ARTICLES

The genome of the model beetle and pest *Tribolium castaneum*

Tribolium Genome Sequencing Consortium*

Tribolium castaneum is a member of the most species-rich eukaryotic order, a powerful model organism for the study of generalized insect development, and an important pest of stored agricultural products. We describe its genome sequence here. This omnivorous beetle has evolved the ability to interact with a diverse chemical environment, as shown by large expansions in odorant and gustatory receptors, as well as P450 and other detoxification enzymes. Development in *Tribolium* is more representative of other insects than is *Drosophila*, a fact reflected in gene content and function. For example, *Tribolium* has retained more ancestral genes involved in cell–cell communication than *Drosophila*, some being expressed in the growth zone crucial for axial elongation in short-germ development. Systemic RNA interference in *T. castaneum* functions differently from that in *Caenorhabditis elegans*, but nevertheless offers similar power for the elucidation of gene function and identification of targets for selective insect control.

By far the most evolutionarily successful metazoans[1], beetles (Coleoptera) can luminesce (fireflies), spit defensive liquids (bombardier beetles), visually and behaviourally mimic bees and wasps, or chemically mimic ants that detect intruders by their foreign odour. Many beetles (for example, boll weevil, corn rootworm, Colorado potato beetle and Asian longhorn beetle) are associated with billions of dollars of agricultural and natural resource losses.

The red flour beetle, *Tribolium castaneum*, found wherever grains or other dried foods are stored, has a highly evolved kidney-like cryptonephridial organ to survive such extremely dry environments. It has demonstrated resistance to all classes of insecticides used against it. Like all beetles, *Tribolium* has elytra (wing covers) that coordinate precisely with folding wings, allowing flight while providing protection.

Tribolium facilitates genetic analysis with ease of culture, a short life cycle, high fecundity, and facility for genetic crosses (see ref. 2), allowing efficient genetic screens by means of chemical mutagens, radiation and binary transposon systems[3]. As in *Caenorhabditis elegans*, RNA interference (RNAi) is systemic in *Tribolium*, facilitating knockdown of specific gene products in any tissue, developmental stage or offspring of double-stranded (ds)RNA-injected females[4,5].

Particularly favoured for developmental studies, *Tribolium* is much more representative of other insects than is *Drosophila*[6]. In contrast to *Drosophila*, *Tribolium* larvae display eyes in a fully formed head and three pairs of thoracic legs (Supplementary Fig. 1). In addition, *Tribolium* develops via short-germ embryogenesis where additional segments are sequentially added from a posterior growth zone (Supplementary Fig. 1). This proliferative mechanism of segmentation differs from the *Drosophila* model, but resembles that of vertebrates and basal arthropods such as millipedes[7].

Genome sequence and organization

Approximately 1.52 million sequence reads ($7.3\times$ coverage) were generated from the highly inbred Georgia 2 (GA2) strain and assembled into contigs totalling 152 megabases (Mb) and scaffolds spanning ~160 Mb of genomic sequence (Supplementary Tables 1–4 and Supplementary Information). Almost 90% of this sequence was mapped to the ten *Tribolium* linkage groups using a genetic map of ~500 markers generated from the GA2 strain[8]. Excluding heterochromatic regions dense in highly repetitive sequences, the genome is well represented and of high quality (see Supplementary Data for details).

G+C content. *Tribolium*, like *Apis*, has a very (A+T)-rich genome (33% and 34% G+C, respectively), but *Tribolium* G+C domains lack the extremes of G+C content present in *Apis mellifera* (Fig. 1 and Supplementary Fig. 3). Despite global G+C similarity to *Apis*, genes in *Tribolium*, as in *Anopheles* and *Drosophila* but not *Apis*, show a bias towards occurring in (G+C)-rich regions of the genome (Fig. 1). Whatever mechanism drives the accumulation of A+T nucleotides in *Tribolium*, it does not affect genes in the manner observed in the honeybee, where perhaps additional mechanisms are present.

Repetitive DNA. Fully one-third of the *Tribolium* genome assembly consists of repetitive DNA, which is also (A+T)-rich. Compared to other insects, there is a paucity of microsatellites (1–6-base-pair (bp) motifs) in *Tribolium*[9]. However, *Tribolium* contains a relative excess of larger satellites, including several with repeat units longer than 100 bp (2.5% of the *Tribolium* genome compared with 0.7% in *Drosophila*). Most (83%) of the microsatellites are found in intergenic regions (63%) or introns (20%), but there is strong over-representation of non-frameshift-causing repeats (3- and 6-bp motifs) due to a dearth of dinucleotide repeats (see Supplementary Information). Of 981 randomly chosen microsatellites, 509 (55.2%) are polymorphic in a sample of 11 *Tribolium* populations from around the world[9], providing an extensive collection of markers for population studies. Preliminary efforts to assess global population structure show a shallow but significant correlation between geographic and genetic distance (Supplementary Fig. 4). This suggests that anthropogenic dispersal may maintain a modest level of gene flow across vast distances in this human commensal.

Transposable elements. Transposable elements and other repetitive DNA accumulate in regions along each linkage group that resemble the pericentric blocks of heterochromatin visible in *Hpa*II-banded chromosomes[10]. These regions are probably composed largely of highly repetitive heterochromatic sequences, and represent most of the 44-Mb difference between the estimated genome size (0.2 pg or

204 Mb[11]) and the current assembly (160 Mb). Indeed, as much as 17% of the *Tribolium* genome is composed of a 360-bp satellite[12] that constitutes only 0.3% of the assembled genome sequence. Several families of DNA transposons, as well as long terminal repeat (LTR) and non-LTR retrotransposons, constituting approximately 6% of the genome, were identified via encoded protein sequence similarity to previously identified elements using TEPipe or BLAST, and are listed in Supplementary Table 5.

Telomeres. *Tribolium* has a telomerase and telomeres containing TCAGG repeats[13], a variant of the standard arthropod TTAGG telomeric repeat. Manual assembly of the proximal regions of multiple telomeres beyond the ends of the assembled scaffolds (Supplementary Information) reveals TCAGG repeats interrupted by full-length and 5′-truncated non-LTR retrotransposons belonging to the R1 clade, best known for insertions in the rDNA locus[14]. *Tribolium* telomeres range in length from 15 kilobases (kb) upwards and probably represent a stage intermediate to the loss of telomeres and telomerase in Diptera compared with the simple canonical structure of the honeybee[15] or the more regular insertion of non-LTR retrotransposons into the simple repeats of the silkmoth[16].

Gene content and the proteome

Comparative gene content analysis. To understand the consensus set of 16,404 gene models in the context of other available insect and vertebrate genomes, all genes were classified according to their degree of similarity using systematic cross-species analysis. Five insects (*Drosophila melanogaster, Anopheles gambiae, Aedes aegypti, T. castaneum, A. mellifera*) and five vertebrates (*Homo sapiens, Mus musculus, Monodelphis domestica, Gallus gallus, Tetraodon nigroviridis*) with similar phylogenetic branching orders were chosen for the comparison. We found the fractions of universal and insect-specific orthologues in *Tribolium* similar to other insect genomes, as expected, whereas the number of genes without similarity is

considerably higher (Fig. 2), possibly attributable to less stringent gene prediction.

Over 47% of *Tribolium* genes (7,579) are ancient, with traceable orthologous relations between insects and vertebrates including 15% (2,403) universal single-copy orthologues. Another 1,462 *Tribolium* genes (9%) constitute the core of what are currently insect-specific orthologues. In comparison, 21% (4,937) of human genes have vertebrate-specific orthologues.

Several hundred ancient genes seem to be under limited evolutionary selection and were independently lost in several species studied (the patchy fraction, defined in Fig. 2). Each new genome uncovers previously invisible ancestral relations among genes—for example, as many as 126 orthologous gene groups shared between *Tribolium* and humans seem to be absent from the other sequenced insect genomes (Fig. 3 and Supplementary Table 10), 44 of which are single-copy genes present in all vertebrates.

The evolutionary emergence of many predicted *Tribolium* genes is not clear. Thousands of genes currently appear to be species-specific as either no sequence similarity to other genes is detectable, or homology but not orthology can be determined. Reassuringly, this fraction is similar in *Tribolium* and *Drosophila*.

We quantified the species phylogeny using a maximum likelihood approach with the concatenated multiple alignment of 1,150 universal single-copy orthologues present in all the organisms studied—an ideal genome-wide data set of essential genes evolving under similar constraints (Fig. 2 and Supplementary Fig. 6). This analysis confirmed previous analyses based on expressed sequence tag (EST) sequences that the Hymenoptera are basal within the Holometabola[17]. The shorter branch length for *Tribolium* implies that the elevated rate of evolution observed in *Drosophila* and *Anopheles* occurred more recently[18].
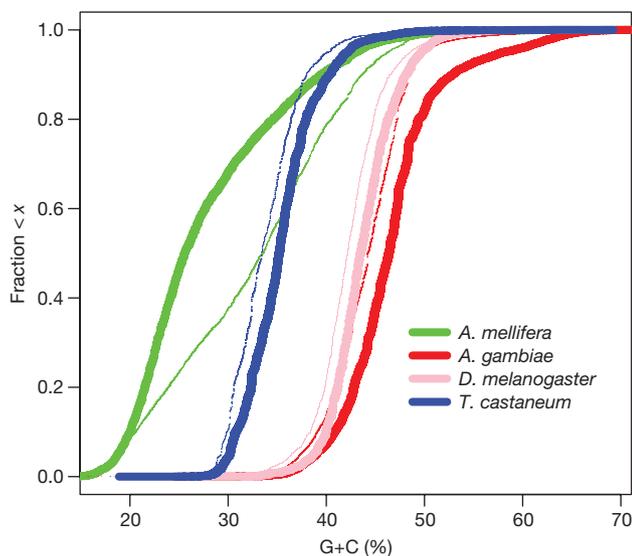
Figure 1 | Cumulative distribution of genic and genomic G+C-content domains in *Apis mellifera, Anopheles gambiae, Drosophila melanogaster* and *Tribolium castaneum*. Cumulative distributions show the fraction of genes (thick lines) or of the entire genome (thin lines) occurring in G+C-content domains less than a given percentage G+C (<*X*). The more (A+T)-rich half of the *T. castaneum* genome contains only 30.8% of all *T. castaneum* genes (31.4% and 33% of *A. gambiae* and *D. melanogaster* genes, respectively), whereas the more (A+T)-rich half of the *A. mellifera* genome contains 77.6% of its genes. At every point on the *T. castaneum, A. gambiae* and *D. melanogaster* curves there are fewer genes present in the fraction of the genome less than a given percentage G+C than would be expected if the genes were randomly distributed. In contrast, *A. mellifera* exhibits the opposite distribution.
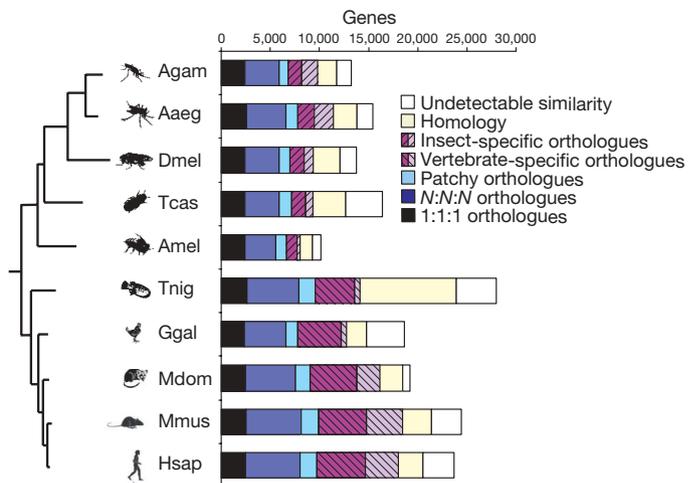
Figure 2 | Insect gene orthology. Comparison of the gene repertoire in five insect and five vertebrate genomes, ranging from the core of metazoan genes (dark blue fraction on the left) to the species-unique sequences (white band on the right). The striped boxes correspond to insect- and vertebrate-specific orthologous genes, where the darker bands correspond to all insects or vertebrates (allowing one loss). *N:N:N* indicates orthologues present in multiple copies in all species (allowing one loss); patchy indicates ancient orthologues (requiring at least one insect and one vertebrate gene) that have become differentially extinct in some lineages. The species tree on the left (shown in detail in Supplementary Fig. 6) was computed using the maximum-likelihood approach on concatenated sequences of 1,150 universal single-copy orthologues. It shows an accelerated rate of evolution in insects and confirms the basal position of the Hymenoptera within the Holometabola[17]. Aaeg, *Aedes aegypti*; Agam, *Anopheles gambiae*; Amel, *Apis mellifera*; Dmel, *Drosophila melanogaster*; Ggal, *Gallus gallus*; Hsap, *Homo sapiens*; Mdom, *Monodelphis domestica*; Mmus, *Mus musculus*; Tcas, *Tribolium castaneum*; Tnig, *Tetraodon nigroviridis*.

Gene family expansion, frequently associated with a particular adaptation pressure, might reveal physiologically and phenotypically unique features of beetles (Supplementary Table 9 and protein family discussions below). Many duplications shaped the gene content of *Tribolium*, most notably among odorant-binding proteins and the CYP450 subfamilies CYP6 and CYP9 (Supplementary Fig. 11), some of which are involved in the development of insecticide resistance in the Diptera[19]. Duplication of genes under copy-number selection in other species is indicative of species-specific neo-functionalization[20]. At least 152 genes duplicated in *Tribolium* have single-copy status in all other insects studied, including sevenfold duplication of genes orthologous to *Drosophila* CG1625, encoding a putative structural constituent of cytoskeleton, and human ENSP00000269392, encoding centrosomal pre-acrosome localization protein 1.

We also analysed the phylogenetic distribution of orthologous gene group members to quantify evolutionary gene losses[21]. Although least affected, dozens of single-copy orthologues seem to be lost in each lineage. Thirty-eight such genes lost in *Tribolium* include rather unique genes, encoding phosphotriesterase-related protein and peroxisome assembly factor 1 (peroxin-2), compared to 59 such genes lost in *Drosophila*. Notably, for the less restricted fractions of orthologues (defined in Fig. 2), several hundred gene orthologues have been lost in each species.

Analysis of specific gene sets

In addition to a global automated analysis of the predicted *Tribolium* gene set, the consortium manually annotated and analysed ~2,000 genes (some additionally subjected to RNAi and expression analysis), focusing on developmental processes and genes of importance for agriculture and pest management.

Development

We identified and analysed homologues of known insect and vertebrate developmental genes to gain novel insights into the molecular basis of developmental differences between *Drosophila* and *Tribolium*. Supplementary Table 11 lists selected *Tribolium* developmental genes and their *Drosophila* and *Apis* orthologues.

Oogenesis. Despite profound differences in ovarian architecture—telotrophic versus polytrophic—we identified *Tribolium* orthologues

Figure 3 | Orthologous genes shared between insect and human genomes. The Venn diagram shows the number of orthologous groups of genes shared between the insect and human genomes. In addition to the majority of Urbilateria (last common ancestor of the Bilateria) genes shared by all the organisms, there are hundreds of genes that have been lost in some lineages (for example, only retained between human and *Tribolium* or human and honeybee, but lost in Diptera). Diptera is represented here by *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster* (with numbers considering only *D. melanogaster* shown in parentheses).

of most *Drosophila* genes required for stem cell maintenance, RNA localization and axis formation. Like *Apis*, however, *Tribolium* lacks a *bag of marbles* orthologue, which is essential for the differentiation of cystoblast versus germline stem cells in *Drosophila*[22]. Interestingly, an orthologue of the gene *gld-1*, which fulfils a similar function in *C. elegans*[22], is present in *Tribolium*.

Anterior–posterior patterning. Analysis of the genome sequence confirmed the absence of a *bcd* orthologue in *Tribolium*. Instead, anterior patterning is synergistically organized by *otd* and *hb* (ref. 23). However, it is still unclear how the posterior gradient of *Tribolium* Caudal is shaped in the absence of Bicoid. Notably, *Tribolium* contains an orthologue of *mex-3*, a factor that translationally represses the *C. elegans cad* homologue[24]. Although the *Tribolium* genome contains orthologues of the *Drosophila* segmentation genes, their functions are not entirely conserved[25–27]. Furthermore, the genome reveals the unexpected polycistronic organization of a novel gap gene, *mille-pattes*[28], the transcript of which encodes several short peptides.

In contrast to the classical protostomian model organisms *Drosophila* and *Caenorhabditis*, *Hox* genes in *Tribolium* map to a single cluster of ~750 kb on linkage group 2. Orthologues of all *Drosophila Hox* genes and the *Hox*-derived genes *ftz* and *zen* are transcribed from the same strand, and we find no evidence for interspersion of other protein-coding genes. Taken together, these results suggest that the evolutionary constraints preserving *Hox* cluster integrity still function in *Tribolium*.

Dorso-ventral patterning. As in *Drosophila*, the dorso-ventral axis of the *Tribolium* embryo depends on a nuclear gradient of Dl, an NF-κB protein, which is established through ventral activation of a Tl receptor[29] (one of four in *Tribolium*). Factors required for localized Tl activation are also present in *Tribolium* (potential Tl ligands: six *spz*-like genes; extracellular proteases: one *gd*, six *snk* and four tandem *ea* genes), suggesting that, as in *Drosophila*, an extra-embryonic signal induces the embryonic dorso-ventral axis.

Tribolium sog inhibition of Dpp/BMP generates a patterning gradient along the dorso-ventral axis[30]. Similar *chordin/sog* function in spiders and a hemichordate suggest that this may represent the ancestral bilaterian condition[30]. Like *Apis*, *Tribolium* lacks an orthologue of *Drosophila scw*, but knockdown of another ligand, *Tribolium gbb1*, affected the embryonic Dpp/BMP gradient. *Tribolium* contains orthologues of all five *Drosophila* TGF-β receptors; however, Dpp signalling moderators that have duplicated and diverged in *Drosophila*, such as *Tol/tok* and *Cv/tsg*, occur as a single copy in *Tribolium*. Most strikingly, *Tribolium* contains homologues of *BMP10* as well as *bambi*, *Dan* and *gremlin* BMP inhibitors, which are all known from vertebrates, but are not found in *Drosophila*.

The growth zone. We identified several members of the Fgf and Wnt signalling pathways. The expression patterns of *Tribolium Fgf8*, Wnt1, Wnt5 and WntD/8 (refs 31, 32) highlight the dynamic organization of the growth zone and underline its role in axis elongation.

Head patterning. Orthologues of 25 out of 30 key regulators of the vertebrate anterior neural plate are specifically expressed in the *Tribolium* embryonic head (Supplementary Table 12). Two orthologues are not expressed in the head neuroectoderm (*barH*, *arx*) and three do not have *Tribolium* or *Drosophila* orthologues (*vax*, *hesx1*, *atx*). Of the canonical *Drosophila* head gap genes, only the late head-patterning function of *otd* is conserved. *ems* function is restricted to parts of the antennal and ocular segments, and knockdown of *btd* seems to have no phenotypic consequences. Thus, analysis of *Tribolium* genes defines a set of genes that is highly conserved in bilaterian head development, and underscores the derived mode of *Drosophila* head patterning.

Leg and wing development. In contrast to *Drosophila*, ventral appendages in *Tribolium* develop during embryogenesis from buds that grow continuously along the proximo-distal axis[33]. Nonetheless, we identified *Tribolium* orthologues for most of a core set of *Drosophila* appendage genes (Supplementary Table 13). On the other

hand, orthologues of genes not found in *Drosophila*, such as *Wnt11*, *gremlin*, *Fgf8* and an F-Box gene, are expressed in the embryonic legs[21,31]. Although their exact function in *Tribolium* appendages is not known, Fgf8 is essential to vertebrate limb development.

A major innovation driving the radiation of beetles was the evolution of a highly modified protective forewing. Expression analysis and RNAi experiments revealed a high degree of conservation between *Tribolium* and *Drosophila* wing gene networks (Supplementary Table 13), supporting the hypothesis that sclerotized elytra evolved from ancestral membranous wings mainly through new interactions between conserved patterning modules and as yet unknown downstream effector genes.

Eye development. *Tribolium* has orthologues of nearly all genes currently known to regulate specification and differentiation in the *Drosophila* retina (Supplementary Table 14). Exceptions are the linker protein Phyllopod and the lens crystallin protein Drosocrystallin, which are restricted to Diptera. Eight of fifty-seven investigated eye developmental genes are duplicated in the *Drosophila* genome but not *Tribolium*, and in four cases the *Drosophila* paralogues have similar function. This suggests a more dynamic evolution of *Drosophila* retina genes and higher genetic complexity, highlighting the value of *Tribolium* as a more ancestral and simply organized model of insect eye development.

Genes relevant to pest and *Tribolium* biology

Tribolium castaneum is a notorious invader of stored grains and grain products. Resultantly, much effort and expense is directed to find better ways to control this and other grain pests. Here we describe established and possible future pesticide targets, as well as genes underlying vision and taste. Finally, we describe genes forming the basis of systemic RNAi in *Tribolium*.

Established insecticide targets

Cys-loop ligand-gated ion channels. Members of this superfamily mediate chemical synaptic transmission in insects and are targets of successful pest control chemicals with animal health and crop protection applications[34]. The *Tribolium* Cys-loop ligand-gated ion channel (Cys-loop LGIC) superfamily contains 24 genes, the largest known so far for insects (*Drosophila* and *Apis* superfamilies comprise 23 and 21 genes, respectively), due in part to the additional nicotinic acetylcholine receptor (nAChR) subunits in *Tribolium*. We also found genes for ion channels gated by γ-aminobutyric acid (γ-aminobutyric acid receptors (GABARs)), glutamate (GluCls) and histamine, as well as orthologues of the *Drosophila* pH-sensitive chloride channel[35]. The molecular diversity of the *Tribolium* Cys-loop LGIC superfamily is broadened by alternative splicing and RNA A-to-I editing, which in some cases generates species-specific receptor isoforms[35]. The *Tribolium* Cys-loop LGIC superfamily is the first complete set of genes encoding molecular targets of several insecticides—imidacloprid and other neonicotinoids (nAChRs), fipronil (GABARs) and avermectins (GluCls)—described for an agricultural pest species.

Cytochrome P450 proteins. Most insect cytochrome P450 proteins (CYPs) are thought to be involved in metabolic detoxification of host plant allelochemicals and toxicants, and several are insecticide resistance genes[36]. Other CYPs act in the synthesis and degradation of lipid signalling molecules, such as ecdysteroids[37]. Similarly to mosquitoes, especially *Aedes*, *Tribolium* has an independently expanded CYP gene family, particularly those involved in environmental response (Supplementary Table 16).

Within the *Tribolium* P450s, the CYP2 and mitochondrial clans have undergone relatively little gene expansion, lack pseudogenes, and are probably reserved for essential endogenous functions in ecdysteroid metabolism and development. In contrast, expansions via tandem duplication produced 85% of *Tribolium* P450s clustered in groups of 2–16 genes, with large expansions of CYP3 and CYP4 clans involved in environmental response. In comparison, *Apis* has

only four CYP4 genes, whereas *Aedes* has relatively similarly sized expansions of CYP3 and CYP4 clans (Supplementary Table 16). We speculate that both mosquito larvae (which are omnivorous scavengers) and *Tribolium* have adapted to diverse chemical environments in part by expansion of CYP gene families involved in detoxification.

Possible future insect control targets

C1 cysteine peptidase genes. *Tribolium castaneum* has successfully exploited cereal grains in spite of the arsenal of defensive allelochemicals, including inhibitors of serine peptidase digestive enzymes. In tenebrionid beetles, cathepsins B, L and serine peptidases such as trypsins and chymotrypsins are part of the digestive peptidase complex in the larval gut[38].

Comparing potential digestive peptidase genes in *Tribolium* with those in other sequenced insects (Supplementary Fig. 12) we found more C1 cysteine peptidase genes in *T. castaneum*. The proliferation of *Tribolium* C1 cysteine peptidase genes reflects expansions into five gene families, corresponding to four major clusters. This expansion is consistent with a trend seen in some beetles relative to other insects: a shift to a more acidic gut, conducive to cysteine peptidase activity.

Tribolium castaneum C1 cysteine peptidase genes encode B and L cathepsins, and include the first-known insect genes similar to O and K cathepsins (Supplementary Table 17). Most of the cathepsin-B-like peptidases lack conserved residues in functional regions and thus may lack peptidase activity, whereas all but two *Tribolium* cathepsin L peptidase genes encode potentially functional enzymes. In vertebrates, O and K cathepsins are lysosomal cysteine peptidases, involved in bone remodelling and resorption. Analysis of *Tribolium* cathepsins may provide insight into this family of proteins whose elevated expression is associated with a significant fraction of human breast cancers and tumour invasiveness.

Neurohormones and G-protein-coupled receptors. Insect neurohormones (neuropeptides, protein hormones and biogenic amines) control development, reproduction, behaviour, feeding and many other physiological processes, often by signalling through G-protein-coupled receptors (GPCRs). We found 20 genes encoding biogenic amine GPCRs in *Tribolium* (compared to 21 in *Drosophila* and 19 in *Apis*) and 52 genes encoding neuropeptide or protein hormone GPCRs (49 in *Drosophila*, 37 in *Apis*[39]). Moreover, we identified the likely ligands for 45 of these 72 *Tribolium* GPCRs. Furthermore, we annotated 39 neuropeptide and protein hormone genes. We found excellent agreement (95%) between the proposed ligands for the *Tribolium* neurohormone GPCRs and the independently annotated neuropeptide and protein hormone genes. Interestingly, the *Tribolium* genome contains a vasopressin-like neuropeptide (TC06626) and a vasopressin-like GPCR gene (TC16363; Supplementary Fig. 14), neither of which has been detected in any other sequenced insect[39]. Vasopressin in mammals is the major neurohormone stimulating water reabsorption in the kidneys[40]. Its presence in *Tribolium* may help the beetle to survive in very dry habitats.

Genes relevant to *Tribolium* biology

Vision. Most of the 21 investigated genes that participate in the *Drosophila* photo-transduction network are conserved in *Tribolium* (Supplementary Table 14). Most notable is the lack of *ninaG* and *inaC*, which may be functionally replaced by closely related paralogues in *Tribolium*.

Tribolium contains only two opsin genes, representing members of the long-wavelength and ultraviolet-sensitivity-facilitating opsin subgroups. In contrast, *Drosophila* contains seven, and there is evidence for minimally three in most other insects. The lack of a blue-light-sensitive opsin gene in *Tribolium* is consistent with the unusual expression of long-wavelength opsin in all photoreceptor cells in this species[41]. The implied reduction in colour discrimination in *Tribolium* is probably a consequence of the widespread cryptic lifestyle of this species group.

Odorant and gustatory receptors. Odorant and gustatory receptors form the insect chemoreceptor superfamily. *Tribolium* has a major expansion of both odorant and gustatory receptors relative to *Drosophila*, *Anopheles* and *Aedes* mosquitoes, silkmoth and honeybee (Supplementary Table 19). We identified and annotated 265 apparently functional odorant receptors, 42 full-length pseudogenes and 34 pseudogene fragments. Most of these *T. castaneum* odorant receptors are in seven species-specific subfamilies, including one containing 150 genes, and most are in tandem gene arrays, created by gene duplication within the *Tribolium* lineage in the last 300 million years[42].

We annotated 220 apparently functional gustatory receptors and 25 pseudogenes (gustatory receptor gene fragments were not assessed). The gustatory receptor families in fruitflies and mosquitoes, but not honeybee, contain several genes that are alternatively spliced, with multiple alternative long first exons encoding at least the amino-terminal 50% of the gustatory receptor spliced into a set of short shared exons encoding the carboxy terminus[43–45]. Most *Tribolium* gustatory receptors are encoded by single genes; however, *T. castaneum* Gr214 is a massive alternatively spliced locus with 30 alternative long 5′ exons (six of which are pseudogenic) spliced into three shared 3′ exons encoding the C terminus. Three *T. castaneum* gustatory receptors are orthologues of highly conserved gustatory receptors in other insects[43–45], two of which form a heterodimeric carbon dioxide receptor[46]. The remainder form many species-specific subfamilies, one of which is expanded to 88 genes (Supplementary Information and Supplementary Fig. 16).

Systemic RNAi. In *Tribolium*, as in *C. elegans* but not *Drosophila*, the RNAi effect spreads systemically from the site of injection to other tissues[5] and from injected females to their offspring[4]. Surprisingly, our survey of genes involved in systemic RNA did not reveal much conservation between *Tribolium* and *C. elegans*.

The SID-1 multi-transmembrane protein, essential for double-stranded RNA (dsRNA) uptake in *C. elegans*, is not found in *Drosophila*, suggesting that the presence or absence of a *sid-1* gene is the primary determinant of whether or not systemic RNAi occurs in an organism. We found three genes in *Tribolium* that encode proteins similar to SID-1. However, their sequences are more similar to another *C. elegans* protein, TAG-130 (also known as ZK721.1), which is not required for systemic RNAi in *C. elegans*[47]. Additionally, the secondary argonaute proteins and RNA-dependent RNA polymerase (RdRP)[48,49], essential for the amplification of the initial dsRNA trigger in *C. elegans*, are absent in *Tribolium*. Therefore, the molecular basis for systemic RNAi in *Tribolium* and other insects might differ from that in *C. elegans* and remains to be elucidated.

Concluding remarks

We observe three trends when comparing *Tribolium* and other insect genomes. First, phylogenetic trees show shorter branch lengths for *Tribolium* (and *Apis*) than *Drosophila*. The accelerated evolution of the *Drosophila* lineage in some cases rendered *Drosophila* atypical for the Insecta. Second, *Tribolium* retains a different set of ancestral genes that have evolved at a moderate rate (for example, *gremlin* and cathepsins), and these may provide insights into the function of their vertebrate orthologues. Third, its own evolutionary path has led to beetle- and perhaps *Tribolium*-specific gene changes (for example, a large increase in odorant receptors).

Expansions of CYP proteins, proteinases, diuretic hormones, a vasopressin hormone and receptor, and chemoreceptors all indicate adaptation to a dry, chemically diverse and toxin-rich microenvironment. Whereas the flour beetle's drought tolerance probably explains the presence of vasopressin, it is more difficult to rationalize a need for such an unprecedented diversity of chemoreceptors. Functions stemming from the diversity of angiosperm-derived chemicals such as distant detection of food sources and avoidance of toxic host plant defence chemicals suggest that this expansion may be common to the Coleoptera. The expansion of odorant receptors is more intriguing

when considered in combination with the reduction of opsin genes. Both trends may reflect the long-term consequences of adaptation to low light biota by *Tribolium*, enforcing selection for increased discrimination of odour reception but not colour perception[41].

Given the chemo-sensing and detoxifying genes described above, it is perhaps no surprise that *Tribolium* has demonstrated resistance to all insecticides used for its control. Given the association of *Tribolium* with human food, knowledge of all possible insecticide targets will aid greater selectivity in pesticide design, thereby mitigating possible side effects. Finally, the true value of this sequence may be the entry it provides into the many and richly diverse facets of beetle biology, physiology and behaviour.

METHODS SUMMARY

Detailed Methods are described in the Supplementary Information. Resources generated by this project can be found at the following locations: genome assemblies, sequences, and automated and manually curated gene model sequences are available from the BCM-HGSC website and ftp site (http://www.hgsc.bcm.tmc.edu/projects/tribolium/). Browser display of the genome sequence, all gene predictions and available tiling array data are available via http://www.genboree.org and Beetle Base (http://www.bioinformatics.ksu.edu/BeetleBase/), a long-term repository for *Tribolium* data.

1.   Hunt, T. *et al.* A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318**, 1913–1916 (2007).
2.   Sokoloff, A. *The Biology of* Tribolium *with Special Emphasis on Genetic Aspects I–III* (Clarendon Press and Oxford Univ. Press, Oxford, 1972, 1974, 1977).
3.   Lorenzen, M. D. *et al. piggyBac*-based insertional mutagenesis in *Tribolium castaneum* using donor/helper hybrids. *Insect Mol. Biol.* **16**, 265–275 (2007).
4.   Bucher, G., Scholten, J. & Klingler, M. Parental RNAi in *Tribolium* (Coleoptera). *Curr. Biol.* **12**, R85–R86 (2002).
5.   Tomoyasu, Y. & Denell, R. E. Larval RNAi in *Tribolium* (Coleoptera) for analyzing adult development. *Dev. Genes Evol.* **214**, 575–578 (2004).
6.   Tautz, D., Friedrich, M. & Schröder, R. in *Development 1994 Supplement* (eds Akam, M., Holland, P., Ingham, P. & Wray, G.) 193–199 (The Company of Biologists Limited, Cambridge, 1994).
7.   Tautz, D. Segmentation. *Dev. Cell* **7**, 301–312 (2004).
8.   Lorenzen, M. D. *et al.* Genetic linkage maps of the red flour beetle, *Tribolium castaneum*, based on bacterial artificial chromosomes and expressed sequence tags. *Genetics* **170**, 741–747 (2005).
9.   Demuth, J. P. *et al.* Genome-wide survey of *Tribolium castaneum* microsatellites and description of 509 polymorphic markers. *Mol. Ecol. Notes* **7**, 1189–1195 (2007).
10. Wang, S. & Brown, S. J. Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. *Genome Biol.* (in the press).
11. Brown, S. J., Henry, J. K., Black, W. C. IV & Denell, R. Molecular Genetic manipulation of the red flour beetle: Genome organization and cloning of a ribosomal protein gene. *Insect Biochem.* **20**, 185–193 (1990).
12. Ugarkovic, D., Podnar, M. & Plohl, M. Satellite DNA of the red flour beetle *Tribolium castaneum*—comparative study of satellites from the genus *Tribolium. Mol. Biol. Evol.* **13**, 1059–1066 (1996).
13. Osanai, M., Kojima, K. K., Futahashi, R., Yaguchi, S. & Fujiwara, H. Identification and characterization of the telomerase reverse transcriptase of *Bombyx mori* (silkworm) and *Tribolium castaneum* (flour beetle). *Gene* **376**, 281–289 (2006).
14. Xiong, Y. & Eickbush, T. H. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. *Mol. Cell. Biol.* **8**, 114–123 (1988).
15. Robertson, H. M. & Gordon, K. H. Canonical TTAGG-repeat telomeres and telomerase in the honey bee, *Apis mellifera. Genome Res.* **16**, 1345–1351 (2006).
16. Fujiwara, H., Osanai, M., Matsumoto, T. & Kojima, K. K. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori. Chromosome Res.* **13**, 455–467 (2005).
17. Savard, J. *et al.* Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Res.* **16**, 1334–1338 (2006).
18. Savard, J., Tautz, D. & Lercher, M. J. Genome-wide acceleration of protein evolution in flies (Diptera). *BMC Evol. Biol.* **6**, 7 (2006).
19. Daborn, P. J. *et al.* A single p450 allele associated with insecticide resistance in *Drosophila. Science* **297**, 2253–2256 (2002).
20. Ciccarelli, F. D. *et al.* Complex genomic rearrangements lead to novel primate gene function. *Genome Res.* **15**, 343–351 (2005).
21. Wyder, S., Kriventseva, E. V., Schröder, R., Kadowaki, T. & Zdobnov, E. M. Quantification of ortholog losses in insects and vertebrates. *Genome Biol.* **8**, R242 (2007).
22. Wong, M. D., Jin, Z. & Xie, T. Molecular mechanisms of germline stem cell regulation. *Annu. Rev. Genet.* **39**, 173–195 (2005).

23. Schröder, R. The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *Tribolium*. *Nature* **422**, 621–625 (2003).
24. Draper, B. W., Mello, C. C., Bowerman, B., Hardin, J. & Priess, J. R. MEX-3 is a KH domain protein that regulates blastomere identity in early *C. elegans* embryos. *Cell* **87**, 205–216 (1996).
25. Bucher, G. & Klingler, M. Divergent segmentation mechanism in the short germ insect *Tribolium* revealed by giant expression and function. *Development* **131**, 1729–1740 (2004).
26. Cerny, A. C., Bucher, G., Schröder, R. & Klingler, M. Breakdown of abdominal patterning in the *Tribolium Kruppel* mutant *jaws*. *Development* **132**, 5353–5363 (2005).
27. Choe, C. P., Miller, S. C. & Brown, S. J. A pair-rule gene circuit defines segments sequentially in the short-germ insect *Tribolium castaneum*. *Proc. Natl Acad. Sci. USA* **103**, 6560–6564 (2006).
28. Savard, J., Marques-Souza, H., Aranda, M. & Tautz, D. A segmentation gene in *Tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell* **126**, 559–569 (2006).
29. Fonseca, R. N. *et al.* Self-regulatory circuits in dorsoventral axis formation of the short germ beetle *Tribolium castaneum*. *Dev. Cell* (in the press).
30. van der Zee, M., Stockhammer, O., von Levetzow, C., Nunes da Fonseca, R. & Roth, S. *Sog/Chordin* is required for ventral-to-dorsal *Dpp/BMP* transport and head formation in a short germ insect. *Proc. Natl Acad. Sci. USA* **103**, 16307–16312 (2006).
31. Beermann, A. & Schröder, R. Sites of FGF signalling and perception during embryogenesis of the beetle *Tribolium castaneum*. *Dev. Genes Evol.* (in the press).
32. Bolognesi, R. *et al. Tribolium* Wnts: evidence for a larger repertoire in insects with overlapping expression patterns that suggest multiple redundant functions in embryogenesis. *Dev. Genes Evol.* (in the press).
33. Beermann, A. *et al.* The *Short antennae* gene of *Tribolium* is required for limb development and encodes the orthologue of the *Drosophila* Distal-less protein. *Development* **128**, 287–297 (2001).
34. Raymond-Delpech, V., Matsuda, K., Sattelle, B. M., Rauh, J. J. & Sattelle, D. B. Ion channels: molecular targets of neuroactive insecticides. *Invert. Neurosci.* **5**, 119–133 (2005).
35. Jones, A. K. & Sattelle, D. B. The cys-loop ligand-gated ion channel gene superfamily of the red flour beetle, *Tribolium castaneum*. *BMC Genomics* **8**, 327 (2007).
36. Ffrench-Constant, R. H., Daborn, P. J. & Le Goff, G. The genetics and genomics of insecticide resistance. *Trends Genet.* **20**, 163–170 (2004).
37. Ono, H. *et al. Spook* and *Spookier* code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. *Dev. Biol.* **298**, 555–570 (2006).
38. Vinokurov, K. S. *et al.* Diversity of digestive proteinases in *Tenebrio molitor* (Coleoptera: Tenebrionidae) larvae. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **145**, 126–137 (2006).
39. Hauser, F., Cazzamali, G., Williamson, M., Blenau, W. & Grimmelikhuijzen, C. J. A review of neurohormone GPCRs present in the fruitfly *Drosophila melanogaster* and the honey bee *Apis mellifera*. *Prog. Neurobiol.* **80**, 1–19 (2006).
40. Bankir, L. Antidiuretic action of vasopressin: quantitative aspects and interaction between V1a and V2 receptor-mediated effects. *Cardiovasc. Res.* **51**, 372–390 (2001).
41. Jackowska, M. *et al.* Genomic and gene regulatory signatures of cryptozoic adaptation: loss of blue sensitive photoreceptors through expansion of long wavelength-opsin expression in the red flour beetle *Tribolium castaneum*. *Front. Zool.* **4**, 24 (2007).
42. Engsontia, P. *et al.* The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* (in the press).
43. Clyne, P. J., Warr, C. G. & Carlson, J. R. Candidate taste receptors in *Drosophila*. *Science* **287**, 1830–1834 (2000).
44. Hill, C. A. *et al.* G protein-coupled receptors in *Anopheles gambiae*. *Science* **298**, 176–178 (2002).
45. Robertson, H. M., Warr, C. G. & Carlson, J. R. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **100** (suppl. 2), 14537–14542 (2003).
46. Jones, W. D., Cayirlioglu, P., Kadow, I. G. & Vosshall, L. B. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* **445**, 86–90 (2007).
47. Tomoyasu, Y. *et al.* Exploring systemic RNA interference in insects; a genome-wide survey for RNAi genes in *Tribolium*. *Genome Biol.* (in the press).
48. Lipardi, C., Wei, Q. & Paterson, B. M. RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell* **107**, 297–307 (2001).
49. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

The *Tribolium* Genome Sequencing Consortium

Project leader: Stephen Richards[1,2]

Principal investigators: Richard A. Gibbs[1,2], George M. Weinstock[1,2]

White paper: Susan J. Brown[3], Robin Denell[3], Richard W. Beeman[4], Richard Gibbs[1,2]

Analysis leaders: Richard W. Beeman[4], Susan J. Brown[3], Gregor Bucher[5], Markus Friedrich[6] Cornelis J. P. Grimmelikhuijzen[7], Martin Klingler[8], Marce Lorenzen[3], Stephen Richards[1,2], Siegfried Roth[9], Reinhard Schröder[10,†], Diethard Tautz[11], Evgeny M. Zdobnov[12,13,14]

DNA sequence and global analysis: DNA sequencing Donna Muzny (leader)[1,2], Richard A. Gibbs[1,2], George M. Weinstock[1,2], Tony Attaway[1,2], Stephanie Bell[1,2], Christian J. Buhay[1,2], Mimi N. Chandrabose[1,2], Dean Chavez[1,2], Kerstin P. Clerk-Blankenburg[1,2], Andrew Cree[1,2], Marvin Dao[1,2], Clay Davis[1,2], Joseph Chacko[1,2], Huyen Dinh[1,2], Shannon Dugan-Rocha[1,2], Gerald Fowler[1,2], Toni T. Garner[1,2], Jeffrey Garnes[16], Andreas Gnirke[17], Alica Hawes[1,2], Judith Hernandez[1,2], Sandra Hines[1,2], Michael Holder[1,2], Jennifer Hume[1,2], Shalini N. Jhangiani[1,2], Vandita Joshi[1,2], Ziad Mohid Khan[1,2], LaRonda Jackson[1,2], Christie Kovar[1,2], Andrea Kowis[1,2], Sandra Lee[1,2], Lora R. Lewis[1,2], Jon Margolis[17], Margaret Morgan[1,2], Lynne V. Nazareth[1,2], Ngoc Nguyen[1,2], Geoffrey Okwuonu[1,2], David Parker[1,2], Stephen Richards[1,2], San-Juana Ruiz[1,2], Jireh Santibanez[1,2], Joël Savard[11], Steven E. Scherer[1,2], Brian Schneider[1,2], Erica Sodergren[1,2], Diethard Tautz[11], Selina Vattahil[1,2], Donna Villasana[1,2], Courtney S. White[1,2], Rita Wright[1,2]; **EST sequencing** Yoonseong Park[18], Richard W. Beeman[4], Jeff Lord[4], Brenda Oppert[4], Marce Lorenzen[4], Susan Brown[3], Liangjiang Wang[3], Joël Savard[11], Diethard Tautz[11], Stephen Richards[1,2], George Weinstock[1,2], Richard A. Gibbs[1,2]; **genome assembly** Yue Liu[1,2], Kim Worley[1,2], George Weinstock[1,2]; **G+C content** Christine G. Elsik[19], Justin T. Reese[19], Eran Elhaik[20], Giddy Landan[20], Dan Graur[20]; **repetitive DNA, transposons and telomeres** Peter Arensburger[21], Peter Atkinson[21], Richard W. Beeman[4], Jim Beidler[22], Susan J. Brown[3], Jeffery P. Demuth[23], Douglas W. Drury[24], Yu-Zhou Du[25], Haruhiko Fujiwara[26], Marce Lorenzen[3], Vincenza Maselli[27], Mizuko Osanai[26], Yoonseong Park[18], Hugh M. Robertson[28], Zhijian Tu[22], Jian-jun Wang[25], Suzhi Wang[3]; **gene prediction and consensus gene set** Stephen Richards[1,2], Henry Song[1,2], Lan Zhang[1,2], Erica Sodergren[1,2], Doreen Werner[29], Mario Stanke[29], Burkhard Morgenstern[29], Victor Solovyev[30], Peter Kosarev[31], Garth Brown[32], Hsiu-Chuan Chen[32], Olga Ermolaeva[32], Wratko Hlavina[32], Yuri Kapustin[32], Boris Kiryutin[32], Paul Kitts[32], Donna Maglott[32], Kim Pruitt[32], Victor Sapojnikov[32], Alexandre Souvorov[32], Aaron J. Mackey[33], Robert M. Waterhouse[14], Stefan Wyder[12], Evgeny M. Zdobnov[12,13,14]; **global gene content analysis** Evgeny M. Zdobnov[12,13,14], Stefan Wyder[12], Evgenia V. Kriventseva[12,34], Tatsuhiko Kadowaki[35], Peer Bork[36,37].

Developmental processes and signalling pathways: Manuel Aranda[11], Riyue Bao[6], Anke Beermann[10], Nicola Berns[10], Renata Bolognesi[3], François Bonneton[38], Daniel Bopp[39], Susan J. Brown[3], Gregor Bucher[5], Thomas Butts[40], Arnaud Chaumot[41], Robin E. Denell[3], David E. K. Ferrier[40], Markus Friedrich[6], Cassondra M. Gordon[3], Marek Jindra[42], Martin Klingler[8], Que Lan[43], H. Michael G. Lattorff[44], Vincent Laudet[38], Cornelia von Levetsow[9], Zhenyi Liu[45], Rebekka Lutz[10], Jeremy A. Lynch[9], Rodrigo Nunes da Fonseca[5], Nico Posnien[10], Rolf Reuter[10], Siegfried Roth[9], Joël Savard[11], Johannes B. Schinko[5], Christian Schmitt[8], Michael Schoppmeier[8], Reinhard Schröder[10], Teresa D. Shippy[3], Franck Simonnet[5], Henrique Marques-Souza[11], Diethard Tautz[11], Yoshinori Tomoyasu[3], Jochen Trauner[8], Maurijn Van der Zee[11], Michel Vervoort[46], Nadine Wittkopp[10], Ernst A. Wimmer[5], Xiaoyun Yang[6]

Pest biology, senses, Medea and RNAi: ligand gated ion channels Andrew K. Jones[47], David B. Sattelle[47]; **oxidative phosphorylation** Paul R. Ebert[48]; **P450 genes** David Nelson[49], Jeffrey G. Scott[50], Richard W. Beeman[4]; **chitin and cuticular proteins** Subbaratnam Muthukrishnan[51], Karl J. Kramer[4,51], Yasuyuki Arakane[4,51], Richard W. Beeman[4], Qingsong Zhu[51], David Hogenkamp[51], Radhika Dixit[51]; **digestive proteinases** Brenda Oppert[4], Haobo Jiang[52], Zhen Zou[52], Jeremy Marshall[3], Elena Elpidina[53], Konstantin Vinokurov[4], Cris Oppert[4]; **immunity** Zhen Zou[52], Jay Evans[54], Zhiqiang Lu[52], Picheng Zhao[52], Niranji Sumathipala[52], Boran Altincicek[55], Andreas Vilcinskas[55], Michael Williams[56], Dan Hultmark[56], Charles Hetru[57], Haobo Jiang[52]; **neurohormones and GPCRs** Cornelis J. P. Grimmelikhuijzen[7], Frank Hauser[7], Giuseppe Cazzamali[7], Michael Williamson[7], Yoonseong Park[18], Bin Li[18], Yoshiaki Tanaka[58], Reinhard Predel[59], Susanne Neupert[59], Joachim Schachtner[60], Peter Verleyen[61]; **neuropeptide processing enzymes** Florian Raible[36], Peer Bork[36,37]; **opsins** Markus Friedrich[6]; **odorant receptors and gustatory receptors** Kimberly K. O. Walden[28], Hugh M. Robertson[28]; **odorant binding and chemosensory proteins** Sergio Angeli[62], Sylvain Forêt[63], Gregor Bucher[5], Stefan Schuetz[5], Ryszard Maleszka[63], Ernst A. Wimmer[5]; **Medea** Richard W. Beeman[4], Marce Lorenzen[4]; **systemic RNAi** Yoshinori Tomoyasu[3], Sherry C. Miller[3], Daniela Grossmann[5] & Gregor Bucher[5]

Affiliations for participants: [1]Human Genome Sequencing Center, and [2]Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [3]Division of Biology, Ackert Hall, Kansas State University, Manhattan, Kansas 66506, USA. [4]Grain Marketing and Production Research Center, Agricultural Research Service, United States Department of Agriculture, 1515 College Avenue, Manhattan, Kansas 66502, USA. [5]Johann Friedrich Blumenbach Institute, Department of Developmental Biology, Georg August University, von-Liebig-Weg-11, 37077 Göttingen, Germany. [6]Department of Biological Sciences, Wayne State University, Detroit, Michigan 48202, USA. [7]Center for Functional and Comparative Insect Genomics, and Department of Cell Biology and Comparative Zoology, Institute of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. [8]Institute for Biology, Department of Developmental Biology, Friedrich-Alexander-University Erlangen, Staudtstrasse 5, 91058 Erlangen, Germany. [9]Institute for Developmental Biology, University of Cologne, 50674 Cologne, Germany. [10]Animal Genetics, Interfaculty Institute for Cell Biology, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany. [11]Department of Genetics, University of Cologne, 50674 Cologne, Germany. [12]Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland. [13]Swiss Institute of Bioinformatics, 1 rue Michel-Servet, 1211 Geneva, Switzerland. [14]Imperial College London, South Kensington Campus, SW7 2AZ London, UK. [15]Children's Hospital Oakland Research Institute, BACPAC Resources, 747 52nd Street, Oakland, California 94609, USA. [16]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [17]AgraQuest, Inc., 1530 Drew Avenue, Davis, California 95616, USA. [18]Department of Entomology, Waters Hall, Kansas State University, Manhattan, Kansas 66506, USA. [19]Department of Animal Science, Texas A&M University, College Station, Texas 77843, USA. [20]Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204, USA. [21]Department of Entomology, University of California, 900 University Avenue, Riverside, California 92521, USA. [22]Department of Biochemistry, Virginia Tech, Blacksburg, Virginia 24061, USA. [23]Department of Biology, University of Texas at Arlington, Arlington, Texas 76019, USA. [24]Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. [25]Department of Plant Protection, Yangzhou University, Yangzhou 225009, China. [26]Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Bioscience Building 501, Kashiwa, Chiba 277-8562, Japan. [27]European School of Molecular Medicine and Telethon Institute of Genetics and Medicine, Via Pietro Castellino 111, 80131 Napoli, Italy. [28]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [29]Institute for Microbiology and Genetics, Department of Bioinformatics, University of Göttingen, Goldschmidtstraße 1, 37077 Göttingen, Germany. [30]Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. [31]Softberry Inc., 116 Radio Circle, Suite 400, Mount Kisco, New York 10549, USA. [32]National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland 20894, USA. [33]GlaxoSmithKline, Collegeville, Pennsylvania 19426, USA. [34]Department of Structural Biology and Bioinformatics, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland. [35]Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan. [36]European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany. [37]Max-Delbruck-Centre for Molecular Medicine, Berlin-Buch, Robert-Roessle-Strasse 10, 13092 Berlin, Germany. [38]Institut de Genomique Fonctionnelle de Lyon, Equipe de Zoologie Moleculaire, ENS Lyon, Universite Lyon 1, CNRS UMR5242, INRA, IFR128, 46 Allee d'Italie, 69364 Lyon cedex 07, France. [39]Zoological Institute of the University Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. [40]Department of Zoology, University of Oxford, Tinbergen Building, South Parks Road, Oxford OX1 3PS, UK. [41]CEMAGREF, Laboratoire d'écotoxicologie, 3bis quai Chauvea, CP220 69336 Lyon cedex 09, France. [42]Institute of Entomology ASCR, Branisovká 31, Ceské Budejovice 370 05, Czech Republic. [43]Department of Entomology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, Wisconsin 53706, USA. [44]Institute of Biology, Molecular Ecology, Martin-Luther-University Halle-Wittenberg Hoher Weg 4, 06099 Halle (Saale), Germany. [45]Department of Molecular Biology and Pharmacology, Washington University in St Louis School of Medicine, 3600 Cancer Research Building, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. [46]Université Paris 7 – Denis Diderot, Centre de genetique moleculaire – CNRS UPR 2167, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette cedex, France. [47]MRC Functional Genetics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK. [48]School of Integrative Biology & School of Molecular and Microbial Sciences, University of Queensland, St Lucia, Queensland 4072, Australia. [49]Department of Molecular Sciences and Center of Excellence in Genomics and Bioinformatics, University of Tennessee, Memphis, Tennessee 38163, USA. [50]Department of Entomology, Daljit and Elaine Sarkaria Professor of Insect Physiology and Toxicology, Cornell University, Ithaca, New York 14853, USA. [51]Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506, USA. [52]Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, Oklahoma 74078, USA. [53]A. N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Leninskie Gory, Moscow 119992, Russia. [54]USDA-ARS Bee Research Laboratory, Beltsville, Maryland 20705, USA. [55]Institute of Phytopathology and Applied Zoology, Interdisciplinary Research Center, Justus-Liebig-University of Giessen, Heinrich-Buff-Ring 26-32, D-35392 Giessen, Germany. [56]Umea Centre for Molecular Pathogenesis, Umea University, Umea SE-90187, Sweden. [57]Institut Biol Moléc Cell, CNRS, Strasbourg 67084, France. [58]National Institute of Agrobiological Science, Division of Insect Science, Tsukuba, Ibaraki 305-8634, Japan. [59]Institute of General Zoology, University of Jena, Erbertstrasse 1, D-07743 Jena, Germany. [60]Department of Animal Physiology, University of Marburg, Karl-von-Frisch Strasse 8, D-35032 Marburg, Germany. [61]Department of Animal Physiology and Neurobiology, University of Leuven, Naamsestraat 59, BE-3000 Leuven, Belgium. [62]Institute for Forest Zoology and Forest Conservation, Büsgenweg 3 D-37077 Göttingen, Germany. [63]Visual Sciences and ARC Centre for the Molecular Genetics of Development, Research School of Biological Sciences, The Australian National University, Canberra, ACT 0200, Australia. †Present address: Bioscience Institute, University of Rostock, Albert-Einstein-Strasse 3, 18059 Rostock, Germany.