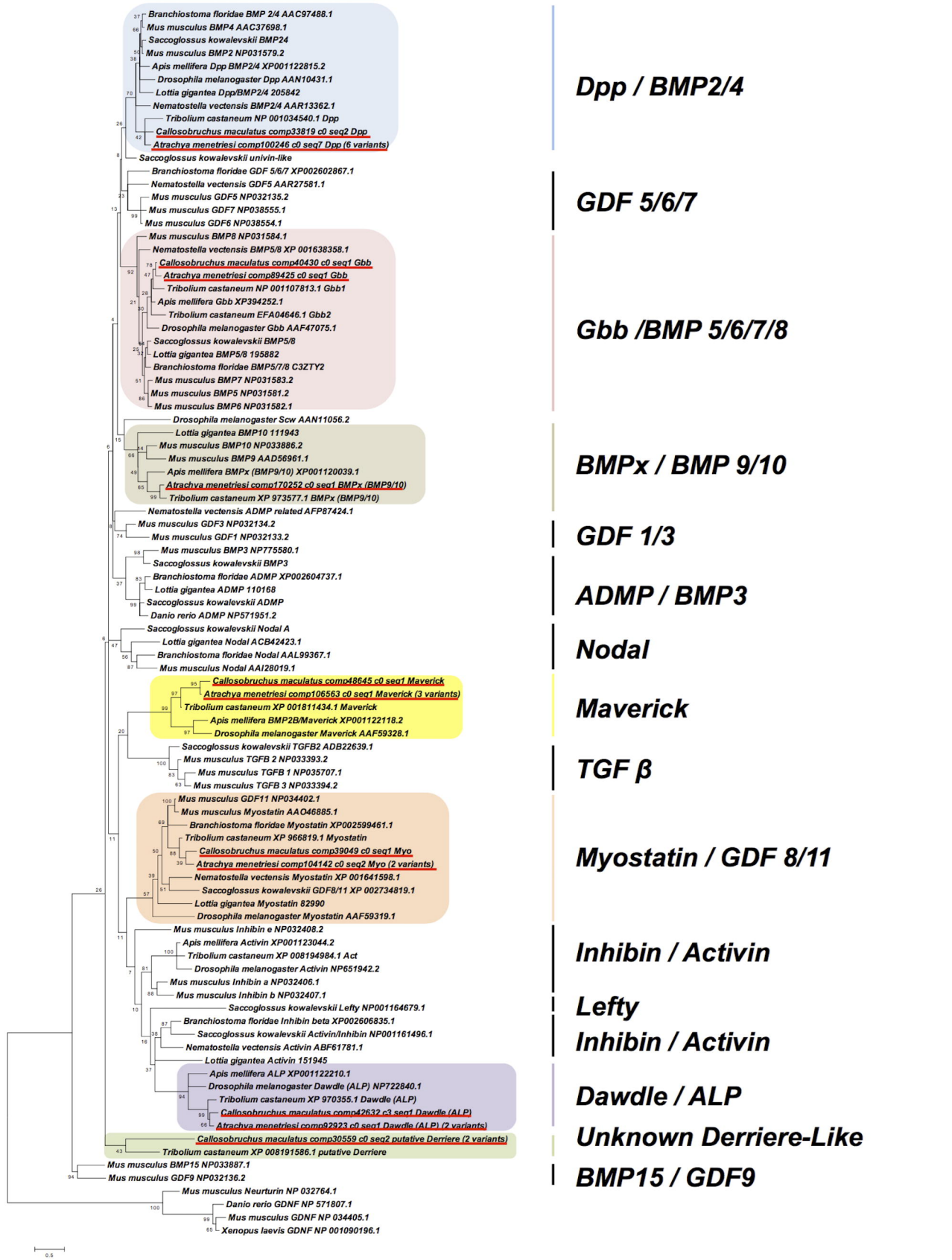

A)
Coleoptera
Tenebrionoidea
Cerylonid Series
Cleroidea
Cucujoidea
Chrysomeloidea
Curculionoidea
Elateriformia
Bostrichiformia
Scarabaeiformia
Staphyliniformia
Hydradephaga
Geadephaga
B)
Chrysomelidae
Alticinae
Galerucinae
Chrysomelinae
Caddidinae
Hispinae
Clytrinae
Cryptocephalinae
Chlamisinae
Eumolpinae
Spilopyrinae
Criocerinae
Donaciinae
Synetinae
Bruchinae
C)
Atrachya menetriesi
2 cm
D)
Callosobruchus maculatus
1 cm

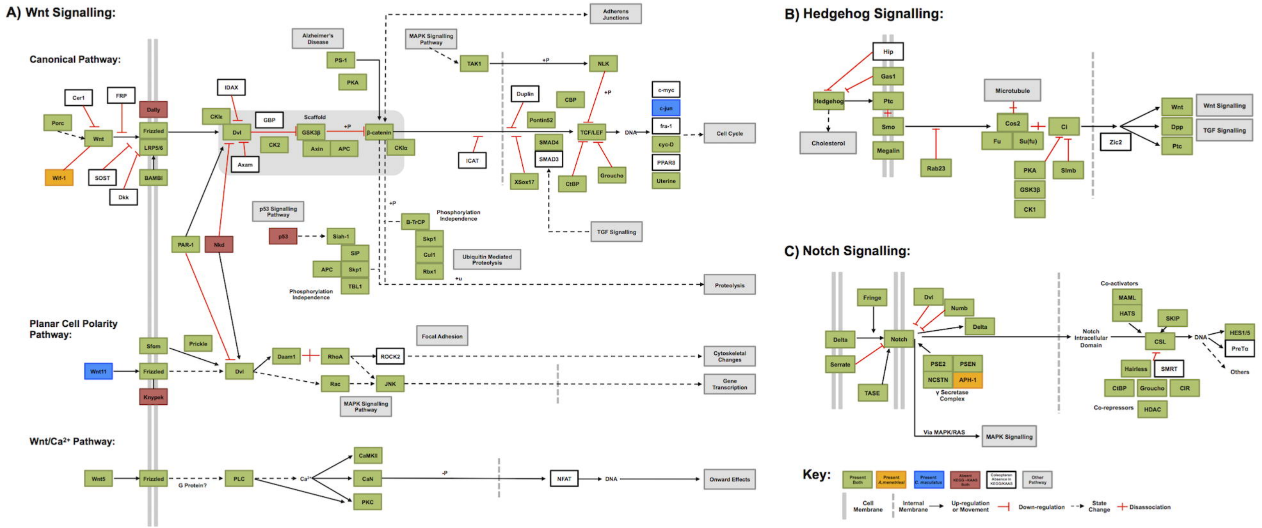

Developing Gamete
Atrachya menetriesi
Callosobruchus maculatus
Ovarian Development
Sample 1: Mixed Ovarian
44,695,199 paired end reads
100,084 contigs
Sample 1: Mixed Ovarian
67,026,913 paired end reads
98,082 contigs
Early Egg to Blastoderm Stage
Sample 2: 0-24 hpf
56,139,626 paired end reads
106,548 contigs
Sample 2: 0-7 hpf
61,060,357 paired end reads
57,879 contigs
Embryo Formation, Gastrulation, Germband Elongation
Sample 3: 24-48 hpf
39,911,618 paired end reads
120,497 contigs
Sample 4: 48-72 hpf
46,217,614 paired end reads
140,868 contigs
Sample 5: 72-100 hpf
45,213,862 paired end reads
126,400 contigs
Sample 3: 7-24 hpf
55,273,927 paired end reads
80,665 contigs
Final, Combined Assemblies
Samples 1-5 Combined:
232,177,919 paired end reads
228,096 contigs
Samples 1-3 Combined:
183,361,197 paired end reads
128,837 contigs


A)
Color Key
0.2 0.4 0.6 0.8 1
Relative Similarity
Atrachya menetriesi
Sample Correlation Matrix
72-100 hour
48-72 hour
24-48 hour
0-24 hour
ovary
ovary
0-24 hour
24-48 hour
48-72 hour
72-100 hour
B)
Color Key
0.2 0.4 0.6 0.8 1
Relative Similarity
Callosobruchus maculatus
Sample Correlation Matrix
7-24 hour
0-7 hour
ovary
ovary
0-7 hour
7-24 hour
C)
Color Key
-10 -5 0 5 10
Relative expression
Atrachya menetriesi
Contig Relative Expression
ovary
0-24 hour
24-48 hour
48-72 hour
72-100 hour
D)
Color Key
-5 0 5
Relative expression
Callosobruchus maculatus
Contig Relative Expression
ovary
0-7 hour
7-24 hour


A)
A. menetriesi
C. maculatus
Number of Top Hits
30000
25000
20000
15000
10000
5000
0
Tribolium castaneum
Dendroctonus ponderosae
Acyrthosiphon pisum
Stegodyphus mimosarum
Bombyx mori
Diaphorina citri
Cerapachys biroi
Camponotus floridanus
Hydra vulgaris
Micropilitis demolitor
Harpegnathos saltator
Danaus plexippus
Acromyrmex echinatior
Zootermopsis nevadensis
Drosophila melanogaster
Nasonia vitripennis
Strongylocentrotus purpuratus
Aedes aegypti
Capitella teleta
Branchiostoma floridae
Other
B)
Atrachya menetriesi
Callosobruchus maculatus
Drosophila melanogaster
Mus musculus
Percentage of GO Category (BP, CC or MF) with Given Annotation
90
80
70
60
50
40
30
20
10
0
Catalytic Activity
Signal Transducer Activity
Receptor Activity
Structural Molecule Activity
Transporter Activity
Binding
Protein Binding
Enzyme Regulator Activity
Oxidoreductase Activity
Transferase Activity
Hydrolase Activity
Ligase Activity
Ion Transmembrane Transporter Activity
Channel Activity
Extracellular Region
Extracellular Space
Cell
Nucleus
Chromosome
Cytoplasm
Membrane
Macromolecular Complex
Transport
Multicellular Organismal Development
Metabolic Process
Catabolic Process
Biosynthetic Process
Cellular Process
Cell Differentiation
Macromolecule Metabolic Process
Secretion
Regulation Of Biological Process
Response To Stimulus
Biological Regulation
Cellular Metabolic Process
Oxidation-Reduction Process
Molecular Function
Cellular Component
Biological Process


Atrachya menetriesi comp103230 c0 seq2 Utx (10 variants)
Callosobruchus maculatus comp41872 c0 seq15 Utx (4 variants)
Tribolium castaneum Utx
Drosophila melanogaster Ubx
Drosophila melanogaster abd-A
Tribolium castaneum abd-A
Atrachya menetriesi comp139998 c0 seq1 abd-A
Branchiostoma floridae Hox8
Branchiostoma floridae Hox6
Branchiostoma floridae Hox7
Drosophila melanogaster Antp
Tribolium castaneum ptl
Atrachya menetriesi comp109961 c0 seq1 ptl (2 variants)
Callosobruchus maculatus comp41872 c0 seq2 ptl (12 variants)
Drosophila melanogaster ftz
Tribolium castaneum ftz
Atrachya menetriesi comp92490 c0 seq1 ftz
Callosobruchus maculatus comp42069 c2 seq12 ftz
Hox6/7/8
Utx/Ubx
Abd-A
Ptl/Antp
Ftz
Branchiostoma floridae Hox5
Tribolium castaneum Cx
Drosophila melanogaster Scr
Atrachya menetriesi comp75090 c0 seq2 Cx
Callosobruchus maculatus comp41872 c0 seq9 Cx (5 variants)
Hox5 / Cx/Scr
Branchiostoma floridae Hox4
Callosobruchus maculatus comp37257 c0 seq2 comp37257 c1 seq1 Dfd
Drosophila melanogaster Dfd
Tribolium castaneum Dfd
Hox4 / Dfd
Branchiostoma floridae Xlox
Branchiostoma floridae Hox3
Tribolium castaneum zen1
Atrachya menetriesi comp102511 c0 seq2 zen (2 variants)
Callosobruchus maculatus comp34354 c0 seq6 zen (4 variants)
Tribolium castaneum zen2
Hox3 / Zen
Drosophila melanogaster Abd-B
Tribolium castaneum Abd-B
Callosobruchus maculatus comp44633 c0 seq26 Abd-B (26 variants)
Atrachya menetriesi comp146889 c0 seq1 Abd-B
Branchiostoma floridae Hox10
Branchiostoma floridae Hox9
Hox9-13 / Abd-B
Branchiostoma floridae Cdx
Drosophila melanogaster cad
Tribolium castaneum Cad2
Tribolium castaneum Cad1
Atrachya menetriesi comp39955 c0 seq1 Cad
Callosobruchus maculatus comp45520 c0 seq1 Cad
Cdx / Cad
Atrachya menetriesi comp42215 c0 seq1 mxp (NB Dfd like)
Branchiostoma floridae Hox2
Callosobruchus maculatus comp37995 c0 seq3 mxp (3 variants)
Tribolium castaneum mxp
Drosophila melanogaster pb
Hox2 / Mxp/Pb
Branchiostoma floridae Hox1
Drosophila melanogaster lab
Tribolium castaneum lab
Atrachya menetriesi comp74830 c0 seq2 lab (2 variants)
Callosobruchus maculatus comp42069 c2 seq15 seq1 lab
Hox1 / Lab
Drosophila melanogaster zen
Drosophila melanogaster zen2
Drosophila melanogaster ind
Branchiostoma floridae Gsx
Callosobruchus maculatus comp26552 c0 seq2 Ind
Tribolium castaneum Ind
Atrachya menetriesi comp108136 c0 seq1 Ind (2 variants)
Gsx / Ind
Callosobruchus maculatus comp38822 c1 seq3 Btn
Drosophila melanogaster btn
Atrachya menetriesi comp68233 c0 seq1 Btn
Tribolium castaneum Btn
Branchiostoma floridae Mox
Mox / Btn
Branchiostoma floridae Evxb
Branchiostoma floridae Evxa
Tribolium castaneum Eve
Drosophila melanogaster eve
Atrachya menetriesi comp98599 c0 seq1 Eve (2 variants)
Callosobruchus maculatus comp39876 c0 seq1 Eve
Evx / Eve
Tribolium castaneum Unpg
Atrachya menetriesi comp127691 c0 seq1 Unpg
Callosobruchus maculatus comp2890 c0 seq2 Unpg
Drosophila melanogaster unpg
Branchiostoma floridae Gbx
Gbx / Unpg
Branchiostoma floridae Mnxa
Branchiostoma floridae Mnxb
Drosophila melanogaster exex
Callosobruchus maculatus comp12077 c1 seq1 Exex
Tribolium castaneum Exex
Atrachya menetriesi comp84723 c0 seq1 Exex
Mnx / Exex
Drosophila melanogaster bcd
Branchiostoma floridae Pitx
Drosophila melanogaster Ptx1
Tribolium castaneum Ptx1
0.2


Branchiostoma floridae BMP 2/4 AAC97488.1
Mus musculus BMP4 AAC37698.1
Saccoglossus kowalevskii BMP24
Mus musculus BMP2 NP031579.2
Apis mellifera Dpp BMP2/4 XP001122815.2
Drosophila melanogaster Dpp AAN10431.1
Lottia gigantea Dpp/BMP2/4 205842
Nematostella vectensis BMP2/4 AAR13362.1
Tribolium castaneum NP 001034540.1 Dpp
Callosobruchus maculatus comp33819 c0 seq2 Dpp
Atrachya menetriesi comp100246 c0 seq7 Dpp (6 variants)
Dpp / BMP2/4
Saccoglossus kowalevskii univin-like
Branchiostoma floridae GDF 5/6/7 XP002602867.1
Nematostella vectensis GDF5 AAR27581.1
Mus musculus GDF5 NP032135.2
Mus musculus GDF7 NP038555.1
Mus musculus GDF6 NP038554.1
GDF 5/6/7
Mus musculus BMP8 NP031584.1
Nematostella vectensis BMP5/8 XP 001638358.1
Callosobruchus maculatus comp40430 c0 seq1 Gbb
Atrachya menetriesi comp89425 c0 seq1 Gbb
Tribolium castaneum NP 001107813.1 Gbb1
Apis mellifera Gbb XP394252.1
Tribolium castaneum EFA04646.1 Gbb2
Drosophila melanogaster Gbb AAF47075.1
Saccoglossus kowalevskii BMP5/8
Lottia gigantea BMP5/8 195882
Branchiostoma floridae BMP5/7/8 C3ZTY2
Mus musculus BMP7 NP031583.2
Mus musculus BMP5 NP031581.2
Mus musculus BMP6 NP031582.1
Gbb /BMP 5/6/7/8
Drosophila melanogaster Scw AAN11056.2
Lottia gigantea BMP10 111943
Mus musculus BMP10 NP033886.2
Mus musculus BMP9 AAD56961.1
Apis mellifera BMPx (BMP9/10) XP001120039.1
Atrachya menetriesi comp170252 c0 seq1 BMPx (BMP9/10)
Tribolium castaneum XP 973577.1 BMPx (BMP9/10)
BMPx / BMP 9/10
Nematostella vectensis ADMP related AFP87424.1
Mus musculus GDF3 NP032134.2
Mus musculus GDF1 NP032133.2
GDF 1/3
Mus musculus BMP3 NP775580.1
Saccoglossus kowalevskii BMP3
Branchiostoma floridae ADMP XP002604737.1
Lottia gigantea ADMP 110168
Saccoglossus kowalevskii ADMP
Danio rerio ADMP NP571951.2
ADMP / BMP3
Saccoglossus kowalevskii Nodal A
Lottia gigantea Nodal ACB42423.1
Branchiostoma floridae Nodal AAL99367.1
Mus musculus Nodal AAI28019.1
Nodal
Callosobruchus maculatus comp48645 c0 seq1 Maverick
Atrachya menetriesi comp106563 c0 seq1 Maverick (3 variants)
Tribolium castaneum XP 001811434.1 Maverick
Apis mellifera BMP2B/Maverick XP001122118.2
Drosophila melanogaster Maverick AAF59328.1
Maverick
Saccoglossus kowalevskii TGFB2 ADB22639.1
Mus musculus TGFB 2 NP033393.2
Mus musculus TGFB 1 NP035707.1
Mus musculus TGFB 3 NP033394.2
TGF β
Mus musculus GDF11 NP034402.1
Mus musculus Myostatin AAO46885.1
Branchiostoma floridae Myostatin XP002599461.1
Tribolium castaneum XP 966819.1 Myostatin
Callosobruchus maculatus comp39049 c0 seq1 Myo
Atrachya menetriesi comp104142 c0 seq2 Myo (2 variants)
Nematostella vectensis Myostatin XP 001641598.1
Saccoglossus kowalevskii GDF8/11 XP 002734819.1
Lottia gigantea Myostatin 82990
Drosophila melanogaster Myostatin AAF59319.1
Myostatin / GDF 8/11
Mus musculus Inhibin e NP032408.2
Apis mellifera Activin XP001123044.2
Tribolium castaneum XP 008194984.1 Act
Drosophila melanogaster Activin NP651942.2
Mus musculus Inhibin a NP032406.1
Mus musculus Inhibin b NP032407.1
Inhibin / Activin
Saccoglossus kowalevskii Lefty NP001164679.1
Lefty
Branchiostoma floridae Inhibin beta XP002606835.1
Saccoglossus kowalevskii Activin/Inhibin NP001161496.1
Nematostella vectensis Activin ABF61781.1
Lottia gigantea Activin 151945
Inhibin / Activin
Apis mellifera ALP XP001122210.1
Drosophila melanogaster Dawdle (ALP) NP722840.1
Tribolium castaneum XP 970355.1 Dawdle (ALP)
Callosobruchus maculatus comp42632 c3 seq1 Dawdle (ALP)
Atrachya menetriesi comp92923 c0 seq1 Dawdle (ALP) (2 variants)
Dawdle / ALP
Callosobruchus maculatus comp30559 c0 seq2 putative Derriere (2 variants)
Tribolium castaneum XP 008191586.1 putative Derriere
Unknown Derriere-Like
Mus musculus BMP15 NP033887.1
Mus musculus GDF9 NP032136.2
BMP15 / GDF9
Mus musculus Neurturin NP 032764.1
Danio rerio GDNF NP 571807.1
Mus musculus GDNF NP 034405.1
Xenopus laevis GDNF NP 001090196.1
0.5


A) Wnt Signalling:
Canonical Pathway:
Cer1
FRP
Porc
Wnt
Wif-1
SOST
Dkk
Dally
Frizzled
LRP5/6
BAMBI
IDAX
CKIε
GBP
Scaffold
Dvl
Axam
GSK3β
CK2
Axin
APC
β-catenin
CKIα
PS-1
PKA
Alzheimer's Disease
MAPK Signalling Pathway
TAK1
NLK
Adherens Junctions
ICAT
Duplin
Pontin52
CBP
SMAD4
SMAD3
XSox17
CtBP
Groucho
TCF/LEF
DNA
c-myc
c-jun
fra-1
cyc-D
PPARδ
Uterine
Cell Cycle
TGF Signalling
PAR-1
Nkd
p53 Signalling Pathway
p53
Siah-1
SIP
APC
Skp1
TBL1
Phosphorylation Independence
B-TrCP
Skp1
Cul1
Rbx1
Ubiquitin Mediated Proteolysis
Proteolysis
Planar Cell Polarity Pathway:
Sfom
Prickle
Wnt11
Frizzled
Knypek
Dvl
Daam1
RhoA
ROCK2
Rac
JNK
Focal Adhesion
MAPK Signalling Pathway
Cytoskeletal Changes
Gene Transcription
Wnt/Ca²⁺ Pathway:
Wnt5
Frizzled
G Protein?
PLC
Ca²⁺
CaMKII
CaN
PKC
NFAT
DNA
Onward Effects
B) Hedgehog Signalling:
Hip
Gas1
Hedgehog
Ptc
Smo
Megalin
Cholesterol
Rab23
Microtubule
Cos2
Fu
Su(fu)
Ci
PKA
GSK3β
CK1
Slmb
Zic2
Wnt
Dpp
Ptc
Wnt Signalling
TGF Signalling
C) Notch Signalling:
Fringe
Delta
Serrate
TASE
Notch
Dvl
Numb
Delta
PSE2
PSEN
NCSTN
APH-1
γ Secretase Complex
Via MAPK/RAS
MAPK Signalling
Notch Intracellular Domain
Co-activators
MAML
HATS
SKIP
CSL
DNA
HES1/5
PreTα
Others
Hairless
SMRT
CtBP
Groucho
CIR
HDAC
Co-repressors
Key:
Present Both
Present A.menetriesi
Present C. maculatus
Absent KEGG-KAAS Both
Coleopteran Absence in KEGG/KAAS
Other Pathway
Cell Membrane
Internal Membrane
Up-regulation or Movement
Down-regulation
State Change
Disassociation