

EST and microarray analysis of horn development in *Onthophagus* beetles

Teiya Kijimoto¹, James Costello^{1,2}, Zuojian Tang³, Armin P. Moczek¹, and Justen Andrews^{1*}

Addresses: ¹Department of Biology, Indiana University, Bloomington, Indiana, 47405, USA, ²School of Informatics, Indiana University, Bloomington, Indiana, 47405, USA, ³Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana, 47405, USA

*Corresponding author

Email: Teiya Kijimoto - tkijimot@indiana.edu; James Costello - jccostel@indiana.edu; Zuojian Tang - tangz@cgb.indiana.edu; Armin P. Moczek - armin@indiana.edu; Justen Andrews* – jandrew@bio.indiana.edu

Running head: Transcriptome analysis of horned beetle

Keywords: *Onthophagus taurus*, horned beetle, expressed sequence tags (ESTs), comparative genomics

Background: The origin of novel traits and their subsequent diversification represent central themes in evo-devo and evolutionary ecology. Here we explore the genetic and genomic basis of a class of traits that is both novel and highly diverse, in a group of organisms that is ecologically complex and experimentally tractable: horned beetles.

Results: We developed two high quality, normalized cDNA libraries for larval and pupal *Onthophagus taurus* including 3,488 EST sequences assembled into 451 contigs and 2,330 singletons. We present a detailed analysis of library composition including GO-term profiles, analysis of sequence conservation, and EST homology analysis. Microarrays developed from the combined libraries were then used to contrast the transcriptome of developing primordia of head horns, thoracic horns, and legs. Our experiments identify a first comprehensive list of candidate genes for the evolution and diversification of beetle horns. We find that developing horns and legs show many similarities as well as important differences in their transcription profiles, suggesting that the origin of horns was mediated partly, but not entirely, by the recruitment of genes involved in the formation of more traditional appendages such as legs. Furthermore, we find that horns developing from the head and thorax differ in their transcription profiles to a degree that suggests that head and thoracic horns are not serial homologs, but instead may have evolved independent of each other.

Conclusions: We have laid the foundation for a systematic analysis of the genetic basis of

horned beetle development and diversification with the potential to contribute significantly to several major frontiers in evolutionary developmental biology.

INTRODUCTION

The origin of novel traits and their subsequent diversification have been central themes in evolutionary biology ever since the discipline's inception over 150 years ago [1, 2]. Specifically, the genetic, developmental, and ecological mechanisms, and the interactions between them, that allow novel phenotypes and functions to arise from pre-existing variation, continue to represent major frontiers in our understanding of phenotypic diversity. With the advent of modern *-omics* approaches, researchers have increasingly departed from a candidate gene or pathway approach and begun to explore organismal development and evolution from a genome, transcriptome, or proteome perspective, focusing in large part on existing genetic model systems such as *Drosophila* or *Caenorhabditis*. However, many key questions in evolutionary biology, including the mechanisms underlying organismal innovation, the role of plasticity in diversification, and the interplay between ecology and developmental evolution, are often difficult to address solely within the confines of classic model systems. Recent efforts have therefore begun to generate genomic and developmental genetic resources for organisms with promise as future model systems in evolutionary developmental biology and ecological genetics (e.g butterflies: [3, 4]; honey bees: reviewed in [5]; red flour beetle: [6]). Here we present and apply the first genomic resources to advance the study of a class of traits that is both novel and highly diverse, in a group of organisms that is ecologically complex and experimentally tractable: beetle horns and horned beetles.

Beetle horns possess many characteristics that make them interesting models for integrating genetic, developmental, and environmental perspectives on the development and evolution of complex, novel traits (reviewed in [7]). First, beetle horns are major structures, often dominating the phenotype of their bearers. Second, beetle horns function as weapons of sexual selection, thus playing a major role in the behavioral ecology of individuals and populations. Third, beetle horns are inordinately variable, both within and between species, including differences in number, size, shape, and location. Moreover, diversity in horn expression is paralleled by amazing species richness. For instance, the genus *Onthophagus* currently contains over 2,400 extant species, making it the most speciose genus in the animal kingdom [8, 9]. Fourth, beetle horns are influenced in their expression by both genetic and environmental factors, ranging from absence of environmental sensitivity to complete determination by nutritional condition. In some cases, both extremes of environmental sensitivity can be found in different horn types expressed by the same individual [10]. Finally, beetle horns lack any obvious homology to structures in other insects or non-insect arthropods. Beetle horns are not modified antennae or mouthparts, but instead horns were "invented" by beetles in addition to their traditional appendages [11], and now provide their bearers with an important new function: a weapon used in male-male competition. Beetle horns and horned beetles therefore offer rich opportunities to explore the mechanisms of organismal innovation and diversification.

Beetle horns are rigid outgrowths of the exoskeleton that originate as epidermal outbuddings of

the head or prothoracic epithelium. Horns lack joints, muscles, and nervous tissue. Several recent studies have begun to shed light on how beetle horns develop and differentiate during ontogeny [11-15]. and showed that two developmentally dissociated processes contribute to the final degree of horn expression seen in adults: a prepupal growth phase late in larval development followed by a pupal remodeling phase just prior to the final, adult molt (reviewed in [10]). As such, the development of horns shows many qualitative similarities to the development of traditional appendages, but also exhibits important differences. For instance, pronotal horn primordia are frequently resorbed during the pupal stage in a sex- and species-specific manner, a phenomenon not usually associated with regular appendages [14].

Furthermore, earlier studies have also begun to question whether horns that develop in different body regions, such as the head vs prothorax, constitute serial homologs, or instead may have evolved and diversified independent of each other [13, 16].

Here we present the first steps toward a systematic analysis of the genetic and genomic basis of horn development and diversification in the genus *Onthophagus*. We first present the results of a comprehensive EST analysis of two normalized cDNA libraries obtained from two disparate developmental stages of *Onthophagus taurus*: larva and pupa. Second, using microarrays developed from our EST libraries we contrast the transcription profiles of the primordia of developing thoracic horns, head horns, and legs right after the transition from larva to pupa. We then use these contrasts to identify candidate genes involved in the development and diversification of beetle horns. Furthermore, we examine two basic questions regarding the

origin and diversification of horns: (a) Are horns highly simplified versions of more traditional appendages such as legs? If so, transcription profiles of developing horn primordia should largely match those of developing legs. If not, transcription profiles of developing horn primordia should only partly match those of developing legs and also include horn specific transcription signatures. (b) Are different horn types produced in different body regions homonomous, i.e. serial homologs of the same ancestral structure? If so, different horn types should exhibit highly similar transcription profiles. However, if different horn types originated and diversified independent of each other, transcription profiles may be predicted to exhibit important horn-type specific differences. We discuss the significance of our findings in the context of the biology of horned beetles in particular, and the origins and diversification of novel traits in general.

RESULTS

Production and analysis of EST sequences

We constructed two normalized, size selected, and directionally cloned cDNA libraries from (i) heads and thoraces dissected from larvae and prepupae (referred to as *OtL* [larval]) and (ii) whole pupae (referred to as *OtP* [pupal]). Individual 5' - sequencing reads were generated from 3,756 randomly selected cDNA clones (1,874 larval and 1,882 pupal). The set of EST sequences were trimmed of vector sequence, adaptor sequence, and poly(A) tails, and filtered to remove

sequences that were either low quality, chimeric, or shorter than 100 nucleotides (Materials and Methods). This yielded 3,488 high-quality sequences (1,783 larval and 1,705 pupal, Table 1) that are available at GenBank (accession numbers FG539013– FG542500). We then used ESTPiper [17] to assemble these sequences into contigs (see Methods). A total of 1,158 of these sequences were assembled into 451 contigs with an average of 2.6 clones per contig and a maximum of 11 clones per contig. The remaining 2,330 sequences did not assemble into contigs and are referred to as singletons. Thus, the 3,488 sequences collapse into 2,781 distinct sequences (451 contigs and 2,330 singletons) that we refer to as “non-redundant” sequences.

It is likely that some of the non-redundant sequences derive from the same transcript but do not overlap, possibly due to 5'-truncated cDNA clones. In order to estimate the magnitude of this redundancy, we aligned the *Onthophagus* non-redundant sequences to *Drosophila* proteins, filtered the alignments for highly similar matches (BLASTx, E-value < 10^{-60}), and then manually examined the alignments for separate *Onthophagus* sequences that align to distinct regions of the same *Drosophila* protein. Among 534 mixed independent *Onthophagus* sequences we found 35 pairs of sequences that aligned to the same *Drosophila* protein. Of these, 19 pairs aligned with highly similar matches to different regions of the same protein, indicating they derive from non-overlapping regions of the same transcript; 12 pairs had co-linear alignments with 95-98% sequence identities, suggesting that they either derive from the same gene with polymorphisms and/or sequencing errors, or derive from highly similar duplicate genes; and 4 pairs appear to be

splice variants. Thus, this sample of 534 *Onthophagus* non-redundant sequences represents approximately 501 distinct genes (93 % unique). While this analysis may be subject to sampling errors, it does indicate that false-negative assemblies are not a pervasive problem, and that the complete set of 2,781 non-redundant sequences represents approximately 2,599 *Onthophagus* unique genes.

Functional annotation of assembled sequences

Given that insects express a broad diversity of genes during embryogenesis and metamorphosis [18], we expected that the larval and pupal ESTs would be a rich source of gene discovery. In order to provide a first pass annotation for the putative function of the *Onthophagus* gene sequences, we annotated the non-redundant sequences using the UniProtKB/TrEMBL protein sequence database (E-value < 10^{-5}). This successfully annotated 71.3% of the mixed sequences. As expected, these annotations covered a wide diversity of biological and molecular functions including the major expected categories such as cellular processes, metabolic processes, biological regulation, multicellular organismal processes, and developmental processes (Supplemental Tables 1 and 2). This, coupled with the low redundancy within the *Onthophagus* libraries, affirms the set of ESTs as a rich source for gene discovery.

Given that the cDNA libraries derive from animals undergoing metamorphosis, which involves dramatic remodeling of the insect body accompanied by complex patterns of gene expression, it

was expected that the EST libraries would include genes involved in a wide range of developmental processes. Indeed the assembled sequences included over 75 genes with close sequence similarity to genes with important functions in the development of other arthropods (Table 3). Briefly, these included the following major groups: 1) genes involved in axis-specification, patterning and morphogenesis, including many transcription factors (*homothorax*, *extradenticle*, *spalt-related*, *bicaudal*, *prothoraxless*, *teashirt-like*, *Sex comb on midleg*, *Cephalothorax*, *Ultrabithorax*, *cut*, *tailup*, *pointed*, *Abdominal B*, *hairy*, *bab2*, *Additional sex combs*); 2) proteins involved in several signaling pathways, including MAPK pathways (*Epidermal growth factor-like protein*, *Star*, *MAP kinase-interacting serine/threonine kinase*, *licorne*, *puckered*, *DRas2*, *misshapen*, *discs large 1*) the Wnt receptor signaling pathway (*frizzled 4*, *shaggy*, *armadillo*, *hyrax*, *Wnt oncogene analog 2*), the Notch signaling pathway (*Notch*, *fringe*, *dishevelled*, *kuzbanian*, *Enhancer of split*, *strawberry notch*), the Hedgehog signaling pathway (*hedgehog*), the TGF-beta related pathway (*bambi*, *cornichon*), and the Toll signaling pathway (*Spatzle-Processing Enzyme*, *pipe*); 3) genes involved in endocrine regulation of development including ecdysone signaling (*ultraspiracle*, *Ecdysone receptor*, *shade*, *disembodied*, *broad*, *Ecdysone-induced protein 78C*, *Ecdysone-induced protein 75B*) and juvenile hormone signaling (*Juvenile hormone epoxide hydrolase 3*, *Juvenile hormone acid methyl transferase*). These clones represent a rich set of annotated genes for future studies investigating the function of the respective pathways in *Onthophagus* development and evolution.

Comparative analysis of the Onthophagus transcriptome

While the beetle order is incredibly species rich and diverse, this study represents only the second systematic study of beetle genes [6]. Comparative analyses of gene content in *Tribolium castaneum* revealed that the proportion of universal and insect specific genes is similar to that in other insects [6]. However, the proportion of genes without similarity to other organisms is higher in red flour beetles than in other insects [6]. In order to explore the conservation of *Onthophagus* genes across metazoans, we aligned the 2,781 non-redundant *Onthophagus* sequences to the protein sequences derived from the annotated genomes of *Tribolium castaneum* (NCBI GenBank), *Drosophila melanogaster* (FlyBase), *Caenorhabditis elegans* (Ensembl), human (Ensembl), as well as non-redundant protein dataset (nr) from GenBank. We also aligned our translated sequences with combined “invertebrate protein datasets” from NCBI (<ftp://ftp.ncbi.nih.gov/refseq/release/invertebrate>).

In order to group the *Onthophagus* sequences according to patterns of conservation and divergence across these datasets we filtered them for those with BLASTx sequence matches with proteins in the various datasets (E-value < 1×10^{-5}), and then clustered them according to the bit scores (Figure 1). A total of 1,086 non-redundant *Onthophagus* sequences (39%) had sequence matches to proteins in all the datasets searched (Figure 1, Group 1). A further 868 of the non-redundant sequences (31%) had matches to proteins in the *Tribolium* dataset, as well as to proteins in one or more of the other datasets (Figure 1, Group 2). Within group 2 there are two

prominent sub-groups. First, there were 300 mixed independent sequences (10.8%) with matches to fly, “invertebrate”, and nr proteins, but no matches to worms and humans, and are thus putatively restricted to insects. While Gene Ontology annotations [19-21] of these sequences show no striking enrichment for specific biological processes, they do include proteins with functions specific to insects, such as cuticle proteins (data not shown). Second, there were 212 (7.6%) mixed independent sequences with matches to proteins in either only *Tribolium*, or *Tribolium* and either/both of “invertebrate” and nr proteins. In order to test if these sequences are indeed restricted to beetles we filtered them to meet the following criteria: (i) the *Onthophagus* sequences did not have a sequence match (BLASTx, E-value < 10^{-20}) with proteins from non-beetle species in the nr dataset; (ii) the *Tribolium* protein to which the *Onthophagus* sequence had the best alignment did not have a sequence match (BLASTx, E-value < 10^{-20}) with proteins from non-beetle species in the nr dataset. In total, 44 of the non-redundant sequences (1.6%) met these criteria and thus are restricted to beetle species among the currently available protein sequence information. This suggests that these genes may have arisen *de novo*, or may be fast evolving, in beetles. A total of 194 translated non-redundant sequences (7%) did not match proteins from *Tribolium*, but did have matches in one or more of flies, worms, humans, invertebrates or nr (Figure 1, group 3). Among these, 43 sequences had matches in *all* other protein datasets, and 33 additional sequences had matches to *Drosophila*, “invertebrate”, and nr. Combined, these data raise the possibility that at least some of these 76 sequences may either have been lost from the *Tribolium* lineage, or alternatively, may be conserved but simply not yet

annotated in *Tribolium*. Indeed, we found 22 cases where the *Onthophagus* sequences matched against regions of the *Tribolium* genome (BLASTn E-value $< 1 \times 10^{-5}$, data not shown) with no gene annotations. These may represent genes that unannotated in the *Tribolium* genome. Thus, the *Onthophagus* expressed sequences reported here would be useful in refining the annotation of the *Tribolium* genome.

Our analysis also identified 633 of the mixed independent sequences (23.2%) to have “no-hit” (Figure 1 Group 4) to any of the proteomes. This is consistent with the finding that approximately 23% of genes annotated in the *Tribolium* genome lack sequence matches in a wide range of other species [6]. However, our estimate of *Onthophagus* specific sequences is likely to be inflated by (i) sequences that are largely, or entirely, within the UTRs of protein coding transcripts, or (ii) sequences that may be non-coding transcripts. In the case of sequences derived from protein coding genes that are unique to *Onthophagus*, we expect that the size distribution of ORFs to be similar to that of the sequences with matches to proteins in other species. Therefore, we compared the size distribution of the longest identified ORF for each independent sequence in the no-hit class, to the longest identified ORF in the independent sequences that hit at least one protein in the other species (2,148 sequences, Supplemental Figure 1). For the hit-class, which is highly likely to include the protein coding sequences, 90% of the sequences had the longest predicted ORFs of greater than 300 nucleotides (100 codons). While size distribution of the no-hit class is skewed towards smaller predicted ORFs, 44% of these

sequences had a longest predicted ORF of greater than 300 nucleotides. This suggests that a significant fraction of the no-hit sequences encode proteins that are unique to *Onthophagus*.

Gene expression profiles in pupal appendage primordia

While our EST analysis identified many genes homologous to interesting *Drosophila* developmental genes, and such an approach to identify candidate genes has been successful in beetles [11, 15, 22, 23], this approach is limited to identifying obvious candidates. Given that *Onthophagus* horns appear to be novel structures invented in beetles, it is highly likely that unexpected, or indeed previously uncharacterized genes may be important in their development. We therefore undertook gene expression profiling of developing horns (early pupal stage) as an unbiased means of identifying such candidates. Since there is evidence that head horns and thoracic horn are quite distinct structures (not simply serial homologs; [13, 16]), we analyzed gene expression in each of these organs separately. Since there is evidence that some, but not all, appendage patterning genes play a role in horn development [15], we included legs in our analysis in order to distinguish similarities and differences between horns and a canonical appendage. Finally, since beetle horns and legs both develop by out-budding of the epithelium, we use non-appendage bearing epithelium (dorsal abdomen) as a common reference sample (Figure 2A). The complete microarray data are available at GEO (<http://www.ncbi.nlm.nih.gov/geo/>, accession number GPL7555).

In order to examine the overall similarities and differences in the patterns of gene expression in head horns, thoracic horns and legs, we used two dimensional hierarchical clustering to cluster the expression data by both array elements (1,367 array elements with signal intensity data from all comparisons) and by samples (Figure 2B).

This revealed that expression patterns of head horns, thoracic horns, and legs are remarkably similar (similar results were obtained when the data were collapsed into non-redundant sequences, data not shown). 83% of the array elements (1,135 out of 1,367) detected enriched or depleted expression in all three tissues compared to abdominal epithelium. Despite the high degree of similarity, the expression patterns in head horns and thoracic horns are still significantly closer to each other than they are to those in legs (reflected in the branch lengths in the sample tree in Figure 2B). 11% of the array elements (150 out of 1,367) detected enriched or depleted expression in both head and thoracic horns and not legs. Thus, in terms of overall patterns of gene expression, head and thoracic pupal horn primordia are similar to but distinct from pupal leg primordia.

If primordial head horns, thoracic horns, and legs differ in terms of expressing distinct regulatory or differentiation programs, then we expected that the Gene Ontology (GO) annotations of the genes preferentially expressed in these tissues would be enriched for specific GO terms reflecting those processes. In order to focus on genes preferentially expressed in these tissues,

we collapsed the data from array elements into non-redundant sequences (contigs and singletons), filtered them for those detecting statistically significant differential expression, and categorized them according to their expression in the three tissues examined (Figure 3A). A total of 1,250 non-redundant sequences detected significant expression difference in one or more tissues (excluding 18 contigs with spots showing inconsistent expression patterns): 44 (3.5%) in head horns only, 183 (14.6%) in thoracic horns only, 93 (7.4%) in head and thoracic horns, and 345 (27.6%) in legs only. By using Blast2GO [20, 21], we tested for statistically enriched Gene Ontology biological process terms (restricted to terms above level 2) among the non-redundant sequences that detected differential expression in either individual tissues or combinations of tissues. None of these sets of sequences showed significant enrichment for GO biological process terms. While it must be noted that the number of genes assayed here is relatively small, this data indicates that, at least at the level of this analysis there are no obvious differences in the types of biological processes expressed in these tissues.

Identifying candidate genes based on expression in horn primordia

We expected that at least some genes involved in horn development would show differential expression in pupal horn primordia. In order to focus on these genes, we filtered the non-redundant sequences for those that were both statistically differentially expressed and showed at least two-fold changes in either head horns, thoracic horn or both types of horns. A total of 306 non-redundant sequences met these criteria (adjusted p-value < 0.05 and fold-change > 2); 73 in

head horns only, 38 in thoracic horns only, and 195 in both head and thoracic horns (Figure 3B).

These 306 non-redundant sequences included 94 with “no-hit” of BLASTx results to protein sequences in UniProtKB and FlyBase (E-value < 10^{-5}) and 212 that we putatively annotated based on matches to proteins from other species (Table 4). The candidate genes for the development of beetle horns can be grouped into three conceptual categories (expected, unexpected, and unknown genes).

First, we identified genes whose expression in the context of horn development could be expected given already existing insights into the developmental biology of horns on one side and knowledge about the function of these genes in other organisms on the other. For instance, the Hox gene *Sex combs reduced* (*Scr*) is enriched in thoracic horn (19.2 fold) and legs (7.9 fold) relative to abdominal epithelium (Table 4). In *Drosophila* and *Tribolium* *Scr* patterns the identity of the labial and first thoracic segment [24-26]. Preliminary results showed that *Onthophagus Scr* executes similar functions during labial and prothoracic development in addition to playing a major role in the regulation of pronotal horn development (Wasik, Rose, and Moczek, unpublished data).

Secondly, we identified genes, albeit functionally well characterized in *Drosophila* or elsewhere, nevertheless could not be readily expected to be expressed in the context of horn development.

Genes in this category include the putative ortholog of *Drosophila doublesex* (*dsx*), enriched

more than 2-fold in the head and thoracic horns relative to abdominal epithelium (Table 4). The known functions of *dsx* in *Drosophila* include the regulation of somatic sex determination and sexual organ development [27, 28], and the expression of the putative *Onthophagus* ortholog in presumptive horns was completely unexpected. Expression and functional studies are now under way to identify the role of *dsx* in the development and diversification of horns.

Similarly, we found that the putative *Onthophagus* orthologs of *yellow-c*, *-e*, and *-f* were enriched more than 2-fold in head and/or thoracic horns relative to abdominal epithelium (Table 4). The functions of yellow family genes are remarkably diverse and include the regulation of pigmentation [29, 30], the production of a major component of royal jelly in the honeybee [31] as well as expression of normal male courtship behavior in *Drosophila* [32]. Combined, these observations suggest that yellow genes may be involved in the regulation of a wide array of sex- or caste-specific functions, at least among insects, though it remains to be determined, what, if any, function the gene family may be executing in *Onthophagus* beetles.

Lastly, we identified 94 genes that were significantly differentially expressed in either head horns or thoracic horns, or both, that lack obvious homology to proteins in *Drosophila* or the UniProtKB non-redundant protein dataset. Of those 94, at least 29 (31%) contained predicted ORFs with longer than 300 nucleotides (100 codons).

DISCUSSION

Horned beetles, most notably in the genus *Onthophagus*, are increasingly being recognized as an emerging model system in evo-devo and eco-devo studies [13, 33-37]. Below we discuss the major findings of our study and their applicability to ongoing and future research efforts in horned beetles and beyond.

***Onthophagus taurus* expressed sequences as a resource**

The expressed sequences and the corresponding cDNAs presented here provide a valuable entry point for studies of gene function in *Onthophagus taurus*. The sequences derived from normalized larval and pupal cDNA libraries had a low level of redundancy. The 3,488 high quality EST sequences from both libraries assembled into 2,781 non-redundant sequences (contigs and singletons) that we estimate represent approximately 2,599 genes. This low level of redundancy resulted in a sample of sequences derived from a wide range of biological functions.

The *Onthophagus* transcriptome

This study provides a first pass survey of genes found in *Onthophagus*. Prior to this study, *Tribolium castaneum* was the only species of beetle for which comprehensive sequence information was available [6, 38]. Comparative analyses indicate that the gene repertoire of *Tribolium* is consistent with the general trends seen across sequenced insects and vertebrates [6,

39]. The *Tribolium* genome is estimated to encode approximately 16,404 genes, of which ~ 35% have orthologs in all species examined, ~ 14 % are insect specific, ~ 1 % are conserved between *Tribolium* and humans but absent from other insects, and ~23% lack similarity to genes in other sequenced organisms [6]. Our estimates of the proportions of *Onthophagus* sequences that are common to other species are consistent with those in *Tribolium*. For instance, we found that 39% of *Onthophagus* sequences had sequence matches to proteins in all the datasets searched, which is consistent with the *Tribolium* genome in which ~ 35% of genes have orthologs in all species examined [6]. Also of interest are genes that are common to beetles and vertebrates but not other insects. Previous studies have found that approximately 0.8% of *Tribolium* genes are shared with humans and not other insects (fruitfly, malaria mosquito, dengue/yellow fever mosquito, and honeybee; [6]). Similarly, we found that 1.1% of *Onthophagus* sequences (31 sequences) are shared between *Onthophagus*, *Tribolium*, and human but not *Drosophila*. Since these genes are not found in other insects, beetles may serve as useful models for functional studies of these genes. Of particular interest are the 23% of *Onthophagus* sequences that lack orthology (Group 4 in Figure 1) to proteins from six proteomes including the non-redundant dataset which is very close to the corresponding estimate of 23% of annotated *Tribolium* genes [6]. About 40% of these Group 4 *Onthophagus* sequences exhibited appreciable putative open reading frames (Supplemental Figure 1) and thus need to be considered potentially protein-coding. This group of genes likely contains genes unique to, or fast evolving in, *Onthophagus* beetles, and studies are under way to further characterize and analyze the

significance of these genes for the evolution, diversification, and radiation of horned beetles.

From ESTs to candidate genes for the evolutionary biology of beetle horns and horned beetles.

Beetle horns and horned beetles are attractive study organisms to address several current frontiers in evolutionary biology. The EST resources and array results presented here provide the first genomic resources to identify candidate genes, pathways, and networks underlying morphological, behavioral, and developmental aspects of the biology of horned beetles, as well as providing insights into their respective evolutionary histories. Below we briefly highlight two broad categories of current research efforts and how they are being advanced by the results presented here.

The origins of horns

Beetle horns have attracted attention because they lack obvious homology to other appendages or outgrowths in the insects. Horns therefore constitute an evolutionary novelty. Understanding how novel traits arise from pre-existing variation remains one of the most challenging and poorly understood questions in evolutionary biology.

One hypothesis that has been proposed toward explaining the origin of horns is based on the observation that horns share many morphological and developmental features with traditional

appendages (e.g. epidermal origin, prepupal growth, dorso-ventral axis formation, or pupal remodeling presumably via programmed cell death; [10]). Furthermore, in several other respects horns are much simpler than legs or mouthparts (e.g. they lack nerves, muscles, or joints). Horns may therefore have evolved via the large-scale co-option of genes ancestrally used to instruct appendage development. Our microarray results suggest that horns and legs are indeed highly similar in gene expression profiles, and support the hypothesis that many genes involved in leg formation may also play a role in horn development. Earlier research has already begun to implicate a small subset of appendage patterning genes in horn development (*Distal-less*, *dachshund*, *extradenticle*, *homothorax*, [11, 13, 15]). The results presented here add a substantial list of gene candidates (Table 4) that may have mediated the origin of horns via co-option from traditional appendage development.

At the same time, horn-specific transcription profiles also included genes not represented in developing legs, suggesting that horns should not be viewed solely as being simplified appendages. While this fraction of genes was small in comparison, it nevertheless highlights a possible class of genes involved in developmental processes of horn formation that are not represented, or at least not to the same degree, during the development of traditional appendages. If correct, this would suggest that the origin of horns was mediated by the co-option of appendage patterning genes alongside integration of genes and pathways unrelated to appendage formation. Combined, these putative candidate genes now offer rich opportunities for comparative genes expression and functional analyses.

The diversification of beetle horns and horned beetles

Beetle horns and horned beetles are attractive study organisms because they permit investigation of the mechanisms underlying phenotypic diversification on many interesting levels. First, species differ in the body region involved in horn expression: horns may extend from the head, thorax, or both, and while their function as weapons in male combat appears to be conserved across species, recent studies suggest that different horn types may have originated and diversified at least in part independent of one another [13]. Our results support this scenario by identifying a list of genes whose expression differs significantly across horn types such as *yellow-e* (head horns), *tailup* (encodes a LIM-homeodomain protein; thoracic horns), or *Scr* (thoracic horns and legs). While the function, if any, of these candidate genes in the context of horn development remains to be explored our results presented here provide an important starting point toward untangling shared, independent, and convergent aspects in the evolution of different horn types across horned beetles.

Substantial diversity in horn expression also exists *within* species in the form of sexual and male dimorphisms. Sexual dimorphisms are brought about via sex-specific regulation of horn growth during the prepupal stage as well as sex-specific pupal remodeling, whereas male dimorphisms are predominantly the product of nutritional differences experienced during larval life (reviewed in [40]). Endocrine factors such as juvenile hormone (JH) are likely to play important roles in

the regulation of both types of diversity [41-43]. Furthermore, the same nutritional or hormonal manipulations affect sexual and male dimorphisms differently in different species and populations, suggesting that evolutionary changes in the interplay between endocrine factors, nutrition, and sexual differentiation have contributed to the diversification of horned beetles [43, 44]. Our EST resources and array results provide an important starting point to begin exploring putative candidate genes that may be associated with sex-specific (such as *doublesex*, *transformer-2* or members of the *yellow* gene family) or nutrition-dependent (e.g. *foxo*) expression of horns. Moreover, the resources presented here should support the development of experiments towards characterizing sex- and morph-specific transcriptomes in *O. taurus* and closely related species in the genus.

In conclusion, the EST resources and microarray results present here provide a first step toward a systematic analysis of the molecular basis of horn development and diversification in beetles with the potential to inform several major frontiers in evolutionary developmental biology.

METHODS

cDNA library construction

Adult *Onthophagus taurus* were collected from pastures near Bloomington, IN and reared as described previously [15]. We constructed two separate libraries from larval and pupal stages. For the larval library we dissected heads and thoraces from mid third instar larvae, late third instar larvae, and early and late prepupal stages. For the pupal library tissues included whole individuals one, two, three and four days after pupation. For both libraries we harvested at least two individuals for each stage and sex, and all samples were frozen in liquid nitrogen, immediately transferred to -80°C for storage until RNA extraction. Total RNA was extracted using TRIreagent (Sigma, MO), precipitated with ethanol and stored at -80°C. The normalized cDNA libraries were each constructed from 1µg of total RNA, using the TRIMMER-DIRECT cDNA normalization kit (Evrogen, Moscow, Russia) for the library normalization, followed by the Creator SMART cDNA library construction kit (Clontech, CA) for cDNA library construction, as described in Zhulidov *et al.* 2004 [45]. We followed the manufacturers protocols with the following modifications and specific conditions: (i) the cycle conditions for the PCR-based double-strand cDNA synthesis were 16 cycles of [95°C for 7 sec, 66°C for 30 sec, and 72°C for 6 min]; (ii) we used 2µl of cDNA mixture for PCR during cDNA library construction and normalization; and (iii) the conditions for the two step amplification of the normalized cDNA were 18 cycles [95°C for 7 sec, 66°C for 30 sec, and 72°C for 6 min] for the first step, and the second amplification was cycled for 12 cycles using the same conditions. Normalized and

amplified cDNA fragments were size-fractionated, digested by *Sfi* I, and ligated with the plasmid vector pDNR-LIB according to manufacturer's instruction. Electro-transformed *E. coli* cells were spread on LB plate containing chloramphenicol (final concentration of 30µg/ml). The estimated titer of both of the libraries were $\sim 1 \times 10^8$. A total of 3,756 colonies were picked at random. Unless stated otherwise standard molecular procedures were used to execute basic molecular analyses [46].

EST sequencing

DNA samples were prepared for sequencing using a Beckman Coulter Biomek FX Laboratory Automation Workstation as described in Burr *et al.* 2006 [47]. Each clone was incubated with 100µl of SOC medium with chloramphenicol (final concentration of 30µg/ml) in 96-well plates overnight at 37°C. 20µl of the cultured cells were mixed with 80µl of water and heat-punctured at 95°C for 10 min. Inserted DNA was PCR-amplified using the standard reactions conditions with 10µl of the cell lysate, the M13fw (5'-GTG TAA AAC GAC GGC CAG TAG-3') and M13rev (5'-AAA CAG CTA TGA CCA TGT TCA C-3') primers, and cycled as follows: 95°C for 5 min then 35 cycles of [95°C for 1 min, 54°C for 1min, and 2 min at 72°C]. The amplified DNA was purified using the Multiscreen-PCR 96-well purification system (Millipore). The purified DNA was subjected to agarose gel electrophoresis against molecular weight standard and visualized using a Kodak 440cf imaging station. Sequencing reactions were performed with the primer pDNRlib30-50 (5'-TAT ACG AAG TTA TCA GTC GAC G-3') and ABI BigDye

chemistry and ABI Prism 3730 sequencer (Applied BioSystems, CA).

EST processing, assembly, and annotation

ESTPiper [17] was used to analyze EST sequences including base calling, data cleaning, de novo assembly, and annotation. Totally 3,756 EST sequences were generated in FASTA format with quality scores after base calling. For data cleaning, ESTPiper first removed low quality and vector sequences using LUCY [48] program with the default parameter settings. PolyA/T tails were then trimmed, where within 50 bp searching range from both ends of the sequences, the minimum length of continuous polyA/T region was set to 9 bp and the maximum number of mismatches within the polyA/T region was set to 3. The potential chimerical clones, which were defined as sequences with at least 30 bp continuous A/T or adaptors occurring in the middle of sequences, were removed. Finally, shorter sequences (< 100 bp) were also removed. A total of 3488 high quality sequences passed data cleaning procedure. We then performed de novo assembly to assemble EST sequences into contigs and singletons. Parameters were set as follows: (i) overlap percent identity cutoff is 95%, (ii) overlap length cutoff is 49, and (iii) maximum number of word matches is 10,000. For annotation, ESTPiper matched contigs/singletons to UniProt database [49] using BLASTX with an E-value cutoff of 1×10^{-5} and only the top match is taken.

Gene Ontology (GO) analysis

The GO terms of the EST sequences were obtained by using Blast2GO (<http://blast2go.bioinfo.cipf.es/>, [20, 21]) with the BLASTx E-value cutoff of 1×10^{-5} and GO term annotation cutoff of 55. In order to analyze the GO term enrichment, we performed Fisher's exact test on the program with FDR cutoff value of 0.05 (refer the manual of Blast2GO for more details about the parameters).

Identifying Open Reading Frames (ORF)

Predictions of open reading frames (ORF) were obtained by using the getorf program from the EMBOSS software suite [50]. All potential ORFs were predicted for the 2,781 assembled EST sequences. The longest predicted ORF for each sequence were used in all analyses.

Microarray printing

The clones analyzed for ESTs were used to print microarrays. We followed the protocol of Indiana University *Drosophila* Genomics Resource Center (Andrews et al. 2006, http://cgb.indiana.edu/publications/alias/dgrc1_prod_protocol/version/current) to print microarrays with a minor revision to post-print washes. Cleaned clones were dried completely, re-dissolved with DGRC spotting solution (1.5M Betaine in 3×SSC), and spotted on GAPSII Microarray Slides (Corning). Using an OmniGrid 300 printing robot we printed 100 microarrays, which were then heated at 85°C for 3hrs and rinsed with 5×SSC/0.1%SDS (55°C), water (twice at RT, once at 95°C, and once again at RT) and then centrifuged to dry. All microarrays were

kept dry at room temperature until use.

Target RNA preparation, hybridization and obtaining data sets

Head horns, thoracic horn, legs, and dorsal abdominal epithelium from 20 male *O. taurus* (day 1 pupae) were collected from our lab colony (also see figure 2A). Dissection and RNA extraction were conducted separately for each sample (RNeasy Mini kit, Qiagen, CA). RNA samples from the same tissue types were combined, resulting in five RNA sample sets (i.e. total RNA from four head horns, four thoracic horns, four sets of legs, and four abdominal epithelium samples). 1 μ g of RNA was reverse transcribed and amplified for each sample set. First strand synthesis was conducted using Oligo(dT)-T7 primer (Ambion) and SuperScriptIII reverse transcriptase (Invitrogen). DNA polymerase and RNase H (Invitrogen) were used for second strand synthesis. After *in vitro* transcription by MEGAscript kit (Ambion), amplified RNA samples were labeled with two kinds of fluorescent dyes (Cy3 and Cy5). Three sets of mixed, amplified RNA samples from head horns, thoracic horns, and legs were labeled with Cy5, while abdominal epithelial tissue samples were labeled with Cy3. The remaining two sets of samples were labeled in the opposite way. We used ULS aRNA Fluorescent Labeling Kit (KREATECH, Amsterdam, The Netherlands) for labeling reactions. After measuring the quantity and labeling efficiency, amplified and labeled RNA samples from test (head horns, thoracic horn, and legs) samples and abdomen (reference sample) were mixed and hybridized onto arrays. aRNA with 50pmol dye from the test sample and reference sample were mixed with KREAblock (ULS aRNA

Fluorescent Labeling kit) and 2×enhanced cDNA hybridization buffer (Genisphere), then heated at 80°C for 10 min. Arrays were pre-treated for more than one hour at 55°C in pre-hybridization buffer (5×SSC, 0.1%SDS, 1% I-block (Applied Biosystems, CA)). Both mixed sample and microarray were kept at 55°C until the hybridization step. Hybridization was performed in a dark humidified chamber at 55°C overnight. The microarray was rinsed in buffer A (2×SSC/0.2%SDS) at 55°C then incubated in buffer A at 65°C for 10min. The microarray was transferred to 2×SSC (room temperature) for 10 min, followed by incubation in 0.2% SSC for 10min at room temperature. The rinsed microarray was dried by centrifuging at 500rcf for 4 min. The hybridized microarrays were scanned by GenePix scanner 4200 (Molecular Devices) to obtain raw data sets. After initial quality check of results using OLIN in Bioconductor (Basic Hybridization Analysis, Costello et al. 2005, <https://dgrc.cgb.indiana.edu/microarrays/support/bha.html>), differential expression was assessed using Limma. The values for each spot were shown as log₂ ratios between the two signal intensities (M-values). The microarray results were submitted to NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>). The accession number of the platform is GPL7555, which contains all microarray sample datasets.

AUTHORS' CONTRIBUTIONS

TK, APM, and JA designed the study. TK developed cDNA libraries and microarrays. TK, JC, ZT and JA analyzed libraries. TK performed microarray experiments. TK, APM, and JA

analyzed the microarray results. TK, JC, ZT, APM and JA wrote the paper. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This manuscript benefited greatly from comments by Amy Cash, Emilie Snell-Rood, and two anonymous reviewers. We thank the Center for Genomics and Bioinformatics at Indiana University and its staff, especially John Colborne, Zhao Lai, Zach Smith, Jade Buchanan-Carter, and Heejung Yang for their advice and expertise in executing this study. Erin Yoder and Sarah Jones provided expert beetle care. Funding for this study, as well as for work in the Center for Genomics and Bioinformatics, was provided in part by the METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment. Additional support was provided by National Science Foundation grants IOS 0820411 to JA and APM.

TABLES

Table 1: Summary of cDNA libraries and EST sequence analysis.

	Larvae (OtL)	Pupae (OtP)	Mixed (OtL and P)
median cDNA fragment size (nt)	715	975	-
average read length (nt)	624	667	645
raw reads	1874	1882	3756
cleaned reads	1783	1705	3488
assembled contigs	217	171	451
singletons	1252	1284	2330
independent sequences	1469	1455	2781
annotated	1104	998	1984

Table 2: Datasets used in the present study.

Genome	Database	Version (date uploaded, YYMMDD)
<i>Drosophila melanogaster</i>	FlyBase	5.2 (070725)
<i>Tribolium castaneum</i>	NCBI	(060410)
<i>Homo sapiens</i>	Ensembl	NCBI 36 release 46 (070803)
<i>Caenorhabditis elegans</i>	Ensembl	Wormpep 180 (070819)
invertebrate	NCBI	(080508)
non-redundant	NCBI	(080514)

Table 3: Examples of putative *Onthophagus taurus* orthologs with known functions in insect development and physiology obtained through random sequencing of larval (OtL) and pupal (OtP) EST libraries (¹contigs consisting of larval ESTs only; ²contigs consisting of pupal ESTs only; ³contigs consisting of larval and pupal ESTs).

Table 4: List of genes exhibiting differential expression in horn- or leg primordial epidermis compared to abdominal epithelium.

Listed are genes exhibiting significant ($p < 0.05$) and at least 2-fold difference in expression intensity in primordial head horns and/or thoracic horns and/or legs. Shown are spot- or contig IDs (if more than one spot within the same contig was detected), tissue type, fold-differences in expression intensity relative to abdominal epithelium, UniProtKB and FlyBase gene description, amino acid sequence identity, E-values, and bit score indicating sequence similarity. Negative values imply depletion of transcript abundance in horns or leg primordia relative to abdominal epithelium. All fold differences were obtained through the analysis of a single spot that indicated the strongest expression difference for a given contig except for contig95 where two spots showed significant and > 2-fold depletion in thoracic horn and leg primordia. Similarity values were obtained from FlyBase, and if not match could be obtained, from UniProtKB.

Abbreviations: H: head horns, T: thoracic horn, HT: head and thoracic horns, HL: head horns and legs, TL: thoracic horn and legs, HTL: head horns, thoracic horn, and legs. Numbers in the misc. column indicate spots or contigs that share the same FlyBase and/or UniProtKB gene description but exhibit distinct sequence differences such that they are unlikely to belong to the same contig.

FIGURE LEGENDS

Figure 1: Overall comparison of *O. taurus* sequences with other protein datasets.

Filtering and clustering analysis of assembled *O. taurus* ESTs based on BLASTx. Shown are bit scores against protein sequences from *Tribolium castaneum* (Tc, NCBI), *Drosophila melanogaster* (Dm, FlyBase), *Caenorhabditis elegans* (Ce, Sanger), invertebrate proteins (inv., NCBI), *Homo sapiens* (Hs, Ensembl), and non-redundant protein dataset (nr, NCBI). Each row represents a single *Onthophagus* sequence, and each column represents sequence matches to proteins from the indicated datasets, where the color intensity is proportional to the bit score (0 = black to 789 = brightest red). The *Onthophagus* sequences are grouped (Groups 1-4) according to the patterns of BLASTx sequence matches with proteins in the various datasets (E-value cut-off = 1×10^{-5}), and clustered according to the bit scores.

Figure 2: Experimental design and clustering analysis of the gene expression pattern in *O. taurus* day 1 pupa.

A. Head horns, thoracic horns, legs, and abdominal epithelium were dissected from *O. taurus* day 1 pupae. Total RNA samples were extracted, amplified, and labeled followed by hybridization on microarrays and scanning. See Materials and Methods for details. **B.** 1,367 spots were clustered based on their M-values when compared to abdominal epithelium. Each row represents a single spot and each column represents the sample. Relative magnitude of gene expression level is indicated by color brightness; red indicates expression relatively stronger

compared to abdominal epithelium whereas green indicates depletion relative to abdominal epithelium. M-values ranged from -4.85 to 4.12. Bootstrap values were obtained after 5000 trials. Branch lengths represent relative distances between the samples. H: head horns, T: thoracic horns, and L: legs.

Figure 3: Expression differences between horn and leg primordia relative to abdominal epithelium. **A.** Distribution of genes that exhibited significantly different expression relative to abdominal epithelium regardless of fold-differences (red = enriched, blue = depleted relative to abdominal epithelium; purple = genes showed changes in the direction of expression differences, i.e. enriched in head horns but depleted in thoracic horn). **B.** Distribution of genes that exhibited significantly different expression relative to abdominal epithelium *and* > 2 -fold difference in expression levels. Color indications are the same as in **A**.

ADDITIONAL FILES

Supplemental Table 1: Summary of BLASTx results of *Onthophagus taurus* ESTs against UniProtKB.

Supplemental Table 2: Summary of BLASTx results of *Onthophagus taurus* ESTs against FlyBase.

Supplemental Figure 1: Size distribution of *O. taurus* sequences.

The size distribution of all 2,781 non-redundant sequences. Indicated by the shaded bars, there are 2,148 sequences that match at least one of the query proteomes; *Tribolium castaneum*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Homo sapiens*, invertebrate, and nr (BLASTx E-value cutoff = 1×10^{-5}). As indicated by the red bars, there were 633 sequences that did not match any of the query proteomes. (A) Distribution of the sequence lengths in nucleotides. (B) Distribution of the longest predicted open reading frame for each sequence shown in nucleotides.

References

1. Raff RA: **The Shape of Life: Genes, Development, and the Evolution of Animal Form**. Chicago: University of Chicago Press; 1996.
2. West-Eberhard MJ: **Developmental Plasticity and Evolution**: Oxford Univ Press; 2003.
3. Beldade P, Rudd S, Gruber JD, Long AD: **A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model**. *BMC Genomics* 2006, **7**:130.
4. Kapan DD, Flanagan NS, Tobler A, Papa R, Reed RD, Gonzalez JA, Restrepo MR, Martinez L, Maldonado K, Ritschoff C, Heckel DG, McMillan WO: **Localization of Mullerian mimicry genes on a dense linkage map of *Heliconius erato***. *Genetics* 2006, **173**(2):735-757.
5. Robinson GE, Fernald RD, Clayton DF: **Genes and social behavior**. *Science* 2008, **322**(5903):896-900.
6. Tribolium Genome Sequencing Consortium: **The genome of the model beetle and pest *Tribolium castaneum***. *Nature* 2008, **452**(7190):949-955.
7. Moczek AP: **The Evolution and Development of Novel Traits, or How Beetles Got Their Horns**. *BioScience* 2005, **55**(11):937-951.
8. Balthasar V: **Monographie der Scarabaeidae und Aphodiidae der palaearktischen und orientalischen Region (Coleoptera: Lamellicornia. Band 2, Coprinae**. Prag: Verlag der Tschechoslowakischen Akademie der Wissenschaften; 1963.

9. Matthews E: **A revision of the Scarabaeinae dung beetles of Australia. I. Tribe Onthophagini.** *Aus J Zool* 1972, (suppl) **9**:1-330.
10. Moczek AP: **Integrating micro- and macroevolution of development through the study of horned beetles.** *Heredity* 2006, **97**(3):168-178.
11. Moczek AP, Rose D, Sewell W, Kesselring BR: **Conservation, innovation, and the evolution of horned beetle diversity.** *Dev Genes Evol* 2006, **216**(11):655-665.
12. Emlen DJ, Corley Lavine L, Ewen-Campen B: **On the origin and evolutionary diversification of beetle horns.** *Proc Natl Acad Sci U S A* 2007, **104 Suppl 1**:8661-8668.
13. Emlen DJ, Szafran Q, Corley LS, Dworkin I: **Insulin signaling and limb-patterning: candidate pathways for the origin and evolutionary diversification of beetle 'horns'.** *Heredity* 2006, **97**(3):179-191.
14. Moczek AP: **Pupal remodeling and the development and evolution of sexual dimorphism in horned beetles.** *Am Nat* 2006, **168**(6):711-729.
15. Moczek AP, Nagy LM: **Diverse developmental mechanisms contribute to different levels of diversity in horned beetles.** *Evol Dev* 2005, **7**(3):175-185.
16. Moczek AP, Cruickshank TE, Shelby A: **When ontogeny reveals what phylogeny hides: gain and loss of horns during development and evolution of horned beetles.** *Evolution* 2006, **60**(11):2329-2341.
17. Tang Z, Choi J, Hemmerich C, Sarangi A, Colbourne JK, Dong Q: **ESTPiper - a web-based analysis pipeline for expressed sequence tags.** *BMC Genomics*, in press.

18. Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP: **Gene expression during the life cycle of *Drosophila melanogaster***. *Science* 2002, **297**(5590):2270-2275.
19. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene Ontology: tool for the unification of biology**. *Nat Genet* 2000, **25**(1):25-29.
20. Conesa A, Gotz S: **Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics**. *Int J Plant Genomics* 2008, **2008**:619832.
21. 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* 2005, **21**(18):3674-3676.
22. Beermann A, Schroder R: **Functional stability of the aristaless gene in appendage tip formation during evolution**. *Dev Genes Evol* 2004, **214**(6):303-308.
23. Tomoyasu Y, Wheeler SR, Denell RE: **Ultrabithorax is required for membranous wing identity in the beetle *Tribolium castaneum***. *Nature* 2005, **433**(7026):643-647.
24. Lewis RA, Kaufman TC, Denell RE, Tollerico P: **Genetic Analysis of the Antennapedia Gene Complex (Ant-C) and Adjacent Chromosomal Regions of *DROSOPHILA MELANOGASTER*. I. Polytene Chromosome Segments 84b-D**. *Genetics* 1980, **95**(2):367-381.
25. Lewis RA, Wakimoto BT, Denell RE, Kaufman TC: **Genetic Analysis of the**

Antennapedia Gene Complex (Ant-C) and Adjacent Chromosomal Regions of DROSOPHILA MELANOGASTER. II. Polytene Chromosome Segments 84A-84B1,2. *Genetics* 1980, **95(2):383-397.**

26. Beeman RW, Stuart JJ, Brown SJ, Denell RE: **Structure and function of the homeotic gene complex (HOM-C) in the beetle, *Tribolium castaneum*.** *Bioessays* 1993, **15**(7):439-444.
27. Burtis KC, Baker BS: ***Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides.** *Cell* 1989, **56**(6):997-1010.
28. Saccone G, Salvemini M, Pane A, Polito LC: **Masculinization of XX *Drosophila* transgenic flies expressing the *Ceratitis capitata* DoublesexM isoform.** *Int J Dev Biol* 2008, **52**(8):1051-1057.
29. Han Q, Fang J, Ding H, Johnson JK, Christensen BM, Li J: **Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes.** *Biochem J* 2002, **368**(Pt 1):333-340.
30. Nash WG, Yarkin RJ: **Genetic regulation and pattern formation: a study of the yellow locus in *Drosophila melanogaster*.** *Genet Res* 1974, **24**(1):19-26.
31. Malecova B, Ramser J, O'Brien JK, Janitz M, Judova J, Lehrach H, Simuth J: **Honeybee (*Apis mellifera* L.) mrjp gene family: computational analysis of putative promoters and genomic structure of mrjp1, the gene coding for the most abundant protein of**

- larval food.** *Gene* 2003, **303**:165-175.
32. Drapeau MD, Radovic A, Wittkopp PJ, Long AD: **A gene necessary for normal male courtship, yellow, acts downstream of fruitless in the *Drosophila melanogaster* larval brain.** *J Neurobiol* 2003, **55**(1):53-72.
33. Brakefield PM, French V: **Evo-devo focus issue.** *Heredity* 2006, **97**(3):137-138.
34. Emlen DJ: **Integrating Development with Evolution: A Case Study with Beetle Horns.** *BioScience* 2000, **50**(5):403-418.
35. Jenner RA, Wills MA: **The choice of model organisms in evo-devo.** *Nat Rev Genet* 2007, **8**(4):311-319.
36. Monteiro A, Podlaha O: **Wings, horns, and butterfly eyespots: how do complex traits evolve?** *PLoS Biol* 2009, **7**(2):e37.
37. Shubin N, Tabin C, Carroll S: **Deep homology and the origins of evolutionary novelty.** *Nature* 2009, **457**(7231):818-823.
38. Park Y, Aikins J, Wang LJ, Beeman RW, Oppert B, Lord JC, Brown SJ, Lorenzen MD, Richards S, Weinstock GM, Gibbs RA: **Analysis of transcriptome data in the red flour beetle, *Tribolium castaneum*.** *Insect Biochem Mol Biol* 2008, **38**(4):380-386.
39. Wyder S, Kriventseva EV, Schroder R, Kadowaki T, Zdobnov EM: **Quantification of ortholog losses in insects and vertebrates.** *Genome Biol* 2007, **8**(11):R242.
40. Moczek A: **Developmental plasticity and the origins of diversity: a case study on horned beetles.** In *Phenotypic Plasticity of Insects: Mechanisms and Consequence.*

Edited by Whitman DW, Ananthakrishnan TN. Plymouth, UK: Science Publishers Inc.;
2009:81-134.

41. Emlen DJ, Nijhout HF: **Hormonal control of male horn length dimorphism in the dung beetle *Onthophagus taurus* (Coleoptera: Scarabaeidae).** *J Insect Physiol* 1999, **45**(1):45-53.
42. Emlen DJ, Nijhout HF: **Hormonal control of male horn length dimorphism in *Onthophagus taurus* (Coleoptera: Scarabaeidae): a second critical period of sensitivity to juvenile hormone.** *J Insect Physiol* 2001, **47**(9):1045-1054.
43. Shelby JA, Madewell R, Moczek AP: **Juvenile hormone mediates sexual dimorphism in horned beetles.** *J Exp Zool B Mol Dev Evol* 2007, **308**(4):417-427.
44. Moczek AP, Nijhout HF: **Developmental mechanisms of threshold evolution in a polyphenic beetle.** *Evol Dev* 2002, **4**(4):252-264.
45. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA *et al*: **Simple cDNA normalization using kamchatka crab duplex-specific nuclease.** *Nucleic Acids Res* 2004, **32**(3):e37.
46. Sambrook J, Russell DW: **Molecular Cloning**, Third Edition. Cold Spring Harbor: Cold Spring Harbor Press; 2001.
47. Burr A, Bogart K, Conaty J, Andrews J: **Automated liquid handling and high-throughput preparation of polymerase chain reaction-amplified DNA for**

- microarray fabrication.** *Methods Enzymol* 2006, **410**:99-120.
48. Chou HH, Holmes MH: **DNA sequence quality trimming and vector removal.** *Bioinformatics* 2001, **17**(12):1093-1104.
49. UniProt Consortium: **The Universal Protein Resource (UniProt).** *Nucleic Acids Res* 2007, **35**(Database issue):D193-197.
50. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**(6):276-277.

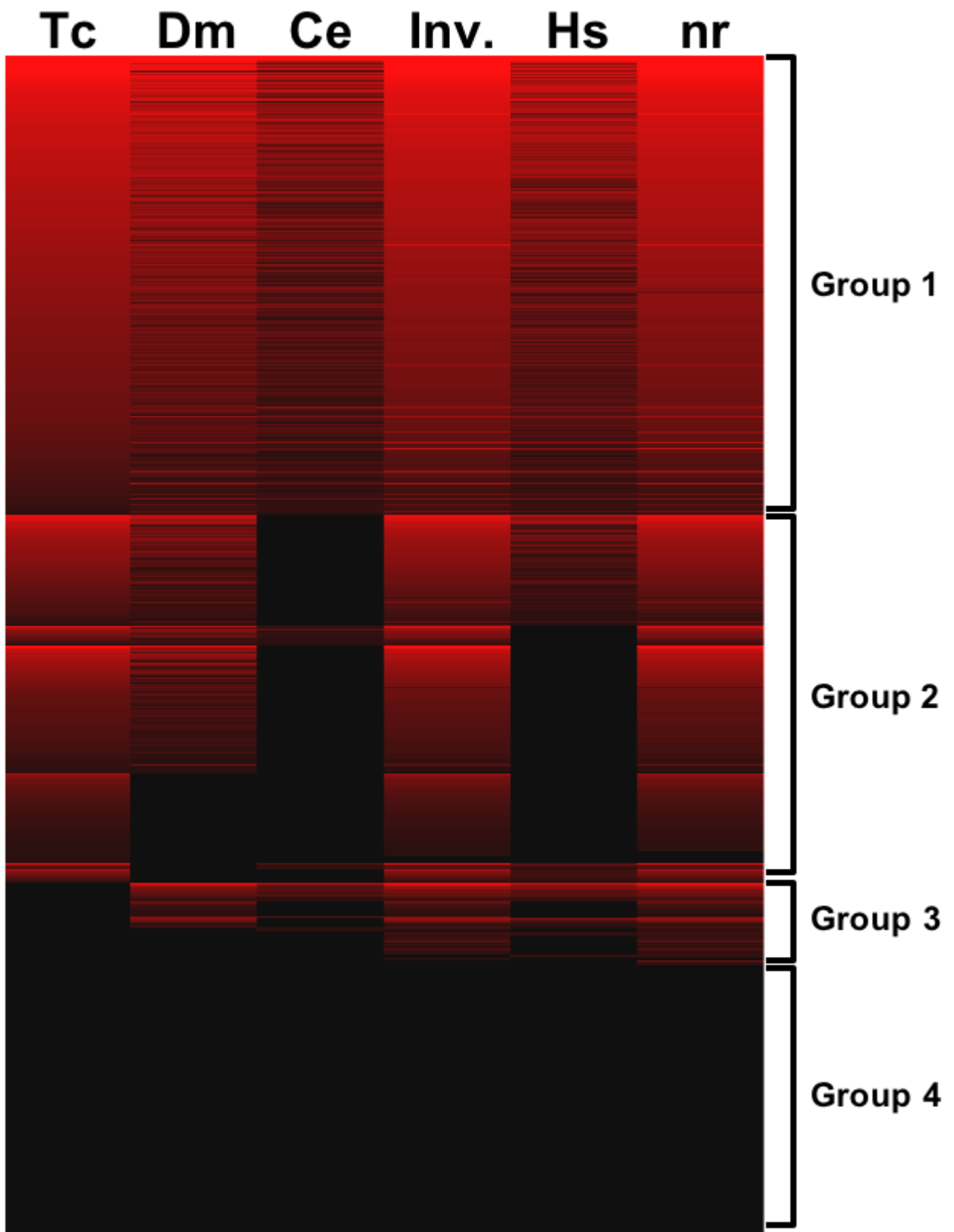


Figure 1

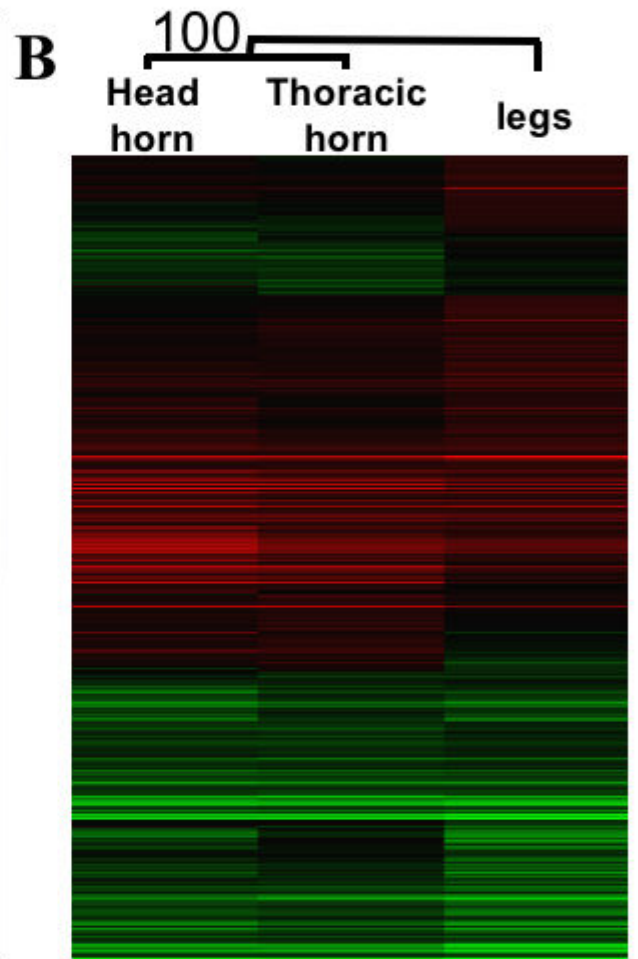
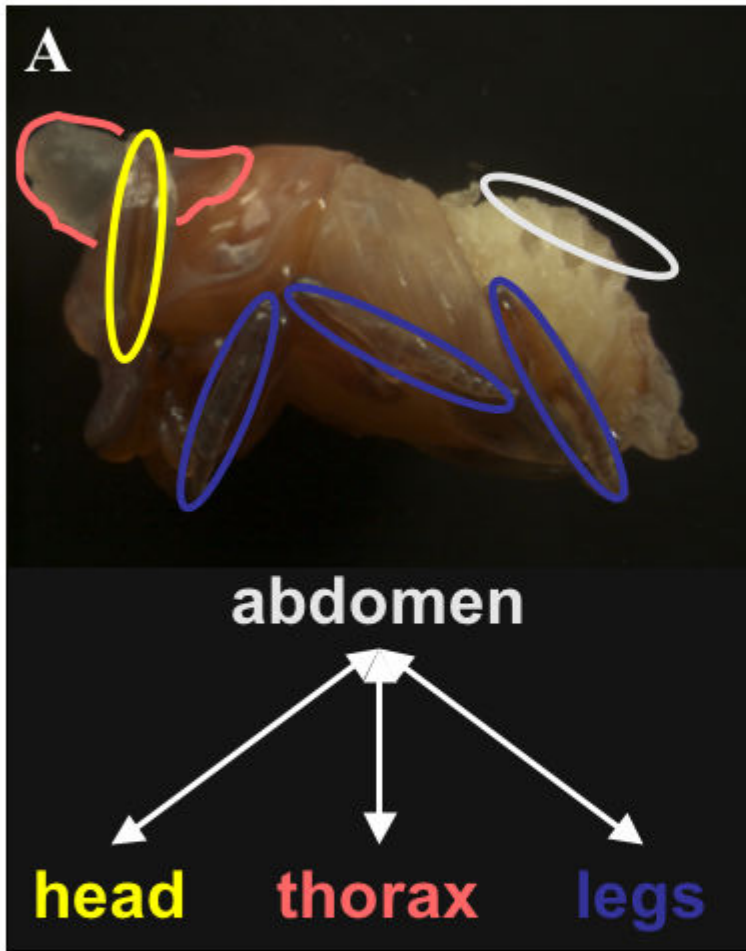
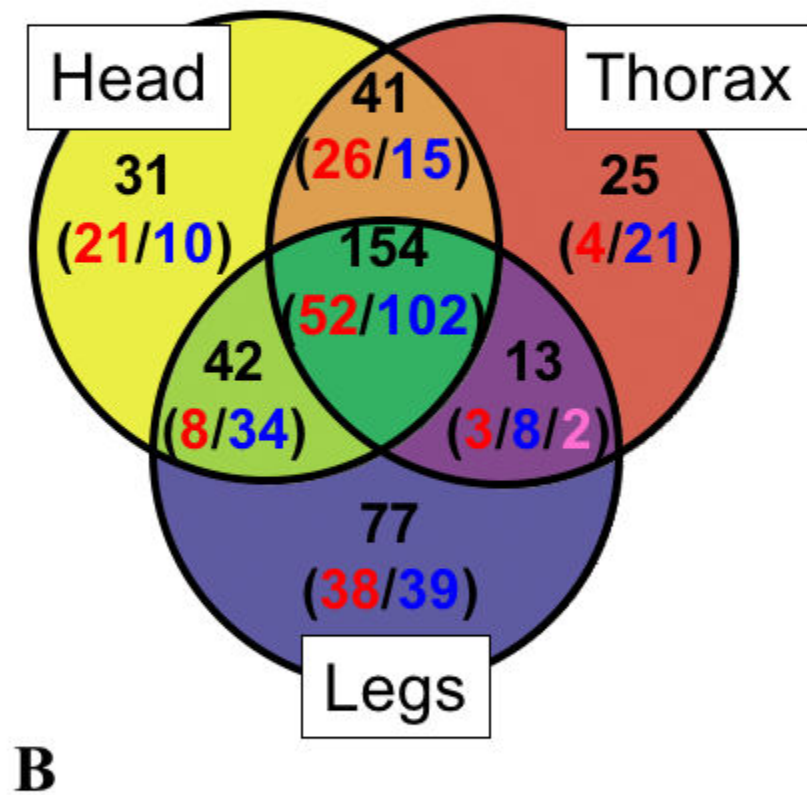
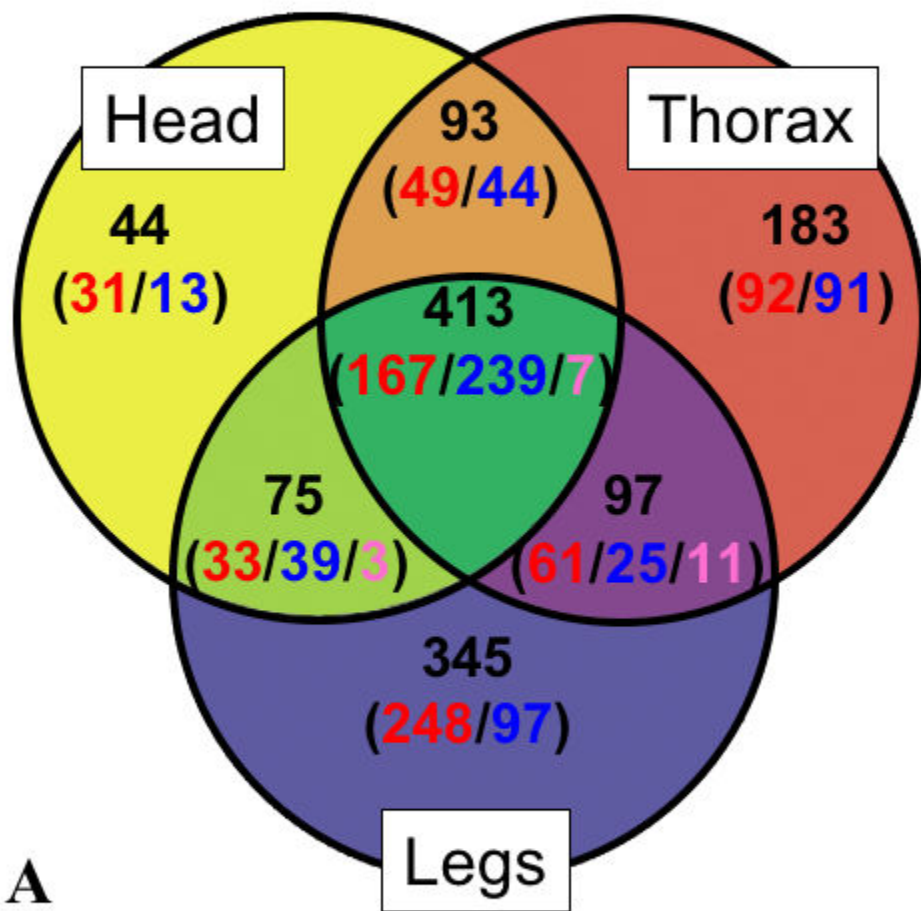


Figure 2



Additional files provided with this submission:

Additional file 1: table3.xls, 33K

<http://www.biomedcentral.com/imedia/9540379542638029/supp1.xls>

Additional file 2: table4arrayresults.xls, 69K

<http://www.biomedcentral.com/imedia/1138750229263802/supp2.xls>

Additional file 3: suppl1bigtablenr.xls, 316K

<http://www.biomedcentral.com/imedia/1813907250263802/supp3.xls>

Additional file 4: suppl2bigtabledm.xls, 473K

<http://www.biomedcentral.com/imedia/1022101838263802/supp4.xls>

Additional file 5: supplfig1.png, 10K

<http://www.biomedcentral.com/imedia/2144098837263802/supp5.png>