Research

Caste- and development-associated gene expression in a lower termite

Michael E Scharf, Dancia Wu-Scharf, Barry R Pittendrigh and Gary W Bennett

Address: Department of Entomology, Purdue University, 901 W State Street, West Lafayette, IN 47907-2089, USA.

Correspondence: Michael E Scharf. E-mail: mike_scharf@entm.purdue.edu

Abstract

Background: Social insects such as termites express dramatic polyphenism (the occurrence of multiple forms in a species on the basis of differential gene expression) both in association with caste differentiation and between castes after differentiation. We have used cDNA macroarrays to compare gene expression between polyphenic castes and intermediary developmental stages of the termite *Reticulitermes flavipes*.

Results: We identified differentially expressed genes from nine ontogenic categories. Quantitative PCR was used to quantify precise differences in gene expression between castes and between intermediary developmental stages. We found worker and nymph-biased expression of transcripts encoding termite and endosymbiont cellulases; presoldier-biased expression of transcripts encoding the storage/hormone-binding protein vitellogenin; and soldier-biased expression of gene transcripts encoding two transcription/translation factors, two signal transduction factors and four cytoskeletal/muscle proteins. The two transcription/translation factors showed significant homology to the *bicaudal* and *bric-a-brac* developmental genes of *Drosophila*.

Conclusions: Our results show differential expression of regulatory, structural and enzyme-coding genes in association with termite castes and their developmental precursor stages. They also provide the first glimpse into how insect endosymbiont cellulase gene expression can vary in association with the caste of a host. These findings shed light on molecular processes associated with termite biology, polyphenism, caste differentiation and development and highlight potentially interesting variations in developmental themes between termites, other insects, and higher animals.

Background

The lower termites possess one of the most ancient forms of insect eusociality, and a social organization that has arisen independently from that of the eusocial Hymenoptera [1,2]. Lower termites of the genus *Reticulitermes* have a worldwide distribution, are the most common termites in the United States and Europe, and are extremely important from both an economic and an ecological perspective [3]. Lower termites, including *Reticulitermes*, undergo incomplete (hemimetabolous) metamorphosis and a caste-differentiation process that is unparalleled among the social insects (Figure 1) [4-6].

Caste differentiation and development in *Reticulitermes* termites can follow imaginal (winged) and apterous (wingless)
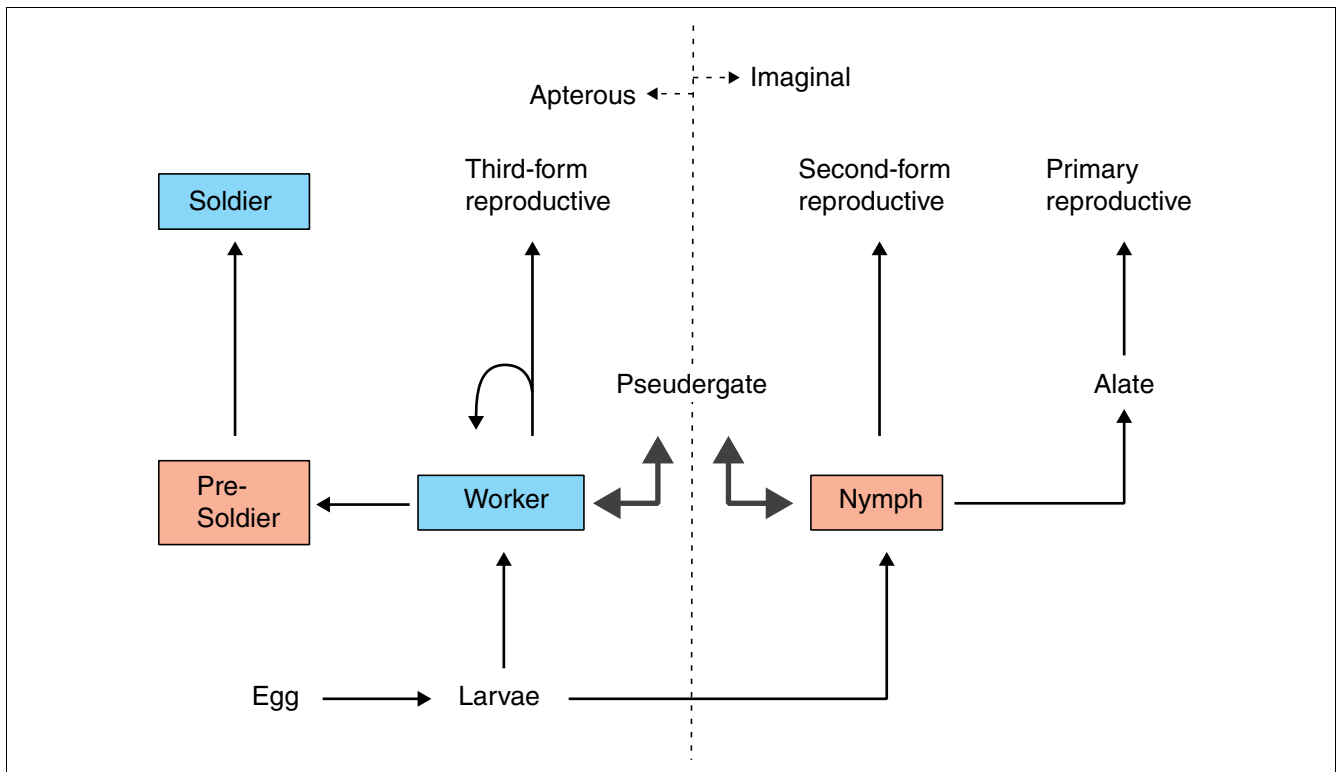
Figure 1
Experimental design and diagram of caste differentiation and development in *Reticulitermes* termites. Castes enclosed in blue boxes (soldier, worker) were compared initially by using cDNA macroarrays prepared from all stages shown except adult reproductives. Next, quantitative PCR was used to compare expression levels of specific gene transcripts between castes enclosed in blue boxes (soldier, worker) versus intermediary developmental stages enclosed in red boxes (nymph, presoldier).

routes (Figure 1). All forms except alates and the three reproductive forms are considered immature in lower termites. The initial developmental branch point occurs in the last larval instar, where either workers or nymphs differentiate. Nymphs have three developmental potentials: regressive molts into worker-like pseudergates (false-workers) which retain the capacity to redifferentiate into nymphs; differentiation into fully winged and eyed adult alates that disperse, mate, and become primary reproductives which head colonies; or differentiation into rudimentary-winged and eyed, non-dispersive second-form reproductives that serve to supplement the reproductive capacity within colonies. Workers have three developmental potentials: *status quo* molts in which they remain as workers; differentiation into the single-instar presoldier stage (followed immediately by molting into the soldier stage); or differentiation into apterous and eyeless third-form reproductives that assume identical functions as second-form reproductives.

Caste differentiation and development in termites has been studied extensively at the organismal level, yet many basic questions remain unanswered [6]. Furthermore, virtually nothing is known about the molecular mechanisms that underlie termite differentiation and development. As with honey-bees and other social insects [7-9], it is logical to assume that termite polyphenism and associated developmental processes occur via differential gene expression [10,11]. Our objective was to use an array-based genomic approach to gain a better understanding of molecular processes associated with termite biology and development through the identification of differentially expressed genes. Here, we report the results of our efforts to identify and characterize levels of gene expression between castes and developmental stages of the most common termite in the United States, *R. flavipes*. Our results provide novel insights into previously uncharacterized aspects of termite polyphenism and caste-associated biology. More generally, these results also demonstrate the applicability of array-based genomic investigations for the study of processes related to developmental polyphenism, as well as relationships between organisms and their endosymbiotic fauna.

Results
Array screens
Macroarrays were probed separately with digoxigenin (DIG)-labeled mRNAs from worker or soldier *R. flavipes* (Figure 2a). Initially, 153 cDNAs with either clear or marginal
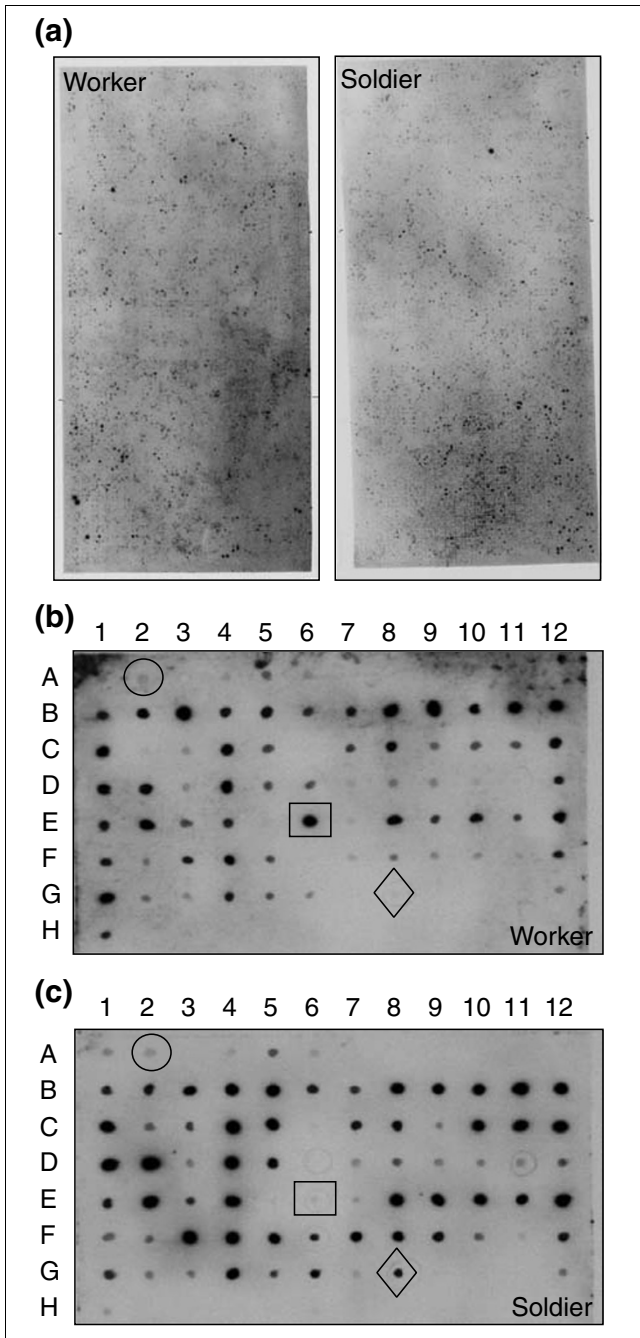
Figure 2
Representative macroarray and dot blot results. **(a)** Duplicate macroarrays were probed separately with mRNAs that were labeled with digoxigenin (DIG) after isolation from worker or soldier *R. flavipes*. Shown are representative autoradiograms from arrays probed with worker (left) or soldier (right) mRNAs. **(b-c)** Plasmid DNA was obtained from array-positive cultures and spotted on duplicate membranes in equal quantities. The resulting dot blots were probed with aliquots of the same DIG-labeled worker or soldier mRNA pools that were used to probe the array membranes originally. The secondary dot blots verified elevated transcript abundance in approximately 50% of instances. Examples of worker- and soldier-associated cDNAs are in positions noted by squares and diamonds, respectively, while a control gene with similar expression (actin-related protein 66B) is circled.

differential expression between soldiers and workers were identified in the array screens. These clones were the basis of further analysis. Clones corresponding to each of the 153 positive array positions were sampled from frozen cultures in 384-well plates and grown as liquid cultures. Plasmid DNA was obtained from each culture and spotted on duplicate membranes in equal quantities. The resulting dot blots were probed with aliquots of the same DIG-labeled worker or soldier mRNA pools that were used to originally probe the array membranes. The secondary dot blots verified elevated transcript abundance in around 50% of instances (Figures 2b,c). In the dot blots shown in Figure 2, examples of worker- and soldier-specific cDNAs are in positions noted by squares and diamonds, respectively.

Dot blots were also used to determine the expression of two potential genes with similar expression levels between castes: these encoded NADH dehydrogenase and actin-related protein 66B (ARP 66B). These two transcripts were identified previously from expressed sequencing tag (EST)-sequencing studies [12] and were used as candidate control genes for expression. Similar expression levels of the two control genes can be seen in the upper left corners of the dot blots, with ARP 66B being circled (Figure 2b,c). In subsequent experiments it was determined that expression levels of these two genes were not robust enough to serve as controls in quantitative PCR. As a result, β-actin was used as the control in all subsequent experiments (see below).

EST sequencing and bioinformatic summaries

Following verification on dot blots, 68 clones were confirmed with clear differences in expression between workers and soldiers. High-quality ESTs were obtained for each of these clones (Additional data file 1). Because the cDNA library was created with 5'-3'-oriented cDNAs, EST sequencing was only performed in one pass from the 5' ends of clones. EST sequencing results are summarized in Table 1 in order of BLASTx E-value significance scores. Lengths of high-quality, vector-trimmed sequence ranged from 66 to 632 nucleotides. On the basis of BLASTx database similarity searches, eight cDNAs occurred more than once in the array screen (Additional data file 1). These multiply-occurring cDNAs, shown in order of their identity-significance scores, correspond to genes encoding: termite salivary cellulases (3 total), termite endoglucanase (10), endosymbiont endoglucanase (3), endosymbiont cellulase (2), mitochondrial cytochrome oxidase I (24), muscle-specific LIM protein (2), mitochondrial 18K protein (2), and unknown #24 (3). For simplicity, only single clones representing the longest sequence are reported in Table 1. Clones 1-16 possess significant E-values of less than 1×10^{-8} and therefore their identities are considered valid (Tables 1, 2). Because clones 17-25 have larger E-values, the identities of these clones, when available, are only considered putative.

Table 1

EST sequence summary of 25 positive *R. flavipes* clones identified from array screens

Clone *	Size (bp)	Identity match by BLASTx (%)	E-value	Gene ontogeny†
1 (D04)	615	98% to *Reticulitermes speratus* salivary cellulase	1e-106	E
2 (C02)	423	89% to *Reticulitermes speratus* endoglucanase 2	1e-79	E
3 (E11)	505	57% to *Pseudotrichonympha grassii* cellulase	7e-69	E
4 (B02)	518	87% to *Blattella germanica* cytochrome oxidase I	7e-54	ET
5 (A10)	449	Unknown (65% to *Anopheles gambiae* unknown)	2e-49	UNK
6 (C07)	505	>65% to *Drosophila melanogaster* β-NAC homolog, 'bicaudal'	2e-45	TF
7 (B09)	585	69% to *Holomastigotoides mirabile* endoglucanase 2	1e-47	E
8 (C11)	509	69% to *Drosophila melanogaster* muscle-specific 'LIM' protein	2e-35	CSP
9 (B11)	569	80% to *Drosophila melanogaster* troponin I 'wings up'	1e-35	CSP
10 (H04)	242	98% to *Bombyx mori* β-tubulin	6e-33	CSP
11 (D10)	499	39% to *Drosophila melanogaster* troponin I 'wings apart', 'heldup'	3e-29	CSP
12 (F10)	527	36% to *Athalia rosae* vitellogenin	4e-20	SP
13 (D03)	387	>64% to *Drosophila melanogaster* 'BAB-I' (BTB/POZ domain containing protein)	3e-18	TF
14 (G12)	296	81% to *Drosophila melanogaster* mitochondrial-ribosomal protein S2	1e-17	MOB
15 (G03)	340	49% to *Mus musculus* SH3-domain kinase binding protein	5e-10	ST
16 (D02)	645	43% to *Homo sapiens* Rab3 GTPase-activating protein	2e-8	ST
17 (H08)	297	39% to *Caenorhabtidis elegans* RNA-dependent DNA polymerase (RTase)	0.001	DRF
18 (B10)	350	Unknown (40% to *Drosophila melanogaster* GC1077 gene product)	0.03	UNK
19 (A05)	630	38% to goldfish mitochondrion hypothetical 18K protein	0.12	MOB
20 (D07)	471	44% to *Oncorhyncus mykiss* cyclic GMP-gated channel	0.39	IC
21 (H06)	632	27% to *Musca domestica nanos* posterior determinant homolog	1.2	TF
22 (E12)	372	34% to *Moraxella bovis* endonuclease Mbo II	3.2	DRF
23 (G06)	370	Unknown (36% to *Anopheles gambiae* unknown)	>5	UNK
24 (A02)	510	Unknown	>5	UNK
25 (E03)	392	Unknown	>5	UNK

Clones are listed according to the significance of their identify-based BLASTx score E-values. * Clone number, with sequence identifier position shown in parentheses. See Additional data file 1 for sequence details. †Gene ontogeny classification based on Adams *et al.* [13]: E (enzyme), CSP (cytoskeletal/structural protein), DRF (DNA replication factor), ET (electron transport), IC (ion channel), MOB (mitochondrial organization and biogenesis), SP (storage protein), ST (signal transduction), TF (transcription factor), UNK (unknown).

Because the six ESTs noted above corresponded to multiple positive array positions, the total number of distinct/unique sequences present in the array-positive EST pool was 25. These 25 cDNAs can be categorized into 10 gene ontogeny categories following the summary presented by Adams *et al.* [13]. These are: enzyme, cytoskeletal/structural, electron transport, transcription factor, mitochondrial organization and biogenesis, signal transduction, storage, DNA replication, ion channel, and five unknown (Table 1).

Relative gene expression

Relative expression patterns for 24 discrete gene transcripts, determined by quantitative PCR, are shown in Table 2. Raw data from quantitative PCR analyses are provided in Additional data file 2. There was only one gene transcript that could not be amplified and evaluated by quantitative PCR. This sequence encodes a putative cyclic GMP-gated ion channel (clone 20) with highly repetitive nucleotide sequences and high CG content [14]. Because of these properties, no PCR primer pairs could be designed that would provide the high-fidelity PCR necessary for quantitative PCR analysis.

Transcripts having highest worker-associated expression were those encoding cellulase and endoglucanase enzymes involved in cellulose digestion. Cellulases and endoglucanases can have origins in either termites or their microorganismal endosymbionts. Cellulases and endoglucanases of apparent endosymbiont origin had substantially higher expression relative to those of apparent termite origin. Other transcripts with elevated expression in workers were either related to mitochondrial organization and biogenesis (mRpS2 and 18K protein-like; also elevated in soldiers), or had unknown identity (clones 18 and 24).

More than half of the transcripts evaluated had exclusive soldier-biased expression (13 of 24). These included two

Table 2

Relative expression levels of *R. flavipes* gene transcripts

Gene ontogeny category and clone identity	Accession number‡	Caste		Developmental stage	
		Worker	Soldier	Presoldier	Nymph
Standard/control					
β-actin†	N/A (N/A)	1.1 ± 0.1	1.0 ± 0.4	1.0 ± 0.1	1.1 ± 0.1
Transcription/translation factor					
6. Rf β-NAC-1, 'bicaudal'	AY258589	5.2 ± 0.2	**13.8 ± 0.3**	1.4 ± 0.1	1.0 ± 0.1
21. Rf PDL, 'nanos'-like	BQ788190	3.0 ± 0.0	**5.0 ± 0.1**	1.0 ± 0.5	2.2 ± 0.1
13. Rf BTB/POZ-1, 'BAB'	AY258590	**2.2 ± 0.3**	1.7 ± 0.0	1.0 ± 0.1	**2.6 ± 0.2**
Signal transduction					
15. SH3-domain kinase-binding	CB518313	85.5 ± 0.1	**131.9 ± 0.2**	1.0 ± 0.2	3.6 ± 1.3
16. GTPase-activating protein	BQ788178	2.3 ± 0.3	**4.5 ± 0.1**	1.0 ± 0.4	1.3 ± 0.0
Enzyme					
1. Termite salivary cellulase	CB518295	**31.5 ± 0.1**	1.6 ± 0.0	1.0 ± 0.0	19.5 ± 0.0
2. Termite endoglucanase 2	CB518296	**32.3 ± 4.8**	2.5 ± 1.5	1.0 ± 0.1	**30.1 ± 1.1**
3. Symbiont cellulase	CB518297	**1939.7 ± 0.0**	1.7 ± 0.2	1.0 ± 0.4	208.9 ± 0.2
7. Symbiont endoglucanase 2	CB518298	**62.7 ± 0.9**	2.7 ± 1.1	1.0 ± 0.6	34.4 ± 0.1
Cytoskeletal/structural protein					
8. Muscle LIM protein	CB518301	1.6 ± 0.3	**5.0 ± 0.2**	1.8 ± 0.3	1.0 ± 1.1
9. Rf1-troponin I	CB518302	2.3 ± 0.1	**3.9 ± 0.4**	1.4 ± 0.2	1.0 ± 0.0
10. β-tubulin	CB518304	1.6 ± 0.3	**5.0 ± 0.2**	1.8 ± 0.3	1.0 ± 1.1
11. Rf2-troponin I	CB518303	1.1 ± 1.3	**11.5 ± 0.6**	1.0 ± 1.2	1.9 ± 3.0
Storage/hormone binding					
12. Vitellogenin	CB518311	1.2 ± 0.6	1.0 ± 0.1	**4.9 ± 0.2**	1.4 ± 1.7
Electron transport					
4. Cytochrome oxidase I	CB518306	3.1 ± 0.6	**6.6 ± 0.1**	2.5 ± 0.2	1.0 ± 0.0
Mitochondrial organization/biogenesis					
14. mRpS2	CB518308	**4.3 ± 0.0**	**4.0 ± 0.0**	1.0 ± 0.7	1.9 ± 0.1
19. 18K protein-like	CB518307	**2.9 ± 0.0**	1.3 ± 0.9	1.0 ± 0.1	**2.3 ± 1.5**
DNA replication factor					
17. Reverse transcriptase-like	CB518309	**1.4 ± 1.5**	**2.1 ± 0.7**	1.2 ± 0.1	1.0 ± 0.7
22. *MboII* endonuclease-like	CB518310	1.1 ± 1.3	**5.3 ± 0.3**	2.8 ± 0.1	2.2 ± 0.4
Unknown					
5. Unknown	CB518314	2.2 ± 0.8	**6.1 ± 0.4**	1.0 ± 0.9	2.6 ± 0.1
18. Unknown	CB518315	**9.5 ± 0.2**	1.2 ± 0.5	1.1 ± 0.1	1.0 ± 0.3
23. Unknown	CB518318	1.0 ± 0.2	**5.9 ± 0.2**	**5.4 ± 2.5**	1.2 ± 2.0
24. Unknown	CB518317	**12.4 ± 0.1**	2.3 ± 0.1	1.0 ± 0.0	1.9 ± 0.1
25. Unknown	CB518319	1.1 ± 0.1	**2.1 ± 0.2**	1.5 ± 0.2	1.0 ± 0.5

Relative expression levels of *R. flavipes* gene transcripts (average ± standard error), identified using cDNA macroarrays and quantified by quantitative PCR. Data shown in bold-face type represent the highest relative expression for a given gene (LSD *t*-test; $P \leq 0.05$). Relative expression calculated using the method described by Pfaffl *et al.* [44] (*n* = 3). See Additional data file 2 for raw data. †All results are normalized to the control gene β-actin. Primers designed from mouse and rat sequences (Accession nos. X03672 and V01207), as described by Giulietti *et al.* [43]. ‡Clone numbers correspond to the order shown in Table 1 (numbers 16 and lower have significant identity scores). See Additional data file 1 and 9, respectively, for sequence and primer details.

transcription factors, two signal transduction factors, four cytoskeletal/musculature proteins, one transcript associated with mitochondrial electron transport, one putative DNA replication factor, and three transcripts of unknown identity. Three additional transcripts had similar expression levels between workers and soldiers, yet had higher expression than the presoldier and/or nymphal intermediary developmental stages (mRpS2,18 K protein-like, and reverse transcriptase-like).

Two soldier-biased troponin I-encoding transcripts had both differential expression and sequence variation (Additional data file 3). The first of these (Rf1-troponin I) aligns with a portion of the 3' region of the *Drosophila* troponin I cDNA whereas the second (Rf2-troponin I) aligns with the 5' region, and includes the ATG start signal. The two troponin ESTs have a 60-nucleotide overlap, but share only 65% amino-acid identity in the overlap region. These termite troponin forms have significantly elevated expression levels in soldiers, which are 1.7- and 10.5-fold elevated for the Rf1 and Rf2 forms, respectively, relative to workers.

Transcripts were also observed with high expression levels in the intermediary presoldier and nymph developmental stages. In all cases, transcripts that were identified with elevated nymphal expression were also elevated in workers (as with cellulases/endoglucanases and Rf BTB/POZ-1) or both workers and soldiers (as with mitochondrial 18 K protein-like). Finally, a transcript encoding the storage/hormone-binding protein vitellogenin had highest expression in presoldiers, a stage incapable of feeding itself, and that is also hypothesized to serve as a sink for juvenile hormone [15].

Putative caste-regulatory genes
Two putative regulatory genes with significant homology to *Drosophila* and honey-bee (*Apis mellifera*) developmental genes were identified. Both these genes are transcriptional/translational regulators. One is *Rfβ-NAC-1*, which is related to the *bicaudal* locus of *Drosophila*, and the second is *Rf BTB/POZ-1*, which is related to the *bric-a-brac* (*BAB*) locus of *Drosophila*. Alignments of the respective translated amino-acid sequences are shown in Additional data files 4 and 5. A third related gene, *Rf-PDL*, has putative identity as a *nanos*-homologous posterior-determinant gene on the basis of weak homology to database sequences; however, hydropathy profiles for the translated amino-acid sequence bear striking similarities to *Drosophila* nanos protein and its homologs from other species (Additional data file 6).

Rfβ-NAC-1 is a *Reticulitermes* homolog of the ribosomal translation factor β-NAC. A partial open reading frame (ORF) of 420 nucleotides (140 amino acids) was included in our cDNA clone. Database searches using the Rfβ-NAC-1 translated amino-acid sequence revealed numerous proteins with significant homology, all of which are associated with ribosomal translational processes. Alignments of Rfβ-NAC-1

revealed 73 invariant amino-acid residues (Additional data file 4). Also observed was significant pairwise homology to β-NAC homologs from *Drosophila* (61.5%), *A. mellifera* (71.5%), and *Caenorhabditis elegans* (60.9%). These findings clearly establish identity of the *Reticulitermes* sequence as a β-NAC homolog.

Complete sequencing of the *Rf BTB/POZ-1* cDNA clone elucidated a complete ORF of 1,320 nucleotides (439 amino acids). Database searches with the translated amino-acid sequence of the entire ORF did not reveal homologs with high degrees of similarity across their entire lengths; however, significant similarity to BTB/POZ domains of many BTB/POZ proteins was identified (Additional data file 5). The presence of the highly conserved BTB/POZ domain indicates that the sequence encodes a zinc-finger protein involved in transcriptional regulation. Alignments of the amino-terminal regions of several BTB/POZ domain-containing proteins revealed an insertion of 60-65 amino acids in the amino-terminal region of Rf BTB/POZ-1 that is not present in any of the homologous sequences, and the presence of three invariant sequence motifs in all BTB/POZ proteins shown (single-letter amino-acid code: LRWN, DVTL, LSACS). Overall identity between the BTB/POZ domains of termite Rf BTB/POZ-1 and *Drosophila* BRcore, *Drosophila* fruitless, *A. mellifera* pipsqueak, and *Drosophila* pipsqueak are 44.5, 43.1, 41.8, and 35.8%, respectively.

A third EST termed *Rf-PDL* (posterior determinant-like) shares modest similarity with *nanos*-homologous posterior determinant genes of *Musca domestica* (AAA87461) and *Xenopus laevis* (X72340) (part a in Additional data file 6). Striking similarities in hydropathy profiles exist, however, for sequenced portions of the termite cDNA versus *nanos* homologs from *Drosophila*, *Musca*, and *Xenopus* (part b in Additional data file 6). The *nanos* gene in *Drosophila* is a transcription factor that controls anterior-posterior polarity and is under direct control of β-NAC. Quantitative PCR results for *Rf-PDL* follow a similar caste and development-associated expression pattern to *Rf β-NAC-1*, implying a link between these two genes.

Two gene transcripts encoding signal transduction factors with established links to well-characterized developmental processes were also identified. The first of these genes, SH3-domain kinase, has significant homology to mammalian kinases occurring at the *Ruk* locus of mouse and rat (Additional data file 7), but interestingly, not to any sequences from insects. The second signal transduction factor, GAP, also has the highest degree of similarity to vertebrate sequences, although modest similarity also exists to a *Drosophila* GTPase-coding sequence (Additional data file 8). Relative to presoldiers, both of these transcripts showed substantial and significantly elevated expression in soldiers (132- and 4.5-fold, respectively), although expression was also significantly elevated in workers (86- and 2.3-fold, respectively).

Discussion

Intuitive agreements between caste-associated physiology and gene expression

Cellulases

Termites and their close relatives the *Cryptocerca* cockroaches comprise most of the invertebrate taxa that have the ability to feed on, digest, and receive nutrition from cellulose. We identified four transcripts encoding enzymes apparently involved in *R. flavipes* cellulose digestion, two each of putative termite and endosymbiont origin, respectively. The assigned identity of these ESTs as either termite or endosymbiont is based on their significant levels of identity to database sequences obtained from endosymbiont-free termite salivary glands [16] or endosymbionts themselves (GenBank accession numbers BAB69425, BAB64564). Abundance of cellulase transcripts was highest in workers and nymphs, which is consistent with cellulose-feeding adaptations in these forms; and is also likely to be attributable to differences in symbiont communities between workers and nymphs versus presoldiers and soldiers. The ability of termites to digest cellulose was long thought to be conferred exclusively by endosymbiotic bacteria and protozoa. However, after cellulase-encoding genes were discovered in the termite genome [16], the main role of protozoan and bacterial gut symbionts was concluded to relate more to anaerobic respiration and nitrogen fixation, than to cellulose digestion [17]. Interestingly, we observed that both endosymbiont- and termite-homologous cellulases were abundant, supporting the contention that cellulases of both endosymbiont and termite origin are important to cellulose digestion [18].

Cytoskeletal, muscular and electron-transport proteins

Differentiation of the soldier caste is associated with a tremendous increase in body mass and musculature; particularly in the head, where a large muscle mass is required to drive the enlarged mandibles. We identified four cytoskeletal/muscular protein-coding transcripts with high expression in soldiers (two troponin I forms, a muscle LIM protein, and β-tubulin), and another encoding mitochondrial cytochrome oxidase I. The elevated expression of cytochrome oxidase I in soldiers would appear to be a result of elevated energy demands associated with their large muscle mass. Similarly, elevated cytochrome oxidase expression levels were observed previously in larval queen honey-bees [19,20], suggesting parallels in caste-associated respiratory requirements between the eusocial Hymenoptera and Isoptera.

Troponin I is a Ca^{2+}-binding protein involved in the interaction of thin (actin) and thick (myosin) muscle filaments, and is therefore central to muscle function [21]. *heldup* mutations arise from aberrations and point mutations in the *Drosophila* troponin I reading frame, resulting in pronounced flight-muscle degeneration [22]. Soldier termites are flightless, unlike their highly developed counterparts, the winged primary reproductives, or alates (Figure 1). Because of the apterous nature of soldiers, the fact that both apterous and

imaginal forms differentiate from last instar larvae, and the existence of apparent aberrations (relative to the *Drosophila* sequence) in the two termite troponin I ESTs, it is possible that Rf1 and/or Rf2-troponin I may be involved in flight-muscle degeneration.

We also identified a termite muscle LIM protein with similarity to muscle LIM protein 60A of *Drosophila*. Muscle LIM proteins are cysteine-rich proteins with zinc-binding motifs that provide cytoarchitectural support and regulate cell growth and differentiation during development [23,24]. Two LIM muscle proteins occur in *Drosophila* (Mlp60A, Mlp84B) [23]. *Drosophila* LIM 60A is expressed throughout muscle fibers, both late in embryogenesis and during pupal metamorphosis [23]. Because soldier termites constitute a developmental end-point, and also because the main role of LIM proteins in *Drosophila* relates to muscle cytoarchitecture [24], we hypothesize that the termite LIM has a primary role in cytoarchitecture. This conclusion excludes any developmental roles for the termite LIM.

Finally, the termite β-tubulin we identified is most similar to constitutively expressed β-tubulins from *Drosophila* and *Bombyx mori* [25]. Tubulins are assembled from α- and β-tubulin dimers, and are involved in cytoskeletal support and mRNA trafficking in cells during development. From the soldier-specific expression of termite β-tubulin, it also appears to have a role in maintaining cytoarchitecture, rather than in caste development-affiliated processes.

Genes with putative caste-regulatory and/or developmental functions

Transcription and translation factors

Three putative regulatory genes were identified with clear caste-associated expression. Two of these, *Rf β-NAC-1* and *Rf BTB/POZ-1*, possess significant sequence homology to highly conserved developmental genes from *Drosophila* that occur at the *bicaudal* [26] and *bric-a-brac* [27,28] loci, respectively. A third gene transcript, *Rf-PDL*, has weak homology but conserved hydropathy profiles with a developmental gene occurring at the *nanos* locus of *Drosophila* [29,30] and its homologs. In *Drosophila*, both *bicaudal* and *nanos* function in embryonic pattern formation; thus, expression of their putative homologs in late and terminal termite developmental stages is highly intriguing. Expression profiles and functions of the termite *bicaudal* and *nanos* (that is, *Rf β-NAC-1* and *Rf BTB/POZ-1*) are presently undefined; therefore, it is not known if their elevated expression in a developmental end-point such as the soldier caste is related to a caste-specific function, or whether it is residual from the presoldier-to-soldier developmental transition.

The *bicaudal* mutation was the first *Drosophila* mutation that was found with a clear effect on embryonic pattern formation [31]. Extreme *bicaudal* mutants have bicaudal embryos with missing heads and thoraces that are replaced

with mirror-image duplications of abdominal segments. Markesich and co-workers [26] mapped the *bicaudal* locus to a region encoding β-NAC, a component of the ribosomal translational machinery and the nascent polypeptide-associated complex. *bicaudal* and a related mutation, *enhancer of bicaudal*, are caused by insertions into the β-NAC transcribed region. Both *bicaudal* mutations have been rescued by wild type copies of β-NAC, which regulates expression of the posterior determinant gene *nanos* (that is, *bicaudal* mutations lead to ectopic expression of *nanos* [26]). The observed high expression of Rf β-NAC-1 in soldiers, and the unique shape of the termite soldier head in relation to the rest of the body, suggest that Rf β-NAC-1 may influence the generalized soldier body plan and possibly direct ectopic expression of the posterior determinant *nanos* (see below).

In insects, *nanos* is an evolutionarily conserved organizer of anterior-posterior polarity, which acts as a translational repressor by directing expression of the *hunchback* gene [30,32]. The mechanism by which *nanos* acts is conserved in insects, and *nanos* sequence divergence among dipterans correlates with evolutionary relationships and natural history. As a result, *nanos* should be more highly diverged in *R. flavipes*. In a study by Curtis *et al.* [30], *nanos* homologs with only 28-63% amino-acid identity from evolutionarily distant dipteran species successfully rescued *nanos* activity in null mutants of *Drosophila*. Although the function encoded by the *Rf-PDL* transcript is unknown, its similar expression patterns to *Rf β-NAC-1*, and its similar hydropathy profiles to known *nanos* homologs, provide support that it may potentially encode *nanos*-homologous functions. Regardless of its true identity, the *Rf-PDL* transcript possesses caste-associated expression that suggests a developmental role.

BTB/POZ domains are conserved amino-acid motifs in the amino terminus of zinc-finger proteins, some of which function as transcriptional regulators. A total of 40 BTB/POZ domain-containing proteins are encoded in the *Drosophila* genome, and therefore these proteins appear to be responsible for a vast number of functions [27,28]. Several lines of evidence establish clear developmental roles for BTB/POZ-domain-containing proteins, however. Godt and co-workers [27] reported that the BTB/POZ protein bric-a-brac (Bab-I), which is most similar to our termite sequence, is involved in pattern formation along the proximal-distal axis of the leg and antenna of *Drosophila*, and that mutants possess legs and antennae with segmentation defects. *bric-a-bac* is also known to be involved in *Drosophila* ovary morphogenesis on the basis of the disrupted ovariole formation and female sterility in *bric-a-brac* mutants [33]. BTB/POZ protein domains can dimerize with themselves [34], possibly leading to self-inactivation. This suggests that the expression of termite Rf BTB/POZ-1 may repress gene expression in workers and nymphs, which have higher BTB/POZ expression than presoldiers and soldiers. Alternatively, elevated BTB/POZ

expression in workers and nymphs may relate to their gonadotropic/reproductive potential.

Signal transduction factors

We also identified two gene transcripts with significant homology to the signal transduction factors SH3-domain kinase and GAP. SH3-domain proteins are adaptor proteins involved in the regulation of signal transduction [35]. The SH3 domains of these proteins can interact with other kinases involved in apoptotic cell death and membrane trafficking. Expression of mouse SH3-Ruk proteins in cells appears to be involved in the regulation of their survival and other intracellular processes [35]. Extremely high levels of expression occurred for the SH3-domain protein in termite soldiers (139-fold relative to presoldiers), but expression was also high in workers (85.5-fold relative to presoldiers), suggesting that dose-dependent factors may be implicated in SH3-domain cascade-dependent processes.

GAPs are involved in regulating the transfer of phosphate groups by GTPases, and are particularly relevant to signal transduction processes. One related protein in *Drosophila*, named Ras1, mediates the modulation of homeotic gene function in cell and segment identity [36] and eye development [37]. Ras1-mediated cell signaling modifies nuclear activity of transcription factors, most notably those encoded by the homeobox genes *proboscopedia* and *ultrabithorax*. Ras1 is active when coupled with GTP but inactive in the presence of GDP, thus implicating GTPases as regulators of Ras1 (and homeobox) action. Ras1 impacts on homeobox gene expression are dose-dependent [36], which may explain high and intermediate (not absent) GAP expression in soldiers and workers, respectively. Together, our signal transduction-related findings suggest interesting possibilities for signal transduction-mediated transcriptional regulation, cascade-dependent processes, and possibly regulation of cellular death in termite caste-differentiation processes.

Conclusions

Here we have presented the first summary of a transcriptome-based survey for caste- and differentiation-associated genes in a termite. We have focused on the lower termite *R. flavipes* because it is the commonest termite in the United States, and because of its ecological and economic importance. Vast differences occurred in sequence composition of this directed array screen versus previous random EST sequencing [12], emphasizing the validity of our approach. Several biologically significant trends were apparent from our findings. First, expression of both termite- and endosymbiont-homologous cellulase genes is correlated with cellulose-feeding adaptations in workers and nymphs. Interestingly, endosymbiotic genes were represented in our arrays, suggesting additional utility of our approach in investigating host-endosymbiont relationships. Second, genes related to transcriptional regulation and signal transduction (some linked

to key developmental processes in *Drosophila* and honey-bees) appear to be important in the development and differentiation of the soldier caste. Third, genes associated with musculature and cytoskeletal architecture (including two troponin I forms) were identified in association with the apterous and highly muscular soldier caste. Fourth, expression of a gene encoding the storage/hormone-binding protein vitellogenin is associated with a stage incapable of feeding and thought to serve as a juvenile hormone sink (the presoldier stage). Finally, several genes with unknown and/or putative identity, but with significant caste and development-associated expression were also identified as having differential expression between castes.

Because the two stages compared in our cDNA arrays represent a developmental end-point (soldiers) and a non-differentiated caste (workers), many important developmental genes were surely not identified. Future sociogenomic-based studies on intermediary developmental stages (nymphs and presoldiers) is likely to yield exciting new information. We observed strong similarities between termite genes and a number of well-characterized *Drosophila* mutants and genes, as well as honey-bee developmental genes. It is unclear if the differences we identified in developmental gene expression between *Drosophila*, honey-bees and termites relates to the holometabolous and hemimetabolous nature (respectively) of these insect groups. Nonetheless, the information on novel termite genes reported here sheds new light on completely uncharacterized aspects of basic termite molecular biology and physiology.

Materials and methods
The authors will provide specific protocols on request.

Termites, cDNA library, array preparation and array screening
Termites of multiple castes and developmental stages were obtained, and a polyphenic cDNA library synthesized as described previously [12]. Total RNA was isolated from groups of whole termites using the RNaqueous total RNA isolation kit (Ambion, Austin, TX), and its integrity and quantity verified by formaldehyde-agarose electrophoresis [38]. Poly(A) RNA was purified from the total RNA (Ambion), cDNA was synthesized using an oligo-dT$_{12}$ adaptor primer, and the library constructed (see [12]). From the cDNA library, the Purdue University Genomics Core Facility prepared multiple identical 'macro' arrays, which are gridded bacterial colony filters. Each 22 × 22 cm filter was printed from 18,432 glycerol stocks contained in forty-eight 384-well plates. Duplicate filters were printed by using a Biorobotics robot (Total Array System, Woburn, MA) to spot samples from plates onto charged nylon membranes overlaid with LB agar containing 100 μg/ml ampicillin. The filters were grown overnight at 37°C, peeled away from the agar, placed on chromatography paper saturated with 5% SDS and 2x SSC for 2 min,

heated in a microwave for 2.5 min, then UV-cross linked in a Stratalinker (Stratagene, La Jolla, CA).

Primary gene-expression comparisons between the soldier and worker castes were made as follows, using soldiers and workers isolated from laboratory colonies. mRNA was obtained separately from whole, pooled worker or soldier termites (Figure 1) using a commercially available protocol (Ambion), then digoxigenin (DIG)-labeled using a separate commercially available protocol (DIG Chem-link Labeling and Detection; Roche, Indianapolis, IN). Individual array membranes were probed with DIG-labeled mRNAs from either workers or soldiers (10 ml of 100 ng/ml). Autoradiograms of the membranes, showing the positions of specific clones, were produced using a commercially available protocol (Roche). Expression levels were compared between soldier and worker array autoradiograms by manual alignment.

For secondary verification, dot blots were performed using plasmid DNA affixed to nitrocellulose membranes (Millipore, Marlboro, MA). Liquid cultures were produced from array-positive cell lines, from which plasmid DNA was obtained using a commercially available protocol (Qiagen, Valencia, CA). Pure plasmid DNA (100-200 ng) was manually spotted onto duplicate membranes [38]. Membranes were probed with labeled aliquots of the same worker- and soldier-derived mRNAs noted above, using an identical procedure, then were manually superimposed to compare relative expression levels between castes.

Sequencing and bioinformatics analyses
Array and dot blot-positive clones were sequenced from their 5' ends (T3 promoter) using a high-throughput EST sequencing procedure. For the putative regulatory genes *Rfβ-NAC-1* and *Rf BTB/POZ-1*, longer sequences shown in alignments (Additional data files 4 and 5) were obtained by sequencing from both the 5' and 3' ends, and by designing internal sequencing primers in the case of BTB/POZ (forward = GCAACGGATGACTACATGGA, reverse = TGGTTGAAGCCTGATTCACA). All sequencing-related procedures were performed by the Purdue University Genomics Core Facility. After contaminating vector sequences were trimmed, the ESTs were used as query sequences to search the GenBank nucleotide and translated databases by BLASTx [39]. Database searches were performed under default settings, with E-values for significance of identity being 1×10^{-10} and smaller [40]. Sequence alignments were performed using clustal analysis in DNAstar-Megalign software (Madison, WI).

Quantitative real-time PCR
Quantitative PCR was performed to precisely quantify relative caste-associated gene expression. cDNA templates for quantitative PCR were synthesized separately from DNase-treated total RNA of whole worker, presoldier, soldier and

nymph termites. Total RNA isolation and DNase treatment were performed using commercially available protocols (Qiagen and Ambion, respectively). The quantity and quality of RNA was assessed before and after DNase treatment by spectrophotometry and formaldehyde-agarose electrophoresis [38]. cDNA was synthesized from 5 μg total RNA using an oligo-dT$_{15}$ primer and 50 U Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA) at 42°C for 50 min.

Primers for use in quantitative PCR were designed from EST sequence data by the MIT Primer 3 program [41]. Primer sequences are provided in Additional data file 9. Quantitative PCR primers were designed to provide products of 100-200 base-pairs (bp) within putative ORFs; and had 45-55% GC content, T_m = 57-62°C, non-complementarity between forward and reverse primers, and pyrimidine nucleotides on their 5' and 3' ends [42].

Quantitative PCR was performed using an iCycler iQ real-time PCR detection system (Bio-Rad, Hercules, CA) with SYBR Green I product tagging. To verify the amplification efficiency for both target and control genes, quantitative PCR conditions were optimized for cDNA template dilution (1:5, 1:10, 1:25, 1:50, and so on) and annealing temperature (Tm ± 5°C). Before experiments, the production of single product bands and absence of 'primer-dimers' was confirmed by viewing products following agarose electrophoresis and ethidium-bromide staining. Simultaneous determination of relative expression levels for specific genes in workers, soldiers, presoldiers and nymphs was determined from quantitative PCR results in relation to the standard gene β-actin [43], using REST software [44] (see Additional data file 2).

Data deposition
Sequences have been deposited in GenBank under accession numbers AY258589, AY258590, BQ788178, BQ788190 and CB518295 through CB518319.

Additional data files
The following additional data are included with the online version of this article: Figures showing a summary of clones and clone sequences (Additional data file 1), the quantitative real-time PCR raw data (Additional data file 2), the troponin alignments (Additional data file 3), the β-NAC alignments (Additional data file 4), the BTB/POZ alignments (Additional data file 5), the *nanos* alignments and hydropathy profiles (Additional data file 6), the SH3-domain kinase alignments (Additional data file 7), the GAP alignments (Additional data file 8), and the real-time PCR primer sequences (Additional data file 9).

Acknowledgements

References

1. Wilson EO: **Caste: Termites.** In *The Insect Societies* Cambridge, MA: Belknap Press; 1971:183-210.
2. Thorne BL: **Evolution of eusociality in termites.** *Ann Rev Ecol Syst* 1997, **28:**27-54.
3. Nutting WL: **Insecta: Isoptera.** In *Soil Biology Guide* Edited by: Dindal DL. New York: Wiley and Sons; 1990:997-1032.
4. Buchli H: **L'origine des castes et les potentialitiés ontogéniques des termites européens du genre *Reticulitermes.*** *Annls Sci Nat Zool Biol Anim* 1958, **20:**261-429.
5. Thorne BL: **Termite terminology.** *Sociobiology* 1996, **28:**253-261.
6. Roisin Y: **Diversity and evolution of caste patterns.** In *Termites: Evolution, Sociality, Symbioses, Ecology* Edited by: Abe T, Bignell DE, Higashi M. Dordrecht: Kluwer; 2000:95-119.
7. Evans JD, Wheeler DE: **Gene expression and the evolution of insect polyphenisms.** *BioEssays* 2001, **23:**62-68.
8. Evans JD, Wheeler DE: **Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera.*** *Proc Natl Acad Sci USA* 1999, **96:**5575-5580.
9. Abouheif E, Wray GA: **Evolution of the gene network underlying wing polyphenism in ants.** *Science* 2002, **297:**249-252.
10. Miura T, Kamikouchi A, Sawata M, Takeuchi H, Natori S, Kubo T, Matsumoto T: **Soldier caste-specific gene expression in the mandibular glands *of Hodotermopsis japonica.*** *Proc Natl Acad Sci USA* 1999, **96:**13874-13879.
11. Miura T: **Morphogenesis and gene expression in the soldier-caste differentiation of termites.** *Insectes Soc* 2001, **48:**216-223.
12. Wu-Scharf D, Scharf ME, Pittendrigh BR, Bennett GW: **Expressed sequence tags from a polyphenic *Reticulitermes flavipes* cDNA library.** *Sociobiology* 2003, **41:**479-490.
13. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, *et al.*: **The genome sequence of *Drosophila melanogaster.*** *Science* 2000, **287:**2185-2195.
14. Vervoort R, Lennon A, Bird AC, Tulloch B, Axton R, Miano M, Meindl A, Meltinger T, Ciccodicola A, Wright AJ: **Mutational hot spot within a new PRGR exon in X-linked retinitis pigmentosa.** *Nat Genet* 2000, **25:**462-466.
15. Henderson G: **Primer pheromones and possible soldier caste influence on the evolution of sociality in lower termites.** In *Pheromone Communication in Social Insects* Edited by: Vander Meer RK, Breed MD, Espelie KE, Winston ML. Boulder, CO: Westview Press; 1998:314-329.
16. Watanabe H, Noda H, Tokuda G, Lo N: **A cellulase gene of termite origin.** *Nature* 1998, **394:**330-331.
17. Bignell DE: **Introduction to symbiosis.** In *Termites: Evolution, Sociality, Symbioses, Ecology* Edited by: Abe T, Bignell DE, Higashi M. Dordrecht: Kluwer; 2000:189-209.
18. Nakashima K, Watanabe H, Saitoh H, Tokuda G, Azuma JI: **Dual cellulose-digesting system of the wood-feeding termite *Coptotermes formosanus.*** *Insect Biochem Mol Biol* 2002, **32:**777-784.
19. Evans JD, Wheeler DE: **Expression profiles during honeybee caste determination.** *Genome Biol* 2000, **2:**research0001.1-0001.6.
20. Corona M, Estrada E, Zurita M: **Differential expression of mitochondrial genes between queens and workers during caste determination in the honeybee *Apis mellifera.*** *J Exp Biol* 1999, **202:**929-938.
21. Squire JM: **Architecture and function in the muscle sarcomere.** *Curr Opin Struct Biol* 1997, **7:**247-257.
22. Beall CJ, Fryberg E: **Muscle abnormalities in *Drosophila melanogaster heldup* mutants are caused by missing or aberrant troponin-I isoforms.** *J Cell Biol* 1991, **114:**941-951.
23. Stronach BE, Siegrist SE, Beckerle MC: **Two muscle-specific LIM proteins in *Drosophila.*** *J Cell Biol* 1996, **134:**1179-1195.
24. Stronach BE, Renfranz PJ, Lilly B, Beckerle MC: **Muscle LIM proteins are associated with muscle sarcomeres and require dMEF2 for their expression during *Drosophila* myogenesis.** *Mol Biol Cell* 1999, **10:**2329-2342.
25. Kawasaki H, Sugaya K, Quan GX, Nohata J, Mita K: **Analysis of α- and β-tubulin genes of *Bombyx mori* using an EST database.** *Insect Biochem Mol Biol* 2003, **33:**131-137.

26.   Markesich DC, Gajewski KM, Nazimiec ME, Beckingham K: *bicaudal* **encodes the *Drosophila* beta NAC homolog, a component of the ribosomal translational machinery.** *Development* 2000, **127:**559-572.
27.   Godt D, Couderc JL, Cramton SE, Laski FA: **Pattern formation in the limbs of *Drosophila* : *bric-a-brac* is expressed in both a gradient and wave-like pattern and is required for specification and proper segmentation of the tarsus.** *Development* 1993, **119:**799-812.
28.   Lehmann M, Siegmund T, Linterman KG, Korge G: **The pipsqueak protein of *Drosophila melanogaster* binds to GAGA sequences through a novel DNA-binding domain.** *J Biol Chem* 1998, **273:**28504-28509.
29.   Wang C, Lehmann R: **Nanos is the localized posterior determinant in *Drosophila*.** *Cell* 1991, **66:**637-647.
30.   Curtis D, Apfeld J, Lehmann R: *nanos* **is an evolutionarily conserved organizer of anterior-posterior polarity.** *Development* 1995, **121:**1899-1910.
31.   Bull A: *bicaudal,* **a genetic factor which affects the polarity of the embryo in *Drosophila melanogaster*.** *J Exp Zool* 1966, **161:**221-242.
32.   Lall S, Ludwig MZ, Patel NH: **Nanos plays a conserved role in axial patterning outside of the Diptera.** *Curr Biol* 2003, **13:**224-229.
33.   Godt D, Laski FA: **Mechanisms of cell rearrangement and cell recruitment in *Drosophila* ovary morphogenesis and the requirement for *bric-a-brac*.** *Development* 1995, **121:**173-187.
34.   Chen W, Zollman S, Couderc JL, Laski FA: **The BTB domain of *bric-a-brac* mediates dimerization *in vitro*.** *Mol Cell Biol* 1995, **15:**3424-3429.
35.   Buchman VL, Luke C, Borthwick EB, Gout I, Ninkina N: **Organization of the *Ruk* locus and expression of isoforms in mouse tissue.** *Gene* 2002, **295:**13-17.
36.   Boube M, Benassayag C, Seroude L, Cribbs DL: **Ras1-mediated modulation of *Drosophila* homeotic function in cell and segment identity.** *Genetics* 1997, **146:**619-628.
37.   Gaul U, Mardon G, Rubin GM: **A putative Ras GTPase activating protein acts as a negative regulator of signaling by the *Sevenless* receptor tyrosine kinase.** *Cell* 1992, **68:**1007-1019.
38.   Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual* 2nd edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
39.   **NCBI BLAST search** 1989 [http://www.ncbi.nlm.nih.gov/BLAST].
40.   Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
41.   **Primer3 software distribution** 1990 [http://www-genome.wi.mit.edu/genome_software/other/primer3.html].
42.   Boeckman F, Brisson M, Tan L: *Real-time PCR: General Considerations.* Hercules: Bio-Rad; 2001.
43.   Giulietti A, Overbergh L, Valckx D, Decallonne B, Bouillon R, Mathieu C: **An overview of real-time quantitative PCR: applications to quantify cytokine expression.** *Methods* 2001, **25:**386-401.
44.   Pfaffl MW, Horgan GW, Dempfle L: **Relative expression software tool (REST®) for group-wise comparison and statistical analysis of relative expression results in real-time PCR.** *Nucleic Acid Res* 2002, **30:**e36.

comment

reviews

reports

deposited research

refereed research

interactions

information