**Lucilia cuprina** genome unlocks parasitic fly biology to underpin future interventions

Clare A. Anstead¹, Pasi K. Korhonen¹, Neil D. Young¹, Ross S. Hall¹, Aaron R. Jex¹, Shwetha C. Murali², Daniel S.T. Hughes², Siu F. Lee³, Trent Perry³, Andreas J. Stroehlein¹, Brendan R.E. Ansell¹, Bert Breugelmans¹, Andreas Hofmann⁴, Jiaxin Qu², Shannon Dugan², Sandra L. Lee², Hsu Chao², Huyen Dinh², Yi Han², Harsha V. Doddapaneni², Kim C. Worley², Donna M. Muzny², Panagiotis Ioannidis⁵, Robert M. Waterhouse⁵, Evgeny M. Zdobnov⁵, Peter J. James⁶, Neil H. Bagnall⁷, Andrew C. Kotze⁷, Richard A. Gibbs², Stephen Richards², Philip Batterham³ & Robin B. Gasser¹

**Lucilia cuprina** is a parasitic fly of major economic importance worldwide. Larvae of this fly invade their animal host, feed on tissues and excretions and progressively cause severe skin disease (myiasis). Here we report the sequence and annotation of the 458-megabase draft genome of **Lucilia cuprina**. Analyses of this genome and the 14,544 predicted protein-encoding genes provide unique insights into the fly’s molecular biology, interactions with the host animal and insecticide resistance. These insights have broad implications for designing new methods for the prevention and control of myiasis.
Insect vectors that transmit viral, bacterial and/or parasitic diseases are of major socioeconomic importance globally. For instance, some dipteran flies are primary parasites of plants or animals, and can also act as vectors of pathogens. In particular, some blowflies, such as *Lucilia* spp., are parasitic and feed on the tissues of animals, such as sheep. The disease caused by blowfly (flystrike or myiasis) is a serious problem in many countries around the world; in Australasia alone, hundreds of millions of dollars are lost annually due to reduced wool and body growth in sheep as well as costs associated with blowfly treatment/ control and animal morbidity. The principal fly involved in flystrike is *Lucilia cuprina* (Insecta, Diptera, Calliphoridae), with the majority of myiasis cases being initiated by this species.

Adult *L. cuprina* females are attracted to odours from the host, particularly those associated with bacterial infections in damp fleece, or areas of fleece or skin soiled by urine or faeces. They lay eggs (~200 eggs per batch per female fly) on skin areas of high humidity. Larvae (maggots) hatch from eggs within 8 h to 3 days and proceed through three stages of development. They use their mouth hooks to abrade the skin and feed on skin secretions, dermal tissues and blood. The resultant damage or ‘strike’ is mainly due to mechanical and chemical effects of larval feeding as well as protease release, which can cause severe disease and, in extreme cases, death.

Although blowfly strike has been the subject of extensive investigations over many years, and some control methods have been developed, an effective and permanent solution to flystrike has not yet been found. A common means of prevention is mulesing, a surgical procedure that removes wool-bearing skin from around the tail and from either side of the breech area of sheep, resulting in an area devoid of wrinkles or skin folds, reducing the accumulation of secretions that attract flies. This controversial practice is heavily scrutinized by animal welfare organizations, because of physical, behavioural and psychological indicators of stress that result from mulesing. Therefore, there is a need for an alternative to this surgical practice. Although immunogens have been studied, no effective vaccine is yet available against blowfly. Insecticides continue to be heavily relied upon to prevent and treat flystrike; however, this reliance is becoming increasingly problematic due to chemical residue problems in animal products and the rapid emergence of resistance in blowflies against many classes of insecticides.

Profound insights into the fundamental, molecular processes in insecticide resistance, expressed a *L. cuprina* gene (Table 1), similar to the findings for the genomes of *D. melanogaster* and *G. morsitans*. We sequenced and characterized the 458-megabase (Mb) draft genome of this fly. We also investigated particular genes involved in insecticide resistance, expressed a *L. cuprina* nicotinic acetylcholine receptor (nAChR) subunit (*Lc*α) gene in *Drosophila melanogaster* and assessed this subunit’s capacity to rescue spinosad resistance in *D. melanogaster* mutants. The present genomic resource for a parasitic fly of major agricultural importance provides a solid foundation for exploring the molecular basis of blowfly development and reproduction, fly–host interactions, the pathogenesis of myiasis and, importantly, insecticide resistance.

### Results

#### Genome assembly and repeat content

We sequenced the genome of *L. cuprina* at ~100-fold coverage (Table 1 and Supplementary Data 1), producing a final draft assembly of 458 Mb (scaffold N50: 744,413 bp; Table 1), with a mean GC content of 29.3%. This genome is more than twice the size of that of *D. melanogaster* (180 Mb), larger than that of *Glossina morsitans* (366 Mb) and smaller than that of *Musca domestica* (691 Mb). We detected 96.0% complete and 100% partial 248 core essential genes by CEGMA, indicating that the assembly represents a substantial proportion of the entire genome. The estimated repeat content of this draft genome is 57.8% (265 Mb), comprising 2.7% DNA transposons, 4.6% retrotransposons, 16.7% unclassified dispersed elements and 5.2% simple repeats (Supplementary Data 2). We identified 78,741 distinct retrotransposons representing at least three categories (16,688 LTRs, 61,619 LINEs and 434 SINEs), with *ERV_classII* predominating for LTRs (*n* = 423) and L3/CRI for non-LTRs (*n* = 6,358). We also identified 60,359 DNA transposons, of which *hAT-Charlie* (*n* = 490) and *TcMar-Tigger* (*n* = 410) predominated (Supplementary Data 2).

#### Gene set and functional annotation

We predicted 14,554 coding genes using *de novo* and homology-based predictions, of which 10,121 were supported by mapping RNA-seq reads (*n* ≥ 5) derived from larval stages (mixed) and adults (both sexes) of *L. cuprina*. Mean gene, exon and intron lengths were 12,197, 432 and 2,560 bp, respectively, with an average of 4.5 exons per gene (Table 1), similar to the findings for the genomes of *D. melanogaster*, *G. morsitans* and *M. domestica*. A total of 4,106 genes are single-copy orthologues (SCOs) shared among the four fly species, and 12,160 genes are shared with at least one other species of Diptera (Fig. 1). In contrast, 2,062 genes (14.2%) are unique to *L. cuprina*, with no homologues detected in any other dipteran for which genome sequence data are currently available (Fig. 1). Of the entire *L. cuprina* gene set, 9,822 genes (67.5%) have an orthologue (E-value cutoff ≤ 10^-5^) linked to one or more of 254 known biological (KEGG) pathways, most of which mapped to those in *D. melanogaster* (see Supplementary Data 3). The completeness of the genome is further supported by the CEGMA results (Supplementary Data 1). By inference, the majority of the *L. cuprina* gene set is represented in the present genomic assembly, and supported by extensive transcriptomic and inferred proteomic data (*n* = 10,121 and 11,553 molecules, respectively) from multiple public databases.

Of the 14,554 protein-encoding genes of *L. cuprina*, 12,160 (83.6%) had homologues in other dipterans; 10,396 (71.5%), 9,023 (62%) and 7,659 (52.7%) had significant matches in the InterProScan, UniProtKB/Swiss-Prot and KEGG BRITE databases, respectively.
Enzymes, channels, pores and transporters. In total, we identified 260 peptidases representing the six main groups (that is, metallo-, cysteine, serine, aspartic, threonine and some of an unknown catalytic type), with the serine (n = 96; 36.9%), metallo- (n = 91; 35.0%) and cysteine (n = 46; 17.7%) peptidases predominating (Supplementary Data 4). Most abundant are serine (n = 74), S28 (n = 4) and S9 ω/β hydrolases, including prolyl oligopeptidase (n = 4), among the serine peptidases; M13 nephrilysin (n = 19), M12 astacin/adamalysin (n = 17) and M14 carboxypeptidase A (n = 12) among the metallo-peptidases; and the C1 papain (n = 8), C14 caspase (n = 7) and C19 ubiquitin-specific peptidase (n = 7) families among the cysteine peptidases. Similar peptidase groups, including families S1, S9, M14, C19 and M13, are represented in the Glossina genome. Interestingly, 254 (97.7%) of the 260 peptidases identified in L. cuprina have homologues in the tsetse fly. We identified at least 167 protein kinases and 199 phosphatases to be encoded in the L. cuprina genome (Supplementary Data 5 and 6). The kinase includes serine/threonine (87.4%) and tyrosine (12.6%) protein kinases. The phosphatome includes principally protein serine/threonine (81.8%) and protein tyrosine (10.1%) phosphatases as well as a small number of haloacid dehalogenase phosphatases (8.1%). In addition, we predicted at least 92 GTPases to be encoded in L. cuprina, including 11 large (heterotrimeric) and 81 small (monomeric) G-proteins representing the Rab (n = 32), Arf/Sar (n = 16), Ras (n = 21), Rho (n = 7) and Ran (n = 3) families as well as some unclassified molecules (Supplementary Data 7). Many of these GTPases, including Ras and Rho, likely coordinate the signal transduction pathways associated with organogenesis and morphogenesis (cell division and differentiation) in the fly. For example, these molecules are involved in the dynamic assembly, disassembly and reorganization of the actin and microtubule cytoskeletons, the interaction of growing axons with other cells and extracellular matrices, the delivery of proteins and lipids to axons through exocytic machinery and/or the internalization of proteins or membranes at the leading edge of the growth cone via endocytosis. Examples of dominant small GTPase homologues are Ras64B, Rab23, GzF, Arl1, Arl2, Rab6, RabX1 and Ras85D whose D. melanogaster orthologues are essential for larval growth and/or development (www.flybase.org). Therefore, we propose that some of these and related enzymes are potential targets for interventions against L. cuprina based on their roles in other organisms such as Drosophila.

In this context, the large complement of receptor, channel, pore and transporter proteins in L. cuprina is also of particular interest, considering that many common insecticides target some of these proteins. We predicted 197 G protein-coupled receptors (GPCRs) to be encoded in L. cuprina, including rhodopsins (n = 73), secretin receptors (n = 18), metabotropic glutamate receptors (n = 9) and some unclassified proteins (Supplementary Data 8). We also predicted 136 ion channel proteins (Supplementary Data 9), the majority of which represent the voltage-gated cation channel superfamily (n = 31), such as the potassium (61.3%) and the calcium (35.5%) channel families, and the epithelial and related channel superfamily (n = 28) including acid-sensing ion channels. We also found channels of the cyclic nucleotide superfamily (n = 24), some of which (for example, nAChRs) are recognized targets of several insecticides in L. cuprina. Molecules involved in chemoreception (n = 93), including a number of gustatory and odorant receptors, were relatively
abundant, some of which (for example, Gr63a) are likely involved in the detection of host carbon dioxide²⁸, and might represent intervention target candidates. In addition, 367 transporters were inferred for *L. cuprina* (Supplementary Data 10), including an abundance of proteins of the solute carrier family (46.4%), major facilitator superfamily (24.3%) and ABC transporters (n = 42), some of which have been shown to relate to insecticide resistance via the active transport of drugs out of cells¹⁷,²⁰. We also identified seven aquaporins (aqp) genes that likely facilitate rapid, highly selective water transport into and out of cells, thus regulating osmotic pressure in cells. On the basis of evidence from other flies²¹, these aquaporins are proposed to play a role in the hydration of saliva during feeding, the reduction in volume of ingesta for the purpose of efficient digestion, the mobilization of water to progeny during oogenesis and to cold and heat tolerance in *L. cuprina*.

**Comparative transcriptomic analyses.** To explore the molecular biology of *L. cuprina*, we compared transcription between male and female adults, and between adults and mixed larval stages. Transcripts in female and male adults were highly enriched (n = 86 and 138, respectively) for gene ontology annotations such as oogenesis and vitelline membrane formation in the females, and sensory perception of chemical stimuli and defence response in the males (Supplementary Data 11 and 12).

The male-enriched transcript set (Supplementary Data 12) represents genes encoding testis-specific serine kinases (proposed to be involved in DNA condensation during post-meiotic chromatin remodelling) as well as three Niemann–Pick type C2 proteins, which are believed to regulate sterol homeostasis and the biosynthesis of 20-hydroxyecdysone, a steroid insect moulting hormone of *Drosophila*²². Niemann–Pick type C2 proteins might play a central role in chemical communication in *L. cuprina*, based on evidence for *Camponotus japonicus* (Japanese carpenter ant)²³. A total of 15 proteins belonging to the sperm-coating protein-like extracellular (SCP/TAPS) protein family were identified based on their characteristic CAP domain (IPR014044). Most SCP/TAPS proteins characterized to date are often secreted and function extracellularly in a variety of physiological processes, such as fertilization or immune responses²⁴,²⁵. For instance, in *Drosophila*, 26 SCP/TAPS genes have been identified, with 70% preferentially expressed in males²⁶, some of which are likely involved in male-specific reproductive processes. Further investigation of these genes and their function is warranted, as SCP/TAPS proteins of helminths can play key roles in reproduction, immunomodulation and/or host invasion²⁷, and might thus represent potential insecticide or vaccine candidates for various edysozoans including blowfly. Proteins phorminic (a defensin)²⁷ and cecropin C²⁸, two antimicrobial peptides of the haemolymph, known to be involved in cell-free immune attack of insects mainly against Gram-positive and/or -negative bacteria, were also represented in the male-enriched transcript set. The crucial role of these two peptides appears to link with a transcription level that is among the highest of any gene and stage of *L. cuprina* (Supplementary Data 12); the extent of male-enriched transcription likely reflects an extensive defence arsenal required to protect male flies from the onslaught of a wide range of microbes of different classes subsisting on diverse food sources/diets (including nectar, honeydew and/or carrion)²⁹.

Among the female-enriched transcripts are various orthologues associated with reproductive processes, including oogenesis/egg laying and eggshell formation (for example, Vm26Aa, Vm34Ca, Vm32E, del and yolk protein (yp) genes; see, for example, FlyBase) and/or female sex-determination (for example, stil) (see, for example, FlyBase), all of which have orthologues in *Drosophila* spp. (Supplementary Data 11). While the vitelline membrane (Vm) genes encode proteins of the first layer of the eggshell produced by the follicular epithelium, the lipase-derived yolk proteins are required for vitelligenesis in *L. cuprina*²⁰. The four yp genes specific to the female blowfly compare with three (yp1, yp2 and yp3) in *Drosophila*, but only one in *Glossina*¹¹; this difference in the number of orthologues is hypothesized to relate to oviparous reproduction in the two dipterans²¹. vis-à-vis adnetropic viviparity in the glossinid fly²⁰. By contrast, transcripts enriched in mixed-stage larvae (n = 256) of *L. cuprina* including those encoding enzymes (for example, cathepsin-D and chymotrypsin) involved in digestion, peritrophin-44 and various proteins linked to growth and development (including Cep84Ab, Lcp1, Lcp2, Lcp65Ab1 and Edg84A) were prominent (Supplementary Data 13). The cluster of genes (Lcp1, Lcp2 and Lcp65Ab1) encoding cuticle proteins is integral to determining characteristics of the cuticle, and orthologue Edg84A likely governs *L. cuprina* metamorphosis, being regulated through transcription factors (TFs) homologous to FTZ-F1 and DHR3 of *D. melanogaster*²³–²⁵. Interestingly, substantial transcription of the peritrophin-44 gene in larvae relative to adults is consistent with an abundance of this protein in the peritrophic membrane of all three larval instars, but trace amounts in adult *L. cuprina*²⁶. Through its binding to chitin, peritrophin-44 likely maintains the structure and porosity of the peritrophic membrane, a semi-permeable chitinous matrix lining the gut, which is proposed to have key roles in maintaining gut structure, protection from microbial invasion and/or the facilitation of digestion, possibly together with cathepsin-D and/or chymotrypsin.

Interestingly, 15% of the 480 transcripts enriched in larvae or either gender of the adult stage had no homologue in any other organism for which the data are currently available in public databases. Most of the 70 orphan (that is, unannotated) transcripts were identified in mixed larvae (n = 37) compared with male (n = 27) and female (n = 6) adults. These findings are consistent with those for other dipterans such as *Glossina* and *Musca*, which have similar complements of orphan genes¹¹,¹²; in a conservative comparison of 28 insect species, similar numbers of orphan genes for individual species were reported³⁷. The presence of a considerable number of orphan genes emphasizes the uniqueness of the biology of *L. cuprina* and encourages in-depth studies of the expression and functions of these unique molecules throughout the fly’s life cycle. Some of them are likely involved specifically in host invasion and/or interactions, and might represent highly selective insecticide or vaccine targets.

**Parasite–host interactions and potential vaccine molecules.** Excretory/secretory (ES) proteins can also play critical roles in the immunobiological relationship between *L. cuprina* larvae and the host animal. Here we predicted the secretome of *L. cuprina* to include 1,004 proteins with a diverse array of inferred functions, of which 234 had homologues in two or more public databases (see Supplementary Data 14). Conspicious were orthologues encoding 58 peptidases, including 47 serine proteases (for example, chymotrypsin and trypsin) and 11 aspartic proteases (for example, cathepsin). In addition, 25 genes encoding hydrolases (for example, chitinase and lipoprotein lipase), 12 mucin-like proteins, seven peritrophin proteins, seven peptidase inhibitors, including serpin B, and 30 cuticle-like proteins as well as 194 orphan molecules were identified. Many secreted peptidases representing the ‘degradome’ (and their respective inhibitors) have central roles in larval establishment,
degradation of blood, skin and various proteins and/or the activation of inflammation and immune responses\textsuperscript{4,38}; some of these peptidases could represent intervention targets in the larval stage of \textit{L. cuprina}. Of the genes encoding the 1,004 predicted ES proteins, 852 were transcribed in larval stages, and 79 were exclusive to these stages. On the basis of comparison with other ecdysozoans, 79 of the 852 (9.3%) ES molecules are predicted to be involved in host interactions and/or are immunogenic (see Supplementary Data 14), and include 11 cuticular proteins, 2 serine peptidases and peritrophin-44. Some of the annotated molecules, such as peritrophins, have already been shown to regulate larval growth and survival\textsuperscript{39} and induce temporary, protective immunity in experimental sheep against challenge infection with \textit{L. cuprina}\textsuperscript{40}. Overall, the present genomic and transcriptomic data sets infer that \textit{L. cuprina} has a major arsenal of ES proteins, including some orphan molecules, which are likely involved in inducing and/or modulating immune responses in the host animal. A detailed understanding of the roles of these molecules could contribute towards developing subunit vaccines against flystrike\textsuperscript{8}.

**Insecticide-resistance genes and functional analysis of \textit{Lc}x6.** Although there is little detailed knowledge of the molecular basis of insecticide resistance in \textit{L. cuprina}, numerous studies\textsuperscript{4} have inferred or proposed a direct or indirect involvement of various genes in such resistance, for both metabolic and target site insensitivity-resistance mechanisms. We have annotated genomic loci for five genes associated with particular resistances, including Ace (acetylcholinesterase, the target for organophosphorus insecticides, OPs), Rdl (resistance to dieldrin), LcaE7 (or Rop1—resistance to OPs; encodes carboxylesterase E3), Scl (transmembrane receptor for intracellular signalling, proposed to be modifier of phenotypes associated with Rop1-mediated OP resistance) and \textit{Lc}x6 (\textit{nAChR} \textalpha{6} subunit).

**Figure 2 | Structures of five insecticide-resistance genes in \textit{Lucilia cuprina.}** Diagrams show the genomic structures of \textit{L. cuprina} genes, which have been implicated in resistance to insecticides used to control \textit{L. cuprina} — Scl (encoding a transmembrane receptor important for intracellular signalling and proposed to modify phenotypes associated with organophosphorus (OP) insecticide resistance conferred by Rop1), LcaE7 (= Rop1, encoding carboxylesterase E3; associated with OP insecticide resistance), Rdl (resistance to dieldrin), Ace (acetylcholinesterase) and \textit{Lc}x6 (\textit{nAChR} \textalpha{6} subunit). Genes Scl, LcaE7, Rdl and Ace were located to scaffolds nos. 120, 113, 568 and 105, respectively (a–d). It was also noted that the current \textit{L. cuprina} assembly did not contain Rdl exon 10 compared with the existing \textit{Lucilia} cDNA sequence (GI: 2565319) (c). The absence of Rdl exon 10 is supported by RNA-seq data. The \textit{Lc}x6 gene (254 kb) was represented on scaffolds nos. 379, 4,253 and 792 (e), and contains 10 exons including four \textit{L. cuprina}-specific \textalpha{6} exons (called 3a, 3b, 8a and 8b; all transcribed). The \textit{Lc}x6 gene is located in a highly repetitive region, and manual sequence analysis of the paired-end reads successfully mapped the scaffold4253 (containing \textit{Lc}x6 exons 2 and 3b) to a 2.1 kb gap within scaffold379. Gene regions are indicated by blue lines; gaps within gene regions are depicted as dashed lines. Red vertical lines/boxes represent exons.
NATURE COMMUNICATIONS | DOI: 10.1038/ncomms8344
ARTICLE

Table 3 | Isoform and RNA editing status of Lc6 expression constructs.

| Construct    | Alternate exons | A-to-I RNA-edited sites* |
|--------------|----------------|--------------------------|
| UAS-Dx6      | 3b, 8b         | 4, 5 and 6               |
| UAS-Lc6      | 3b, 8a         | 4, 5, 6 and 7            |

*RNA editing sites numbered according to Perry et al.44
†GenBank accession no. KP260561.

resistance) and Lc6 (nACHr 26 subunit) (Fig. 2). Importantly, previously, we had characterized full-length L. cuprina complementary DNA (cDNA) sequences, which assisted direct cDNA–gDNA alignments to support the definition of exon–intron boundaries in the present study. Using the genomic and transcriptomic data sets for L. cuprina, we identified these genes in long genomic scaffolds and established their structures (Fig. 2), which should provide a foundation for functional studies of insecticide resistance in L. cuprina and other pests.

From previous studies41–43, we know that resistance to the widely used insecticide spinosad is due to loss-of-function (LOF) mutations in the gene encoding the nACHr 26-like subunit. Mutations in 26-like receptors in D. melanogaster, Plutella xylostella and Frankliniella occidentalis led to high levels of spinosad resistance, which suggests a common mechanism across insect species41–43. The model insect D. melanogaster proved to be very useful to explore this aspect. LOF mutations in the D. melanogaster orthologue of this gene (Dx6) confer high levels of resistance, suggesting that spinosad exerts its lethal effect by binding to this subunit. Introducing a Dx6 orthologue from various insect pest species into this LOF background has been shown to render D. melanogaster susceptible to spinosad, indicating that the introduced receptor subunit is functional and binds spinosad when expressed in D. melanogaster44. Therefore, we proposed that 26 LOF mutations confer high-level resistance to spinosad in various insect pests.

To examine whether 26-based spinosad resistance might evolve in L. cuprina, we performed heterologous expression of Lc6 in D. melanogaster (Fig. 3), and assayed for functional rescue and insecticide susceptibility in transgenic flies. Utilizing the D. melanogaster GAL4-UAS system45, we cloned Lc6 into either a dαmx or a d26W337t spinosad-resistant background (61- and 1,176-fold)44 and expressed Lc6 in the elav-GAL4 driver line of D. melanogaster (Fig. 3). Rescue experiments showed that Lc6 restored spinosad susceptibility in D. melanogaster (Fig. 3); no significant mortality in the D. melanogaster line 8ΦX-86Fb46 was observed using 0.1, 0.3 and 0.5 p.p.m. of spinosad in a dαmx background, and low mortality (9.4% ± 6.8) was seen only at 0.5 p.p.m., but not at the two lower doses in a d26W337t background. The UAS-Dx6 insertion line was susceptible to spinosad at all three doses, whereas the UAS-Lc6 line was susceptible only at 0.5 p.p.m. (due to ‘leaky expression’ at the aptP landing site47). The driver line elav-GAL4 expressing Dx6 was highly susceptible at all three doses. Although transgenics with the Lc6 subunit responded significantly at all doses, mortality at 0.1 p.p.m. was significantly lower than Dx6 in both the backgrounds (dαmx and d26W337t) when driven by elav-GAL4, showing that rescue was not as efficient as for Dx6.

Prospects for new insecticides. Clearly, the excessive use of various chemicals against L. cuprina has led to major insecticide-resistance problems3. Unfortunately, limited progress has been made in discovering new classes of insecticides effective against this parasite3. Genomic-guided drug target or drug discovery provides a promising approach to support screening and repurposing48; the goal of such discovery is to identify genes or gene products whose inactivation by one or more insecticides selectively kill fly larvae but do not harm the host animal. As gene-specific perturbation by double-stranded RNA interference is not yet practical for the direct evaluation of gene functions on a genome-wide scale in L. cuprina, gene essentiality can be predicted from functional genomic data (for example, lethality) for D. melanogaster, and this approach has already yielded credible insecticidal targets and provided insight into the mechanisms of resistance48. In L. cuprina, we inferred 988 genes with essential homologues/orthologues in D. melanogaster linked to lethal or semi-lethal phenotypes on gene silencing (Supplementary Data 15). We assigned highest priority to insecticide or vaccine target candidates inferred to be encoded by single genes, reasoning that lower allelic variability in L. cuprina populations would less likely give rise to resistance. We predicted 251 druggable genes/proteins using ChEMBL, of which 79 had interacting ligands that satisfy the Lipinski rule-of-threes and ‘rule-of-five’, and are considered ‘MedChem-friendly’ (Supplementary Data 16); one of them (Rpd3) is linked to lethal phenotypes in D. melanogaster (Supplementary Data 15). Conspicuous among the 79 druggable molecules are seven transporters and four ion channels that could represent primary targets for multiple classes of natural or synthetic insecticidal compounds. Other candidates among the 79 druggable proteins include 19 kinases, five peptidases, five growth factor receptors and seven TFs, some of which have been suggested as targets for proteinase inhibitors49, genetically modified baculoviruses50 or Bacillus thuringiensis endotoxins57.

Interestingly, in L. cuprina, we identified an SCO of ladybird late (lbI), a homeobox-containing gene encoding a TF that plays an essential role in regulating developmental processes, such as embryonic neurogenesis, myogenesis and/or cardio genesis in D. melanogaster52. The sequence of lbI is relatively conserved due to its crucial regulatory functions in vertebrates and invertebrates52,53; we propose that Lc-lbI plays a key role in regulating the expression of reporter gene products in the adult female accessory gland of L. cuprina, as reported for Drosophila52. Given that female accessory glands perform essential reproductive functions (for example, fertilization and egg hatching), we believe that Lc-lbI could be critical for successful reproduction, which is consistent with evidence for some other insects, such as Drosophila and Glossina54,55. Gene sequence conservation among (some) insects and evidence of serious phenotypes (for example, reduced larval growth or abortion) on gene perturbation in selected dipterans53,55 indicate that this TF gene should be an important focus for comparative functional genomic explorations of developmental processes in both embryonic and adult female L. cuprina, and might serve as an intervention target in this fly.

Discussion
The present genomic and transcriptomic exploration provides a global insight into the molecular biology of L. cuprina. We have elucidated molecules likely involved in host–fly interactions and immune responses, and studied transcriptional differences between stages and/or sexes of this parasitic fly. Over the years, there has been a major emphasis on the development of various control strategies to combat the blowfly, including mulesing, experimental vaccines, genetic transformation technologies and effective insecticides4. Although the use of insecticides against the blowfly has been successful, resistance in this insect has emerged to almost all currently used compounds.

The present investigation shows, for the first time, the structures of five genes related to resistance. For example, the Lc6 gene is relatively large and complex, as in D. melanogaster,
and spans several scaffolds in the draft genome of *L. cuprina*. The genomic sequences match well with the previously cloned *Lc*6, including all of its alternative exons. Several features of this gene from other species, such as alternative splicing and RNA editing, are also conserved between *L. cuprina* and *D. melanogaster*.

Susceptibility to spinosad was restored in transgenic *D. melanogaster* (Δ6 mutant backgrounds) expressing the *Lc*6 subunit. All individuals of the F1 generation have copies of the driver and the construct. First instar larvae were placed in sets of 50 on culture medium containing 0.1, 0.3 and 0.5 p.p.m. spinosad (0.5 p.p.m. not tested), while the individuals each were tested for each dose. *—Not tested at that dose.

![Figure 3 | Experimental design for the expression of the *L. cuprina* nAChR subunit gene (*Lc*6) in *Drosophila melanogaster* and rescue of spinosad resistance in *D. melanogaster* (Δ6) mutants. (a) Virgin female elav>Gal4;Δ6 flies were crossed with male Δ6>UAS-Lc6 flies. The elav-driver produces Gal4 in neuronal cells, and the Gal4 binds to the UAS site to express the Δ6 subunit that can be assembled into nAChRs. All individuals of the F1 generation have copies of the driver and the construct. First instar larvae were placed in sets of 50 on culture medium containing 0.1, 0.3 and 0.5 p.p.m. spinosad (0.5 p.p.m. not tested), while the individuals each were tested for each dose. *—Not tested at that dose.
through the investigation of essential, fly-specific molecules using functional genomic tools. In particular, various gene-silencing platforms, including double-stranded RNA interference and clustered regularly interspaced short palindromic repeats technology, could provide unique opportunities to systematically investigate essential orthologues as intervention targets in L. cuprina and to explore in-depth the functions of orphan genes/gene products in this fly. Understanding the functions of essential genes, particularly those involved in reproduction, could pave the way to the development of a sterile insect technique for the control of L. cuprina, a proposal supported by the success in eradicking the flesh-eating blowfly Cochliomyia hominivorax (New World screwworm) from the USA, Central America and some other regions of the world.

Clearly, we are now at a point of being able to use the present future. These resources will also support comparative biological questions, and to facilitate the development of improved tools for blowfly prevention and control in the future. These resources will also support comparative investigations of a range of parasitic dipterans.

Methods

Blowfly inbreeding and propagation. A laboratory strain of L. cuprina (designated L. cuprina) was maintained for more than 20 years in the laboratory of P.J.J. using an established culture method, employing bovine liver as a medium for ovipositing and larval rearing. Originally, this strain was isolated from the Australian Capital Territory before the use of organophosphate (OP) insecticides and has since had no exposure to insecticides. For this study, five lines were established and indexed for six generations to reduce genetic variability. In each generation, mating pairs of adult L. cuprina from each line were kept at 28°C and 80% relative humidity in separate cages. Each pair was given water and cubed sugar ad libitum, and provided with bovine liver on days 1, 2 and 4, to mature ovaries and stimulate ovipositing. The largest egg masses from each line were selected, and the resultant larvae reared to adulthood on liver without fly-proof containers, with a minimum larval nutrition and mating procedure was used for producing successive generations, with 8–10 mating pairs (depending on availability) selected from adults emerging from each egg mass (n = 50 pairs) until the fifth or sixth generation.

Genomic sequencing and assembly. L. cuprina is one of the 30 species whose genome has been sequenced as a part of the pilot project to sequence 5000 arthropod genomes at the Baylor College of Medicine Human Genome Sequencing Center. In the i5k programme, an enhanced Illumina ALLPATHS-LG sequencing and assembly strategy has been develop to allow the genomes of multiple species to be sequenced in parallel at substantially reduced cost. For the sequencing of the L. cuprina genome, we isolated high molecular weight genomic DNA from individuals of each of the mixed larval stages and adults (both sexes) using an established protocol. We constructed four genomic DNA libraries of nominal insert sizes of 180 bp, 500 bp, 3 kb and 8 kb at coverages of 83.6, 36.5, 75.1, 31.1-times, respectively (assuming a genome size of 470 Mb). To construct the 180 and 500-bp libraries, we used a gel-excision, paired-end (PE) library protocol. In brief, 1 μg of genomic DNA was sheared using a Covaris S-2 system (Covaris Inc., Woburn, MA) using the 180- or 500-bp protocols. Sheared PE DNA fragments were purified with beads (Agencourt AMPure XP system, Beckman Coulter), end-repaired, da-tailed and ligated to universal adapters (Illumina). Following ligation, DNA fragments were further size-selected on agarose gel and PCR-amplified for six to eight cycles using the primers P1 and Index (illumina). Amplified DNA fragments were subjected to analysis using MAKER2 to provide a consensus set of genes for L. cuprina. Genes inferred to encode peptides of ≥30 amino-acids in length were preserved. To remove extraneous sequences of mammalian, bacterial, mycotic, protistan and/or plant origin(s), scaffolds were broken into contigs at points of indeterminate sequence (Ns). For individual contigs, GC content and average read depth were measured and plotted, then clusters of contigs with high GC content and low read depth were quarantined, following the verification (via BLASTn) of the origin(s) of extraneous sequences. After this filtering step, genes predicted de novo (encoding ≥ 150 a.a.) by Annotation Edit Distance (AED ≥ 1) were preserved, resulting in the final gene set for L. cuprina. Predicted genes were represented by their coding and inferred amino-acid sequences.

Functional annotation of all predicted protein sequences. First, conserved protein domains of individual inferred amino-acid sequences were identified using the programmes InterProScan 5 and InterPro (http://www.ebi.ac.uk/interpro/), employing the default settings. Second, amino-acid sequences were subjected to BLASTp (E-value cutoff ≤ 10⁻⁵) against proteins in the following databases: FlyBase (Drosophila melanogaster, D. mojavensis, D. grimshawi, D. pseudoobscura, D. virilis and D. willistoni; http://flybase.Indiana.edu); VectorBase (Anopheles gambiae and Musca domestica; http://www.vectorbase.org); Ensembl Genomes (Megabusta; http://www.ensembl.org/index.html), UniprotKB/Swiss-Prot (http://www.expasy.org/docs/swiss-prot_guideline.html), KEGG (release 58; http://www.genome.jp/kegg/) and NCBI protein nr (release September 2013; http://www.ncbi.nlm.nih.gov). These protein sequences were then used to construct individual libraries (using known KEGG orthology terms by BLASTp analysis (E-value cutoff ≤ 10⁻⁵)). Homologues were clustered to known protein families using the KEGG BRITE hierarchy employing a custom
GALA driver line of *D. melanogaster* was crossed separately into a background of *dbx* or *dtn* spinosad-resistant alleles (chromosome 2) and made homozygous to create elav > GALA driver lines for expression experiments. The UAS-Dtx6 line has been reported previously. The landing-site strain expressing the FSC1-integrase (FX-86Bf) was provided by the Basler Laboratory, University of Zurich, with the second chromosome pair substituted with chromosomes carrying a resistant allele. The fly line with UAS-Lc63/2 integrated on the third chromosome was created by microinjection into *FX; dtn*; 86Bf; or *FX; dtx*; 86Bf lines. The spinosad bioassay for survival to eclosion was performed on standard culture medium, and experimental data were corrected for control mortality using Abbott’s formula, adapted for the calculation of 95% confidence intervals.

**Addition analyses.** Data analysis was conducted in a Unix environment or Microsoft Excel 2007 using standard statistical tools. Routines required to facilitate data analysis were designed using mainly the Python 2.6 scripting language and are available via http://research.vet.unimelb.edu.au/gasserlab/.

**References.**

1. Githiko, A. K., Lindsay, S. W., Confalonieri, U. E. & Patz, J. A. Climate change and vector-borne diseases: a regional analysis. *Bull. WHO 78*, 1136–1147 (2000).

2. Hall, M. & Wall, R. Myiasis of humans and domestic animals. *Adv. Parasitol.* 35, 357–334 (1993).

3. Fischer, O. A. et al. Blowflies Calliphora vicina and Lucilia sericata as passive vectors of Mycobacterium avium subsp. avium, *M. a. paratuberculosis* and *M. a. hominisuis*. *Med. Vet. Entomol.* 18, 116–122 (2004).

4. Sanderman, R. M. et al. Control of the sheep blowfly in Australia and New Zealand—the future? Int. J. Parasitol. 41, 879–892 (2011).

5. Tellam, R. L. & Bowles, V. M. Control of haematophagous insect vectors of *M. a. paratuberculosis* in sheep: current strategies and future prospects. *Int. J. Parasitol.* 27, 261–273 (1997).

6. Beveridge, W. I. B. The Mules operation—prevention of blowfly strike by surgical measures. *Aust. Vet. J.* 11, 97–104 (1935).

7. Sneddon, J. & Rollin, B. Mulesing and animal ethics. *J. Agric. Environ. Ethics* 23, 371–386 (2010).

8. Elkington, R. A. & Mahony, T. J. A blowfly strike vaccine requires an understanding of host-pathogen interactions. *Vaccine* 25, 5133–5145 (2007).

9. i5k Consortium. The i5k initiative: advancing arthropod genomics for human health, agriculture, and the environment. *J. Hered.* 104, 595–600 (2013).

10. Adams, M. D. et al. The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195 (2000).

11. International Glossina Genome Initiative. Genome sequence of the tsetse fly (*Glossina morismontia*); vector of African trypanosomiasis. *Science* 275, 380–386 (2010).

12. 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).

13. Hall, A. & Lalli, G. Rho and Ras GT-Pases in axon growth, guidance, and branching. *Cold Spring Harb. Perspect. Biol.* 2, a001810 (2018).

14. Sung, P. J. et al. Cystolic Ras supports eye development in *Drosophila*. *Mol. Cell. Biol.* 30, 5649–5657 (2010).

15. Pe, T., Tang, W. & Zhang, X. Involvement of Rabs in the regulation of phagocytosis against virus infection in invertebrates. *J. Proteome Res.* 11, 4834–4846 (2012).

16. Bloomquist, J. R. Ion channels as targets for insecticides. *Annu. Rev. Entomol.* 41, 163–190 (1996).

17. Dermaw, W. & Van Leeuwen, T. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. *Insect Biochem. Mol. Biol.* 45, 89–110 (2014).

18. Dederer, H., Berger, M., Meyer, T., Werr, M. & Ilg, T. Structure-activity relationship of the Large Gap assembly algorithm. PE read data from the 500 bp genomic library were used when multiple scaffolds constituted a gene (for example, scaffold nos. 379, 4233 and 792 for *Lc63*). Intronic regions were confirmed using transcrimptomic (RNA-seq read) data, and intron–exon junctions were confirmed by manual inspection. If required, a reference-guided BWA-MEM alignment (http://www.genecodes.com) was performed to verify the presence or absence of exons in scaffolds and the draft genome assembly.

**Cloning of *Lc63***

The full-length coding region of the *Lc63* gene was PCR-amplified using oligonucleotide primers LuciferCycL2 (5′-CGTCTTATTTCCTG CATT-3′) and LuciferCycR (5′-TATCGCCACTTGGAGATG-3′) with a high-fidelity Taq polymerase (Expand High Fidelity; Roche) from cDNA (SuperScript III, Invitrogen), synthesized from RNA isolated (TRIZol) from *Lc6* adult heads. The product was cloned into the p-GEM-T-Easy vector (Promega), sequenced (Macrogen) and then shuttled into the Norl site of plasmid pUASTattB (Promega) to produce the construct designated UAS-Lc63.

**Heterologous expression of *Lc63* in *D. melanogaster***

Flies homozygous for *dtx* or *dtn* are 61-fold and at least 1,176-fold more resistant to spinosad compared with the spinosad-susceptible parental line *Armenia* 2, an isofemale line derived from the *Drosophila* Genetic Resource Centre stock no. 103394 (ref. 44). To allow expression in the *dtxw* or *dtnw* spinosad-resistant background, the P[w + mW] = GwB]dax145 (Bloomington *Drosophila* Stock Centre; BL458)
24. Niderman, T. et al. Pathogenesis-related PR-1 proteins are antifungal. Plant Physiol. 108, 17–27 (1995).

25. Cancedda, C. M. A portrait of the “SCP/TAPS” proteins of eukaryotes – developing a framework for fundamental research and biotechnological outcomes. Biotechnol. Adv. 27, 376–388 (2009).

26. Kovalick, G. E. & Griffin, D. L. Characterization of the SCP/TAPS gene family in Drosophila melanogaster. Insect Biochem. Mol. Biol. 35, 825–835 (2005).

27. Lambert, J. et al. Insect immunity: isolation from immune blood of the dipteran Phormia terraenovae of two insect antibacterial peptides with sequence homology to rabbit lung macrophage bactericidal peptides. Proc. Natl Acad. Sci. USA 86, 262–266 (1989).

28. Hultmark, D. et al. Insect immunity. Attacins, a family of antibacterial proteins from Hyalophora cecropia. EMBO J. 2, 571–576 (1983).

29. Mackerras, M. J. Observations on the life-histories, nutritional requirements and fecundity of blowflies. Bull. Entomol. Res. 24, 353–362 (1933).

30. Hens, K., Macours, N., Gayes, L., Francis, C. & Huybrechts, R. Cloning and expression of the yolk protein of the tsetse fly Glossina morsitans morsitans. Insect Biochem. Mol. Biol. 34, 1281–1287 (2004).

31. Markow, T. A., Beall, S. & Matzkin, L. M. Egg size, embryonic development time and ovooviparity in Drosophila species. J. Evol. Biol. 22, 430–434 (2009).

32. Charles, J. P., Chihara, C., Nejad, S. & Riddiford, L. M. A cluster of cuticle protein genes of Drosophila melanogaster at 65A: sequence, structure and evolution. Genetics 147, 1213–1224 (1997).

33. Murata, T., Kageyama, Y., Hirose, S. & Ueda, H. Regulation of the EDG84A gene by FTZ-F1 during metamorphosis in Drosophila melanogaster. Dev. Growth Differ. 44, 419–425 (2002).

34. Elvin, C. M. et al. Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of Lucilia cuprina. J. Biol. Chem. 271, 8925–8935 (1996).

35. Wissler, L., Gadau, J., Simola, D. F., Helmkampf, M. & Bornberg-Bauer, E. The homeodomain protein ladybird late regulates a conserved regulator of appendicular myogenesis. Development 140, 1519–1533 (2013).

36. Heath, A. C. G., Broadwell, A. H., Chilcott, C. N., Wigley, P. J. & Shoemaker, C. B. Efficacy of native and recombinant Crys1B protein against experimentally induced and naturally acquired ovine myiasis (fly strike) in sheep. J. Econ. Entomol. 97, 1797–1804 (2004).

37. Bischof, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. 32, 347–355 (2014).

38. Attardo, G. et al. The homeodomain protein ladybird late regulates synthesis of milk proteins during pregnancy in the tsetse fly (Glossina morsitans). PLoS Negl. Trop. Dis. 8, e2645 (2014).

39. Grauso, M., Reeman, R. A., Caleito, E. & Sattelle, D. B. Novel putative nicotinic acetylcholine receptor subunit genes, Dalph5a, Dalph6a and Dalph7a, in Drosophila melanogaster identify a new and highly conserved target of adenosine deaminase acting on RNA-mediated A-to-I pre-mRNA editing. Genetics 160, 1519–1533 (2002).

40. Barnard, A.-C., Nijhof, A. M., Fick, W., Stutzer, C. & Maritz-Olivier, C. RNAi in arthropods: insight into the machinery and applications for understanding the pathogen-vector interface. Genes 3, 702–741 (2012).

41. Sander, J. D. & Young, J. R. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. 32, 347–355 (2014).

42. Baumhower, A. et al. Screwworm control through release of sterilized flies. J. Econ. Entomol. 48, 462–466 (1955).

43. Knipping, E. F. Possibilities of insect control or eradication through the use of sexually sterile males. J. Econ. Entomol. 48, 459–462 (1955).

44. Krafur, E. S., Whitten, C. J. & Novy, J. E. Screwworm eradication in North and Central America. Parasitol. Today 3, 131–137 (1987).

45. Kostas, A. C., Sales, N. & Barchia, I. M. Dehydrocinnamoyl tolerance associated with monoxygenase activity in field strain larvae of the Australian sheep blowfly (Diptera: Calliphoridae). J. Econ. Entomol. 90, 15–20 (1997).

46. Greenberg, B. & George, J. in Handbook of Insect Rearing. (eds Singh, P. & Moore, R. F.) Elsevier, 1985).

47. Sambrook, J. & Russell, D. W. Molecular Cloning: A Laboratory Manual 3rd edn. (Cold Spring Harbor Laboratory Press, 2001).

48. Morris, R. F.) (Elsevier, 1985).

49. Mortazavi, A. et al. Scaffolding a Caenorhabditis nematode genome with RNA-seq. Genome Res. 20, 1740–1747 (2010).

50. Elbeck, K., Moore, B., Holt, C. & Yandell, M. Quantitative measures for the management and comparison of annotated genomes. BMC Bioinformatics 10, 67 (2009).

51. Lin, D. M. & Goodman, C. S. Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance. Neuron 13, 507–523 (1994).

52. Perry, T. et al. Effects of mutations in Drosophila nicotinic acetylcholine receptor subunits on sensitivity to insecticides targeting nicotinic acetylcholine receptor antagonists. Pesticide Biochem. Physiol. 102, 56–60 (2012).

53. Rosenheim, J. A. & Hoyer, A. Confidence intervals for the Abbott’s formula correction of bioassy data for control response. J. Econ. Entomol. 82, 331–335 (1989).

**Acknowledgements**

Funding from Australian Wool Innovation (to P.B. and R.B.G.) and the Australian Research Council (to R.B.G.) is gratefully acknowledged. Support to R.B.G.’s laboratory by YourGene Bioscience and Melbourne Water Corporation is gratefully acknowledged. We thank the staff at the Baylor College of Medicine Human Genome Sequencing Center for their contributions. Genome sequencing and assembly was funded by a grant U54 HG003273 from the National Human Genome Research Institute (NHGRI to R.A.G.). This project was also supported by a Victorian Life Sciences Computation Initiative (VLCI: grant number VR0007) on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government (R.B.G.). C.A.A. holds an INSERC Postdoctoral Research Fellowship. N.D.Y. holds an NHMRC Early Career Research Fellowship. P.K.K. is the recipient of a scholarship (STRAP) from the University of Melbourne. R.M.W. holds a Marie Curie International Outgoing Fellowship PIOF-GA-2011-303312. R.M.W. holds a Marie Curie International Outgoing Fellowship PIOF-GA-2011-303312. P.I. holds a Swiss National Science Foundation (SNF) Award 31003A-145396 (to E.M.Z.) We would specifically like to acknowledge all scientists who developed programmes used in this study. Given the restriction on the number of papers that could be cited in this article, we were not able to include all original publications. Instead, we have included links to respective websites. We would also like to thank Q. Lang for the use of his L. cuprina image.

**Author contributions**
P.J.J. provided L. cuprina and A.C.K. and N.H.B. purified nucleic acids for sequencing. S.R. and P.B. coordinated the sequencing. S.R. and S.C.M. undertook the assembly, with
inputs from S.D., S.L.L., H.C., H.D., Y.H., H.V.D., K.C.W and D.M.M. and C.A.A., P.K.K., N.D.Y. and R.S.H. conducted the annotations and analyses of genomic and transcriptomic data. C.A.A., P.K.K., N.D.Y., R.S.H., S.F.L., T.P., A.R.J., A.J.S., B.R.E.A., B.B. and A.H, planned or undertook additional, detailed bioinformatic analyses. C.A.A., R.B.G., S.R., S.F.L., T.P., P.I., R.M.W., E.M.Z. and P.B. drafted and edited the manuscript, tables, figures and supplementary information, with inputs from other authors. P.B. conceived and planned the project, and P.B., S.R., R.A.G. and R.B.G. supervised and coordinated the research.

Additional information
Accession codes: This whole genome shotgun project is available from the DDBJ/EMBL/GenBank databases under the accession JRES00000000 (version JRES01000000). Raw sequences have been deposited in the NCBI short read archive (SRA) under accession code SRX579209. The *L. cuprina* genome sequence is available from NCBI under BioProject accession codes PRJNA248412 and PRJNA203545, as well as BioSample accession codes SAMN02794242, SAMN02947403 to SAMN02947405 and SAMN2422564.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Anstead, C. A. et al. *Lucilia cuprina* genome unlocks parasitic fly biology to underpin future interventions. *Nat. Commun.* 6:7344 doi: 10.1038/ncomms8344 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/