ARTICLE

Genomic landscape and genetic manipulation of the black soldier fly *Hermetia illucens*, a natural waste recycler

Shuai Zhan [1,2], Gangqi Fang[1,2], Minmin Cai[3], Zongqing Kou[1], Jun Xu[1], Yanghui Cao[1], Liang Bai[1], Yixiang Zhang[1,2], Yongmao Jiang[1,2], Xingyu Luo[1,2], Jian Xu[1,2], Xia Xu[1,2], Longyu Zheng[3], Ziniu Yu[3], Hong Yang[4], Zhijian Zhang[5], Sibao Wang [1,2], Jeffery K. Tomberlin[6], Jibin Zhang[3] and Yongping Huang [1,2]

The black soldier fly (BSF), *Hermetia illucens* (Diptera: Stratiomyidae), is renowned for its bioconversion of organic waste into a sustainable source of animal feed. We report a high-quality genome of 1.1 Gb and a consensus set of 16,770 gene models for this beneficial species. Compared to those of other dipteran species, the BSF genome has undergone a substantial expansion in functional modules related to septic adaptation, including immune system factors, olfactory receptors, and cytochrome P450s. We further profiled midgut transcriptomes and associated microbiomes of BSF larvae fed with representative types of organic waste. We find that the pathways related to digestive system and fighting infection are commonly enriched and that Firmicutes bacteria dominate the microbial community in BSF across all diets. To extend its potential practical applications, we further developed an efficient CRISPR/Cas9-based gene editing approach and implemented this to yield flightless and enhanced feeding capacity phenotypes, both of which could expand BSF production capabilities. Our study provides valuable genomic and technical resources for optimizing BSF lines for industrialization.

INTRODUCTION

Increases in the global human population mean that an ever expanding volume of organic waste must be managed. Organic wastes produced in urban environments are treated in three major methods, burn, landfills, and compost, whereas waste resulting from confined animal facilities is held in lagoons or placed on land as fertilizer. Unfortunately, most methods employed for processing these wastes do not ameliorate environmental quality and may cause secondary pollution.

The black soldier fly (BSF), *Hermetia illucens* (L.) (Diptera: Stratiomyidae), is one of the most promising insect species being mass produced globally, due to its ability to convert a variety of organic wastes into insect biomass that can be used as feed for many aquaculture species as well as poultry (Fig. 1). Up to now, BSF is the only insect species approved globally for use as a feed ingredient in aquaculture and poultry.

More importantly, BSF has the ability to recycle many types of organic waste efficiently and effectively. Through this recycling process, waste streams are converted into valuable products such as protein for animal feed,[1–4] fat for bioenergy,[5,6] and compost that can be used as fertilizer.[7] Using BSF to recycle food and animal wastes has numerous other benefits. Recycling of these nutrients results in reduction of noxious odors,[8] carbon dioxide emissions,[9] pathogenic bacteria,[10] and antibiotics.[11] Thus, because of its utility and unique features, this species is quickly used for insect farming and as a model organism for basic research.[12–14]

BSF is naturally distributed throughout the tropics and subtropics, but can also be reared indoors under controlled conditions; consequently, it is now distributed throughout the world. Currently, optimizing BSF for recycling particular types of waste is particularly challenging because nothing is known about its genetics, preventing the use of current molecular technology to optimize its traits for waste recycling. Genome references and efficient genetic manipulation systems are foundational for molecular and genetic research of high quality. Next generation sequencing has driven the growth of more genomic and transcriptomic resources that significantly advance the studies in non-model organisms.[15] The recent development of genome editing techniques, such as CRISPR/Cas9, provides the capacity for genetic manipulation in a variety of organisms.[16] Here, we utilized a comprehensive omics approach, including genomics, transcriptomics, metagenomics, together with the establishment of genetic manipulation system, to explore the genetic bases underlying the key aspects of BSF biology.

[1]CAS Key Laboratory of Insect Developmental and Evolutionary Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China; [2]CAS Center for Excellence in Biotic Interactions, University of Chinese Academy of Sciences, Beijing 100049, China; [3]State Key Laboratory of Agricultural Microbiology, National Engineering Research Center of Microbial Pesticides, College of Life Science and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China; [4]Hubei Key Laboratory of Genetic Regulation and Integrative Biology, Institute of Entomology, School of Life Sciences, Central China Normal University, Wuhan, Hubei 430079, China; [5]School of Economics, Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China and [6]Department of Entomology, Texas A&M University, 2475 TAMU, College Station, TX 77845, USA
Correspondence: Shuai Zhan (szhan@sibs.ac.cn) or Jeffery K. Tomberlin (jktomberlin@tamu.edu) or Jibin Zhang (zhangjb@mail.hzau.edu.cn) or Yongping Huang (yphuang@sibs.ac.cn)
These authors contributed equally: Shuai Zhan, Gangqi Fang, Minmin Cai, Zongqing Kou
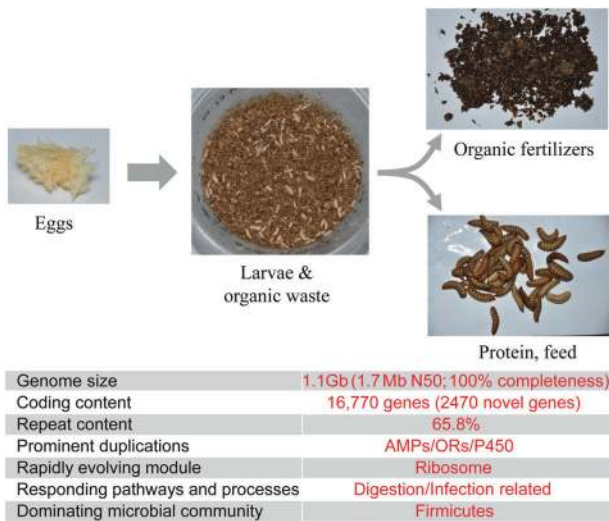
Fig. 1 Biology and genomic features of the black soldier fly. N50, the minimum length of scaffolds that cover 50% of the genome assembly; AMP antimicrobial peptide, OR olfactory receptor, P450 cytochrome P450s

RESULTS AND DISCUSSION

Characteristics of the BSF genome

Many dipteran genomes have been sequenced, yet little genomic information is available for species belonging to Stratiomyidae. We sequenced the genome of a 10-generation inbred line of BSF to ~300× coverage of Illumina sequencing data, including both paired-end libraries of short inserts and mate-pair libraries of long inserts (Supplementary information, Table S1, Fig. S1). The final genome assembly contains 1102 Mb of assembled scaffolds with a 1.69 Mb N50 length (Fig. 1 and Supplementary information, Table S2). Completeness of the assembly determined using BUSCO[17] was 99.5% and that determined using CEGMA[18] was 100%, suggesting a near-complete representation of the BSF genome (Fig. 1 and Supplementary information, Table S2). Analyses of GC content and sequencing coverage revealed a normal distribution among assembled scaffolds (Supplementary information, Fig. S2), suggesting very few contaminations in the assembly.

The BSF genome (1102 Mb) is relatively large compared to those of other dipteran species, and it is larger than any Brachycera (a more recently evolved Diptera) genome (ranging from 90 to 750 Mb; Supplementary information, Table S2). Consistent with the idea that variation in genome size is likely due to the relative amounts of transposable elements and other repetitive non-coding DNA,[19,20] approximately two-thirds of the BSF genome is repeated, which partially accounts for the large size (Supplementary information, Table S3). We generated an official set of 16,770 protein-coding genes by combining information on homologs from six dipteran species, transcriptome data from 12 continuous BSF developmental stages, and three sets of ab initio gene predictions (Supplementary information, Table S4). The number of genes in the BSF genome is comparable to those of other dipteran species (Supplementary information, Table S5). The mean intron size is the second longest among dipteran species that we investigated in this study (Supplementary information, Table S5).

Comparison of the BSF genome with those of other dipterans

We first compared the gene repertoire of BSF with those of other dipteran species. Ortholog analysis indicated that half of the BSF genes are common to all dipteran species analyzed in this study (Fig. 2a). Inferring phylogeny across all examined dipteran species using single-copy orthologs placed BSF ancestrally within the
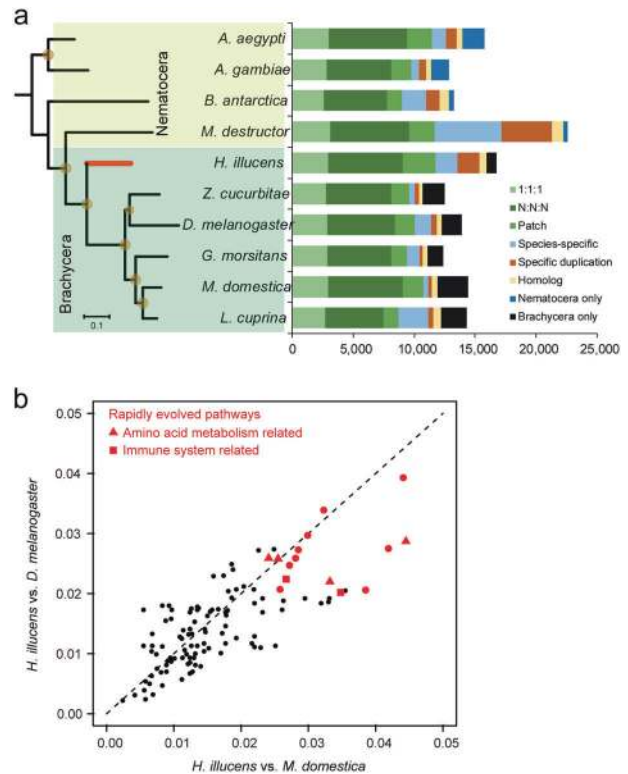
Fig. 2 Relationships of Diptera species and pathways that differ between BSF and parallel lineages. a Phylogenetic relationship and assigned orthology across ten dipteran species. The maximum likelihood phylogenomic tree was calculated based on 814 single-copy universal genes. Nodes labeled with a brown circle indicate those with high bootstrap support (at least 82 of 100 replicates). The colored histogram indicates category of orthology as follows: "1:1:1", single-copy universal genes; "N:N:N", multi-copy universal genes; "Brachyera only", genes specific to Brachyera species; "Nematocera only", genes specific to Nematocera species; "Species-specific", genes without an ortholog in any other species; "Specific duplication", genes with species-specific duplication; "Patchy", orthologs in some species; "Homology", homologs detected in other species (E < 1e−5) but not grouped in the orthology analysis. The species used in the analysis were: *Aedes aegypti*, the yellow fever mosquito; *Anopheles gambiae*, the African malaria mosquito; *Belgica antarctica*, the Antarctic midge; *Mayetiola destructor*, the hessian fly; *H. illucens*, BSF; *Zeugodacus cucurbitae*, the melon fly; *Drosophila melanogaster*, the fruit fly; *Glossina morsitans*, the tsetse fly; *Musca domestica*, the house fly; and *Lucilia cuprina*, the blow fly. b Identification of pathways that have rapidly evolved in BSF. dN/dS ratios were calculated independently in two parallel evolutionary lineages, *M. domestica* and *D. melanogaster*, using BSF as the common ancestor. Each dot indicates the median dN/dS ratios of all related genes in the corresponding pathway. Significantly enriched (FDR-adjusted $P < 0.05$), rapidly evolving genes in KEGG pathways are highlighted in red

Brachycera sublineage (short-horned Diptera) (Fig. 2a). Thus, the BSF genome fills a gap between the Nematocera, the earliest diverging suborder of Diptera, and more recent flies.

We estimated the nonsynonymous-to-synonymous substitution (dN/dS) ratios between BSF and two parallel derived lineages, the house fly (*Musca domestica* (Diptera: Muscidae))[21] and the fruit fly (*Drosophila melanogaster* (Diptera: Drosophilidae)).[22] We identified 342 genes with higher ratios of dN/dS in BSF. These rapidly evolving genes are significantly enriched in only one biological module, the ribosome (ko03010; FDR $P = 5.2 \times 10^{-17}$) (Supplementary information, Tables S6 and 7), genes of which contribute to central aspects of the translation mechanism and protein synthesis. Given the relative conservation of this gene family,[23] the

associated ribosomal RNA (rRNA) genes have been explored as cytogenetic markers to study the evolutionary history of a species. The rapid evolution of rRNA genes in BSF was thus unexpected. The 16 pathways with higher dN/dS ratios in BSF than other examined species included four amino acid metabolism-related pathways and two immune-related pathways (Fig. 2b and Supplementary information, Table S8), which may also contribute to the exceptionally rich protein content and strong adaptation to high pathogen loads in BSF.

The BSF genome encodes 797 Brachycera-specific genes, the fewest across Brachycera, and 1798 species-specific duplicated genes, the most across Brachycera (Fig. 2a and Supplementary information, Table S4). BSF-specific duplicated genes were generally expressed at low levels, indicating recent origins.[24] These duplicated genes were clustered as three main groups based on expression profiles, whose high-expression stages were continuous agreeing well with the developmental process (Supplementary information, Fig. S3). Interestingly, most of these genes were expressed during the late larval stage (L-D8 and L-D12; Supplementary information, Fig. S3). Given that BSF recycles organic waste in the larval stage, these BSF-specific gene duplication events may help shape these key aspects of the BSF biology.

We also categorized and compared gene families across dipteran gene repertoires based on annotated InterPro domains. The 20 most expanded gene families in BSF are related to detoxification (*cytochrome P450*, *GST*, and *ABC*), chemoreception (*OR* genes), the immune system (*AMP*), and some regulation modules (Supplementary information, Table S9). In summary, several lines of genome-wide evidence suggest connections between the environmental adaptation of BSF and rapidly evolving, adaptive functional modules.

Expansions in gene families are related to BSF environmental interactions

We next focused manual annotations on gene families with potential relationships to environmental adaptation (Fig. 3). BSF larvae live in intimate interaction with various pathogens. Previous study revealed that the larval extract of BSF possessed a broad-spectrum antibacterial activity.[25] Thus, we expected that their immune system has adapted to potentially pathogenic microbes. This may be similar to the genome of the house fly, another fly species that lives in septic environments and is reported to have higher copy numbers of genes encoding recognition and effector components of the immune system than *D. melanogaster*.[21] We annotated the full set of immune-related genes in the BSF genome (Supplementary information, Table S10), providing a more complete set than a previous study using only the transcriptome.[14] Although the BSF genome encodes a similar number of genes in most signal transduction pathways (e.g., IMD, Toll, and JAK-STAT pathways) compared to other sequenced dipteran genomes, it has notable expansions in both recognition and effector molecules (Fig. 3a). The genomes of BSF and the house fly, respectively, encode 31 and 20 secreted peptidoglycan recognition proteins (PGRPs), which regulate signaling pathways during bacterial infection.[26,27] These numbers are substantially higher than those in other dipteran species (13 or fewer). The most predominant PGRPs in BSF are PGRP-LBs (20 in BSF versus 4 or fewer in other dipterans), which negatively regulate the IMD pathway, and PGRP-SAs (6 versus 3 or fewer in other dipterans), which activate the Toll pathway.[26,27] Interestingly, the house fly has eight genes encoding PGRP-SC2 proteins (the most in Diptera), which are completely absent in BSF. Expansion of recognition components may also facilitate responses to diverse pathogens.[21] Among these are gram-negative binding proteins (GNBPs), hemolymphatic proteins that participate in the activation of the Toll pathway. We identified 16 GNBP-coding genes in the BSF genome (Fig. 3b), strikingly more than that in any other dipteran species (the second most is 7 in the mosquito (Diptera: Culicidae)).

The BSF genome also encodes 50 antimicrobial peptides (AMPs), making the largest AMP family yet identified in insects. The majority of AMPs in BSF belong to the cecropin family: we identified 36 cecropin-coding genes in a BSF-specific expanded lineage (Fig. 3c). By comparison, although the house fly genome encodes a similar number of AMPs (33), only 12 of them are cecropins. We also found substantial gene expansions in genes encoding lysozymes, another ubiquitous type of immune effector,[26,27] with 36 and 34 lysozyme genes in BSF and the house fly versus ~10 in other dipteran species. Altogether, these results suggest that expansions of genes involved in the immune response underlie the adaptation of BSF to a pathogen-rich environment. Furthermore, these expansions occurred in parallel in two diverse dipterans that live in ecologically similar niches.

Dipterans colonize a wide range of habitats.[15] Evolution of chemoreception systems may play important roles in host specialization given that insects sense their environment largely by smell and taste.[28] We manually annotated the three main chemosensory receptor families in BSF: olfactory receptors (ORs), gustatory receptors (GRs), and ionotropic glutamate receptors (IRs). Unlike the considerable gene family sizes of GRs and IRs, we identified a total of 153 genes encoding ORs in BSF, twice the number of OR-coding genes found in the house fly, which has the second largest number of these genes previously reported in Diptera. Within this massive expansion, 91 ORs are potential BSF-specific pheromone receptors (Fig. 3d). These ORs may be involved in BSF-specific recognition of environmental cues or mating/social behaviors. The other expansions resided in three specific ORs. Or56a is reported to elicit an avoidance behavior in the presence of harmful microbes in *Drosophila*;[29] we placed 13 BSF ORs in a monophyletic sister cluster to *DmOr56a*, which may increase the plasticity of microbe detection in BSF. Or67c is sensitive to ethyl lactate and many alcohol odorants in *Drosophila*;[30] we found that 12 BSF ORs co-exist in a cluster with *DmOr67c* and *DmOr92a*. We also found a co-expansion of genes related to *DmOr46a* in both BSF and the house fly, where we identified 17 and 3 genes for these two species, respectively. Previous studies revealed that this Or46a, which is predicted to be a phenol-responsive OR,[31,32] is expressed in adults.

We further annotated five classic groups of enzymes commonly associated with detoxification of xenobiotics. Of these, we found substantial expansions of cytochrome P450s in genomes of both BSF and the house fly (Fig. 3e), with at least twice the numbers observed in other fly species. The most predominant P450s in BSF were clustered in clan 3, a clade of CYPs commonly associated with detoxification.[33,34] Unlike the house fly, which shows a rapidly evolving resistance to insecticides,[35] BSF is believed to remain highly sensitive to insecticides.[36] This co-expansion of CYP3 P450s thus challenges the assumption that an evolutionary increase in the number of detoxification-related genes necessarily contributes to insecticide resistance.[37,38]

Intestinal transcriptome of BSF larvae fed on organic waste

BSF is a promising natural recycler for the bioconversion of organic waste into feed for livestock and aquaculture. The larvae can thrive on diverse substrates including manure[39] and even food waste.[40] In insects, the midgut directly interacts with these pathogen-dense substrates.[41] We reared BSF larvae on diets supplemented with representative forms of organic wastes, including food waste and common manure types (e.g., poultry, dairy cow, and swine), and dissected the midgut at four time points (days 4, 6, 8, and 12) to profile gene expression using RNA-seq (Supplementary information, Table S11). Correlation analysis failed to distinguish clear clusters based on diet or time point (Supplementary information, Fig. S4), implying that BSF larvae utilize a common set of genes in response to different types of organic waste. We identified 9417 genes expressed in at least one sample, half of which were expressed across all diets (Fig. 4a).
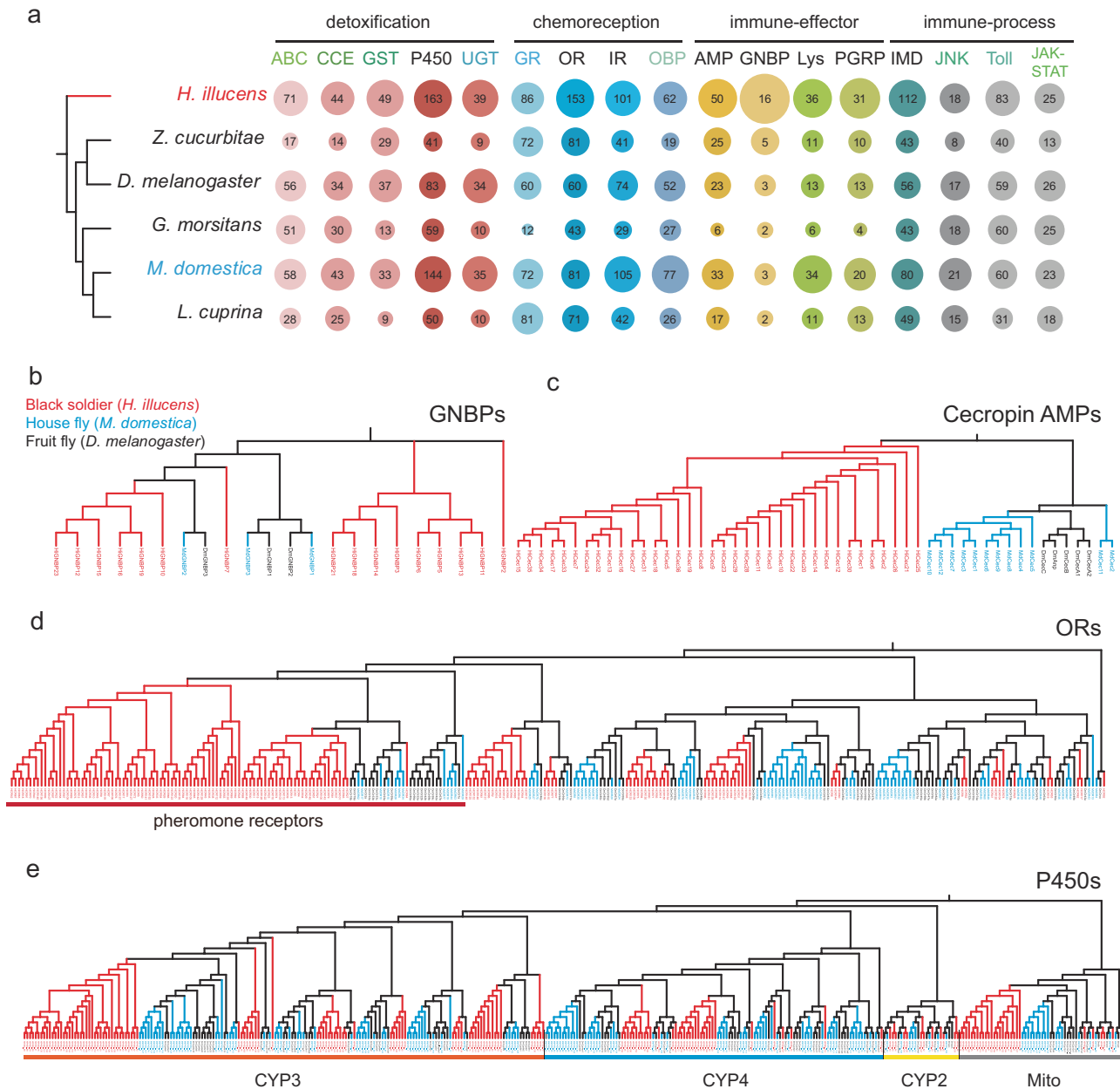
Fig. 3 Expansions in gene families related to BSF environmental adaptation. **a** Number of gene copies in the indicated families related to environmental adaptation in dipteran species. The area size of each pie indicates the relative gene number in each family. **b–e** Phylogenetic relationships across three dipteran species for gene families with prominent expansions in BSF: gram-negative binding proteins (**b**), cecropin antimicrobial peptides (**c**), Olfactoery receptors (**d**), cytochrome P450s (**e**). Phylogenetic trees were estimated using the maximum likelihood method

Principle component analysis based on the profiles of all expressed genes separated larvae fed dairy manure from other diets (Fig. 4b). This is probably explained by the fact that dairy cows are fed a specialized diet quite different from diets formulated for poultry or swine.[42]

To characterize the core set of the BSF genes important in digestion, we selected the 500 most highly expressed genes (approximately the top 5%) from larvae fed each diet (Fig. 4c). A total of 326 genes were expressed by larvae fed all four diets (Supplementary information, Table S12), supporting our hypothesis that BSF utilizes a common gene repertoire in responding to different types of organic waste. These 326 genes were significantly enriched in 13 biological pathways (Fig. 4d and Supplementary information, Table S13). Surprisingly, the most

prominent pathway was still the ribosome (ko03010; FDR $P = 1.1 \times 10^{-74}$), which is also the most significantly enriched pathway of rapidly evolving genes in BSF (Supplementary information, Tables S6 and 7). Ribosomes serve as the site of biological protein synthesis. BSF is able to colonize a multitude of decomposing organic substrates and yield high load of proteins. Signatures of genome evolution and expression both pinpointed the ribosome as the extreme outlier, revealing the strong association between the ribosome of BSF and its unique ability to utilize organic wastes. Other significantly enriched pathways were related to digestive systems and infectious diseases in human (pathways in red in Fig. 4d). We also identified a total of 150 BSF-specific genes that were highly expressed (Transcripts Per Million (TPM) ≥ 200) (Supplementary information, Table S14) but whose functions have

a

b

c

d

e
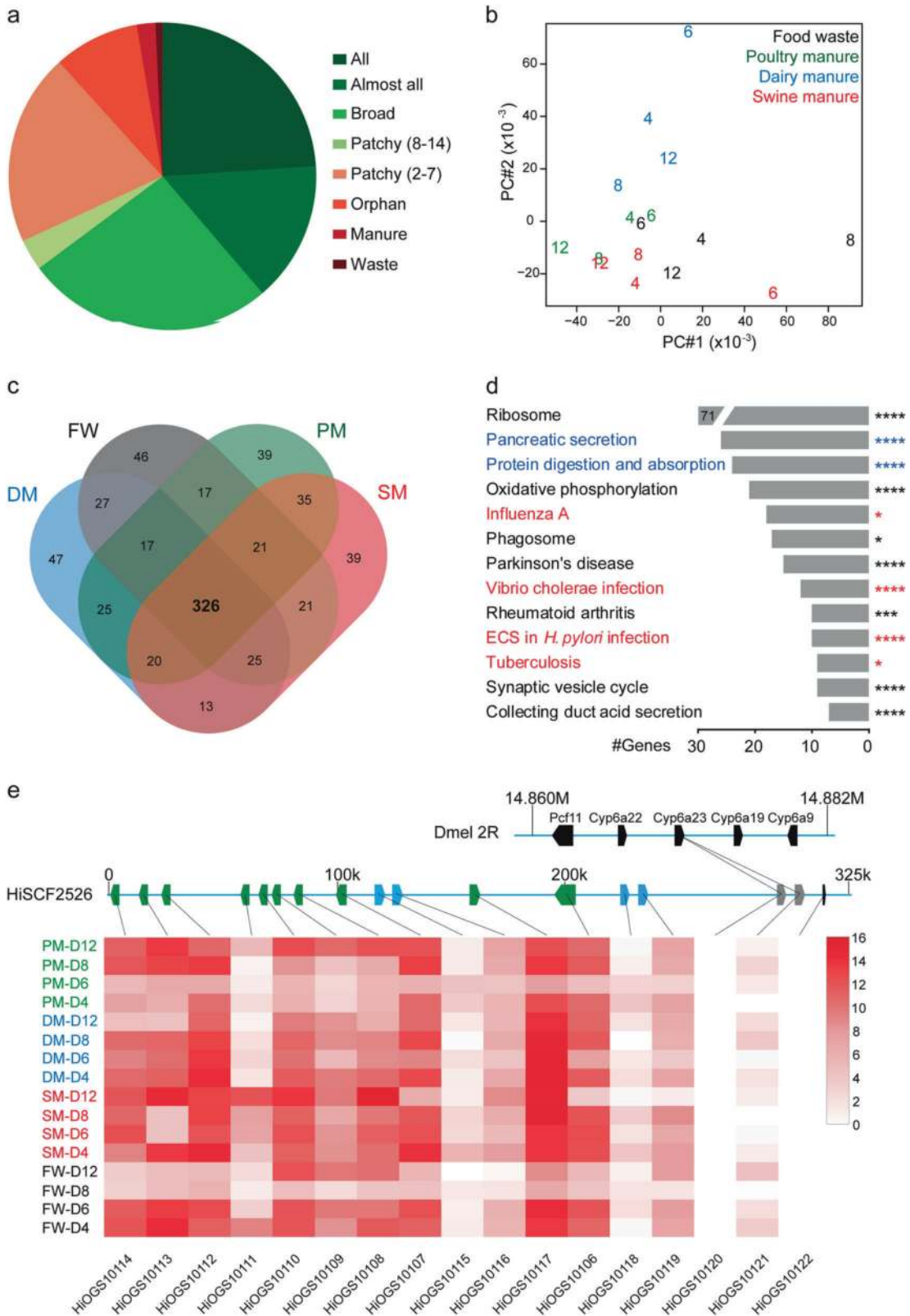
yet to be defined. These genes may uniquely contribute to the unique adaptive traits of BSF, which deserve further functional studies. Additionally, many genes that encode factors involved in the immune system were globally expressed across diets, although a fraction of them were expressed in any sample

(Supplementary information, Fig. S5). Interestingly, some of these novel genes were located in clusters on the same scaffold and co-expressed at extremely high levels (Fig. 4e and Supplementary information, Fig. S6), suggesting their biological importance, which may have driven their recent duplication in the evolution

Fig. 4 Intestinal transcriptome in BSF larvae fed with organic waste. Midguts of BSF larvae fed with food waste (FW), poultry manure (PM), dairy manure (DM), or swine manure (SM) were sampled on days 4, 6, 8, and 12 of feeding with the indicated diet. The samples were subjected to RNA-seq. **a** Distributions of expressed genes ($n = 9417$) across 16 samples: Genes expressed at each time point under each type of diet are labeled "All"; those expressed in 15 out of 16 samples are labeled "Almost all"; genes commonly expressed under each diet but not at every time point are labeled "Broad"; genes only expressed in one sample are labeled "Orphan"; genes only expressed by larvae fed with manure are labeled "Manure"; and genes only expressed in larvae fed with food waste are labeled "Waste". **b** Principal component analysis of intestinal samples based on their overall expression profiles. The first two eigenvectors that explained 34.2% and 20.4% of the variance are plotted. **c** Venn diagram of the 500 most highly expressed genes (~5% of all expressed genes), selected for each type of diet based on the average expression values across all time points. A total of 326 genes were expressed by larvae fed all four diets. **d** The 326 genes expressed by larvae fed all four diets were subjected to KEGG enrichment analysis. Pathways in blue belong to digestive systems, and pathways in red indicate those related to infectious diseases. Gene counts are presented as histograms. Hypergeometric test (FDR-adjusted): *$P < 0.05$, ***$P < 0.005$, ****$P < 0.001$. **e** A representative gene cluster specific to BSF and highly expressed in larvae fed with organic waste. Genomic organization in BSF and the homologous region in *D. melanogaster* are shown. Homolog pairs between these species are linked by lines. Genes in green and blue indicate BSF-specific genes that belong to two ortholog groups. These 14 genes do not have homology to genes of any other sequenced invertebrate species. Note that this cluster is located in the end of an assembled BSF scaffold. The heatmap shows the expression pattern of corresponding genes in BSF larvae fed with the other diets at each of the four time points

of BSF. The role of these gene clusters in digestion could be explored by genetic manipulation and bioassays.

Microbiota of BSF larvae fed on organic wastes
In addition to the metabolism performed by the midgut itself, the intestinal microbiota is fundamental for bioconversion processes in insects.[43] Although a few studies have reported preliminary investigation of the BSF intestinal microbiota in limited conditions,[10,44–49] a full landscape of BSF microbiota present in response to common types of manure and feeding time dynamics is still absent. Consequently, we investigated how different organic wastes influenced the microbiota in the midgut of BSF. Briefly, we performed 16S rRNA gene-based community profiling to probe microbial load and diversity (Supplementary information, Table S15) and analyzed the microbial diversity induced by each diet based on the operational taxonomic unit (OTU) richness. BSF larvae fed with dairy and swine manure yielded a higher microbiota complexity than those fed with poultry manure (Fig. 5a). Canonical analysis of principal coordinates based on between-sample diversity showed a strong clustering of microbiota communities based on diet (Fig. 5b), which explained 43.2% of the overall variance ($P < 0.001$). Unlike the undifferentiated pattern of midgut transcriptome, both the between-sample diversity and Bray-Curtis dissimilarity analysis suggested that diet greatly influences the bacterial community in the BSF midgut (Supplementary information, Fig. S7). The most likely explanation would be that some members of the midgut flora were directly imported along with the diet rather than induced to proliferate.

We generated a full catalog of microbiota composition across different types of diets and time points (Fig. 5c and Supplementary information, Fig. S8). A total of 16 phyla of OTUs were detected independently of diet or time point; these phyla are the core microbiome associated with the digestion of organic waste in BSF. Firmicutes were the most dominant bacterial phylum with an abundance of 59% in larvae fed with swine manure and 74% in larvae fed with dairy manure (Fig. 5c). Firmicutes, independent of diet, were present at similar levels of OTU richness, suggesting that balancing the Firmicutes community is critical for the digestion process in BSF larvae. Firmicutes have an important role in digestion of animal manure as these bacteria secrete a variety of proteases and pectinases and are involved in degradation of indigestible carbohydrates in straw-related compost.[50,51] Firmicutes have also been linked with obesity in human where a large Firmicutes population is capable of converting food to energy at greater rates.[52] The dominance of Firmicutes further reinforces the economic significance of BSF in recycling wastes and producing a high fat load for use as feed. Within Firmicutes, sequences belonging to the classes Bacilli and Clostridia dominated the BSF community (Supplementary information, Table S16). Two other abundant phyla were Bacteroidetes and

Proteobacteria, which dominated the midgut fed with food waste and poultry manure, respectively. Bacteroidetes are specialists in degradation of high molecular weight organic matter.[53] Further, an increased abundance of Proteobacteria has been proposed as a potential microbial signature of disease in human.[52] Thus, the toleration of the BSF midgut to a high load of Proteobacteria provides a potentially informative model for Proteobacteria-related diseases.

Genetic manipulation to facilitate the utilization of BSF larvae
Despite its great potential for consuming organic waste, some features of BSF need to be improved for industrial use. The many kinds of omics data presented in this study, as well as the relative genetic conservation between BSF and the widely used model insect *Drosophila*, provide a strong foundation to screen molecular targets for optimizing key features of BSF. Thus, we developed a CRISPR/Cas9-based genome editing approach in BSF and implemented this technology to test the function of some modified BSF genes in vivo.

Flies mainly feed during the larval stage. The efficiency of consuming organic waste would be improved if the larval stage of BSF could be stably extended. Metamorphosis in insects is controlled by a cascade of hormones and neuropeptides.[54] By screening genes involved in this peptidergic signaling pathway, we focused on a gene encoding the prothoracicotropic hormone (PTTH), which contributes to molting and metamorphosis by initiating the signaling cascade that results in the biosynthesis and release of ecdysone.[55] Knockout of *Ptth* in BSF dramatically delayed the pupation process of BSF larvae. The genome of BSF encodes a single copy of *Ptth*, with marginal conservation with the *Drosophila* ortholog except for the N-terminal end (Supplementary information, Fig. S9). We designed two sgRNAs, targeted to the second and fourth exons, to disrupt *HiPtth* substantially in vivo (Fig. 6a, b). Upon CRISPR/Cas9-mediated ablation of *HiPtth*, the average duration of the last larval instar increased from 4–5 days in controls to > 85 days in mutant larvae of any mosaic forms of disrupted *HiPtth*. We also found that both the body size (Fig. 6c) and weight (Fig. 6d) of the *Ptth* mutants were significantly larger than wild type. It has been proposed that *Ptth* does not mediate the growth rate in *Drosophila*.[56] This suggests that the increased body size of *Ptth* mutants probably results from the prolonged feeding. An increased feeding capacity will likely benefit the utilization of BSF larvae by increasing consumption efficiency per insect.

Another characteristic that may lighten the burden of maintaining large numbers of BSF insects is to restrict the movement of the adults. Loss of flight ability is a classic phenotypic change in the domestication of animals (e.g., silkworm, chicken). To develop such phenotypes in BSF, we annotated orthologs of *Drosophila* genes involved in wing development and used CRISPR/Cas9 to test their function in BSF. *Vestigial* (*Vg*) encodes a selector gene that specifies
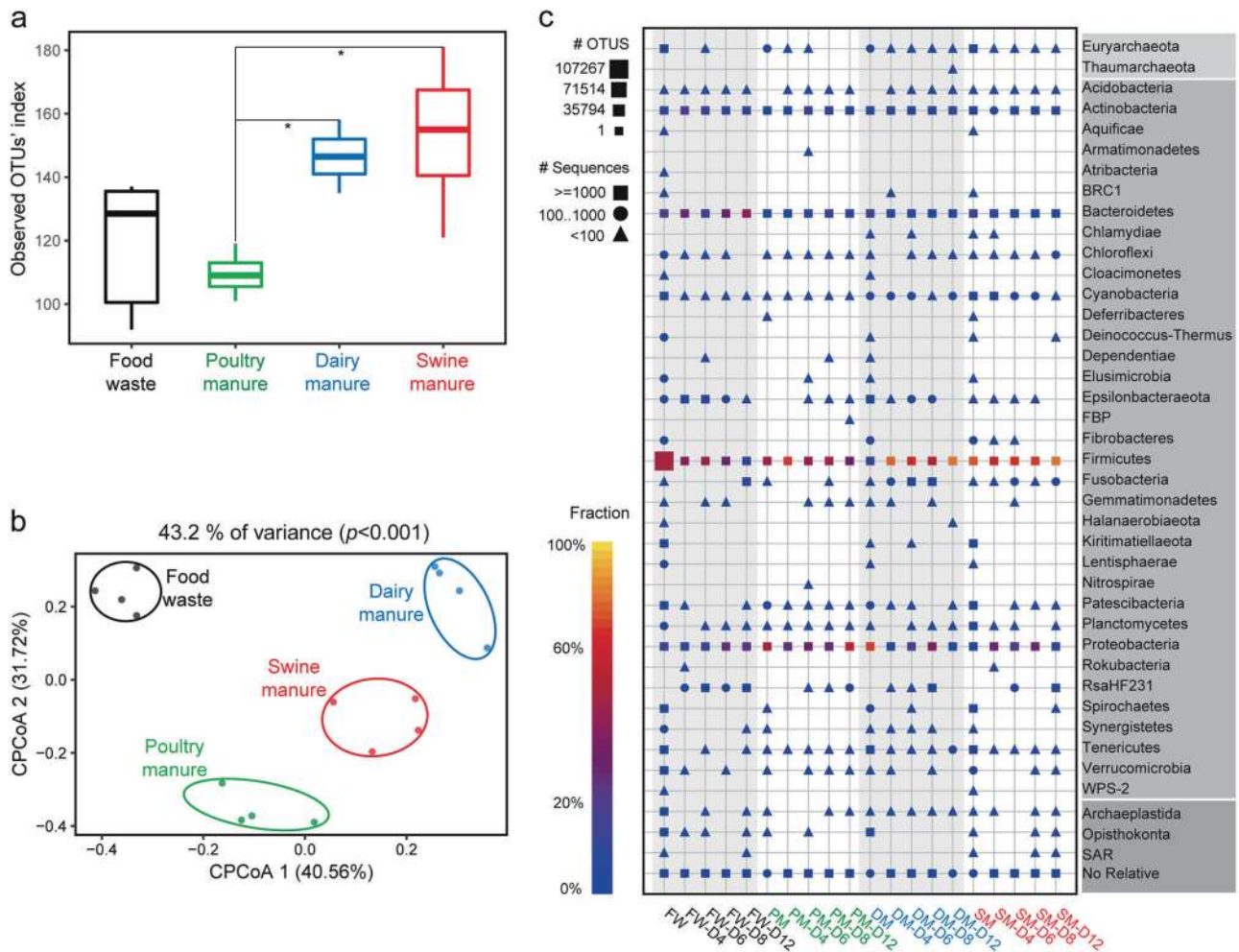
Fig. 5 Microbiome of BSF larvae fed with different types of organic waste. **a** Within-sample diversity estimates of the bacterial communities in larvae fed with the indicated diets. **b** Constrained principal coordinate analysis of between-sample diversity. Bray-Curtis distances between samples constrained by diets plotted for the first two CPCoAs. **c** The dynamic landscape of OTUs across all communities at a phylum level. OTU richness is indicated by the area of corresponding symbols. Symbols indicate counts of contained sequences. Colors indicate the fraction of target OTUs relative to all OTUs of the corresponding sample

the size and shape of *Drosophila* wings.[57,58] Alignment analysis revealed that *Vg* in BSF is conserved with the *Drosophila* copy (Supplementary information, Fig. S10). We found that somatic mosaic mutants of *Vg* (any type of deletions in Fig. 7a, b) were viable but completely lacked wings without exhibiting any other morphological or developmental abnormalities (Fig. 7c). Since mosaic mutagenesis was enough to lead to a deficient phenotype, this flightless BSF line could be maintained in the laboratory by outcrossing with the lines expressing Cas9 and sgRNA, respectively, in the future. Through this work, we could potentially reduce the BSF colony foot print and enable the development of an industrial insect in an urban environment with a greater production value.

SUMMARY AND PERSPECTIVE
In this study, we generated a high-quality genome assembly of BSF with a full set of gene annotations. Comparative genomic analyses revealed multiple gene duplication events in families related to septic adaptation including those involved in the immune system and various classes of digestion. We also characterized the core gene catalog and the microbiota community in larvae fed with food waste and three kinds of animal manure. We established a high-efficiency CRISPR/Cas9-based gene manipulation system for BSF, and generated two types of BSF mutant lines with improved

characteristics for potential industrial application, including prolonged feeding duration accompanied by increased size, and defective flight. Recent study evaluated the passive transmission of animal parasites by feeding of BSF and indicated issues of potential contamination.[59] Our study provides a list of essential genes responding to a variety of organic wastes (Supplementary information, Tables S12–S14). Genetic modification of these potential targets, using the genome editing system that we established, should be forthcoming to ensure that the recycling process by BSF is safe.

This study represents the most comprehensive molecular study on BSF to date. Data generated through this publication will allow accelerated research on BSF and its uses in agriculture globally. While this study will facilitate efforts to improve the attributes of BSF as a natural recycler and promote the use of BSF as a model to study adaptation to septic environments, it will also serve to establish BSF as a model organism for conducting basic research.

MATERIALS AND METHODS
Genome sequencing
The quality of de novo assembly is sensitive to genomic heterozygosity. For genome sequencing, we used a laboratory-maintained line of BSF, which was originally sampled in Wuhan,
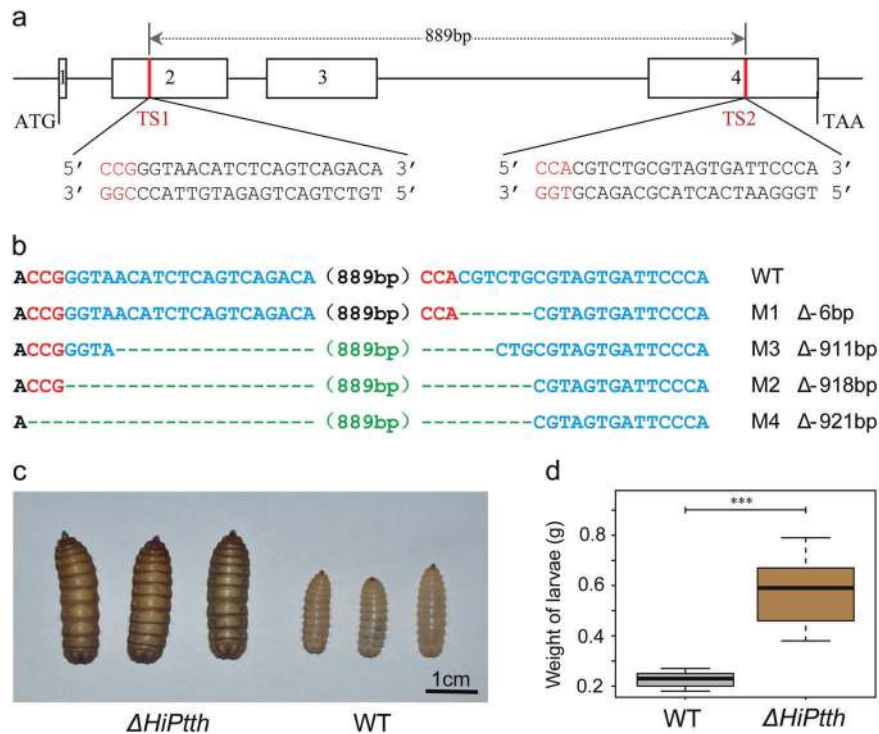
Fig. 6 Mutagenesis of *Ptth* leads to increased feeding capacity in BSF larvae. The CRISPR/Cas9 system was used to induce mutations at the *HiPtth* locus in *H. illucens*. **a** Schematic representation of the exon/intron boundaries of the *HiPtth* gene. Exons are shown as boxes; thin lines represent introns; numbers are fragment lengths in base pairs (bp). Target site (TS) locations are noted and PAM sequences are shown in red. **b** Sequences of the targeted region in the *HiPtth* locus in the mutants. The PAM sequence is in red. The numbers of nucleotides deleted in each line are indicated on the right. **c** Morphology of *HiPtth* mutants showing their greater size relative to wild type (WT) controls. **d** Average body weights of mutants and control (*n* = 30; mean values ± SEM)

China (30.6°N, 114.4°E) and underwent inbred crossing for ten generations by Dr Ziniu Yu's Lab. The line was kept under a 16:8 L: D photoperiod at 25 °C and relative humidity of 35%–40% in plastic cages and was fed wheat diet. Genomic DNA was isolated from pupae using standard protocols as described.[60] We note that BSF is commonly susceptible to infection by entomopathogenic nematodes and microorganisms due to its special living environment. Pupae of BSF are of relatively less genetic contamination than other developmental stages. We employed Illumina sequencing platforms to generate genomic reads of high coverage and libraries with stepwise-increased insert size (Supplementary information, Table S1). The MiSeq platform was employed to generate a library using DNA from a single pupa with fragment size of ~450 bp and relatively long paired-end reads (250 bp at each end); thus paired ends could be bridged into single long reads that were used to build initial contigs. The HiSeq platform was employed to generate standard paired-end and mate-pair reads (150 bp at each end). Libraries of increasing insert size, ranging from 800 bp to 13 Kb, were used to assemble scaffolds. Genomic DNA used for long read libraries required a large amount of DNA; thus, DNA from brothers or sisters of the individual that was used for MiSeq sequencing was combined. All libraries were constructed following Illumina standard protocols. Library construction and sequencing were performed by Berry Genomics Co. Ltd.

Genome assembly
K-mer analysis was initially performed to estimate the basic characteristics of the BSF genome. Adaptors and low-quality bases were trimmed using Seqtk v1.0 (https://github.com/lh3/seqtk) as described previously. Kmers were counted using jellyfish version global 1 with 21-mers.[61] Heterozygosity and other characteristics were determined by GenomeScope v1.0.0.[62] The genome size was

determined as ~1 Gb. A total of 314 Gb sequencing data, which equals > 300-fold genome coverage, was generated to perform de novo assembly of the BSF reference genome. Details of sequencing data are presented in Supplementary information, Table S1. MiSeq read pairs were utilized to assemble contigs using DiscovarDeNovo (v52488; http://software.broadinstitute.org/software/discovar) with default parameters. Initial contigs were processed by redundans v0.11c[63] to remove potential redundant sequences. The paired-end read information from the long libraries was used step by step from 800-bp to 13-kb insert size to join contigs into scaffolds using SSPACE v3.0.[64] The remaining gaps within scaffolds were iteratively filled with paired-end reads of 250-bp and 800-bp inserts using GapCloser v1.12 available in SOAPdenovo.[65] The resulting draft assembly had a final scaffold N50 size of 1.7 Mb (spanning 1102 Mb). The completeness of the assembly was evaluated using two classic pipelines for the assessment of genome assembly, CEGMA v2.4 (Core Eukaryotic Genes Mapping Approach)[18] and BUSCO v3 (Benchmarking Universal Single-Copy Orthologs).[17] Both pipelines revealed a near-complete quality of the assembled BSF reference (Supplementary information, Table S2).

Genome annotation
Repetitive sequences and transposable elements were identified using RepeatMasker v4.0.5 (http://www.repeatmasker.org). The arthropod set of Repbase v1.40a,[66] as well as a de novo repeat library that was built by RepeatModeler v1.0.4 (http://www.repeatmasker.org), were both subjected to repeat searching. Non-interspersed repeat sequences were identified by TRF v4.04.

To aid annotation of protein-coding genes, two biological replicates each from twelve continuous developmental stages (see detailed sampling time points in Supplementary information, Fig. S3) of regularly reared BSF (25 °C, 16:8 L:D, 35% of humidity) were taken
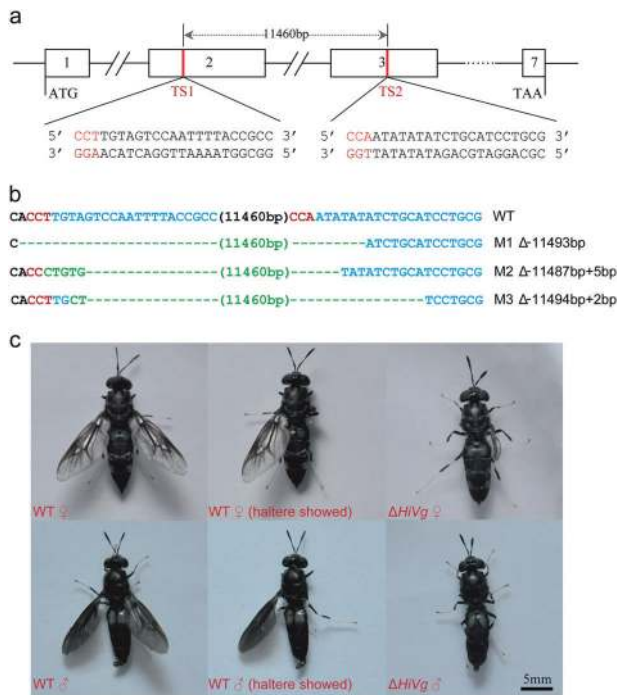
Fig. 7 Mutagenesis of *Vg* in BSF eliminates wings in adults. **a** Schematic representation of the exon/intron boundaries of *HiVg*. Exons are shown as boxes and thin lines represent the introns. Target site (TS) locations are noted and PAM sequences are shown in red. **b** Sequences of the targeted region in the corresponding loci of *Vg* mutants. The PAM sequence is in red. The numbers of nucleotides deleted in each line are indicated on the right. **c** Phenotypic images show that *Vg* mutants lack wings in the adult stage

for RNA-seq using the Illumina platform. We used HISAT2 v2.0.0-beta[67] to map RNA-seq reads to the reference genome and StringTie v1.3.4d[68] to predict exons. The official gene set (Supplementary information, Table S4) was generated from the GLEAN consensus model[69] by combing transcriptome evidence, homolog alignments, and ab initio gene annotation sets (Supplementary information, Table S4). Homolog alignments were generated using GeneWise v2.2.0[70] with protein inputs from six dipteran species (*Anopheles gambiae*,[71] *Drosophila melanogaster*,[22] *Glossina morsitans*,[72] *Lutzomyia longipalpis* (https://www.vectorbase.org/organisms/lutzomyia-longipalpis), *Musca domestica*,[21] and *Stomoxys calcitrans* (https://www.vectorbase.org/organisms/stomoxys-calcitrans)) as well as the UniProt database.[73] Three independent gene predictors were applied to generate ab initio signatures, including AUGUSTUS v3.1,[74] SNAP v2006-07-28,[75] and Genscan.[76] All pipelines were run under the default settings and subjected to feeding with GLEAN to generate a consensus set. Genes without transcriptome evidence or homology were finally removed. A total of 16,770 protein-coding genes were included in OGS 1.0. Of them, 99.3% were supported by transcriptome evidence and 85.3% were supported by homology evidence. We note that the BSF genome is of risk of contamination by symbiont or parasitic microorganism and nematodes, in particular the latter, which shares the similar DNA properties with insects and is of few public sequences. By evaluating the official gene set, we identified an extremely low fraction (~0.8%) of genes exhibiting higher sequence identity to *Caenorhabditis elegans* than to *D. melanogaster*, indicating a low percentage of nematode contamination in the BSF genome assembly.

Approximately 500 genes of biological interest were manually annotated. Some gene families, such as chemosensory receptor genes, which are difficult to identify from automated predictions, were identified directly in the genome assembly using an iterative searching approach. In brief, TBLASTN searches with dipteran homologs as queries were used to determine genomic loci with significant hits (E < 10^{-5}); then gene structures were predicted using GeneWise v2.2.0.[70] Genes were also functionally clustered by conserved domains or biological pathways based on the KO annotation of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.[77] For comparison of gene family expansion and contraction, a local InterProScan v5.26–65.0 was performed for each dipteran genome. Expression profiling was determined using salmon v0.12.0[78] with the parameter "–validateMappings". Normalized expression values, expressed as TPM, were used to compare expression levels across samples.

Comparative genomics
Published dipteran genomes were selected for ortholog analysis. As inputs, we removed proteins of short length (< 30 aa) and redundant splicing isoforms for each protein set. All-against-all protein comparisons were performed using BLASTP with E < 10^{-5}, then HSPs were processed using orthomclSoftware-v2.0.2. MCL v10-201[79] was subsequently used to define the final orthologs, inparalogs, and co-orthologs, following the suggested parameter values. To infer the phylogenomic relationships across dipteran species, 814 strict single-copy universal ortholog groups were utilized. Multiple alignments of protein sequences for each group were performed using Muscle v 3.8.31[80] and then processed by Gblocks v 0.91b to identify conserved blocks.[81] Conserved blocks were finally concatenated to 10 super genes with 255,475 amino acids, which were used to quantify the maximum likelihood phylogeny using RAxML v8.2.10.[82] The JTT model with 100 bootstraps was used for the analysis. We also used pal2nal v13[83] to process the Muscle alignments to calculate synonymous (*d*S) and non-synonymous (*d*N) substitution rates. Codeml from the PAML package v4.3 was used to calculate *d*N/*d*S ratios under the F3X4 codon frequency.[84] Functional enrichment analyses were performed via an online OMICSHARE cloud platform (http://www.omicshare.com/tools/Home/Soft/pathwaygsea).

Analysis of the BSF intestinal transcriptome
BSF was fed with wheat bran and reared under standard conditions until the sixth day of the larval stage. The same colony of larvae at the same developmental stages were treated in parallel with food waste, fresh poultry manure, fresh dairy manure, and fresh swine manure for 12 days. During this process, midguts were dissected and sampled at four time points, days 4, 6, 8, and 12, after exposure. Total RNA of each sample was independently prepared using Trizol and stored at −80 °C. Construction of cDNA libraries and subsequent sequencing using the Illumina Hiseq 4000 platform under the 2 × 150 bp mode were conducted by Berry Genomics Co. Ltd. Statistics of sequence data are shown in Supplementary information, Table S11. Each sample was independently mapped to the reference genome and subjected to expression profiling using the mode "quant" of salmon v0.12.0[78] with the parameter "-validateMappings". All independent profiles were finally merged to a "TPM" matrix using the mode "quantmerge" of salmon v0.12.0. Expression profile-based principle component analysis was performed using the built-in R function "prcomp". Highly expressed genes were selected based on the TPM's rank of each diet treatment group; the most highly expressed 500 genes were determined. Pathway enrichment analyses were performed via the online platform "OMICSHARE" (https://www.omicshare.com/). FDR-adjusted multiple tests were added to the hypergeometric test.

Metagenomic analyses of BSF intestinal microbiota
Samples described above for transcriptome sequencing were also used to explore gut microbiota by 16S rRNA sequencing. Microbial

DNA was prepared from midguts using the Gentra Puregene Yeast/Bact Kit B (Qiagen), following the manufacturer's protocol. As controls, before feeding to BSF, DNA was independently isolated from various diets, food waste, poultry manure, dairy manure, and swine manure, using a QIAamp PowerFecal DNA Kit and the total 16S rRNA was kept frozen at −80 °C for further use. Standard libraries (V3 + V4 regions) were constructed and subjected to paired-end sequencing on an Illumina Hiseq 2500 platform under the 2 × 250 bp mode. Clean read pairs were merged using the built-in command "join_paired_ends.py" from QIIME v1.9.14.[85] OTU analyses were performed by VSEARCH v2.13.05.[86] Within- and between-sample diversities were estimated by the built-in QIIME scripts "alpha_diversity.py" and "beta_diversity.py", respectively. The dynamic landscape of OTUs was generated using the online platform, SILVAngs (https://www.arb-silva.de/ngs).[87]

Mutagenesis of BSF target genes

We predicted and verified the HiPtth and HiVg ORFs based on manual annotations. Orthologs in other species were searched using BLASTP. Multiple alignment was performed using Clustal Omega v1.2.4.[88] With the PAM sequences in consideration, newly designed sgRNAs should follow the NNN19GG rule.[89] Based on our annotations and sequence identity, we identified two 23-bp sgRNA targeting sites named S1 and S2. sgRNA templates were transcribed using a T7 promoter and synthesized in vitro using the MAXIscript T7 Kit (Ambion, Austin, TX, USA) according to the manufacturer's instruction. The Cas9 protein was purchased from Thermo Fisher.

Fertilized eggs were collected within 1 h and microinjection was performed within 2 h of oviposition. Cas9 protein (200 ng/μL) with the sgRNA-1 (100 ng/μL) and sgRNA-2 (100 ng/μL) molecules were co-injected into preblastoderm embryos. Injected eggs were incubated in a humidified chamber at 25 °C for 3–4 days until hatching. Hatched larvae were reared on wheat food at 25 °C. To identify somatic mutations induced by the treatment combinations, first instar larvae were selected for genomic DNA preparation. Fragments covering the two targeting sites were amplified with the following primers: HiVg-TS1-F, GACATCTGCAAGGAT CAGGT; HiVg-TS1-R, GCCAGAACATGGTGAAAGTAT; HiVg-TS2-F, CA CTATATGGTGCCTAGGACT; HiVg-TS2-R, GGATCTTACGAGGACTTC CT; HiPtth-TS-F, ATGAGGCCTTGGGTAAGTCAG; and HiPtth-TS-R, TT AGAAAGAGCAAAAGCAACCAGTTG. The amplified fragments were cloned into a pJET1.2 vector (Fermentas) and sequenced on the Sanger platform. The positive statistics of injection were listed in Supplementary information, Table S17.

Data availability

All raw reads and assembled sequence data have been uploaded to NCBI under BioProjectID PRJNA547968 and SRA under SRR10158821.

AUTHOR CONTRIBUTIONS

Y.H. conceived and supervised the project. Y.H., S.Z. and J.Z. designed the study. J.Z., Z.Y., L.Z. and M.C. provided samples and performed treatments. Y.C., L.B., M.C., Z.K., JuX., Y.J., X.L., JiX. and X.X. performed experiments. S.Z., G.F. and Y.Z. analyzed the data. S.Z., Y.H., J.K.T., J.Z., JuX., L.Z., Z.Y., H.Y., Z.Z. and S.W. interpreted the data. S.Z., H.Y. and J.K.T. wrote the manuscript. Affiliations 1 and 3 contributed equally to this study.

REFERENCES

1. St-Hilaire, S. et al. Fly prepupae as a feedstuff for rainbow trout, *Oncorhynchus mykiss*. *J. World Aquacult. Soc.* **38**, 59–67 (2007).
2. Bondari, K. & Sheppard, D. Soldier fly larvae as feed in commercial fish production. *Aquaculture* **24**, 103–109 (1981).
3. Bondari, K. & Sheppard, D. Soldier fly, *Hermetia illucens* L., larvae as feed for channel catfish, *Ictalurus punctatus* (Rafinesque), and blue tilapia, *Oreochromis aureus* (Steindachner). *Aquac. Res.* **18**, 209–220 (1987).
4. Hale, O. M. Dried *Hermetia illucens* larvae (Diptera: Stratiomyidae) as a feed additive for poultry. *Ga. Entomol. Soc. J.* **8**, 16–20 (1973).
5. Li, Q. et al. Bioconversion of dairy manure by black soldier fly (Diptera: Stratiomyidae) for biodiesel and sugar production. *Waste Manag.* **31**, 1316–1320 (2011).
6. Surendra, K., Olivier, R., Tomberlin, J. K., Jha, R. & Khanal, S. K. Bioconversion of organic wastes into biodiesel and animal feed via insect farming. *Renew. Energy* **98**, 197–202 (2016).
7. Choi, Y.-C. et al. Potential usage of food waste as a natural fertilizer after digestion by *Hermetia illucens* (Diptera: Stratiomyidae). *Int. J. Indust. Entomol.* **19**, 171–174 (2009).
8. Beskin, K. V. et al. Larval digestion of different manure types by the black soldier fly (Diptera: Stratiomyidae) impacts associated volatile emissions. *Waste Manag.* **74**, 213–220 (2018).
9. Perednia, D. A., Anderson, J. & Rice, A. A comparison of the greenhouse gas production of black soldier fly larvae versus aerobic microbial decomposition of an organic feed material. *Res. Rev. J. Ecol. Environ. Sci.* **5**, 10–16 (2017).
10. Erickson, M. C., Islam, M., Sheppard, C., Liao, J. & Doyle, M. P. Reduction of *Escherichia coli* O157: H7 and *Salmonella enterica* serovar enteritidis in chicken manure by larvae of the black soldier fly. *J. Food Prot.* **67**, 685–690 (2004).
11. Cai, M. et al. Systematic characterization and proposed pathway of tetracycline degradation in solid waste treatment by *Hermetia illucens* with intestinal microbiota. *Environ. Pollut.* **242**, 634–642 (2018).
12. Müller, A., Wolf, D. & Gutzeit, H. O. The black soldier fly, *Hermetia illucens* — a promising source for sustainable production of proteins, lipids and bioactive substances. *Z. Naturforsch. C. Biosci.* **72**, 351–363 (2017).
13. Wang, Y.-S. & Shelomi, M. Review of black soldier fly (*Hermetia illucens*) as animal feed and human food. *Foods* **6**, 91 (2017).
14. Vogel, H., Müller, A., Heckel, D. G., Gutzeit, H. & Vilcinskas, A. Nutritional immunology: diversification and diet-dependent expression of antimicrobial peptides in the black soldier fly *Hermetia illucens*. *Dev. Comp. Immunol.* **78**, 141–148 (2018).
15. Wiegmann, B. M. & Richards, S. Genomes of Diptera. *Curr. Opin. Insect Sci.* **25**, 116–124 (2018).
16. Xu, J., Xu, X., Zhan, S. & Huang, Y. Genome editing in insects: current status and challenges. *Natl. Sci. Rev.* **6**, 399–401 (2019).
17. Waterhouse, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2017).
18. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
19. Boulesteix, M., Weiss, M. & Biémont, C. Differences in genome size between closely related species: the *Drosophila melanogaster* species subgroup. *Mol. Biol. Evol.* **23**, 162–167 (2005).
20. Kapusta, A., Suh, A. & Feschotte, C. Dynamics of genome size evolution in birds and mammals. *Proc. Natl Acad. Sci. USA* **114**, E1460–E1469 (2017).
21. Scott, J. G. et al. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biol.* **15**, 466 (2014).
22. Adams, M. D. et al. The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
23. Nakajima, R. T., Cabral-de-Mello, D. C., Valente, G. T., Venere, P. C. & Martins, C. Evolutionary dynamics of rRNA gene clusters in cichlid fish. *BMC Evol. Biol.* **12**, 198 (2012).
24. Li, Z.-W. et al. On the origin of de novo genes in *Arabidopsis thaliana* populations. *Genome Biol. Evol.* **8**, 2190–2202 (2016).
25. Park, S. I., Chang, B. S. & Yoe, S. M. Detection of antimicrobial substances from larvae of the black soldier fly, *Hermetia illucens* (Diptera: Stratiomyidae). *Entomol. Res* **44**, 58–64 (2014).
26. Buchon, N., Broderick, N. A. & Lemaitre, B. Gut homeostasis in a microbial world: insights from *Drosophila melanogaster*. *Nat. Rev. Microbiol.* **11**, 615–626 (2013).
27. Early, A. M. et al. Survey of global genetic diversity within the *Drosophila* immune system. *Genetics* **205**, 353–366 (2017).

28. McBride, C. S. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc. Natl Acad. Sci. USA* **104**, 4996–5001 (2007).

29. Stensmyr, M. C. et al. A conserved dedicated olfactory circuit for detecting harmful microbes in *Drosophila*. *Cell* **151**, 1345–1357 (2012).

30. Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).

31. Song, E., de Bivort, B., Dan, C. & Kunes, S. Determinants of the *Drosophila* odorant receptor pattern. *Dev. Cell* **22**, 363–376 (2012).

32. Ray, A., Van Naters, W. G. & Carlson, J. R. Molecular determinants of odorant receptor function in insects. *J. Biosci.* **39**, 555–563 (2014).

33. Mao, W., Schuler, M. A. & Berenbaum, M. R. CYP9Q-mediated detoxification of acaricides in the honey bee (*Apis mellifera*). *Proc. Natl Acad. Sci. USA* **108**, 12657–12662 (2011).

34. Feyereisen, R. INSECT P450 ENZYMES. *Annu. Rev. Entomol.* **44**, 507–533 (1999).

35. Busvine, J. Mechanism of resistance to insecticide in houseflies. *Nature* **168**, 193–195 (1951).

36. Tomberlin, J. K., Sheppard, D. C. & Joyce, J. A. Susceptibility of black soldier fly (Diptera: Stratiomyidae) larvae and adults to four insecticides. *J. Econ. Entomol.* **95**, 598–602 (2002).

37. Daborn, P. J. & Le Goff, G. The genetics and genomics of insecticide resistance. *Trends Genet.* **20**, 163–170 (2004).

38. Liu, N., Li, M., Gong, Y., Liu, F. & Li, T. Cytochrome P450s — their expression, regulation, and role in insecticide resistance. *Pestic. Biochem. Physiol.* **120**, 77–81 (2015).

39. Myers, H. M., Tomberlin, J. K., Lambert, B. D. & Kattes, D. Development of black soldier fly (Diptera: Stratiomyidae) larvae fed dairy manure. *Environ. Entomol.* **37**, 11–15 (2008).

40. Nguyen, T. T. X., Tomberlin, J. K. & Vanlaerhoven, S. Ability of black soldier fly (Diptera: Stratiomyidae) larvae to recycle food waste. *Environ. Entomol.* **44**, 406–410 (2015).

41. Lehane, M. Peritrophic matrix structure and function. *Annu. Rev. Entomol.* **42**, 525–550 (1997).

42. Chen, S. et al. *Value-added chemicals from animal manure. No. PNNL-14495.* (Pacific Northwest National Lab., Environmental Molecular Sciences Laboratory, Richland, WA (US), 2003).

43. Engel, P. & Moran, N. A. The gut microbiota of insects — diversity in structure and function. *FEMS Microbiol. Rev.* **37**, 699–735 (2013).

44. Jeon, H. et al. The intestinal bacterial community in the food waste-reducing larvae of *Hermetia illucens*. *Curr. Microbiol.* **62**, 1390–1399 (2011).

45. Bruno, D. et al. The intestinal microbiota of *Hermetia illucens* larvae is affected by diet and shows a diverse composition in the different midgut regions. *Appl. Environ. Microbiol.* **85**, e01864–18 (2019).

46. Wynants, E. et al. Assessing the microbiota of black soldier fly larvae (*Hermetia illucens*) reared on organic waste streams on four different locations at laboratory and large scale. *Microb. Ecol.* **77**, 913–930 (2019).

47. Liu, Q., Tomberlin, J. K., Brady, J. A., Sanford, M. R. & Yu, Z. Black soldier fly (Diptera: Stratiomyidae) larvae reduce *Escherichia coli* in dairy manure. *Environ. Entomol.* **37**, 1525–1530 (2008).

48. Yu, G. et al. Inoculating poultry manure with companion bacteria influences growth and development of black soldier fly (Diptera: Stratiomyidae) larvae. *Environ. Entomol.* **40**, 30–35 (2011).

49. Zheng, L. et al. A survey of bacterial diversity from successive life stages of black soldier fly (Diptera: Stratiomyidae) by using 16S rDNA pyrosequencing. *J. Med. Entomol.* **50**, 647–658 (2013).

50. Sun, L., Pope, P. B., Eijsink, V. G. & Schnürer, A. Characterization of microbial community structure during continuous anaerobic digestion of straw and cow manure. *Microb. Biotechnol.* **8**, 815–827 (2015).

51. Zhang, L. et al. Enhanced growth and activities of the dominant functional microbiota of chicken manure composts in the presence of maize straw. *Front. Microbiol.* **9**, 1131 (2018).

52. Rizzatti, G., Lopetuso, L., Gibiino, G., Binda, C. & Gasbarrini, A. Proteobacteria: a common factor in human diseases. *Biomed. Res. Int.* **2017**, 9351507 (2017).

53. Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut bacteroidetes: the food connection. *Front. Microbiol.* **2**, 93 (2011).

54. Truman, J. W. Hormonal control of insect ecdysis: endocrine cascades for coordinating behavior with physiology. *Vitam. Horm.* **73**, 1–30 (2005).

55. Fellner, S. K., Rybczynski, R. & Gilbert, L. I. Ca2+ signaling in prothoracicotropic hormone-stimulated prothoracic gland cells of *Manduca sexta*: evidence for mobilization and entry mechanisms. *Insect Biochem. Mol. Biol.* **35**, 263–275 (2005).

56. McBrayer, Z. et al. Prothoracicotropic hormone regulates developmental timing and body size in *Drosophila*. *Dev. Cell* **13**, 857–871 (2007).

57. Williams, J. A., Bell, J. B. & Carroll, S. B. Control of *Drosophila* wing and haltere development by the nuclear vestigial gene product. *Genes Dev.* **5**, 2481–2495 (1991).

58. Zecca, M. & Struhl, G. Control of *Drosophila* wing growth by the vestigial quadrant enhancer. *Development* **134**, 3011–3020 (2007).

59. Müller, A., Wiedmer, S. & Kurth, M. Risk evaluation of passive transmission of animal parasites by feeding of black soldier fly (*Hermetia illucens*) larvae and prepupae. *J. Food Prot.* **82**, 948–954 (2019).

60. Wu, N. et al. Fall webworm genomes yield insights into rapid adaptation of invasive species. *Nat. Ecol. Evol.* **3**, 105 (2019).

61. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

62. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).

63. Pryszcz, L. P. & Gabaldón, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113 (2016).

64. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2010).

65. Li, R. et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).

66. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).

67. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

68. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511 (2010).

69. Elsik, C. G. et al. Creating a honey bee consensus gene set. *Genome Biol.* **8**, R13 (2007).

70. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).

71. Holt, R. A. et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129–149 (2002).

72. Initiative, I.G.G. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science* **344**, 380–386 (2014).

73. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2016).

74. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).

75. Korf, I. Gene finding in novel genomes. *BMC Bioinforma.* **5**, 59 (2004).

76. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).

77. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

78. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Meth.* **14**, 417–419 (2017).

79. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

80. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

81. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).

82. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

83. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).

84. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

85. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Meth.* **7**, 335 (2010).

86. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

87. Pruesse, E. et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).

88. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).

89. Wang, Y. et al. The CRISPR/Cas system mediates efficient genome engineering in *Bombyx mori*. *Cell Res.* **23**, 1414–1416 (2013).