

Table 1. Differential expression of EST contigs during the *A. gambiae* life cycle

	La	Lb	Lc	Ld	Le	P	F	M
E	663 (492)	678 (513)	672 (513)	816 (626)	627 (489)	818 (626)	624 (505)	1,009 (759)
La		14 (12)	19 (18)	68 (59)	81 (65)	279 (232)	272 (232)	506 (384)
Lb			0 (0)	49 (42)	61 (51)	294 (241)	281 (238)	572 (435)
Lc				25 (24)	37 (31)	226 (192)	214 (191)	454 (359)
Ld					13 (11)	333 (296)	313 (273)	541 (434)
Le						172 (159)	273 (238)	466 (377)
P							265 (224)	245 (206)
F								179 (156)

Numbers represent differentially expressed T-contigs in each pair-wise comparison (ANOVA Tukey test, $P \leq 0.001$). The number of contigs that display at least 2-fold difference is shown in parentheses.

We performed three biological and one technical (dye-swap) replicates. Embryonic, pupal, female and male adult replicates, and each of the adult tissues, displayed high reproducibility of T-contig expression (SI Fig. 5C). Expression profiles at different larval periods tended to cluster by mosquito rearings rather than time period, possibly reflecting physiological differences between generations and a predominant, inherent similarity of development in larval instars. This is consistent with the continuous growth of larval and imaginal disk tissues and with the recurrence of developmental phases within each larval instar (e.g., a shift from growth to hormonally induced epithelial retraction from the old cuticle and formation of a new cuticle). We did not attempt to define separately within instar phases.

Gene Expression Differences During Development. Statistically significant expression differences between developmental stages were established for 2,421 T-contigs with a one-factor ANOVA (SI Fig. 5D, $P = 0.001$). Of these, 1,571 displayed at least 2-fold differences between their respective minimum and maximum expression (Table 1 and SI Data Set 1) and were used thereafter. Embryos were the most distinct stage, displaying 624–1,009 differentially expressed contigs in pair-wise comparisons (Tukey test) with each of the other stages. Adult males were also very distinctive differing from larvae by 454–572 contigs, as compared with 214–313 differentially expressed contigs between females and larvae and 245–265 between pupae and adults. Approximately 5% of the pupae/adults differentially expressed T-contigs encode proteasome components (reflecting the extensive histolysis of larval tissues during pupal life), whereas 5% encode components of the cuticle (which is obviously different between pupae and adults). Larval periods were mostly indistinct (0–81 differences) but early larvae (La-c) tended to cluster apart from late larvae (Ld-e). It is known that precursors of adult organs (imaginal discs) develop continuously in larvae, with slow cell divisions at early instars and faster cell divisions subsequently. Most of the T-contigs that differed between larval stages encode proteins implicated in nucleic acid binding, protein metabolism, or cuticular constituents.

Developmental Programs. The expression profiles of the 1,571 contigs (encompassing 1,065 Ensembl genes and 783 *Drosophila* orthologs) were grouped by self-organizing maps (SOM) into 30 coexpression clusters with 5×6 node geometry (Fig. 1). Additional nodes did not reveal novel patterns, and fewer nodes yielded loose clusters. We considered each cluster individually, but, because several were closely related, we describe them in a consolidated manner according to their constituents and broad developmental dynamics (see also SI Table 3 and SI Data Set 2).

The embryo high program (EH) encompasses four clusters with strong expression in embryos. Six major functional gene classes predominate: replication, transcription, mRNA processing and regulation, cell cycle, signal transduction, cell growth,

and metabolism. *Drosophila* orthologs of many EH contigs are well known transcriptional embryonic regulators, or are implicated in mRNA splicing and downstream processing. The presence of ubiquitin domain sequences suggests a role for ubiquitination in mosquito embryonic development (8). Two EH clusters, 2 and 12, are deficient in Ensembl gene models.

Six clusters share low embryonic expression and are further

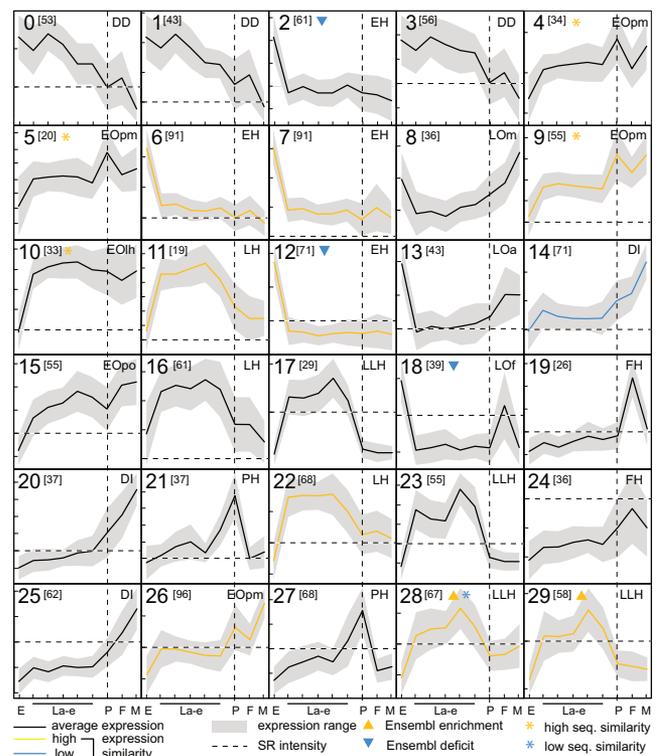


Fig. 1. Coexpression clusters and developmental transcription programs. During the mosquito life cycle, 1571 T-contigs that display at least 2-fold difference between their maximum and minimum expression are grouped into 30 SOM coexpression clusters. Developmental program designations are at the top right of each cluster. Numbers in brackets refer to clustered contigs, solid lines refer to average expression, and gray areas refer to range of expression. y axis scale shows increments of 0.5 in log₁₀-transformed expression values, horizontal dashed lines indicate SR signal levels, and vertical dotted lines point to the pupal stage. Arrowheads indicate clusters enriched (yellow) or deficient (blue) in Ensembl gene models. Average expression is plotted in black if the mean expression similarity of mosquito–fruitfly orthologous contigs within a cluster is comparable with that of all orthologous pairs and is plotted in yellow or blue if it is statistically above or below that standard, respectively. Yellow or blue asterisks show clusters of genes displaying CDS similarity to their fruitfly orthologs that is statistically above or below that of all orthologous pairs, respectively.

distinguished by subtler features. The single-cluster embryo low/pupa low (EOpo) program shows reduced expression in pupae, whereas EOpm (four clusters) is highly expressed in pupae and adult males, the latter suggesting expression in the testis (9). EOlh (one cluster) is different in showing high expression in larvae continuing into the pupa. In all EO programs, genes involved in metabolic reactions are prominent (overall total 25%), as are numerous proton- or electron-transport components; other contigs encode components of immune responses or cytoskeletal elements and regulators.

Two programs, LH and LLH (three clusters each), are named for their high larval or late larval expression. They differ from EOlh because of more detailed features. LH accelerates in early instars but declines gradually between Ld and pupa, reflecting growth in the larval body but not in imaginal discs, where growth remains modest. However, LLH peaks at Ld and then sharply declines between Le and pupa. LH is associated with 30% prevalence of proteins implicated in metabolic reactions (carbohydrate and lipid metabolism and proteolysis), whereas LLH is enriched in metabolic (especially proteolytic) enzymes but also cuticle components and putative defense proteins.

The female high program (FH, two clusters) is defined by prominent expression in adult females but not in males. It is enriched in putative immune components, suggesting adaptation to increase survival of the almost completely monogamous females (polygamous males are more dispensable).

The developmentally increasing program (DI, three clusters) displays low expression in embryos and larvae, followed by a strong progressive increase in pupae and adults (especially males). After emergence from the pupal cuticle, adult organs such as salivary glands, midgut epithelium, and flight and orientation organs continue to develop. Adult maturation begins earlier in males than in females, explaining the observed sex-specific expression difference (10, 11). DI encompasses proteins involved in sensory perception including odorant-binding proteins and members of the rhodopsin signaling pathway, as well as antimicrobial peptides, other immune-related proteins, and digestive enzymes.

The pupa high expression program (PH, two clusters) peaks in pupa, consistent with the restructuring of the mosquito body at metamorphosis, when many larval tissues histolyze, whereas adult structures develop, some of them acquiring an adult cuticle. Indeed, many PH genes encode structural and enzymatic components of the cuticle, including *yellow* family members that control pigmentation; others are implicated in ubiquitination and proteolysis, suggesting a role in histolysis.

Genes engaged in the developmentally declining program (DD, three clusters) display an overall progressive, steady decline in expression (including a detectable trough at the pupal stage); this is sharply different from the abrupt decline seen in EH. Almost 20% of DD sequences encode components of protein biosynthesis, modification, and folding; others control DNA, RNA, and nucleotide synthesis.

Finally, the larva low program, LO, is characterized by high expression in embryos and adults and encompasses three distinct expression clusters: LOa (equal expression in both adult sexes), LOm (higher in males), and LOf (higher in females). Many components of LOf have *Drosophila* orthologs often associated with maternal effects on embryos or involvement in asymmetric mRNA or protein localization.

Developmental Expression of Gene Functional Categories. We grouped T-contigs by functional categories (GO or INTERPRO domain annotation) and, for each category, determined the percentage of contigs displaying top or bottom expression at each time period. Bottom percentages were subtracted from top percentages, and resulting data sets (representing the expression

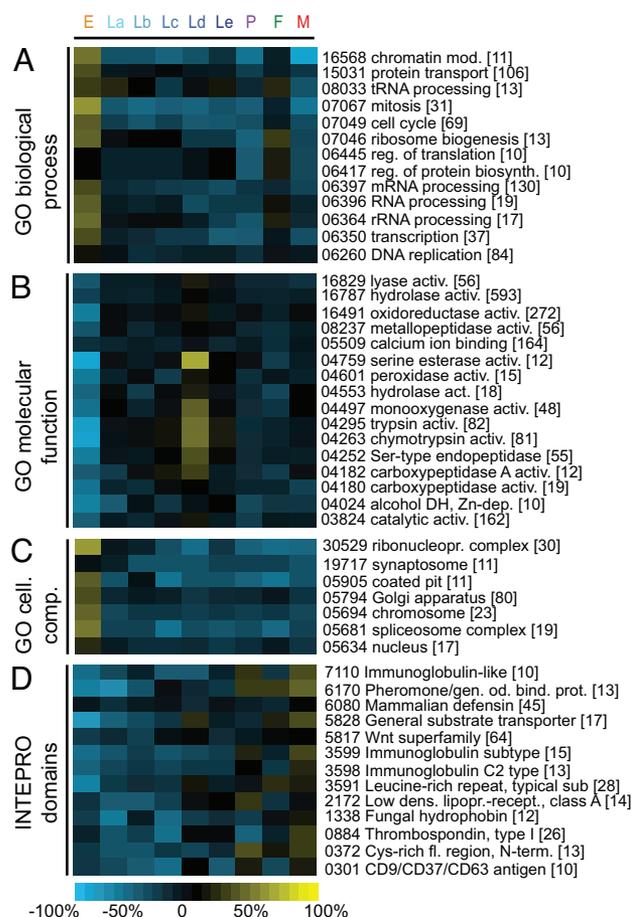


Fig. 2. Coordinated expression of related functional gene groups. Selected *k*-means clusters from the analysis of GO biological processes (A), INTERPRO domains (B), GO molecular functions (C), and GO cellular components (D) are presented. For each functional group, the percentage of contigs showing the lower 25% of the expression range at each time period was subtracted from the respective percentage of contigs showing the upper 25%, and resulting values (ranging from blue to yellow) were used for clustering. Numbers on the right (omitting the GO or IPR prefixes) denote functional group identifier; numbers in brackets indicate the size of each functional group in contigs.

tendency of each functional category) were subjected to *k*-means clustering (Fig. 2 and SI Fig. 6).

Many contigs associated with nuclear processes, i.e., DNA replication, transcription and RNA processing, or cell-cycle processes display top expression in embryos but bottom expression in much of the remaining life cycle. These contigs mostly map to the EH program and highlight fundamental postfertilization events such as rapid succession of cell cycles associated with chromatin replication and initiation of transcription and translation for embryo patterning.

Numerous catabolic reactions show bottom expression in embryo, pupa, and adults and top expression in the Ld period, as do contigs mapping to specific subcellular organelles, e.g., microsomes, peroxisomes, and lysosomes. Catabolic clusters probably serve histolysis of larval tissues at the onset of metamorphosis. Several proteins implicated in immune reactions are associated with bottom expression in embryos and larvae but top expression in pupae and adults, indicating enhanced immune system activity during and after metamorphosis. They map mainly to the DI and EO programs, presumably reflecting adaptation to the increased infection risk of pupae and adults. Some hydrolytic functions are associated with immunity factors such as GNBP (hydrolase activity) and CLIPs (trypsin and

Table 2. Differential expression of EST contigs in adult female tissues

	Carcass	Gut	Ovaries
Head	295 (271)	172 (156)	370 (322)
Carcass		226 (212)	332 (302)
Gut			177 (161)

Numbers represent differentially expressed T-contigs in each pair-wise comparison (ANOVA Tukey test, $P \leq 0.001$). The number of contigs that display at least 2-fold difference is shown in parentheses.

chymotrypsin-like activities), suggesting that immunity may have evolved in catabolic, gut-associated components that were in persistent contact with gut biota. Catabolic and numerous immunity genes are notable in the LH and, especially LLH, programs.

Coexpression Patterns in Adult Female Tissues. We identified 898 T-contigs exhibiting differential expression between at least two tissues (SI Fig. 5D, Table 2, and SI Data Set 3). Of these, 829 contigs exceeding 2-fold difference were subjected to *k*-means clustering and grouped into 10 coexpression patterns; six patterns are surprisingly specific, containing sequences overexpressed in practically single tissue (Fig. 3, SI Table 3 and SI Data Set 4).

The insect head carries the major sensory organs, the vision center and endocrine glands. Expression in the head includes two distinctive patterns (0 and 1) encompassing contigs of three major functional categories: rhodopsins and visual perception, odorant-binding proteins and pheromone-related proteins. A *Drosophila* allatostatin homolog, an adult brain peptide that

blocks the synthesis of the developmental juvenile hormone, is included.

The midgut is the primary organ for nutrient absorption, synthesis and secretion of digestive enzymes, and formation of the gut-lining peritrophic membrane. It also has an endocrine role and contributes to diuresis, e.g., after a bloodmeal when blood cells are concentrated before being digested. Indeed, a quarter of T-contigs in the two midgut-specific patterns (2–3) are associated with metabolic reactions; some with previously identified midgut specific expression, and others with domains involved in vasoconstriction and diuresis.

Aside from genes putatively involved in defense reactions, annotation information is sparse for the carcass-enriched patterns 4 and 5. The surprising duality of head and carcass association in cluster 5 may be related to the presence of fat body, stationary hemocytes, and hypodermis in both body parts.

Patterns 6 and 7 encompass contigs that are strongly expressed in ovaries but differ in that cluster 7 also shows substantial expression in the carcass. Many of these contigs are associated with transcriptional programs FH and LOF; 25% are implicated in transcription regulation, translation and mRNA processing. Some contigs encode odorant-binding proteins, suggesting unorthodox functions that merit further analysis.

The triple-tissue (midgut, carcass, and ovary) expression pattern 8 is enriched in genes with metabolic functions. Annotation suggests housekeeping processes, because 20% of the genes are implicated in general polysaccharide and fatty acid metabolic reactions and 16% in protein synthesis and degradation.

Finally, the four-body-part pattern 9 differs from pattern 8 in showing the most pronounced expression in the head. It contains housekeeping genes from diverse functional classes, e.g., electron and proton transport, polysaccharide metabolism, signal transduction, etc.

Comparative Transcriptomics of the *Anopheles* and *Drosophila* Life Cycles.

Half of the genes in the *Anopheles* and *Drosophila* genomes are identified as 1:1 orthologs (12). This, in conjunction with an earlier transcriptional study of *Drosophila* development (13), allowed us to compare the developmental expression profiles of orthologs in the two insects, after normalization of the respective experimental designs to create comparable notional developmental phases. Pearson and smooth correlation analysis of the expression similarity of 1,039 orthologous genes showed a drastic shift toward positive correlation (Fig. 4A). A similar shift was detected with the nonparametric Spearman coefficient (data not shown). When the same data set was randomly rearranged 100 times to generate nonorthologous gene pairs, no shift was detected, and both average distributions were largely symmetric.

We queried whether the expression similarity of orthologs was due to specific gene sets, such as particular developmental programs, tissue patterns, or functional groups. As shown in Fig. 4B and SI Tables 4–10, coexpression clusters belonging to six developmental programs or tissue patterns (EH, EOpm, DI, LH, LLH, and ovary-enriched) and two functional groups (nuclear localization and protein folding) showed significant positive or negative deviations from the median expression similarity (Wilcoxon *U* test ≤ 0.005). Next, we examined whether the degree of expression similarity between orthologs varies in parallel with their coding sequence (CDS) similarity. Such a global connection was not detected (SI Fig. 7). However, when the analysis matrix was disarticulated into coexpression clusters and functional categories (SI Tables 4–10) insights were revealed. The EOpm developmental cluster 9 displayed coherent positive deviations of both CDS and expression similarities, whereas the LLH developmental cluster 28 and the nuclear localization group showed opposite deviations, negative for CDS and positive for expression similarity.

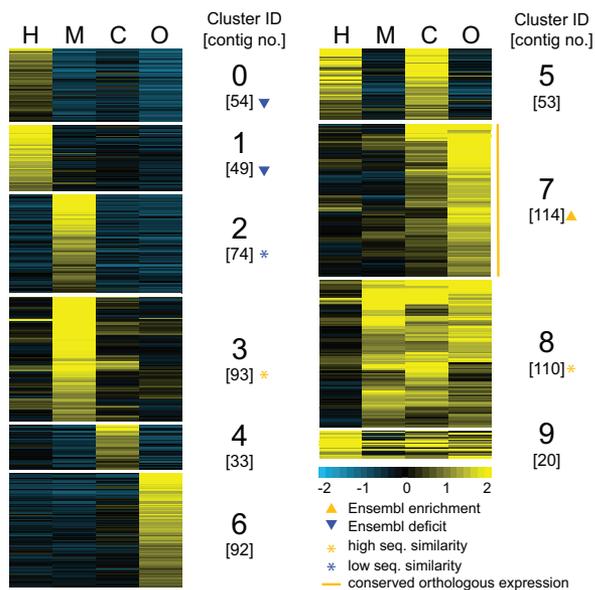


Fig. 3. Tissue expression patterns in adult females. *k*-means clusters of 829 contigs showing at least 2-fold regulation between two or more tissues are presented; 477 correspond to Ensembl genes, and 318 have *D. melanogaster* orthologs. Numbers in brackets indicate cluster size (in contigs) and the scale bar represents \log_2 -transformed gene expression values. Yellow and blue arrowheads indicate clusters enriched or deficient in Ensembl gene models, respectively. The ovary-enriched cluster 7 contains contigs with mean expression similarity to their fruitfly orthologs that is statistically above the mean of all orthologous pairs. Yellow or blue asterisks show clusters of genes with CDS similarity to their fruitfly orthologs that is statistically above or below that of all orthologous pairs, respectively.

evolutionary pressures exerted on CDS and expression properties of orthologs during the 250 myr separating these two insects.

The degree of orthologous expression similarity varies between different developmental programs or functional groups. In general, orthologs engaged in early developmental events such as egg fertilization, embryo formation, and body patterning display highly similar expression, whereas some later events engage genes of significantly lower similarity in expression. The success of insects is based in part on their great diversification for different niches in adult life. Retention of successful genetic solutions for early development may have provided a stable background, on which innovations for the adult life could be accepted with minimal disruption. However, this trend allows for exceptions: orthologs in the late EOpm program show high expression similarity as well as high CDS similarity. Apparently, adult male differentiation (probably for development of the male gonad) is highly conserved in insects. Our study points toward a new concept of defining corresponding genes: expression-based “orthoregulation” alongside sequence-based orthology.

Materials and Methods

Biological Material. Approximately 400 laboratory-reared *A. gambiae* mosquitoes of the G3 strain were fed at days 3 or 8 of adulthood on CD1 mice and produced the experimental generations P₁ and P₂. The P₃ generation was the progeny of P₁ mated females fed on mice 3 days after emergence. Adult head, gut, and carcass (freed of head, gut, wings, and legs) tissues were collected from 1-day-old females; ovaries were from 5- to 6-day-old females that were bloodfed 48 h earlier.

Analysis of Functional Groups. Gene Ontology (GO) terms and INTEPRO domains represented in at least 20 distinct T-contigs were assembled into functional groups. For each group and time period, the percentage of T-contigs showing bottom expression (the lower 25% of each contig’s overall expression range) was subtracted from the percentage of T-contigs showing top (upper 25%) expression, and resulting values were subjected to *k*-means clustering.

Comparative Transcriptomic Analysis. The raw microarray data from a *D. melanogaster* life cycle study (13) were downloaded from [http://genome.med.yale.edu/Life cycle/Data_download](http://genome.med.yale.edu/Life%20cycle/Data_download) and

analyzed by using the same criteria as for the *A. gambiae* data set, except for the negative spike-in control criterion. Expression profiles of ESTs were averaged to the respective gene, and data were normalized as for *Anopheles*. The 3,571 genes with reliable measurements in at least 130 of 151 hybridizations and with a *t* test *P* value <0.05 in at least 1 of 75 developmental time periods were considered further.

The *Anopheles* and *Drosophila* data sets were divided into seven notional developmental periods. *Anopheles* male and female stages were compared with 24-h-old *Drosophila* adult males (Am24h) and females (Af24h), respectively. Based on correlation analysis (data not shown), the mosquito embryo period was compared with the average of *Drosophila* embryonic time points E056–E0112 and the mosquito pupa period with the average of *Drosophila* metamorphosis time points M04–M12. By using a sliding window procedure, three (early, middle, and late) comparable phases of larval development were defined from the *Anopheles* (La–Lc, Lb–Ld, and Lc–Le) and the *Drosophila* (L24–L57, L43–L84, and L67–L105) studies.

Orthologous gene pairs were constructed from best reciprocal hits and information from syntenic regions (12). The combined expression matrix of orthologs was normalized to the median of each gene and the 50th percentile of each notional time period. Correlation analysis with Pearson, smooth, and Spearman coefficients was performed in GeneSpring.

Methods for EST construction, sequencing and clustering, microarray construction, preparation of experimental and SR RNA samples and hybridizations, imaging, and data analysis of microarrays are provided in *SI Materials and Methods*.

We thank E. Furlong for the updated version of the *Drosophila* microarray annotation, M. Gross for suggestions about microarray analyses, and L. Ettwiller for discussions about regulatory sequence-finding algorithms. This work was supported by National Institute of Allergy and Infectious Diseases–National Institutes of Health (NIAID/NIH) Grants U01AI48846 and 2P01AI044220-07, European Commission Research and Training Networks Grant HPRN-CT-2000-00080 and the European Molecular Biology Laboratory. It was also part of the activities of the BioMalPar European Network of Excellence supported by European Grant LSHP-CT-2004-503578 from the Priority 1 “Life Sciences, Genomics and Biotechnology for Health” in the 6th Framework Program. This work was greatly facilitated by the AnoEST Project, which was funded by NIAID/NIH VectorBase Contract NIAID-DMID-04-34.

- Monnerat AT, Machado MP, Vale BS, Soares MJ, Lima JB, Lenzi HL, Valle D (2002) *Mem Inst Oswaldo Cruz* 97:589–596.
- Goltsev Y, Hsiong W, Lanzaro G, Levine M (2004) *Dev Biol* 275:435–446.
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, et al. (2002) *Science* 298:129–149.
- Dimopoulos G, Christophides GK, Meister S, Schultz J, White KP, Barillas-Mury C, Kafatos FC (2002) *Proc Natl Acad Sci USA* 99:8814–8819.
- Dana AN, Hong YS, Kern MK, Hillenmeyer ME, Harker BW, Lobo NF, Hogan JR, Romans P, Collins FH (2005) *BMC Genomics* 6:5.
- Marinotti O, Nguyen QK, Calvo E, James AA, Ribeiro JM (2005) *Insect Mol Biol* 14:365–373.
- Kriventseva EV, Koutsos AC, Blass C, Kafatos FC, Christophides GK, Zdobnov EM (2005) *Genome Res* 15:893–899.
- Daniel JA, Torok MS, Sun ZW, Schieltz D, Allis CD, Yates JR, III, Grant PA (2004) *J Biol Chem* 279:1867–1871.
- Belyakin SN, Christophides GK, Alekseyenko AA, Kriventseva EV, Belyaeva ES, Nanayev RA, Makunin IV, Kafatos FC, Zhimulev IF (2005) *Proc Natl Acad Sci USA* 102:8269–8274.
- de Meillon B, Sebastian A, Khan ZH (1967) *Bull World Health Org* 36:7–14.
- Haddow AJ, Gillett JD, Corbet PS (1959) *Ann Trop Med Parasitol* 53:123–131.
- Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, Copley RR, Christophides GK, Thomasova D, Holt RA, Subramanian GM, et al. (2002) *Science* 298:149–159.
- Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP (2002) *Science* 297:2270–2275.
- Vlachou D, Schlegelmilch T, Christophides GK, Kafatos FC (2005) *Curr Biol* 15:1185–1195.
- Ma L, Chen C, Liu X, Jiao Y, Su N, Li L, Wang X, Cao M, Sun N, Zhang X, et al. (2005) *Genome Res* 15:1274–1283.
- Jiao Y, Ma L, Strickland E, Deng XW (2005) *Plant Cell* 17:3239–3256.
- Stearman RS, Dwyer-Nield L, Zerbe L, Blaine SA, Chan Z, Bunn PA, Jr, Johnson GL, Hirsch FR, Merrick DT, Franklin WA, et al. (2005) *Am J Pathol* 167, 1763–1775.
- McCarroll SA, Murphy CT, Zou S, Pletcher SD, Chin CS, Jan YN, Kenyon C, Bargmann CI, Li H (2004) *Nat Genet* 36:197–204.
- Goodisman MA, Isoe J, Wheeler DE, Wells MA (2005) *Evol Int J Org Evol* 59:858–870.
- Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS (2002) *Bioinformatics* 18:405–412.
- Tan PK, Downey TJ, Spitznagel EL, Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC (2003) *Nucleic Acids Res* 31:5676–5684.
- Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, Bumgarner RE, Bushel PR, Chaturvedi K, Choi D, et al. (2005) *Nat Methods* 2:351–356.
- Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, et al. (2005) *Nat Methods* 2:345–350.
- Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J (2005) *Nat Methods* 2:337–344.