

Research article

Open Access

## Analysis and functional annotation of expressed sequence tags from the fall armyworm *Spodoptera frugiperda*

Youping Deng<sup>1,2</sup>, Yinghua Dong<sup>1</sup>, Venkata Thodima<sup>2</sup>, Rollie J Clem<sup>1</sup> and A Lorena Passarelli\*<sup>1</sup>

Address: <sup>1</sup>Molecular, Cellular, and Developmental Biology Program, Division of Biology, Kansas State University, Manhattan, Kansas 66506, USA and <sup>2</sup>Department of Biological Sciences, The University of Southern Mississippi, Hattiesburg, MS 39402, USA

Email: Youping Deng - youping.deng@usm.edu; Yinghua Dong - yinghua@ksu.edu; Venkata Thodima - venkata.thodima@usm.edu; Rollie J Clem - rclem@ksu.edu; A Lorena Passarelli\* - lpassar@ksu.edu

\* Corresponding author

Published: 19 October 2006

Received: 28 March 2006

BMC Genomics 2006, 7:264 doi:10.1186/1471-2164-7-264

Accepted: 19 October 2006

This article is available from: <http://www.biomedcentral.com/1471-2164/7/264>

© 2006 Deng et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Little is known about the genome sequences of lepidopteran insects, although this group of insects has been studied extensively in the fields of endocrinology, development, immunity, and pathogen-host interactions. In addition, cell lines derived from *Spodoptera frugiperda* and other lepidopteran insects are routinely used for baculovirus foreign gene expression. This study reports the results of an expressed sequence tag (EST) sequencing project in cells from the lepidopteran insect *S. frugiperda*, the fall armyworm.

**Results:** We have constructed an EST database using two cDNA libraries from the *S. frugiperda*-derived cell line, SF-21. The database consists of 2,367 ESTs which were assembled into 244 contigs and 951 singlets for a total of 1,195 unique sequences.

**Conclusion:** *S. frugiperda* is an agriculturally important pest insect and genomic information will be instrumental for establishing initial transcriptional profiling and gene function studies, and for obtaining information about genes manipulated during infections by insect pathogens such as baculoviruses.

### Background

The nucleotide sequences from numerous animal genomes have been derived and include examples from both vertebrate and invertebrate organisms. In addition, determination of the genomic sequences of many more organisms are in progress, yielding a broad picture of the diversity and common pathways among differing organisms. Genome sequences for the insects *Apis mellifera* (honeybee), *Anopheles gambiae* (mosquito), *Drosophila melanogaster* (fruit fly), and *Bombyx mori* (silkworm) have been reported [1-5], and additional insect genome

sequences including *Acyrtosiphon pisum* (pea aphid), *Aedes aegypti* and *Culex pipiens* (mosquitoes), several *Drosophila* species, *Nasonia vitripennis* (parasitoid wasp), *Rhodnius prolixus* (insect vector for *Trypanosoma cruzi*), and *Tribolium castaneum* (red flour beetle) are anticipated. In addition, a number of EST databases derived from Lepidoptera are available (NCBI dbEST). The Lepidoptera (moths and butterflies) are the second largest order of the class Insecta. As such, they are a diverse group of insects distributed worldwide and throughout different climates ranging from that in Siberia to the tropics. Thus, it is

important to compare the genomes of a number of species within the order and to that of other insects.

Lepidoptera are viewed as being among the most beautiful insects, yet their larvae are major pests to economically important crops and forests. Among the Lepidoptera, the silkworm, *B. mori*, has been studied intensively, since it is a commercially important insect. In addition, the Lepidoptera are valued as models to examine insect-plant and insect-pathogen interactions. Their study and comparative genomic analyses will render valuable tools for insect pest management and the improvement of widely used lepidopteran pathogens, the baculoviruses, as foreign gene expression vectors.

In this study, we report the establishment of an expressed sequence tag (EST) database of 1,195 unique sequences from the cell line IPLB-SF-21 (SF-21) [6], a cell line derived from immature ovaries of pupae of the fall armyworm, *Spodoptera frugiperda*. This and other EST databases can serve as the starting point from which other *S. frugiperda* genome clones can be surveyed or to explore gene expression profiles using microarray assays. More importantly, these and additional ESTs can be the basis for comparative genomic analyses among Lepidoptera or other genomes within the Insecta.

**Results**

**General sequence survey**

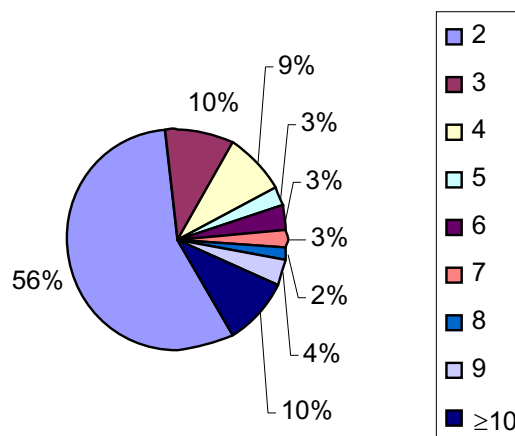
To start characterizing SF-21 sequences, two independent directionally cloned cDNA libraries that had been previously constructed for use in yeast two-hybrid screens using the plasmid vectors pB42AD and pYES2 (Invitrogen) were utilized. Initially, about 200 clones were partially sequenced from each library to assess library quality. Both libraries yielded acceptable results. The library cloned in pYES2 was selected for further sequencing. In all, a total of 3,365 cloned inserts were subjected to single-pass sequencing from their 5' ends, 192 clones in the vector pB42AD and 3,173 clones in the vector pYES2. The 3,365 sequences were trimmed of vector sequences, poly A/T tails, low quality, adaptor, and contaminating bacterial sequences, and screened for a minimum length of 200 bp. This resulted in a total of 2,367 high quality ESTs with an average length of 610 bases (156 sequences from pB42AD and 2,211 sequences from pYES2). No attempt was made to carry out sequencing to saturation. These ESTs were assembled using the CAP3 [7] program and verified using the Phrap [8] program. Both programs assemble overlapping ESTs to commence forming contigs. A total of 1,417 ESTs were assembled into 244 contigs, leaving 951 sequences as singlets. Contigs and singlets together resulted in 1,195 unique sequences that putatively represent different transcripts. The number of ESTs in the 244 contigs varied from 2 to 63; 56% of contig sequences had

two ESTs, 10% had three ESTs, and 10% had greater than 10 ESTs (Fig. 1). The average length of the assembled contigs was 854 bases; longer than the average length of singlets (617 bases). The longest contig sequence, contig 138, was 2,361 bases.

**Highly redundant genes**

A total of 14 contigs containing 307 ESTs were highly redundant. This accounted for more than 13% of the total high quality ESTs. The minimum number of ESTs that made up these most highly redundant contigs was 13 (Table 1). Distribution of ESTs in each contig can be accessed via the ESTMD database [9] using the contig viewer search function. The best matched genes of 6 contigs, totaling 138 ESTs, are from *S. frugiperda*, verifying the source of the cDNAs. Nearly half (6) of the highly redundant contigs, totaling 90 ESTs, had significant homology to various ribosomal proteins, indicating high transcript abundance of ribosomal protein genes, as expected. Four contigs totaling 89 ESTs had matches similar to sequences derived from mitochondrial cytochrome b or cytochrome oxidase subunits (Table 1). The most redundant contig was composed of 63 ESTs and had significant homology to NADH dehydrogenase subunit -1 (ND-1) from *S. frugiperda* [10].

Percentage of contigs with number of ESTs



**Figure 1**  
**Distribution of *S. frugiperda* ESTs.** Percentage distribution of contig sequences with number of ESTs. The color-coded legend indicates the number of ESTs in the contig sequences.

### Comparative sequence analysis of *S. frugiperda* cDNA data

We used the 1,195 unique ESTs to search non-redundant protein databases using BLASTX (Table 2). A total of 724 sequences (60.6%) matched known proteins at a cut-off expectation (E)-value of  $10^{-5}$  or below. Eleven sequences (0.9%) had hits with E-values at  $E < 10^{-150}$ , 53 sequences (4.4%) had hits with E-values between  $10^{-150}$  and  $10^{-100}$ , 283 sequences (23.6%) had hits with E-values between  $10^{-100}$  and  $10^{-50}$ , 237 sequences (19.8%) had hits with E-values between  $10^{-50}$  and  $10^{-20}$ , and 140 sequences (12%) had hits with E-values between  $10^{-20}$  and  $10^{-5}$ . The main matched E-value, between  $10^{-50}$  and  $10^{-20}$ , included 510 searched unique sequences, which was more than 70% of the 724 matched sequences. The remainder of the unique sequences (39.4%) had no meaningful matches ( $E > 10^{-5}$ ).

Given that *Drosophila* is the most thoroughly annotated insect genome [11], we compared *S. frugiperda* unique sequences with *Drosophila* genes using BLASTX. A total of 677 sequences had hits with *Drosophila* genes at  $E < 10^{-5}$ , that is, 56.7% of the 1,195 unique sequences (Table 2). A subset of 53 unique sequences (4.4%) matched to *Drosophila* genes with the cut-off equal to  $E < 10^{-100}$ . A total of 274 sequences (22.9%) had matches with *Drosophila* genes between  $10^{-100}$  and  $10^{-50}$ , 224 sequences (18.7%) had matches between E-values of  $10^{-50}$  and  $10^{-20}$ , and 116 sequences (9.7%) had matches between E-values of  $10^{-20}$  and  $10^{-5}$  (Table 2).

We compared our unique sequences from the SF-21 cell line with ESTs obtained from another *S. frugiperda*-derived cell line, Sf9 [12], using BLASTN [13]. A total of 419 sequences (35%) matched the ESTs from Sf9 cells with an E-value equal to 0 (Table 5). A total of 241 sequences (20.2%) were similar but not exact matches with ESTs from Sf9 with E-values  $> 0$  and  $< 10^{-5}$ . In addition, almost half of our ESTs (535 sequences or 44.8%) had no significant match with the Sf9 ESTs. Therefore, 776 of our sequences (65%) were not previously reported in the Sf9 EST project.

We also compared our sequences with about 4,000 *S. frugiperda* midgut-specific ESTs available in NCBI dbEST database using BLASTN. Only 88 sequences (7.36%) matched with E-value equal to 0 (Table 5). The remaining sequences matched midgut ESTs to different extents: 19 sequences (1.59%) had E-values between 0 and  $10^{-150}$ , 20 sequences (1.67%) had E-values of  $10^{-150}$  and  $10^{-100}$ , 24 sequences (2.01%) had E-values of  $10^{-100}$  and  $10^{-50}$ , 31 sequences (2.59%) had E-values of  $10^{-50}$  and  $10^{-20}$ , and 76 sequences (6.36%) had E-values of  $10^{-20}$  and  $10^{-5}$ . A total of 937 sequences (78.4%) had no hits with the available midgut ESTs.

In addition, we compared our unique sequences with those of the silkworm *B. mori*. We used BLASTN to search the all *B. mori* EST sequences available using a BLAST search site [14], given that the genome sequence is not fully annotated. A total of 492 from the 1,195 unique sequences (41.17%) had hits with silkworm sequences at  $E < 10^{-5}$  (Table 6). Of these, 133 unique ESTs (27% of the 492 sequences) had E-values between  $10^{-100}$  and  $10^{-50}$ . A total of 703 sequences (58.8%) had no matches with silkworm sequences.

### Conserved *S. frugiperda* and *Drosophila* gene sequences

We found 11 highly conserved sequences between *S. frugiperda* and *Drosophila* genes based on BLASTX analyses. All of the 11 sequences were from contigs, with one, contig 134, having an E-value of  $10^{-154}$ . Six sequences had matches with their homologous *Drosophila* genes at an E-value of 0.

We chose contig 134 for phylogenetic analysis given it was the most conserved sequence between *Spodoptera* and *Drosophila*. The sequences from heat shock 70 cognate 4 proteins were aligned with CLUSTALW and only similar sequences with complete coding sequences (CDS) were included in the alignment as described in Methods. Alignments of the heat shock protein 70 cognate 4 (contig 134) with similar ones in the Class Insecta (Fig. 2) showed that the heat shock protein 70 cognate 4 of *S. frugiperda* formed a single clade with *Trichoplusia ni*, *Manduca sexta*, *Bombyx mori*, and *Lonomia oblique*, as expected since all these organisms belong to the order Lepidoptera. This clade shares a common ancestor with members of other orders, Diptera, Orthoptera, and Hymenoptera, and with insects in other clades (e.g., *Ceratitidis capitata*, *Chironomus tentans*, *Drosophila melanogaster*, *Anopheles gambiae*, *Locusta migratoria*, and *Cotesia rubecula*).

### Functional classification of *S. frugiperda* ESTs

Gene Ontology (GO) has been widely used to characterize gene function annotation and classification [15]. GO describes gene function using controlled vocabulary and hierarchy including molecular function, biological processes, and cellular communication. In this report, we used well annotated GO information of *Drosophila melanogaster* to interpret the gene function of our ESTs. Each unique sequence from *S. frugiperda* was assigned the same gene function of the best BLASTX hit ( $E \leq 10^{-5}$ ) with *Drosophila* sequences based on the annotated GO of *Drosophila* [15]. This method has been successfully used to annotate bee brain EST function [16].

The major GO categories for the unique sequences included those outlining gene molecular function (Additional file 1- Table 7), biological processes (Additional file 2- Table 8), and cellular components (Additional file 3-

**Table 1: Most abundantly represented transcripts in the *Spodoptera frugiperda* cDNA library.**

Contig	ESTs	GI#	Bit score	E-value	Identities	Gene descriptions	Organism
Contig 190	13	40363707	238	1e-61	129/215	cytochrome oxidase II	<i>Glyphodes bicolor</i>
Contig 98	13	18314310	241	5e-96	122/159	Cytochrome c oxidase subunit 3	<i>Ostrinia furnacalis</i>
Contig 116	13	16566722	458	e-128	229/244	ribosomal protein S3A	<i>S. frugiperda</i>
Contig 146	13	18253045	138	5e-32	76/112	60s acidic ribosomal protein P2	<i>S. frugiperda</i>
Contig 61	14	7302066	249	6e-65	129/243	CG11522-PB, isoform B	<i>D. melanogaster</i>
Contig 225	14	18253043	147	1e-34	78/111	60s acidic ribosomal protein P1	<i>S. frugiperda</i>
Contig 70	15	54609281	452	e-126	233/307	ribosomal protein SA	<i>B. mori</i>
Contig 139	17	27260896	411	e-113	204/218	ribosomal protein S2	<i>S. frugiperda</i>
Contig 160	17	22094837	389	e-106	199/283	Cytochrome b	<i>Samia cynthia ricini</i>
Contig 23	18	18253041	550	e-155	283/315	60 Saccadic ribosomal protein PO	<i>S. frugiperda</i>
Contig 134	21	12585261	1119	0	568/608	Heat shock 70 kDa cognate 4	<i>Manduca sexta</i>
Contig 239	30	39752635	480	0	234/241	elongation factor-1 alpha F2	<i>D. melanogaster</i>
Contig 141	46	1438928	685	0	364/504	Cytochrome oxidase subunit I	<i>Feltia jaculifera</i>
Contig 19	63	552886	226	e-111	102/104	ND-I protein gene	<i>S. frugiperda</i>

Table 9). The highest final child GO term for molecular function was the hydrogen transporting two-sector ATPase in the nucleotide binding category. The highest final GO term for biological processes was protein biosynthesis, under the protein metabolism and biosynthesis categories, which had 84 unique sequences accounting for 7% of the total unique sequences matched in this category. The largest number for a final GO term in cellular components was cytosolic large ribosomal subunit under both the ribosome and cytosol categories. Seventy-eight unique sequences belonged to this GO term, which accounted for 6.5% of the total unique sequences annotated for cellular components.

We found 13 unique sequences (1.1%) showing significant similarity with *Drosophila* signal transduction factors (Table 3). Among these, 6 sequences belonged to the receptor binding category and the remaining 7 sequences belonged to receptor and receptor signaling proteins.

Based on GO, we also found one sequence for an apoptosis-related gene, pyes2-ct\_017\_g10.p1ca, which showed

similarity to the *Drosophila Aac11* gene. Two additional sequences, pb42ad-1\_001\_f09.pb42 primer and pyes2-ct\_010\_g11.p1ca, showed significant similarity to *Drosophila Gnbp3*, a gene involved in defense and immunity.

**Pathway analysis based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification**

KEGG has been widely used for pathway mapping [17]. Enzyme commission (EC) numbers were used to judge which sequences pertained to a specific pathway. Twenty-nine unique sequences including 8 contigs and 21 singlets accounted for 2.4% of unique sequences and matched enzymes with an EC number. Within these 29 sequences, 11 unique sequences (1% of total) containing 4 contigs and 7 singlets were mapped to KEGG biochemical pathways (Table 4). Genes involved in amino acid metabolism had the highest mapping sequences (5) and 6 mappings. Genes involved in carbohydrate metabolism included 5 sequences and 8 mappings. Other genes included those involved in nucleotide metabolism (2 sequences), translation (2 sequences), energy metabolism (1 sequence), lipid metabolism (1 sequence), and metabolism of other

**Table 2: Comparative analysis of *Spodoptera* ESTs to *Drosophila*<sup>1</sup> and other sequences.**

Homology	Contig		All matches		Total		Contig		Drosophila		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
E ≤ 10 <sup>-150</sup>	11	5	0	0	11	2	10	5	0	0	10	1
E ≤ 10 <sup>-100</sup>	30	15	23	4	53	7	30	16	23	5	53	8
E ≤ 10 <sup>-50</sup>	92	45	191	37	283	39	88	46	186	38	274	40
E ≤ 10 <sup>-20</sup>	48	24	189	36	237	33	45	24	179	37	224	33
E ≤ 10 <sup>-5</sup>	22	11	118	23	140	20	17	9	99	20	116	17
Total matched	203	91	521	55	724	61	190	85	487	50	677	57
No match	41	9	430	45	471	39	54	22	484	50	518	43
<b>Total</b>	<b>244</b>	<b>100</b>	<b>951</b>	<b>100</b>	<b>1195</b>	<b>100</b>	<b>244</b>	<b>100</b>	<b>951</b>	<b>100</b>	<b>1195</b>	<b>100</b>

<sup>1</sup> [11]

**Table 3: Signal transduction sequences of *Spodoptera frugiperda* compared with *Drosophila* genes.**

<i>S. frugiperda</i> sequences	Flybase number	Hit length	Bit score	E-value	Identities	<i>Drosophila</i> gene	Gene description
pyes2-ct_019_b03.plca	FBgn0039541	836	275	1.00e-74	138/266	<i>Cg12876</i>	Signal transduction activity
pyes2-ct_006_f12.plca	FBgn0035771	753	360	e-104	164/231	<i>Cg8583</i>	Signal recognition particle binding
pyes2-ct_021_a12.plca	FBgn0027363	689	84	1.00e-17	35/59	<i>Stam</i>	Signal transducing adaptor molecule
Contig 14	FBgn0003963	1191	139	2.00e-33	86/223	<i>ush</i>	Involved in torso signaling pathway
pyes2-ct_005_g01.plca	FBgn0035771	753	222	1.00e-58	104/149	<i>cg8583</i>	Involved in signal recognition particle complex
pyes2-ct_006_f09.plca	FBgn0037277	2228	314	3.00e-86	152/242	<i>Cg17735</i>	Ligand-dependent nuclear receptor interactor activity
pyes2-ct_028_g06.plca	FBgn0020618	318	199	1.00e-51	96/107	<i>Rack1</i>	Receptor of activated protein kinase C I
Contig 140	FBgn0020618	318	578	e-165	274/319	<i>Rack1</i>	Receptor of activated protein kinase C I
pyes2-ct_030_g06.plca	FBgn0004569	444	67	4.00e-12	29/43	<i>argos</i>	Receptor antagonist activity
p42ad_2_001_b07.plcb.exp	FBgn0037113	1258	133	3.00e-32	71/131	<i>cg33291</i>	Putative protein binding
pyes2-ct_003_e12.plca	FBgn0013984	2144	73	1.00e-13	74/313	<i>lnR</i>	Insulin like receptor
Contig 220	FBgn0031547	406	125	3.00e-43	65/191	<i>CG3212</i>	Scavenger receptor activity involved in defense response
Contig 226	FBgn0037357	773	102	7.00e-25	52/62	<i>sec23</i>	Putative GTPase activator activity

amino acids (1 sequence). Contigs 35 and 97 each had 4 mappings to different pathways under a different metabolism category, which was the highest number for a single sequence. Contig 120, contig 152, pyes2-ct\_008\_e11.plca, and pyes2-ct\_012\_c04.plca mapped to two pathways. The remaining pathway-assigned sequences only mapped to one pathway.

**EST database**

To efficiently manage and retrieve information in the ESTs analyzed in this project, we developed an EST model database (ESTMD version 1) [9]. The EST model database is a highly efficient, web-accessible, user-friendly relational database. It provides several tools to search raw, cleaned, and assembled EST sequences, genes and GO, as well as pathway information. The user may input and submit keywords or IDs to the server using the web interface. ESTs and annotated function data are in the relational database

**Table 4: KEGG pathway mapping for *Spodoptera frugiperda* unique sequences.**

KEGG pathway	Sequence ID	Number of sequences	Percentage of total
Carbohydrate metabolism		4	36
Glycolysis/gluconeogenesis	Contig 35, contig 97, pyes2-ct_008_e11.plca	3	27
Fructose and mannose metabolism	pyes2-ct_027_b04.plca	1	9
Pyruvate metabolism	Contig 35, contig 97	2	18
Propanoate metabolism	Contig 35	1	9
Starch and sucrose metabolism	pyes2-ct_010_g11.plca	1	9
Energy metabolism		1	9
Carbon fixation	Contig 97	1	9
Lipid metabolism		1	9
Sphingoglycolipid metabolism	pyes2-ct_010_a06.plca	1	9
Nucleotide metabolism		2	18
Purine metabolism	Contig 97, contig 120	2	18
Amino acid metabolism		5	45
Alanine and aspartate metabolism	Contig 120	1	9
Arginine and proline metabolism	pyes2-ct_026_h11.plca	1	9
Glycine, serine and threonine metabolism	pyes2-ct_012_c04.plca	1	9
Cysteine metabolism	Contig 35	1	9
Phenylalanine, tyrosine and tryptophan biosynthesis	Contig 152, pyes2-ct_008_e11.plca	2	18
Metabolism of other amino acids		1	9
Selenoamino acid metabolism	pyes2-ct_021_f10.plca	1	9
Translation		2	18
Aminoacyl-tRNA biosynthesis	Contig 152, pyes2-ct_012_c04.plca	2	18

**Table 5: Comparative analysis of *Spodoptera frugiperda* SF-21-derived ESTs with Sf9- and midgut-derived ESTs<sup>1</sup>.**

Homology	<i>S. frugiperda</i> Sf9ESTs						<i>S. frugiperda</i> midgutESTs					
	Contig		Singlets		Total		Contig		Singlets		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
E ≤ 0	243	99.59	176	18.51	419	35.06	7	2.87	81	8.52	88	7.36
0 < E ≤ 10 <sup>-150</sup>	0	0.00	37	3.89	37	3.10	2	0.82	17	1.79	19	1.59
10 <sup>-150</sup> < E ≤ 10 <sup>-100</sup>	0	0.00	47	4.94	47	3.93	4	1.64	16	1.68	20	1.67
10 <sup>-100</sup> < E ≤ 10 <sup>-50</sup>	0	0.00	58	6.10	58	4.85	0	0.00	24	2.52	24	2.01
10 <sup>-50</sup> < E ≤ 10 <sup>-20</sup>	0	0.00	49	5.15	49	4.10	1	0.41	30	3.15	31	2.59
10 <sup>-20</sup> < E ≤ 10 <sup>-5</sup>	0	0.00	50	5.26	50	4.18	4	1.64	72	7.57	76	6.36
Total matched	243	99.59	417	43.85	660	55.23	18	7.38	240	25.24	258	21.59
No match	1	0.41	534	56.15	535	44.77	226	92.62	711	74.76	937	78.41
Total	244	100	951	100	1195	100	244	100	951	100	1195	100

<sup>1</sup> SF-21 ESTs, this report; Sf9 ESTs [12]; midgut ESTs (NCBI dbEST)

and results are sent back to the user in proper formats in response to a query. It also provides a clear contig view and BLAST searches, data submission, and download pages. In both the GO and KEGG pathway search pages, ESTMD allows users to search GO and pathways not only by single gene name, symbol, or ID, but also by using a file that contains a batch of sequence IDs or FlyBase IDs. All the sequence function classification based on GO and KEGG pathway in this study was done using ESTMD. The file search provision in ESTMD allows users to obtain information regarding the possible function of many ESTs or genes at one time instead of searching them individually (Fig. 3).

**Discussion**

Single-pass sequencing was performed on 3,365 cDNA clones derived from two SF-21 cDNA libraries. Using this sequence data, we have established an EST database comprised of 1,195 unique sequences from the SF-21 cell line, derived from the lepidopteran insect *S. frugiperda*. A total of 667 unique sequences (57%) had homology to

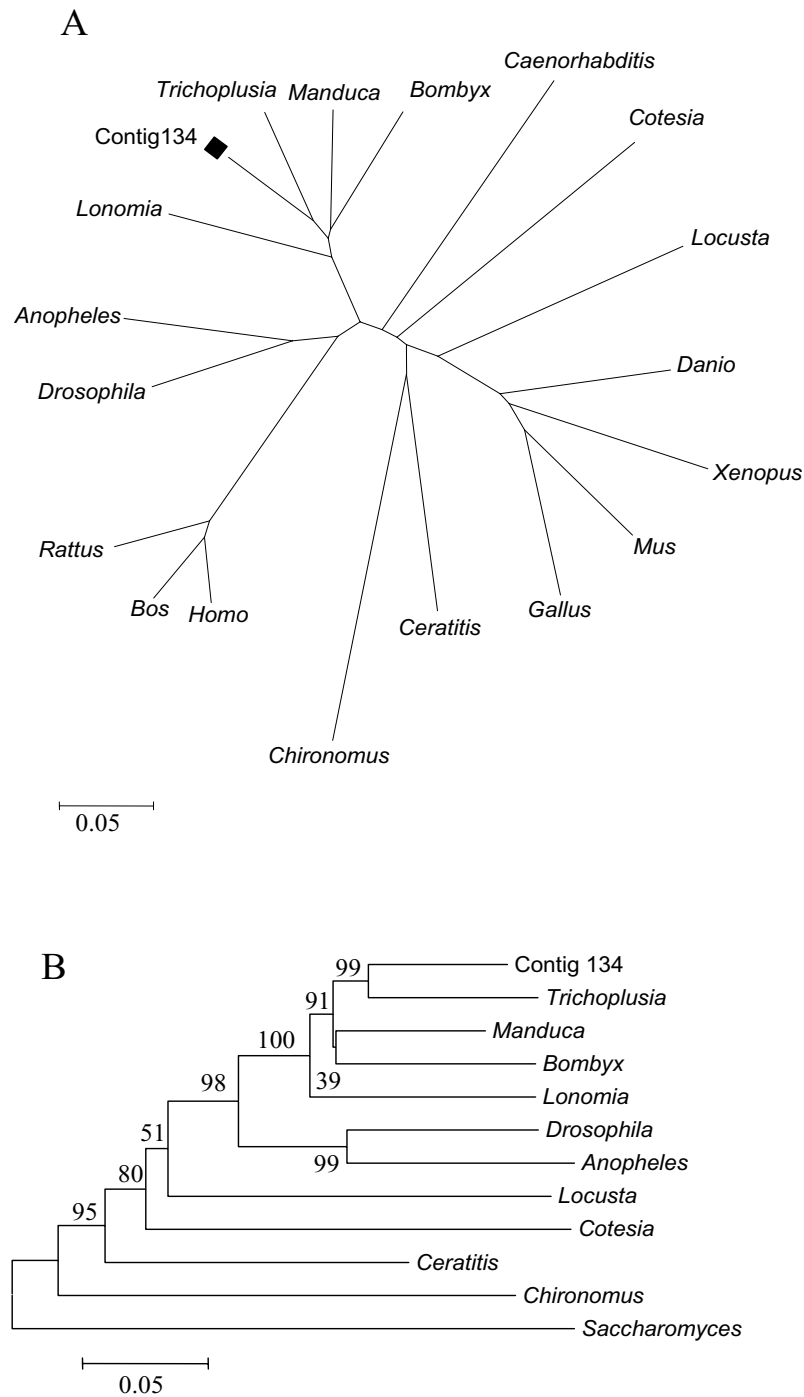
sequences found in *Drosophila*. These sequences will be useful for comparative genomics within and outside the Lepidoptera, establishing microarrays, and as probes to either clone or down-regulate genes of interest by RNA interference in order to perform studies related to *Spodoptera*, other closely related Lepidoptera, or their pathogens.

This is one of two published annotated EST studies available for *S. frugiperda*. A prior report included a similar scale project as that reported here with 5,937 ESTs of which 1,855 were unique sequences obtained from a clonally-derived cell line of SF-21, Sf9 [12]. The majority of unique sequences in the previous study consisted of the highly abundant ribosomal protein genes and these were found to have low codon usage bias [12]. Our data provides 776 novel *S. frugiperda* sequences. A small percentage of our sequences (20.2%) had similarities, whether these reflect cell line specific differences is not clear at this time. Together these two studies and other available *S. frugiperda* ESTs constitute seminal work on the genome sequence of *S. frugiperda*. The sequences reported in this

**Table 6: Comparative analysis of *Spodoptera frugiperda* SF-21 ESTs with silkworm ESTs.**

Homology	<i>Bombyx mori</i> ESTs					
	Contig		Singlets		Total	
	N	%	N	%	N	%
E ≤ 0	53	21.72	7	0.74	60	5.02
0 < E ≤ 10 <sup>-150</sup>	17	6.97	12	1.26	29	2.43
10 <sup>-150</sup> < E ≤ 10 <sup>-100</sup>	51	20.90	60	6.31	111	9.29
10 <sup>-100</sup> < E ≤ 10 <sup>-50</sup>	44	18.03	89	9.36	133	11.13
10 <sup>-50</sup> < E ≤ 10 <sup>-20</sup>	13	5.33	79	8.31	92	7.70
10 <sup>-20</sup> < E ≤ 10 <sup>-5</sup>	13	5.33	54	5.68	67	5.61
Total matched	191	78.28	301	31.65	492	41.17
No match	53	21.72	650	68.35	703	58.83
Total	244	100	951	100	1195	100

<sup>1</sup> SF-21 ESTs, this report; Sf9 ESTs [12]; midgut ESTs (NCBI dbEST)



**Figure 2**  
**Phylogenetic analysis with neighbor-joining tree.** A. The heat shock proteins of 10 insects and 8 other organisms (*Cotesia rubecula*, *Ceratitidis capitata*, *Chironomus tentans*, *Manduca sexta*, *Locusta migratoria*, *Drosophila melanogaster*, *Anopheles gambiae*, *Lonomia oblique*, *Bombyx mori*, *Trichoplusia ni*, *Bos taurus*, *Gallus gallus*, *Rattus norvegicus*, *Danio rerio*, *Xenopus laevis*, *Caenorhabditis elegans*, *Mus musculus*, *Homo sapiens*) along with Contig 134 (heat shock 70 cognate 4 protein) are presented in unrooted phylogenetic analyses. B. Phylogenetic tree showing heat shock proteins with *Saccharomyces cerevisiae* as the outgroup. The bootstrap values (percentages) are indicated at the corresponding node.

study have been made available for incorporation into Spodibase [18].

Many insects within the Lepidoptera, including the fall armyworm *S. frugiperda* (family Noctuidae), are pests that cause significant annual damage to a number of field crops and tree foliage worldwide. Deciphering their genomic sequences will aid in developing improved pest control agents, such as baculoviruses and polydnviruses/parasitic wasps. Although these pathogens are being used or sought as biological control agents, there remains ample room for improvement of their entomopathogenic properties.

Finally, molecular tools have been used in the study of Lepidoptera or have been derived from Lepidoptera. The transposable element *piggyBac* was discovered in the lepidopteran *T. ni* (cabbage looper) and has been used to create somatic and germline transformations in a number of organisms including crickets, butterflies, *Plasmodium falciparum*, and more recently, mice [19-22]. Lepidoptera are also amenable to down-regulation of genes by RNA interference and transgenic techniques [23]. Thus, knowledge of the genomics of Lepidoptera will aid in their manipulation or use as molecular tools.

## Conclusion

We have established an EST database from the *S. frugiperda*-derived cell line SF-21, containing 1,195 unique sequences. Lepidoptera are among the most diverse insects and as such, sequences and EST databases from various genomes will be instrumental in assessing species-specific genes, phylogeny, and parallels within species of the same order. In addition, comparative analyses with available genomes of other insects including *A. mellifera*, *D. melanogaster*, *A. gambiae*, *Ae. aegypti*, and *T. castaneum* will yield additional insights since these include members of distinct orders (Hymenoptera, Diptera, and Coleoptera), providing a more accurate picture of the conserved pathways and the order-specific gene elements in the Insecta.

## Methods

### cDNA library construction

Two independently constructed cDNA libraries were used for sequencing. For both libraries, mRNA isolated from log phase SF-21 cells was used for cDNA synthesis, and the libraries were directionally cloned into plasmid vectors. One library was custom made by Clontech using the plasmid vector pB42AD. The pB42AD library had a titer of  $3.6 \times 10^{13}$  colony forming units per ml. The second library was constructed using the SuperScript™ Plasmid System (Invitrogen) and the plasmid vector pYES2/CT (Invitrogen) that had been modified by addition of a *Sal* I linker at the *Bam* HI site. The pYES2/CT library had a titer of  $1.2$

$\times 10^{12}$  colony forming units per ml. The average insert size for both libraries was 1.5 kbp.

### EST sequencing

Initially, approximately 200 randomly selected clones from each library were subjected to single-pass sequencing using 5' vector primers. DNA sequencing was performed by MWG Biotech (High Point, NC). Although both libraries yielded acceptable sequence quality, the pYES2/CT library appeared to yield slightly longer sequences; consequently, the remainder of the sequencing was performed using clones from the pYES2/CT library.

### Sequence processing

Sequence information was stored in chromatograph trace files, and Phred [24] was used to perform Base-calling [24]. Flanking vector and adaptor sequences were trimmed using Cross-match [25] and Lucy [26], while low quality bases (quality score < 20) were cleaned at both sequence ends by using our custom program. RepeatMasker [27] was used to mask repeated sequences, and the masked sequences were further screened to remove contaminating sequences from bacteria and viruses using BLASTN [13]. High quality ESTs were assembled using CAP 3 [7] and verified with Phrap [8], which perform similar tasks. After assembly, Consed [24] was used to assess contig quality, and assembled ESTs were chosen for further analysis. Contigs flagged for possible miss-assembly were manually edited in Consed and potential chimeric ESTs or other suspect ESTs were removed from the pool of traces.

### Sequence annotation

High quality assembled ESTs were annotated using BLASTX through NCBI and our local BLAST server. We searched several databases including the NCBI non-redundant and *Drosophila* protein databases. The BLAST results were automatically extracted and transferred into a relational database. The sequences reported in this study (2,367 ESTs) have been deposited in GenBank under accession numbers [GenBank: [DY792773](#) to [DY795139](#)].

### Functional classification

Functional classification of unique sequences from *S. frugiperda* was based on GO [15]. Unique sequences, including contigs and non-overlapping singlets, were used to search *Drosophila* predicted protein databases using BLASTX. The *Drosophila* genes corresponding to the best hits at a threshold of E-value  $\leq 10^{-5}$  with known GO term were assigned to the query "*Spodoptera* sequences". All the matched GO information was stored in our local MySQL database.



# ESTMD

## EST Model Database

[Introduction](#) | [Search in Details](#) | [Search by Keyword](#) | [Gene Ontology](#) | [GO Classification](#) | [Pathway](#) | [Blast](#) | [Downloads](#) | [Data Submission](#) | [Help](#) | [Contacts](#)

### GO Sequence Search Result

SequenceID	GO_ID	GO_Term	Type	HitID	Gene_Symbol
<a href="#">Contig1</a>	GO:0003735	structural constituent of ribosome	molecular_function	FBgn0004413	RpS26
<a href="#">Contig1</a>	GO:0006412	protein biosynthesis	biological_process	FBgn0004413	RpS26
<a href="#">Contig1</a>	GO:0005843	cytosolic small ribosomal subunit (sensu Eukarya)	cellular_component	FBgn0004413	RpS26
<a href="#">Contig1</a>	GO:0005840	ribosome	cellular_component	FBgn0004413	RpS26
<a href="#">Contig10</a>	GO:0006412	protein biosynthesis	biological_process	FBgn0011272	RpL13
<a href="#">Contig10</a>	GO:0005842	cytosolic large ribosomal subunit (sensu Eukarya)	cellular_component	FBgn0011272	RpL13
<a href="#">Contig10</a>	GO:0005840	ribosome	cellular_component	FBgn0011272	RpL13
<a href="#">Contig10</a>	GO:0003735	structural constituent of ribosome	molecular_function	FBgn0011272	RpL13
<a href="#">Contig106</a>	GO:0004772	sterol O-acyltransferase	molecular_function	FBgn0031327	CG5397
<a href="#">Contig35</a>	GO:0006096	glycolysis	biological_process	FBgn0001258	ImpL3
<a href="#">Contig35</a>	GO:0004459	L-lactate dehydrogenase	molecular_function	FBgn0001258	ImpL3

**Figure 3**  
**Gene Ontology search results.** The search results used a sequence file by choosing all three ontologies.

#### Pathway assignments

Pathway assignments were carried out according to KEGG mapping [17]. EC numbers [28] were assigned to unique sequences that had BLASTX scores with a cut off value of  $E = 10^{-5}$  or less upon searching SWIR protein databases. The sequences were mapped to KEGG biochemical pathways according to the EC distribution in the pathway database.

#### Phylogenetic analysis

Proteins were aligned with CLUSTALW using only CDS in the alignment. The alignment was then used to generate phylogenetic trees by the Neighbour-Joining method using the MEGA version 2.1 program. The bootstrap values for the nodes were determined by analyzing 500 bootstrap replicate data sets to estimate the strength of the groupings.

#### Database implementation

A web-based interface of the database was created using HTML and JavaScript to evaluate the validation of the input on the client side and to reduce the burden on the server side. Apache 2.0 was used as the http web server, while Tomcat 4.1 was the servlet container. Both of these programs were developed and maintained on UNIX, Linux, and Windows NT, ensuring that ESTMD was transplantable and platform-independent. ESTMD is currently hosted on Red Hat 9, and it can be implemented in

MySQL 4.0 or higher version. The main tables were on clones, ESTs, uniSequence, uniHit, FlyBase and FlyBase-Details. The server-side programs were implemented by Java technologies. Servlet and JavaServer Pages were used to communicate between users and databases and to implement a query. XML and XSLT technologies were used to describe, generate, and express GO trees.

#### Authors' contributions

ALP and RJC contributed to the conception and design of the study, coordinated the study, and were involved in drafting the manuscript. YDeng participated in data analysis and drafting of the manuscript. YDong and VT were involved in data analysis. All authors read and approved the final manuscript.

#### Additional material

##### Additional file 1- Table 7

Table 7. Distribution of molecular functional categories based on gene ontology for Spodoptera frugiperda unique sequences

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-264-S1.pdf>]

**Additional file 2- Table 8**

Table 8. Distribution of biological processes based on gene ontology for *Spodoptera frugiperda* unique sequences

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-264-S2.pdf>]

**Additional file 3- Table 9**

Table 9. Distribution of cellular component categories based on gene ontology for *Spodoptera frugiperda* unique sequences

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-264-S3.pdf>]

**Acknowledgements**

We thank the late Lois K. Miller and Casey W. Wright for providing the cDNA libraries, Yonghua Li for help with the database, Vijayaraj Nagarajan for help with phylogenetic analysis, and Kuan Yang for help with sequence comparisons.

This work was supported in part by the NIH National Center for Research Resources awards P20 RR16443, P20 RR107686, P20 RR16475, and P20RR016476. Y. Deng was also supported by the Dean's Research Initiative award of the University of Southern Mississippi. This is contribution number 06-273-J from the Kansas Agricultural Experiment Station.

**References**

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, S. R. Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YHC, Blazej RG, Champe M, Pfiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Miklos GLG, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Bernam BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Candra I, Cherry JM, Cawley S, Dahlke C, Daenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Days AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabriellian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarrey C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RDC, Scheeler F, Shen H, Shue BC, Sidenkiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao A, Ye J, Yeh RF, Zaveri JS, Zhang M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Runbin GM, Venter JC: **The genome sequence of *Drosophila melanogaster***. *Science* 2000, **287**:2185-2195.
- Biology analysis group: Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, Li C, Shen Y, Lan X, Yuan L, Li T, Xu H, Yang G, Wan Y, Zhu Y, Yu M, Shen W, Wu D, Xiang Z, Genome analysis group: Yu J, Wang J, Li R, Shi J, Li H, Li G, Su J, Wang X, Li G, Zhang Z, Wu Q, Li J, Zhang Q, Wei N, Xu J, Sun H, Dong L, Liu D, Zhao S, Zhao X, Meng Q, Lan F, Huang X, Li Y, Fang L, Li C, Li D, Sun Y, Zhang Z, Yang Z, Huang Y, Xi Y, Qi Q, He D, Huang H, Zhang X, Xi Y, Qi R, He D, Huang H, Zhang X, Wang Z, Li W, Cao Y, Yu Y, Yu H, Li J, Ye J, Chen H, Zhou Y, Liu B, Wang J, Ye J, Ji H, Li S, Ni P, Zhang J, Zhang Y, Zheng H, Mao B, Wang W, Ye C, Li S, Wang J, Wong GKS, Yang H: **A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*)**. *Science* 2004, **306**:1937-1940.
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anhouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscut D, Barnstead M, Cai S, Center A, Chaturvedi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarrey C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL: **The genome sequence of the malaria mosquito *Anopheles gambiae***. *Science* 2002, **298**:129-149.
- Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Kamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin-i T, Abe H, Shimada T, Morishita S, Sasaki T: **The genome sequence of silkworm, *Bombyx mori***. *DNA Res* 2004, **11**:27-35.
- Project HBG: [<http://www.hgsc.bcm.tmc.edu/projects/honeybee/>].
- Vaughn JL, Goodwin RH, Tompkins GJ, McCawley P: **The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera: Noctuidae)**. *In Vitro* 1977, **13**:213-217.
- Huang X, Madan A: **CAP3: A DNA Sequence Assembly Program**. *Genome Research* 1999, **9**:868-877.
- Nickerson DA, Tobe VO, Taylor SL: **PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing**. *Nucleic Acids Res* 1997, **25**:2745-2751.
- database ESTMD: [<http://www.bioinformatics.ksu.edu:8080/estweb/index.html>].
- Pashley DP, Ke LD: **Sequence evolution in mitochondrial ribosomal and ND-1 genes in Lepidoptera: Implications for phylogenetic analyses**. *Mol Biol Evol* 1992, **9**:1061-1075.
- Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnick SE, Smith CD, Tupy JL, Whitfield EJ, Bayraktaroglu L, Berman BP, Bettencourt BR, Celniker SE, de Grey AD, Drysdale RA, Harris NL, Richter J, Russo S, Schroeder AJ, Shu SQ, Stapleton M, Yamada C, Ashburner M, Gelbart WM, Rubin GM, Lewis SE: **Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review**. *Genome Biol* 2002, **3**:RESEARCH 0083.1-83.22.
- Landais I, Oligastro M, Mita K, Nohata J, López-Ferber M, Duonon-Cerutti M, Shimada T, Fournier P, Devauchelle G: **Annotation pattern of ESTs from *Spodoptera frugiperda* Sf9 cells and analysis of the ribosomal protein genes reveal insect-specific features and unexpectedly low codon bias**. *Bioinformatics* 2003, **19**:2343-2350.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403-410.
- Genomes BLASTSAS: [<http://pistol.ab.a.u-tokyo.ac.jp/kanzen/blast.html>].
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nat Genet* 2000, **25**:25-29.
- Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE: **Annotated expressed**

- sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res* 2002, **12**:555-566.
17. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto Encyclopedia of Genes and Genomes**. *Nucleic Acids Res* 1999, **27**:29-34.
  18. Spodobase: [<http://bioweb.enscm.inra.fr/spodobase/>].
  19. Balu B, Shoue DA, Fraser Jr. M], Adams JH: **High-efficiency transposition of Plasmodium falciparum by the lepidopteran transposable element piggyBac**. *Proc Natl Acad Sci USA* 2005, **102**:16391-16396.
  20. Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T: **Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice**. *Cell* 2005, **122**:473-483.
  21. Marcus JM, Ramos DM, Monteiro A: **Germline transformation of the butterfly *Bicyclus anynana***. *Proc R Soc Lond B Biol Sci* 2004, **271**:S263-S265.
  22. Shinmyo Y, Mito T, Matsushita T, Sarashina I, Miyawaki K, Ohuchi H, Noji S: **piggyBac-mediated somatic transformation of the two-spotted cricket, *Gryllus bimaculatus***. *Dev Growth Differ* 2004, **46**:343-349.
  23. Bettencourt R, Terenius O, Faye I: **Hemolin gene silencing by ds-RNA injected into *Cecropia* pupae is lethal to next generation embryos**. *Insect Mol Biol* 2002, **11**:267-271.
  24. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities**. *Genome Res* 1998, **8**:186-194.
  25. Crossmatch: [<http://www.sanger.ac.uk/Software/>].
  26. Chou HH, Holmes MH: **DNA sequence quality trimming and vector removal**. *Bioinformatics* 2001, **17**:1093-1104.
  27. RepeatMasker: [<http://ftp.genome.washington.edu/>].
  28. IUBMB: **Enzyme nomenclature: Recommendations of the nomenclature committee of the international union of biochemistry and molecular biology**. San Diego , Academic Press; 1992.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

