

# SCIENTIFIC REPORTS



OPEN

## The microbiomes of blowflies and houseflies as bacterial transmission reservoirs

Ana Carolina M. Junqueira<sup>1,10</sup>, Aakrosh Ratan<sup>2</sup>, Enzo Acerbi<sup>1</sup>, Daniela I. Drautz-Moses<sup>1</sup>, Balakrishnan N. V. Premkrishnan<sup>1</sup>, Paul I. Costea<sup>3</sup>, Bodo Linz<sup>4</sup>, Rikky W. Purbojati<sup>1</sup>, Daniel F. Paulo<sup>5</sup>, Nicolas E. Gaultier<sup>1</sup>, Poorani Subramanian<sup>6</sup>, Nur A. Hasan<sup>6,7</sup>, Rita R. Colwell<sup>6,7,8</sup>, Peer Bork<sup>3</sup>, Ana Maria L. Azeredo-Espin<sup>5</sup>, Donald A. Bryant<sup>1,9</sup> & Stephan C. Schuster<sup>1</sup>

Blowflies and houseflies are mechanical vectors inhabiting synanthropic environments around the world. They feed and breed in fecal and decaying organic matter, but the microbiome they harbour and transport is largely uncharacterized. We sampled 116 individual houseflies and blowflies from varying habitats on three continents and subjected them to high-coverage, whole-genome shotgun sequencing. This allowed for genomic and metagenomic analyses of the host-associated microbiome at the species level. Both fly host species segregate based on principal coordinate analysis of their microbial communities, but they also show an overlapping core microbiome. Legs and wings displayed the largest microbial diversity and were shown to be an important route for microbial dispersion. The environmental sequencing approach presented here detected a stochastic distribution of human pathogens, such as *Helicobacter pylori*, thereby demonstrating the potential of flies as proxies for environmental and public health surveillance.

Interactions between hosts and microorganisms are increasingly recognized as a ubiquitous principle in nature<sup>1</sup>, capable of modulating animal physiology, fitness, and host social behaviour<sup>2</sup>. Most of the microbiome research has been performed in humans, but massively parallel sequencing and recent advances in computational analyses have made it possible to extend the same level of taxonomic resolution to hosts from other phyla. More recently, mammals<sup>3</sup>, plants<sup>4</sup>, and insects<sup>5</sup> were also targeted by metagenomics. In insects, the microbiomes of termites, bees, ants, flies, mosquitoes, and the triatome bug<sup>6</sup> have recently been analysed. However, many of these studies have been restricted to 16S sequencing<sup>7</sup>, which generally limits the identification of microbial taxa by the experimental shortcomings and biases of PCR-based and/or cultivation-based methods, yet providing a record of the phylum- and genus-level diversity.

Despite their medical, forensic, and sanitary importance, an assessment of the microbiome of mechanical vectors, such as blowflies (Diptera: Calliphoridae) and houseflies (Diptera: Muscidae), has not been undertaken by large-scale metagenomic approaches. Although transmission of microbiota between hosts has been reported between humans and pets in a household<sup>8</sup>, potential interchange of the microbiota of synanthropic insects and their surroundings is under-characterized.

<sup>1</sup>Present address: Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, 637551, Singapore. <sup>2</sup>Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia, Charlottesville, VA, 22908, USA. <sup>3</sup>European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg, 69117, Germany. <sup>4</sup>Center for Vaccines and Immunology, College of Veterinary Medicine, University of Georgia, Athens, 30602, GA, USA. <sup>5</sup>Centro de Biologia Molecular e Engenharia Genética, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, 13083-875, Brazil. <sup>6</sup>CosmosID Inc, Rockville, MD, 20850, USA. <sup>7</sup>Center for Bioinformatics and Computational Biology, University of Maryland. Institute for Computational Biology, University of Maryland College Park, College Park, MD, 20742, USA. <sup>8</sup>Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA. <sup>9</sup>Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, 16802, USA. <sup>10</sup>Present address: Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941-902, Brazil. Correspondence and requests for materials should be addressed to A.C.M.J. (email: [anacmj@gmail.com](mailto:anacmj@gmail.com)) or S.C.S. (email: [scschuster@ntu.edu.sg](mailto:scschuster@ntu.edu.sg))

Blowflies and houseflies are the first organisms to arrive on carcasses, decaying organic matter and faeces on which they feed, breed, and lay eggs<sup>9</sup>. It is likely that they acquire a significant part of their microbiome from such environments, which they transport and subsequently deposit onto other hosts, such as humans, animals, and plants. Because they are ubiquitous and synanthropic, these flies play an important role in the transport and dispersal of microbes in urban and natural environments. It has been suggested that flies serve as a vector for a number of pathogens<sup>10,11</sup>, but they have rarely been implicated in any specific medical condition. This absence of cause-effect linkage has prevented more targeted studies of mechanical vectors, compared to biological vectors, for which host-microbe associations have been intensively analysed<sup>12</sup>.

In this study, we have investigated 116 microbiomes by whole-genome shotgun (WGS) sequencing of 63 samples of blowflies of the species *Chrysomya megacephala* and 53 individual houseflies of the species *Musca domestica*, describing the complexity of the pan-microbiomes of both mechanical vectors at the species level. Flies were collected on three continents in urban, rural, and natural settings. Despite inhabiting similar ecological niches, they showed different compositions and abundances for the organisms in their associated microbial communities. Host-specific microbial taxa were observed, but blowflies and houseflies share more than 50% of their microbiome that is likely acquired from similar habitats where they feed and breed. To depict the microbiomes of both fly species, we analysed the entire dataset using three approaches with different degrees of stringency, which we combined for high confidence of taxa assignment. This allowed us to reach a species-level identification of the microbial communities in these mechanical vectors and thus provide a better understanding of the roles of these flies as potential agents of pathogen transmission.

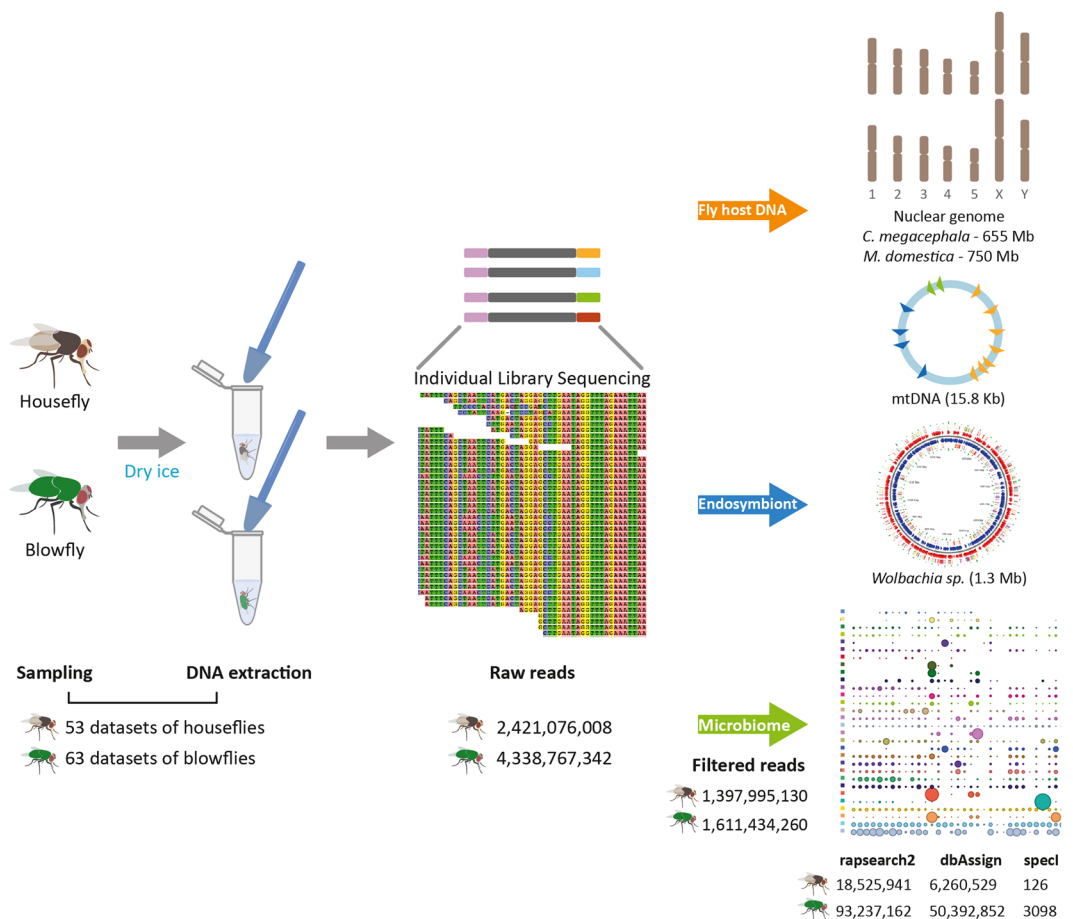
## Results and Discussion

**Sampling and environmental sequencing approach.** Blowflies and houseflies were caught in different environments on three continents using a protocol that prevented sample contamination while preserving nucleic acids found in or on the insect body. Cross-contamination was avoided by using dry-ice, which immediately suffocated the animals and ensured instant preservation of the entire specimen at low temperature at the sampling sites under sterile conditions (see Material and Methods). Using the entire body of a fly, a novel metagenomic approach was used to sequence DNA molecules of the insect and its microbiome together. Therefore, individual flies may serve as environmental proxies that provide information on the locations they have recently visited. We believe this versatile approach could be applicable to environmental and public health surveillance as well as forensics. In contrast to previous studies that focused only on cultivatable microorganisms, insect midguts or 16S rRNA<sup>7,13,14</sup>, our metagenomics strategy resulted in coverage of the host and microbiome genomes. Specifically, the genomes of *M. domestica* and *C. megacephala* were sequenced to a depth of 3.2-fold and 6.6-fold respectively, the host mitochondrial DNA (mtDNA) was sequenced to a depth of 7000-fold<sup>15</sup>, and the *Wolbachia* spp. endosymbiont genome was covered to a depth of >2000-fold. The remaining >93 million reads were effectively assigned to the microbiomes of the respective hosts (Fig. 1).

**Microbial assignment of the metagenomic datasets.** We generated a total of 116 separate metagenomic datasets (blowflies = 62; houseflies = 53; lab-reared pooled control = 1) from 3 different continents. The blowfly datasets contained approximately 70 million reads per sample (control excluded) and the housefly datasets had approximately 45 million reads per sample (Table S1 for an average of reads per sample). A total of 6,759,843,350 reads were generated. After the *in silico* removal of the fly genomic sequences using Bowtie2<sup>16</sup>, the remaining 3,009,429,390 reads (44%; Fig. 1 and Table S1) were used for downstream metagenomics analyses with three different bioinformatics methods: (1) rapsearch2, (2) dbAssign and (3) specI (Table S1 for summary, Table S2 for extended information). When collapsed into super kingdom taxonomy (Fig. 2A), these large-scale datasets showed minimal traces of Archaea. Most of the reads assigned to Eukaryotes belong to the order Diptera, indicative of the incompleteness of the reference genome for these species (Figure S1). Sequences assigned to the domain Bacteria are the most prevalent in all datasets, except in the housefly sample AJ155 (identified with an asterisk on Fig. 2A), in which viral DNA was highly abundant. An in-depth analysis of this sample revealed the presence of the *M. domestica* Salivary Gland Hypertrophy Virus (MdSGHV). The alignment of viral reads against the MdSGHV reference genome<sup>17</sup> (NC\_01067) gave a mean coverage of 12,596-fold (detailed in Fig. 2A). MdSGHV is a double-stranded DNA virus that is orally transmitted to houseflies and causes the inhibition of ovarian development, thus leading to a shutdown of egg production in infected females. Flies also show hypertrophy of the salivary gland as a symptom<sup>18</sup>. The other viruses observed in these datasets were mainly bacteriophages (Figure S2).

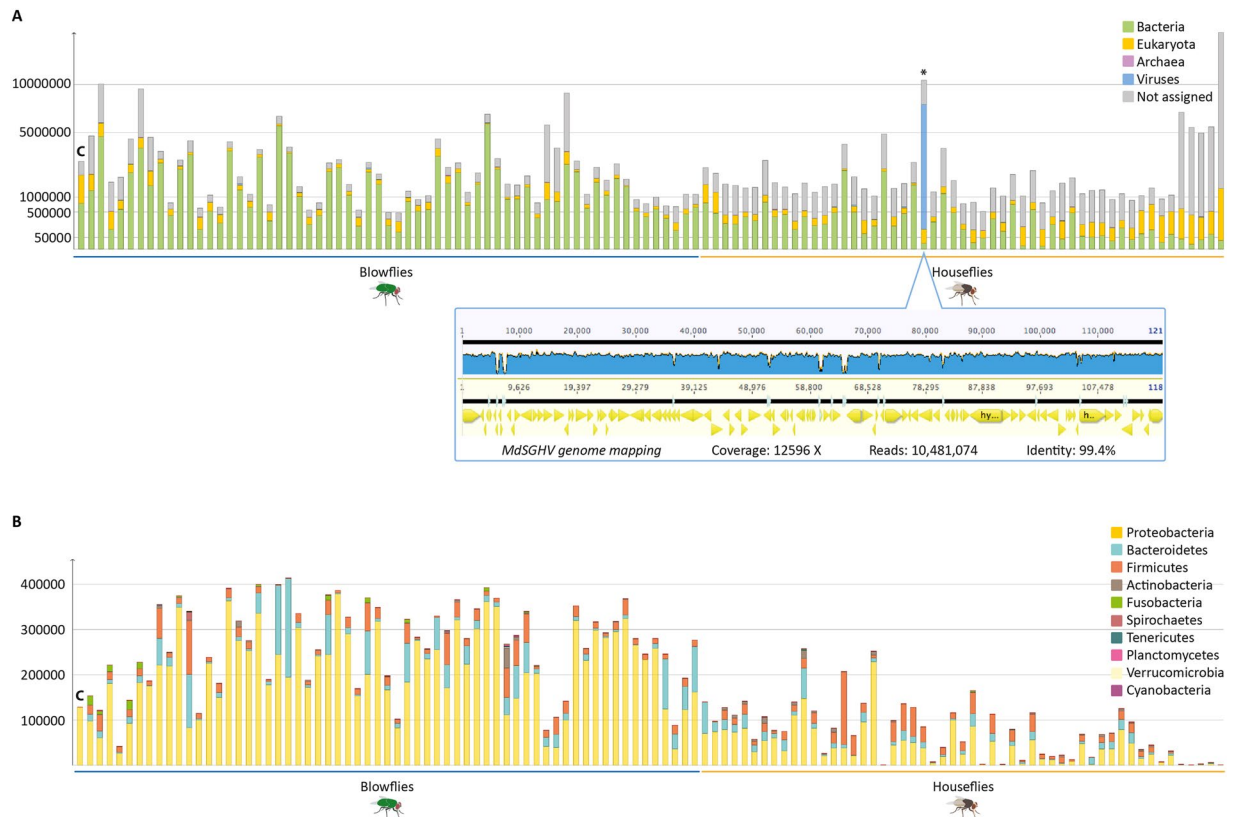
Taxa assignments were performed with normalized datasets (see Methods), which showed that members of the phyla *Proteobacteria*, *Bacteroidetes* and *Firmicutes* are the most abundant organisms in the microbiomes of both blowflies and houseflies (Fig. 2B and Figure S3). This result corroborates previous findings for the green bottle fly<sup>7</sup>, houseflies<sup>19</sup>, bees, cockroaches, fruit flies and mosquitoes<sup>20</sup>, except for the low representation of *Actinobacteria* in our datasets. This difference is likely due to that fact that insect studies use cultivation and amplification of the 16S rRNA, while this study was undertaken with amplification-free metagenomics based on WGS. The preponderance of *Proteobacteria* in blowflies is mainly associated with the presence of the endosymbiont *Wolbachia* (Alphaproteobacteria), while in houseflies it is related to the dominance of *Psychrobacter* (Gammaproteobacteria). However, other members of the *Proteobacteria* are also major components of the blowfly and housefly microbiomes, including the genera *Enterobacter*, *Escherichia*, *Klebsiella*, *Proteus*, *Morganella*, *Hafnia*, *Pseudomonas*, *Aeromonas*, *Acinetobacter*, *Providencia* and *Serratia* (Fig. 3). Some of these bacterial genera have previously been identified by cultivation or amplification assays in different species of flies<sup>7,21</sup>.

The DNA-protein (rapsearch2) and DNA-DNA (dbAssign and specI) alignment approaches allowed species-level identification of the bacteria in the microbiome (Fig. 3 and Tables S3, S4 and S5 for the normalized number of reads assigned by dbAssign, rapsearch2 and the specI, respectively). All three methods showed that the prevalent microorganisms associated with the blowfly *C. megacephala* are similar to the endosymbionts



**Figure 1.** Summary of sampling datasets, data generation and analyses. Blowflies ( $n = 62$ ; 1 control) and houseflies ( $n = 53$ ) were collected in individual vials and immediately placed on dry ice until DNA extraction. Samples were individually sequenced in a multiplexed run, generating a total of 6,759,843,350 reads for both fly species. The blowfly draft genome generated in this study and the housefly reference genome (RefSeq number GCF\_000371365.1) were used as filters to remove host-related reads. Final metagenomic dataset included a total of 3,009,429,390 reads for 116 flies. See also Tables S1 for a summary of reads generated and assigned to blowflies and houseflies, and Table S2 for the detailed information of each individual sample. Reads were processed with three different bioinformatics methods and assigned to bacterial taxa using the rapsearch2 algorithm against the NR database (April 2015 version), the dbAssign in-house script (<https://github.com/aakrosh/dbAssign>) against a database with 5,614 complete and chromosome-level assembled microbial genomes (April 2016 version) and a BWA approach against specI clusters (Tables S3, S4 and S5 for detailed information).

*Wolbachia pipientis* of *Culex quinquefasciatus* and *Wolbachia* sp. *wRI* (isolated from *Drosophila simulans*). Both *Wolbachia* strains were collapsed into *Wolbachia* spp. in Fig. 3 and represent 60.9%, 62.5% and 87.1% of the total bacterial assigned reads analysed with rapsearch2, dbAssign and the specI pipelines, respectively. The broad distribution of *Wolbachia* spp. in all *C. megacephala* samples analysed confers a higher read count of bacteria in blowflies than houseflies (Fig. 2A and B). The genus *Wolbachia* is one of the most pandemic bacterial genera, described in more than 60% of insect species worldwide<sup>22</sup> and its presence in the blowfly *C. megacephala* has previously been reported<sup>23</sup>. However, unlike former studies, our analysis not only indicates the presence of the endosymbiont, but simultaneously allowed the quantitative assessment of the relative distribution of *Wolbachia* across the samples. A total of 31,883,141 reads were assigned to the *Wolbachia* endosymbiont in blowfly samples, which provided >2000-fold coverage of its genome. The pairwise identity of the *Wolbachia* surface protein gene (*wsp*), commonly used for typing *Wolbachia* strains<sup>24</sup>, shows that *C. megacephala* and the butterfly *Eurema hecabe* (AB094396.1) have strains with 100% identity. However, the genome-wide mapping of assigned reads indicates similarity to the *Wolbachia* sp. from the mosquito *Culex quinquefasciatus* (NC\_010981.1). Based on these lines of evidence, we hypothesize that *C. megacephala* hosts a specific strain of *Wolbachia* classified in the supergroup B branch. This intracellular symbiont plays important roles in manipulating invertebrate reproductive biology by killing males, feminization, parthenogenesis or cytoplasmic incompatibility, and it has been extensively studied as a biological control agent for arthropods<sup>22</sup>. Evidence of horizontal transmission of *Wolbachia* within and between species has been reported<sup>22</sup> and, more recently, it was also shown that some *Wolbachia* spp. strains can influence the resistance of different hosts to viruses<sup>25</sup>.

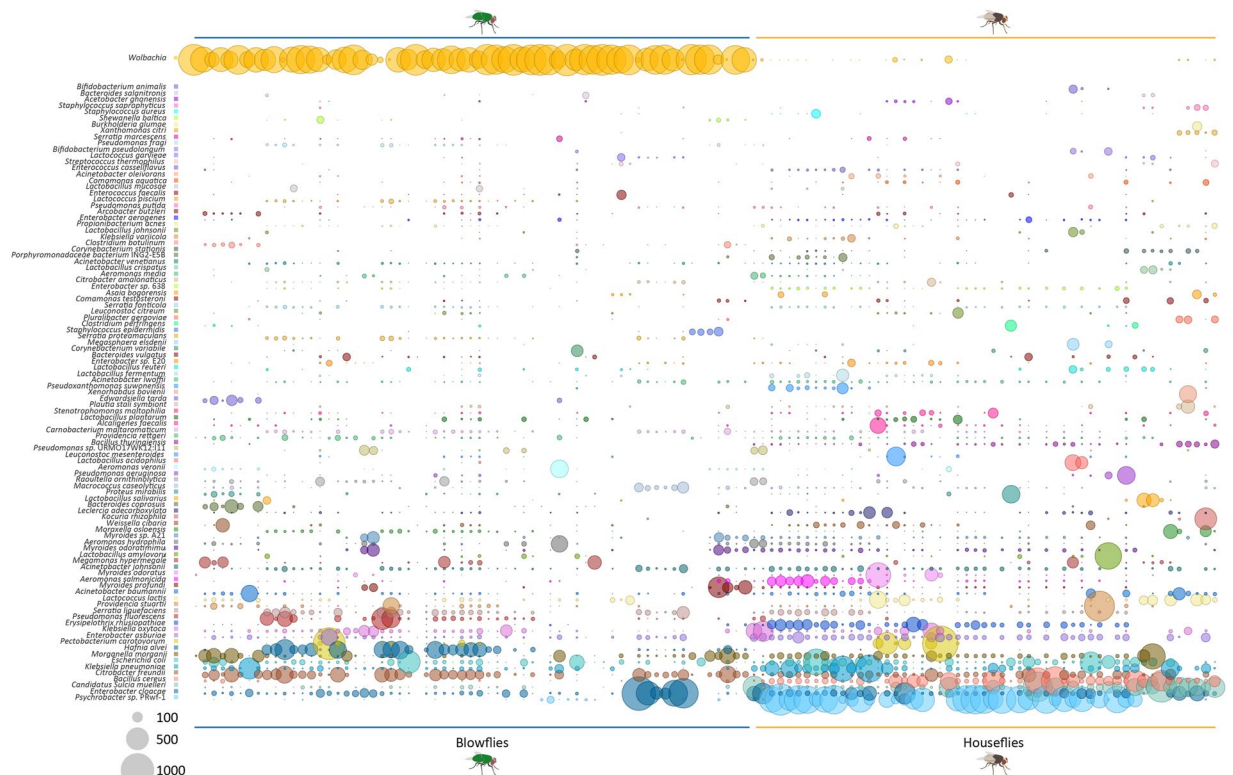


**Figure 2.** Higher rank taxonomy of the microbiome of blowflies and houseflies. **(A)** Super kingdom classification of the metagenomic reads, indicating bacteria are the main component of the microbiome of fly mechanical vectors. Reads assigned to Eukaryota are mostly assigned to insects (Diptera, in particular. See Supplementary Figure S1 for detailed analysis of the eukaryote reads). The sample marked with an \* shows a high virus load that was identified as the MdsGHV DNA virus that infects houseflies. The genome mapping of viral reads against the MdsGHV reference genome showed that the metagenomic dataset was spread across the viral genome with >12,000-fold coverage on average. **(B)** Bacterial counterpart of the metagenomic reads at phylum-level taxonomic rank. *Proteobacteria* dominates the microbiome of blowflies and houseflies, followed by *Bacteroidetes* and *Firmicutes*. Most of the proteobacterial reads are assigned to the endosymbiont *Wolbachia* sp. in blowflies and to *Psychrobacter* sp. PRwf-1 in houseflies. Detailed results also show the presence of the *Wolbachia* endosymbiont in housefly samples collected in three different countries. Sample marked with “C” indicates the lab-reared pool sample serving as a control.

In houseflies, there are sparse reports of *Wolbachia* spp. infection<sup>23,26</sup> and this is the first unbiased report of the bacteria in houseflies, that is, without specific cultivation or PCR amplification (Fig. 3). A total of 4,016 reads were assigned to the *Wolbachia* endosymbiont in 29 housefly samples. The reads covered 23.8% (353,027 bp) of the *Wolbachia* sp. reference genome (NC\_010981) with a pairwise identity of 99.2%. However, the strain could not be identified due to low coverage of both the genome and the *wsp* gene.

After subtracting reads that map to fly or mitochondrial DNA, we analysed the remaining microbial read abundances. The most prevalent bacterium detected in the microbiome of houseflies is the *Psychrobacter* sp. PRwf-1, present in 73.6% of the individuals analysed (Fig. 3) and represents 9.5%, 2.4% and 1.5% of all reads assigned by dbAssign, rapsearch2 and specl, respectively. Members of the genus *Psychrobacter* are extremophilic bacteria associated with cold environments and permafrost soils with remarkable capabilities to adapt to subzero temperatures. However, the strain PRwf-1 that accounts for 25% of the housefly microbiome (Fig. 3), shows physiological adaptation to survival in warmer temperatures and has been previously associated with food spoilage<sup>27</sup>.

Figure 3 compiles the top 100 bacterial species (vertical axis) assigned to each individual fly (horizontal axis) using dbAssign, representing 97.3% of all reads assigned. The normalized datasets of 116 flies resulted in the identification of 431 bacterial taxa at the species-level (Table S3). The species distribution patterns seen in Fig. 3 suggest that blowflies and houseflies share parts of their microbiomes, but also have host-specific microbial occurrences. The alpha-diversity calculated for rapsearch2 and dbAssign (Figure S4) shows that these values depend on the reference database composition and the stringency used for and after the alignment of short-read sequences to these databases. Rapsearch2, based on the translated nucleotide sequence similarity search against the NR protein database, shows a slightly higher microbial diversity of blowflies when compared to houseflies (median of 5.18 and 5.04, respectively), while the dbAssign analysis, based on the assignment of reads to reference genomes in a database, showed the opposite (blowflies = 1.69; houseflies = 3.15). The OTU (operational taxonomic unit)

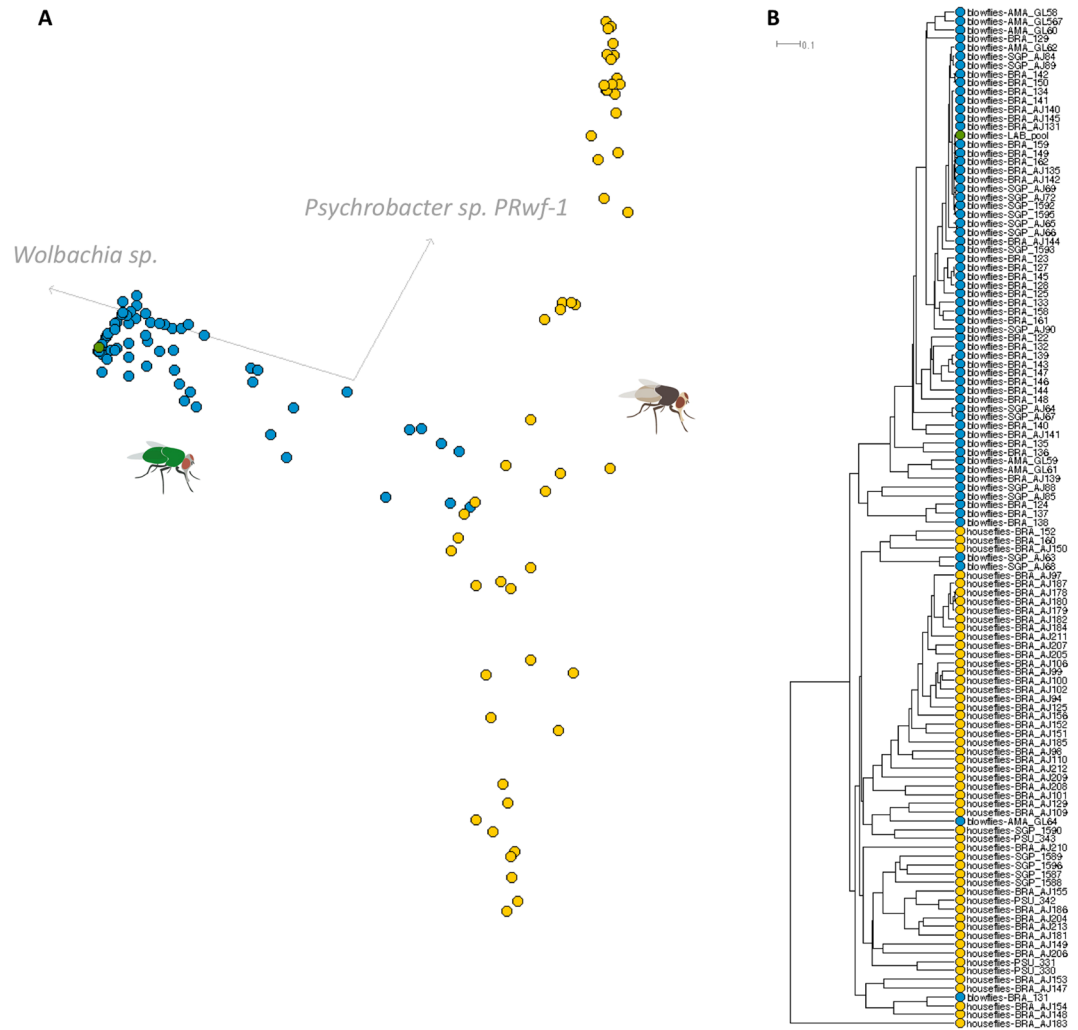


**Figure 3.** Microbiome of blowflies and houseflies. The bubblechart shows the top 100 bacterial species assigned to 116 fly sample analysed by the dbAssign tool kit. The size of the bubbles is square-root-scaled (scale in grey) and refers to the normalized number of reads assigned to each bacterial reference genome (listed in the Y axis) in each of the samples (listed in the X axis). See Figure S4 for alpha-diversity indices.

richness analysis showed an average number of  $551 \pm 373$  bacterial species for blowflies and  $228 \pm 157$  bacterial species for houseflies using rapsearch2 (Figure S5), and  $37 \pm 22$  OTUs for blowflies and  $42 \pm 20$  OTUs for houseflies using dbAssign (Figure S5). Species richness and sequence depth can also be evaluated with rarefaction curves (Figure S5), which showed that sequence depth plateaus at 25,000 reads for rapsearch2 and around 500 reads for dbAssign. These findings underline the importance of database size and diversity for metagenomic analyses and affected both approaches used in this study (nr database = 131,807,364 records; full microbial genomes database = 5,614 records).

**Newly emerged lab-reared flies as an environment control.** The control sample comprised of a pool of 98 newly emerged adults of blowflies (*C. megacephala*) reared in cages under controlled conditions. The pool of adults showed a low alpha-diversity (3.0 for rapsearch2 and 0.58 for dbAssign) and is mainly composed of *Wolbachia* sp., with 99.5% of the microbiome reads assigned to the endosymbiont. Only four additional taxa are present in the control: *Morganella morgani*, *Myroides odoratus*, *Providencia rettgeri*, and *Providencia stuartii*, all with very low read counts. The reduced microbiome is likely due to the fact that holometabolous insects (i.e. those undergoing complete metamorphosis) expel their degenerated larval and pupal guts soon after eclosion<sup>28</sup>. They also replace the cuticle that will develop into an exoskeleton<sup>29</sup>. These developmental changes may eliminate not only the gut bacteria, but also those attached to the outer body surface, thus substantially reducing the endogenous fly microbiome diversity and complexity. The gut microbiomes of newly emerged adults were also of similarly low complexity for different mosquito species and worker bees of *Apis mellifera*<sup>30,31</sup>.

**Shared and unique microbiomes.** Using dbAssign, 316 bacterial taxa were identified in blowflies, compared to 351 in houseflies (Table S6). A comparison of the summarized microbiomes of houseflies and blowflies showed that more than 55% (237 taxa) of the microbial species are shared between the two hosts (Figure S6 and Table S6). The large overlap between microbiomes is likely related to the similar environments that both species frequent, from which they acquire a major part of the microorganisms they transport. The analysis of the shared microbiome showed that seven taxa are present in at least 80% of all flies analysed (Figure S7), including *E. coli* and *Enterobacter cloacae* that could be detected at the species level. Both species are gram-negative, facultatively anaerobic bacteria that are found as part of the normal gut flora of humans and animals. However, strains of *E. cloacae* can play an important role in nosocomial infections of the urinary and respiratory tract<sup>32</sup>, while *E. coli* serotypes are typically responsible for food poisoning and gastrointestinal infections<sup>33</sup>. All other microbial taxa of the core microbiome could be assigned only to genus level or a higher taxonomic rank. This includes the genera *Klebsiella* and *Acinetobacter* and additional organisms classified at the class, order or family level (Figure S7).



**Figure 4.** Segregation of fly host species based on their microbiome. **(A)** PCoA using Bray-Curtis index (stress = 0.12) showing that *C. megacephala* (blue dots) and *M. domestica* (yellow dots) individuals can be separated based on their full microbiome complexity, despite the fact that they share >50% of their microbiome (see also Figure S6, Figure S7, and Table S6). A few blowflies clustered with housefly individuals and deeper analyses showed an abnormal low amount of *Wolbachia* sp. endosymbiont in these blowflies. **(B)** UPGMA dendrogram visualization of the clustering analysis, showing the samples of houseflies and blowflies that clustered together. The detailed microbiome profiles of samples GL64, BRA\_131, AJ63 and AJ68 are shown in Figure S8.

Despite the fact that blowflies and houseflies share more than half of their microbiomes, host-species specific bacterial taxa were also detected with a cutoff of a minimum of 500 reads per species. A total of 114 microbial species were uniquely found in houseflies and 79 in blowflies. The host-specific microbiome is most likely driven by the fly species, but may also occur as a result of stochastic events. As shown in Fig. 3, random occurrence of microbial taxa can be observed in one or a few individuals of flies at significant numbers. In fact, principal coordinate analysis (PCoA) show segregation of the two fly species, with minimal overlap, based on their microbiomes (Fig. 4). The few blowfly individuals with microbiomes that overlap with those of houseflies display a remarkably low abundance of *Wolbachia* spp., as well as higher read counts that were assigned to *Enterobacter cloacae*, *Klebsiella pneumoniae* or *Pectobacterium carotovorum* (Figure S8).

To test the effect of geographic location on the observed variation of the microbiome, analysis of variance (ANOVA) and permutational multivariate analysis of variance (PERMANOVA) were performed. ANOVA indicates no significant discrimination of the microbiome of flies collected in different countries ( $p > 0.05$ ). PERMANOVA suggests that the geographic origin has a minor effect on the beta-diversity of the microbiome ( $R^2 = 0.057$ ,  $p$ -value = 0.002 for rapsearch2;  $R^2 = 0.074$ ,  $p$ -value = 0.001 for dbAssign), thus explaining a small proportion of the variation observed. The PCoA plots in Figure S9 show that samples from different continents do not segregate and that most of the microbiome variation is likely found between individuals and not between populations.

**Intersection of three metagenomic analyses.** The complete assignment of reads to microbial taxa, using the three bioinformatic methods employed in this study, is shown in Tables S3, S4 and S5. The number of OTUs assigned to species level by each method is summarized in the Venn diagram in Figure S10 (cutoff of >500 reads assigned per bacterial species) and intersection of OTUs detected by all three methods is given in Table S7. Because of the high stringency that allowed only unique paired-end reads with >97% identity to be considered, the specI method for microbial identification yielded only 50 OTUs assigned to the species-level. The dbAssign algorithm, which employs coverage of reads uniquely mapped across prokaryotic genomes, identified 316 bacterial species. Results of rapsearch2 analysis, employing relaxed stringency (DNA-protein comparison using blastx), indicated 1,655 species. The number of microbial species assigned in Figure S10 reflects both size and diversity of the respective databases. The specI method used a compilation of 1,753 prokaryotic species clusters<sup>34</sup> whereas the dbAssign toolset used a database consisting of 5,614 prokaryotic genomes that had been fully sequenced and curated. The rapsearch2 method used the complete and non-redundant NCBI protein database. The number of OTU assignments is therefore a function of mapping stringency as well as reference database size and curation.

To assess false positive identifications, the results of all three methods were compared, yielding a shortlist of 33 bacterial species identified by all three methods. Table S7 lists the species of these 33 microbes. Strain-level identification was provided by the specI mappings that cluster all reference genomes into species groups based on the similarity of 40 universal marker genes. The cross-validated set of 33 species identified in blowfly and housefly microbiomes were categorized according to the habitat from which they have been isolated and assigned to the potential disease association record of the Pathosystems Resource Integration Center (PATRIC)<sup>35</sup>. Three species are associated with insects, including different species of the endosymbiont *Wolbachia* spp. and *Bacillus thuringiensis* (Bt). The latter produces crystal proteins used as a biopesticide for insect pest control of genetically modified crops. The Bt strains identified in the fly samples are most closely related to serovar chinensis CT-43, toxic for moths, butterflies, and flies<sup>36</sup>. The remaining of the 33 species are either environmental isolates or have been described as associated with plants, animals or humans.

Potential disease association indicated that the pathogens identified in the fly microbiomes may colonize plants, animals and humans. Three species of *Pectobacterium* related to soft-rot disease were identified. These bacteria are necrotrophs, infecting a range of hosts, degrading the cell wall and promoting the death of host cells. They are reported to consume cellular nutrients and degrade plant tissues, turning the host plant to liquid mush<sup>37</sup>. *P. atrosepticum* almost exclusively infects potatoes<sup>38</sup> and was identified as being associated with the Brazilian fly samples by rapsearch2. Brazil is the second largest producer of potatoes in South America, with cultivation restricted mainly to the southern and southeastern regions of the country where most of the fly samples were collected (Food and Agriculture Organization of the United Nations - FAOSTAT, 2008).

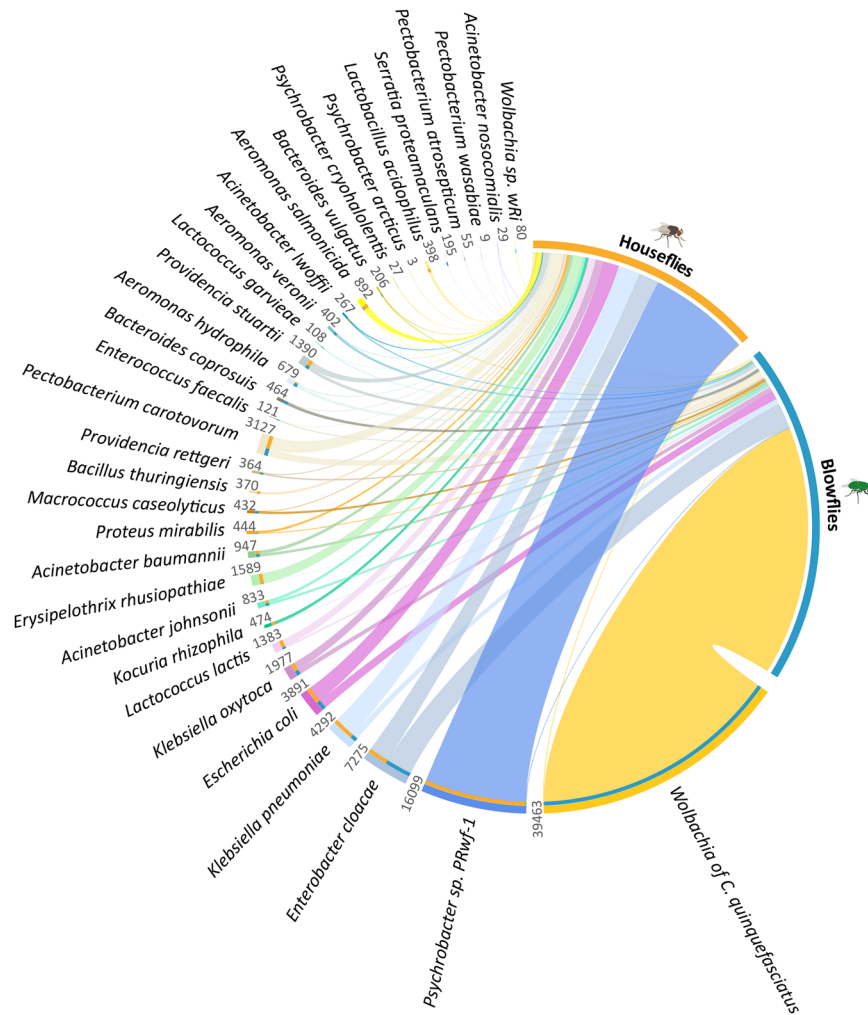
*Aeromonas salmonicida*, an animal-associated species, was identified in the housefly microbiome associated with flies collected at a farm in Brazil where there was a lake used for recreational fishing. *A. salmonicida* is a salmonid fish pathogen causing furunculosis, a bacterial septicaemia, even at low infectious dose<sup>39</sup>. *Lactococcus garvieae*, also a fish pathogen, is the etiological agent of lactococcosis<sup>40</sup>, a form of hyperacute, haemorrhagic septicaemia in fishes that was initially described as affecting Japanese yellowtail (amberjack), but later detected also in fishes in Mediterranean and Eastern countries. Interestingly, *L. garvieae* was identified as present only in significant amounts (>1000 reads) in blowflies collected in Singapore near a canteen (samples SGP\_1593, SGP\_AJ63, SGP\_AJ65, SGP\_AJ67 and SGP\_AJ68).

The 33 species that had been identified by all the methods employed in this study (Table S7) also comprised opportunistic pathogens of humans. Commonly described as of environmental origin or as a component of the normal skin or gut flora, these bacteria have been reported to cause a variety of nosocomial infections, including diarrhea, septicemia, bacteremia, pneumonia, skin and soft tissue infections, osteomyelitis, and urinary tract infections. For example, *Acinetobacter* spp. notably have been reported as nosocomial pathogens with long-term survival in hospital environments and transmitted directly between patients or via fomites<sup>41</sup>.

Potential human pathogens identified in the blowfly and housefly microbiomes are species associated with both nosocomial and generic infections, such as bacteraemia, septicaemia, and gastroenteritis. In a few cases, these species are etiological agents of specific disease, like in the case of erysipelas (in animals) and erysipeloid (in humans) caused by *Erysipelothrix rhusiopathiae*<sup>42</sup>. Similarly, *Proteus mirabilis* identified in the microbiomes of the blowfly and housefly is often associated with urinary tract infections<sup>43</sup>. It has also been reported to be present in the salivary glands of the green bottle fly, *Lucilia sericata*, and the larvae of the screwworm fly, *Cochliomyia hominivorax*, interestingly as producing antibiotic molecules<sup>44</sup>. *P. mirabilis* generates volatile swarming signals in decaying proteinaceous material that induces attraction and oviposition in flies<sup>45</sup>. The odorant that is produced is hypothesized to serve as a lure for carrion-feeding flies, enabling transport and dispersal of the bacteria.

Clearly, the opportunistic and the potentially pathogenic bacteria identified in the blowfly and housefly microbiomes are not necessarily associated with a clinical condition or infection of a specific host, whether animal, plant, or human. Rather, it is important to note that mechanical vectors disperse a range of bacterial species to a variety of hosts. The risk of infection ultimately depends on host susceptibility and contact with the agent transported by the insect vector, which moves from one reservoir to another. Figure 5 shows occurrence and abundance of the 33 microbial species identified by all of the bioinformatics methods (Table S7) used in this study and for all fly samples collected, showing host-specific species as well as the shared microbiome of both the housefly and the blowfly. The host-specific patterns of distribution for the *Wolbachia* sp. and for *Psychrobacter* sp. PRwf-1, together with the more extensive incidence of *E. cloacae*, *E. coli*, and *P. carotovorum*, suggest that the microbiomes of carrion-feeding flies represent a combination of host-inherited microorganisms and those acquired from a similar habitat and/or migrations.

**The microbiome of fly body parts.** Contamination of surfaces upon which the fly lands and traverses occurs by both physical and passive mechanisms, and thus a mechanical vector can potentially infect susceptible



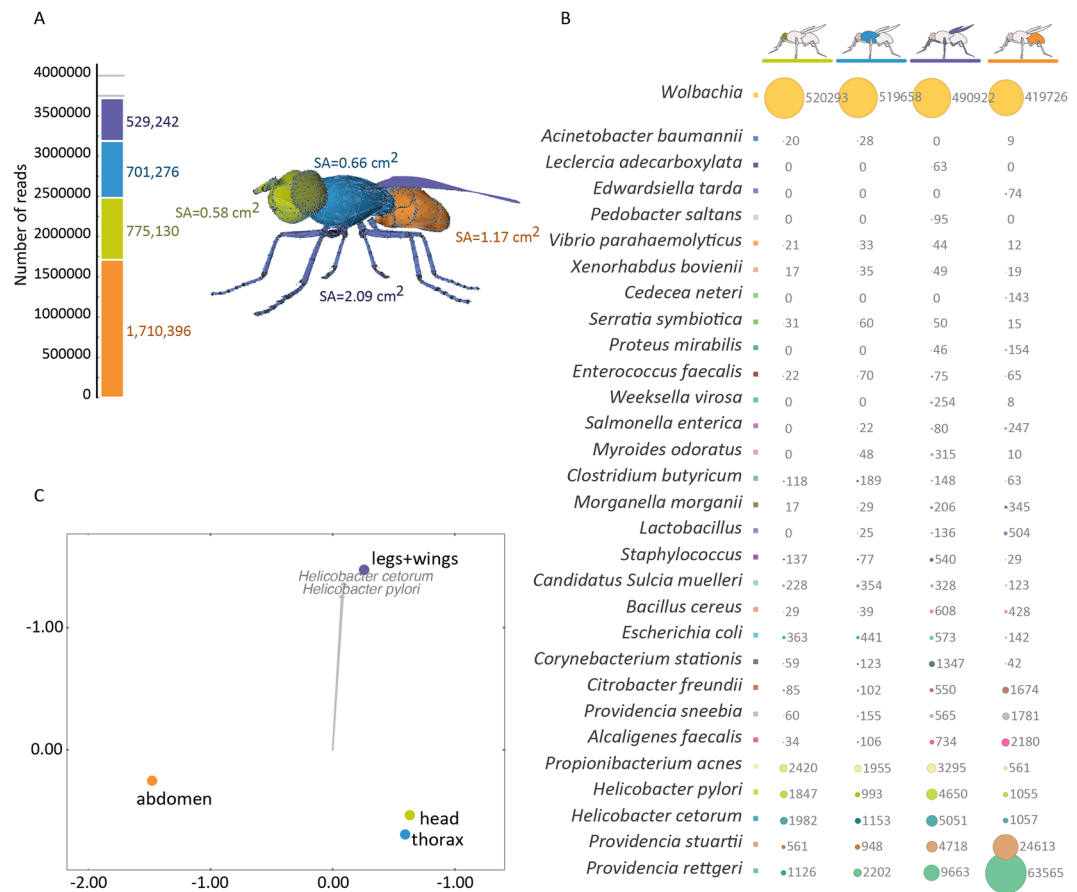
**Figure 5.** Occurrence and abundance of OTUs in blowflies and houseflies shows both shared and specific microbiomes. Circular visualization of the distribution of 33 bacterial species identified by all three bioinformatics methods. Thickness of the ribbon refers to number of reads of each bacterial species assigned to a fly host by dbAssign. The outer ring summarizes the total number of reads assigned to each bacterial species analysed (values in grey) whereas the relative abundance of reads present in the fly hosts is represented in the innermost ring by orange (housefly) and blue (blowfly) bars. Part of the microbiome is shared by both fly species (environment-acquired), and part is species specific (host-selected). See Table S7 for information about disease-association of 33 bacterial taxa identified by all methods. Data was parsed with Circos table viewer<sup>72</sup>.

humans, animals, and plants. Most studies to date have investigated the gastrointestinal tract of insects, without addressing the role of the outer body of flies. It can be hypothesized that the fly feet, wings, mouthparts and other body surfaces constitute the main route of microbial dispersal by mechanical vectors. Therefore, we investigated the bacterial abundance and diversity of four distinct body parts of a blowfly: head, thorax, abdomen, and legs + wings. Based on a computer-generated three-dimensional model of a fly, the relative ratio of surface areas (SA) of a blowfly was estimated to be 1.0:1.1:2.0:3.6 for head, thorax, abdomen, and legs + wings, respectively (Fig. 6A). The abdomen accounted for more than twice the number of reads generated relative to other body fractions, likely because this body part covers most of the gastrointestinal tract. Despite a small body mass, the legs + wings fraction yielded the highest diversity of bacterial species (Figure S11) and likely plays a significant role in dispersing bacteria by the fly from one landing site to another.

The bubble chart in Fig. 6B shows the 30 species of bacteria most frequently associated with a given body part. The endosymbiont *Wolbachia* sp. is the most abundant and ubiquitous organism for all body parts of the blowfly. The fact that all body parts have shown similar abundances of *Wolbachia* spp. is an indirect evidence of somatic localization of this endosymbiont in the blowfly *C. megacephala*. This suggests that not only germline cells are being colonized, but somatic tissues as well<sup>46</sup>.

The microbiomes of the four body parts are essentially similar to the range of diversity previously described for the 62 blowflies in Fig. 3, indicating that most of the microorganisms associated with flies are not restricted to their gastrointestinal tract. In fact, the PCoA of the four body parts, seen in Fig. 6C, indicates that the microbiome of the head and thorax are more similar to each other, while the microbiomes of the abdomen and legs + wings





**Figure 6.** Microbiome of the four body parts. **(A)** A 3-dimensional model of a fly (composed of 11,421 triangles) was used to calculate the approximate surface areas (SA) of four body parts, indicating that the legs + wings, despite small volume and mass, have the largest surface area. The stacked bars indicate total number of reads for each dissected body fraction, showing the abdomen generated more reads than other body parts, most likely because it contains the major part of the intestinal tract and other organs. **(B)** Bubble chart of the top 30 bacterial species assigned to each of the four body parts of a blow fly. The endosymbiont *Wolbachia* sp. was detected in all tissues. Despite the tiny amounts of biological material recovered from legs + wings, the microbial diversity of this body part was found to be higher than other tissues. The bubble chart indicates presence of *Helicobacter* spp. in the blowfly, most prominently in legs + wings. This could be a potential route of dispersion of this pathogen to humans and animals. See Figure S11 for alpha-diversity of the four body parts, Table S8, and Figure S13 for results obtained with Cosmos ID metagenomics software package. **(C)** PCoA of the microbiome of the head, thorax, abdomen and legs + wings, showing separation of legs + wings dataset is mainly driven by presence and abundance of *Helicobacter* species. See Figure S12 for genome-wide coverage of *H. pylori* with metagenomics reads of the four body parts. PCoA was generated using Bray-Curtis ecological index (stress = 0.80).

clustered separately. *Providencia* spp. are more abundant on the abdomen and have been isolated from whole insects and from the gut of blowflies, stable flies, houseflies and fruit flies, also showing a varied virulence and mortality in *D. melanogaster*<sup>47</sup> by being capable of avoiding their detection by the insect immune system<sup>48</sup>. Interestingly, the dissimilarity that drives the separation of the microbiome of the legs + wings is the abundance of *Helicobacter ceterum* and *Helicobacter pylori* (Fig. 6B and C). Both species are closely related and their proteomes showed an average similarity of >75%<sup>49</sup>. It is estimated that *H. pylori* and *H. ceterum* diverged 450,000 years ago<sup>50</sup> and are part of the same branch together with *H. acinonychis*<sup>51</sup>.

**Detection of *Helicobacter pylori*.** The surprisingly high incidence of *Helicobacter pylori* motivated us to explore the presence of this important human pathogen on the blowfly body in greater detail. *H. pylori* is known to colonize the human stomach and chronic infection with *H. pylori* can result in peptic ulcers, increased risk of mucosa-associated lymphoid tissue lymphoma, and even gastric adenocarcinoma<sup>52</sup>. Transmission of *H. pylori* and its colonization of humans are not fully understood, but oral-oral and fecal-oral routes of transmission have been proposed. Houseflies were reported to be an alternative reservoir, after it was shown that *H. pylori* survives in the gut of houseflies artificially infected in the laboratory<sup>53</sup>. However, the presence of virulent strains of *H. pylori* captured from the natural environment has never been reported. In the study reported here, several individual wild-caught blowflies were found to carry *H. pylori* DNA. The *H. pylori* genome coverage was

evaluated using genome-wide assignment methods. Figure S12 shows 5,890 reads of DNA extracted from body parts which mapped across the *H. pylori* reference genome (NC\_000915.1). A total of 408,693 bp were covered, with an average pairwise identity of 97.5%. The reads are distributed across the genome, covering 25.3% of the reference sequence. Specific virulence factors were identified, showing that both the *cag* pathogenicity island (cytotoxin-associated genes) and *vacA* gene (vacuolating cytotoxin autotransporter) are present. The *cag* pathogenicity island is responsible for unique virulence factors of *H. pylori* strains with enhanced virulence<sup>54</sup>, while *vacA* is a multi-functional toxin with polymorphisms linked to increased risk of gastric cancer<sup>55</sup>. Further analysis was performed using the Cosmos ID metagenomics software package as described previously<sup>56</sup>, which identified *H. pylori* with relative abundance of 0.57% on the abdomen, 0.6% on the thorax, 2.29% on the head, and 7.21% on the legs + wings (Figure S13 and Table S8). In addition, CosmosID detected four virulence factors associated with *H. pylori*, including *cag11*, *cag18*, *flgC*, and *fur* (Figure S13).

*Helicobacter* sp. reads were found in 15 blowflies of the 116 sampled in this analysis and all had been collected in Brazil. Most likely, the flies acquired this bacterium from untreated, open sewage sites or outside latrines. Whether *H. pylori* survives and persists on the outer body of the fly has not been determined. However, *H. pylori* remains viable for up to 12 hours on agar plates exposed to air and can be ingested and excreted by houseflies infected in laboratory conditions<sup>53</sup>. Also, the acidic pH of 2.9–3.5 of the midgut of blowflies<sup>57</sup>, may play a role in the viability of *H. pylori*. These findings strongly indicate further study of alternative routes of transmission of *H. pylori* is needed, notably those that may be mediated by flies in both urban and rural environments.

**Mechanisms of dispersal.** It has been reported that flies transmit microorganisms by regurgitation, excretion, and via contact with a body part<sup>58</sup>. To investigate mechanisms of bacterial transmission, blowflies were introduced into a Petri dish and allowed to land on an *E. coli* lawn (marker strain) that had been inoculated onto LB agar. Subsequently, the flies walked across the surface of sterile agar plates and the pattern of *E. coli* observed on the plates, after incubation, matched the footprints of their walk (Fig. 7A). This simple experiment, not unique to our study, shows clearly that flies can disperse bacteria from one landing site to another and that the inoculum can persist over time and after a series of many individual contacts. Unique to this experiment was that there was little evidence of bacterial growth consistent with contact of the abdomen or mouthparts of the fly with the agar, indicating that legs of flies are the main source for rapid and extensive dispersion of bacteria. Blowflies and houseflies show anatomical features, such as bristles and pads, coated with substances that increase the adhesion of particles to their legs and feet<sup>59</sup>. These structures maximize surface area and may promote retention of bacteria that can then be dispersed to other surfaces. By sequencing the DNA extracted from bacterial colonies on agar from the footprints of the fly, we verified that the *E. coli* strains at the point of landing and beginning of the fly walk on agar were identical (data not shown). These results demonstrate the vector capacity of flies for bacterial dispersal and show that flies can transport viable, cultivatable bacteria from one place to another, thus confirming the DNA-based evidence discussed above. Further evidence of bacterial viability on the outer body of a fly was verified with bacterial growth from sterile buffer used for washing individual flies collected in urban environments. The washes were spread on LB agar plates and yielded up to 30,000 colony-forming units (CFUs) per fly (Fig. 7B). However, the epidemiological relevance of these findings will depend on determining that viability of bacteria on the body of flies persists over extended periods of time and/or if those bacteria have a significant role in the host life cycle, as has been demonstrated for *P. mirabilis*<sup>45</sup>.

**Intersection of synanthropic microbiome datasets.** It is notable that a portion of the blowfly and housefly pan-microbiomes overlaps (Figure S14) with the gastrointestinal microbiome of humans<sup>60</sup> and the urban microbiome<sup>61</sup>, emphasizing the role of carrion flies as a shuttle between source and sink that are ecologically associated with humans and animals. The precision of our taxonomic identification of bacterial species allows the conclusion that carrion flies are a proxy for the environment from which a significant part of their microbiome is acquired. It highlights the importance of surveillance of fly microbiomes, especially in densely populated areas. If included in public health surveillance programs, it will be possible to predict and prevent routes of transmission of microbes and potential pathogens mediated by these mechanical vectors.

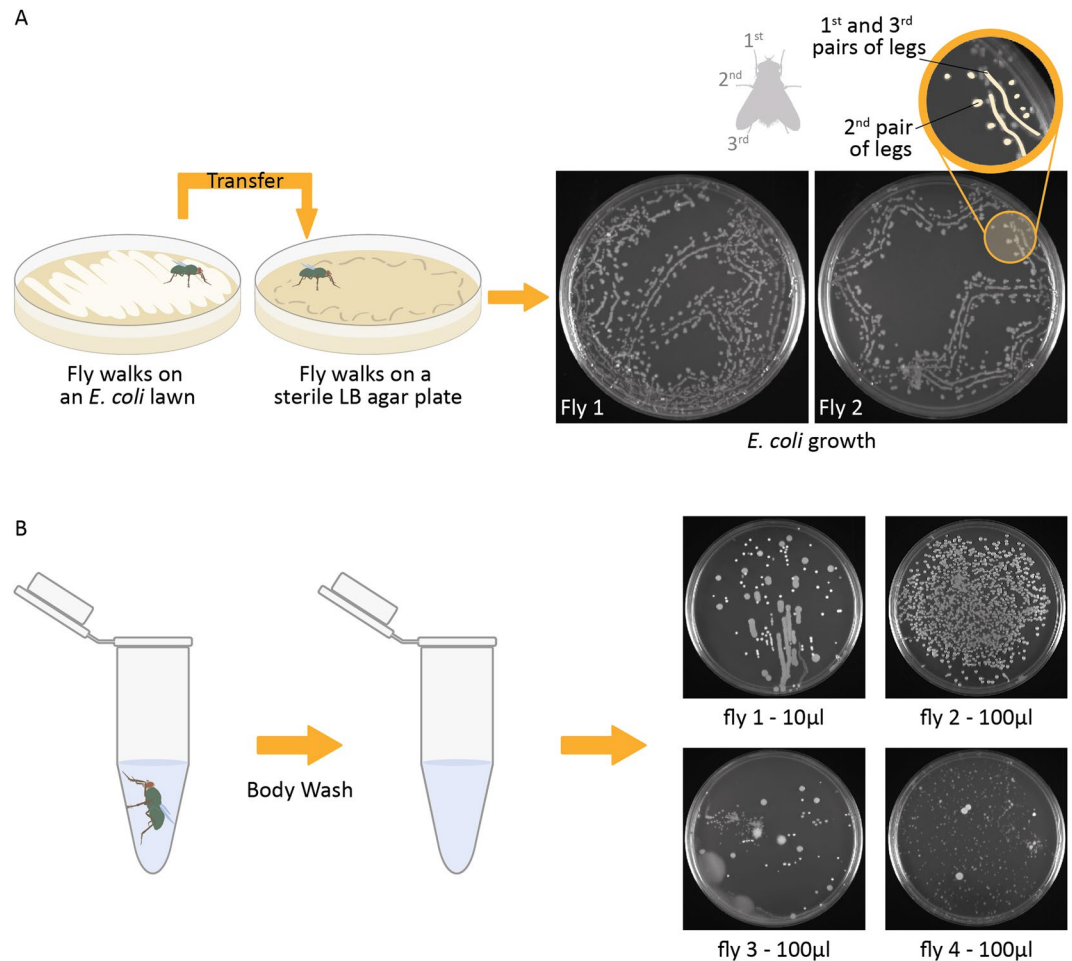
## Conclusions

By sequencing the holobiomes of individual insects, we demonstrate that houseflies and blowflies contain a host-specific microbiome. Nevertheless, a significant shared microbiome, comprising approximately half of the bacterial species, crosses the host species boundary and is present in the microbiomes of the two carrion-flies. As described for biological vectors, our data suggest that flies also harbour a stochastic component in their microbiome, serving as mechanical vectors of bacteria largely derived from decaying material in the environment. Lastly, it is also concluded that adhesion of bacteria from the environment to the outer surface of the exoskeleton, notably the legs and wings, is a significant route of transmission of bacteria from surface to surface, including plant, animal, and human pathogens. The environmental sequencing approach presented in this study can be an effective tool for vector control programs and public health environmental surveillance.

## Materials and Methods

**Sampling.** Blowflies (*C. megacephala*) and houseflies (*M. domestica*) were collected with an entomological net using decomposing fish as bait. Flies were immediately placed in individual microcentrifuge tubes and maintained on dry ice for further taxonomic identification and photodocumentation. All samples were stored at –80 °C until DNA extraction was performed.

Flies were collected from urban, rural and natural sites in three countries: Brazil, the United States, and Singapore. Urban sites included a food market, the emergency entrance of a public hospital, a public park in a



**Figure 7.** Microbial transport by a mechanical vector. **(A)** Blowflies were exposed to a Petri dish with an *E. coli* lawn and then allowed to walk on a fresh, sterile agar plate. The path of the fly walking on the agar can be seen as footprints after incubation. The line of growth indicates that dispersal of bacteria by the blowfly occurs mainly via the legs. The track pattern matched the arrangement of the three pairs of legs, with the first and the third pair resulting in nearly continuous, linear bacterial growth and the second pair of legs generating separate circular colonies on the outside of the lines (see inset for detail). In a few instances, it was possible to observe bacterial growth between the lines of growth probably from mouthparts. **(B)** Growth experiments to evaluate viability of bacteria on the outer body surface of blowflies. The flies were sampled in urban environments near a food court and washed with PBS sterile buffer for 10 minutes. The buffer was then spread in LB agar plates and incubated at 37 °C overnight. Up to 30,000 CFUs were estimated from body washes, showing that bacteria can be transported on the exterior surfaces of flies in a viable state that they can be cultivated.

metropolitan setting, a sanitary landfill for household waste, and a hawker centre/food court. The natural environment comprised of a protection area at the Instituto Nacional de Pesquisas Amazônicas in the Amazon rainforest. Rural sites included a mixed farm with animal stables for pigs, horses, chickens, and cattle, and a poultry farm. Table S2 lists the location, coordinates, and metadata associated with the 116 samples analysed in this study.

**Blowfly colony.** A laboratory strain of the blowfly *C. megacephala*, derived from a colony maintained for 20 generations at the Laboratory of Animal Evolutionary Genetics (University of Campinas, Brazil), was reared as an environmental control. One female fly was kept in a cage with mouse carcass to stimulate oviposition. After the eggs hatched, larvae were reared to pupae in mouse carcasses provided by the animal facility at University of Campinas and maintained under controlled conditions (temperature 25 °C ± 1 °C; 70% relative humidity; 12/12 hour photoperiod) in a rearing chamber Fitotron (Eletrolab, model EL011). Pupae were individualized in tubes and kept in the same conditions until the emergence of adults. Ninety-eight adult flies newly emerged were pooled and their DNA was used as an environment control.

**DNA extraction and sequencing.** Total DNA of individual flies was extracted with DNeasy Blood & Tissue kit (Qiagen). For the blowflies, the individuals were usually split (if weighing >25 mg) and two separate DNA extractions were performed for each individual. After DNA precipitation, the DNA from two extractions was combined in a single tube. For houseflies, usually weighing <25 mg, the DNA extraction was performed with an entire, individual fly. One individual blowfly was also dissected into head, thorax, abdomen and legs + wings

using bleached fine-tipped forceps and a sterile scalpel. The total DNA was then extracted from each of the four body parts. Some modifications were made to the DNA extraction protocol suggested by the manufacturer. Briefly, flies (or body parts) were ground and homogenized with a microcentrifuge teflon pestle in ATL (animal tissue lysis) buffer containing 20 µl of Proteinase K. The mixture was vortexed and incubated at 56 °C overnight to complete the tissue lysis. For maximum DNA yield, elution from the column was conducted twice with 50 to 100 µl of AE buffer (provided with the kit). DNA samples were quantitated by Qubit fluorescence assay (Thermo Fischer) and DNA integrity was determined on a 2100 Bioanalyzer (Agilent Technologies). Total DNA was sheared to 300 bp using a Covaris S220 focused-ultrasonicator (Covaris Inc.). Automated library construction was either performed on a SPRI-TE Nucleic Acid Extractor (Beckman Coulter), using the SPRIworks Fragment Library Kit I or on a Bravo NGS Workstation (Agilent Technologies) with Illumina TruSeq DNA sample preparation reagents, following the manufacturers' recommendations. For all library preps barcoded, TruSeq DNA LT adapters (Illumina Inc.) were used to enable library pooling for sequencing. For samples processed on the SPRI-TE, the DNA was size-selected on a Pippin Prep (Sage Science) electrophoresis system prior to library preparation to remove small DNA fragments. The size-selection was performed on 2% agarose ethidium free cassettes (Sage Science), followed by manual DNA purification with Agencourt AMPure XP magnetic beads (Beckman Coulter). Finished libraries were quantitated with Quant-iT™ Picogreen® (Invitrogen) and concentrations validated by qPCR, following KAPA SYBR® FAST qPCR kit instructions (Kapa Biosystems). Equimolar amounts (8–12 pM) of libraries were pooled for multiplexed sequencing on a HiSeq. 2000 or HiSeq. 2500 (Illumina Inc.) platform with 101-bp, 151-bp or 251-bp paired-end protocols. Table S2 shows the read length produced for each sample. Sequencing was performed at the Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University (Singapore) and at the Center for Comparative Genomics and Bioinformatics, The Pennsylvania State University (University Park, PA, USA).

**Metagenomic analyses.** Raw fastq files were trimmed for both adapter and low-quality bases using cutadapt 1.0<sup>62</sup>. A maximum error rate of 0.2 during the search for the adapter sequences was allowed, and a quality cutoff of 20 was used to trim low-quality ends from reads before adapter removal. High-quality reads were aligned against an in-house draft genome of *C. megacephala* (unpublished) and the *M. domestica* reference genome<sup>63</sup> (AQPM00000000.1) to filter out the host reads, thus reducing the size of datasets and the computational resources required for the metagenomic analyses. The alignments were done using bowtie2 v2.1.0<sup>16</sup> with parameters selected for higher level of sensitivity. The remaining non-host fraction of the datasets was used for the individual microbiome analyses of 62 blowflies, 53 houseflies, and 1 pool of lab-reared flies used as environmental control. The non-host sequence reads used for metagenomic analyses are available under NCBI BioProject ID PRJNA385554.

The resulting non-host filtered metagenomic datasets were analysed with three separate methods: (i) alignment of the translated reads against the non-redundant (NR) database using rapsearch2<sup>64</sup>; (ii) alignment of reads against a customized prokaryotic genomic database using dbAssign; and (iii) alignment of short-read datasets against specI representative genomes<sup>34</sup>. Specific bit-score cutoffs were implemented for read assignment to microbial taxa, based on the read length (Table S2).

The rapsearch2 (v2.15) analyses were performed with the nucleotide fastq format as input in a multithreaded run. Reads were translated in six frames and aligned against the NCBI NR protein database (downloaded on 12 April 2015) with BLASTX using default parameters.

dbAssign is an in-house python toolset (<https://github.com/aakrosh/dbAssign>) that can be used to assign fragments to multiple reference sequences from a customized database. We created a database that included 5614 prokaryote genomes available on NCBI (downloaded on 25 February 2016 - complete and chromosome-level genomes). Reads were aligned against the database using BWA v0.7.4<sup>65</sup> with the default parameters to find all alignments that exceeded a score threshold. dbAssign was used to filter and keep matches with a minimum coverage of 90% of the bases and a minimum identity of 95% over those bases. dbAssign then permutes all likely alignments of the fragment using the filtered alignments of the reads and assigns the fragment to a particular reference if both reads of the fragment can be aligned to it. It calculates a bit-score for the alignments and the details are output in the BLAST tabular format. NCBI taxonomy IDs were recorded for use as an input to MEGAN 5<sup>66</sup>, which was restricted to show only unique paired-read matches assigned to a species-level bacterial taxon. Matches to multiple 'species' were assigned to the higher taxonomical level.

For the alignment of reads against a database of 1,753 representative genomes, we used specI (<http://vm-lux.embl.de/~mende/specI/>)<sup>34</sup>. This method groups reference genomes into species clusters, based on 40 universal, single-copy marker genes and allows for species-level identification. The microbial fraction of paired-end reads were competitively mapped using BWA v0.7<sup>65</sup>, using default parameters. Reads were counted only when they were mapped with greater than 97% identity. We computed horizontal and vertical coverage per sample over each of the genomes using qaCompute (<https://github.com/CosteaPaul/qaTools>).

The alignment results were imported into MEGAN 5 and taxa assignments were done with strict parameters of the LCA-assignment algorithm and considering the read length generated for each sample (Table S2), with the following settings: Max Expected = 0.01, Top Percentage = 10.0, Min Support = 25, Min Complexity = 0.33, Paired Reads = On.

All metagenomic analyses were performed for individual datasets and further normalized to perform comparative analyses of microbiomes. The datasets were normalized to the dataset with the smallest number of reads of the 116 samples to represent relative abundances of assigned bacterial taxa and presented as a bubblechart in Fig. 3. The bubbles were plot with a log-squared scale. The calculation of principal coordinate analysis (PCoA) was performed using Bray-Curtis dissimilarity<sup>67</sup> and the tree was generated with UPGMA method to show hierarchical clustering of fly species based on their microbiome. Species diversity index Shannon-Weaver<sup>68</sup> was calculated using 'species' level of NCBI taxonomy to generate a table with MEGAN 5<sup>66</sup>. PERMANOVA and ANOVA

statistical analysis were performed with phyloseq<sup>69</sup>, as implemented in the tool MicrobiomeAnalyst<sup>70</sup>. Venn diagrams were generated with BioVenn<sup>71</sup> using the 'species' level table as an input to generate area-proportional representation.

The alignment of assigned reads against full reference genomes of the *M. domestica* Salivary Gland Hypertrophy Virus (NC\_01067) and *Helicobacter pylori* (NC\_000915.1) was conducted using bowtie 2<sup>16</sup> with default parameters.

**Bacterial growth.** Individual blowflies were collected and enclosed in a chamber with a Petri dish containing an LB-agar plate covered with a lawn of *Escherichia coli* K12. The flies walked on the lawn for about 30 seconds before the plates were exchanged with a fresh sterile LB agar plate. After the flies had walked on the sterile plate for 30–60 seconds, they were incubated overnight at 37°C and photographed. To investigate the presence of bacteria on the outer body of flies, sixteen whole adult blowflies were incubated in sterile PBS buffer at room temperature for 10 minutes with gentle shaking. Volumes of 1, 10 and 100 µl of the fly body washes were plated on LB-agar Petri dishes and incubated at 37°C overnight. Plates with colony-forming units were counted and photographed.

**Data and software availability.** The full datasets used for metagenomic analyses (non-host datasets) were deposited in the SRA database and are available under NCBI BioProject ID PRJNA385554.

The source code for dbAssign is available at <https://github.com/aakrosh/dbAssign>. dbAssign is a new python toolset developed to assign short DNA fragments to multiple reference sequences from a customized database.

## References

- Morgan, X. C., Segata, N. & Huttenhower, C. Biodiversity and functional genomics in the human microbiome. *Trends in Genetics* **29**, 51–58 (2013).
- Archie, E. A. & Tung, J. Social behavior and the microbiome. *Current Opinion in Behavioral Sciences* **6**, 1–7 (2015).
- Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science (New York, N.Y.)* **320**, 1647–1651 (2008).
- Schlaeppli, K. & Bulgarelli, D. The plant microbiome at work. *MPMI* **28**, 212–217 (2015).
- Lewis, Z. & Lizé, A. Insect behaviour and the microbiome. *Current Opinion in Insect Science* **9**, 86–90 (2015).
- Engel, P. & Moran, N. A. The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol. Rev.* **37**, 699–735 (2013).
- Singh, B. *et al.* A metagenomic assessment of the bacteria associated with *Lucilia sericata* and *Lucilia cuprina* (Diptera: Calliphoridae). *Appl Microbiol Biotechnol* **99**, 869–883 (2014).
- Song, S. J. *et al.* Cohabiting family members share microbiota with one another and with their dogs. *eLife* **2**, e00458–22 (2013).
- Amendt, J., Krettek, R. & Zehner, R. Forensic entomology. *Naturwissenschaften* **91**, 51–65 (2004).
- Blazar, J., Lienau, E. K. & Allard, M. Insects as vectors of foodborne pathogenic bacteria. *Terrestrial Arthropod Reviews* **4**, 5–16 (2011).
- Gupta, A. K. *et al.* Molecular phylogenetic profiling of gut-associated bacteria in larvae and adults of flesh flies. *Med Vet Entomol*, <https://doi.org/10.1111/mve.12054> (2014).
- Azambuja, P., Garcia, E. S. & Ratcliffe, N. A. Gut microbiota and parasite transmission by insect vectors. *Trends in Parasitology* **21**, 568–572 (2005).
- Brits, D., Brooks, M. & Villet, M. H. Diversity of bacteria isolated from the flies *Musca domestica* (Muscidae) and *Chrysomya megacephala* (Calliphoridae) with emphasis on vectored pathogens. *African Entomology* **24**, 365–375 (2016).
- Yun, J. H. *et al.* Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology* **80**, 5254–5264 (2014).
- Junqueira, A. C. M. *et al.* Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. *Sci. Rep.* 1–13, <https://doi.org/10.1038/srep21762> (2016).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
- García-Maruniak, A., Maruniak, J. E., Farmerie, W. & Boucias, D. G. Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of *Musca domestica*, MdSGHV. *Virology* **377**, 184–196 (2008).
- Promptboon, P. *et al.* *Musca domestica* salivary gland hypertrophy virus, a globally distributed insect virus that infects and sterilizes female houseflies. *Applied and Environmental Microbiology* **76**, 994–998 (2010).
- Bahrndorff, S., de Jonge, N., Skovgård, H. & Nielsen, J. L. Bacterial communities associated with houseflies (*Musca domestica* L.) sampled within and between farms. *PLoS ONE* **12**, e0169753–15 (2017).
- Douglas, A. E. Multiorganismal insects: diversity and function of resident microorganisms. *Annu. Rev. Entomol.* **60**, 17–34 (2015).
- Sukontason, K. L. *et al.* Comparison between *Musca domestica* and *Chrysomya megacephala* as carriers of bacteria in northern Thailand. *Southeast Asian J. Trop. Med. Public Health* **38**, 38–44 (2007).
- Werren, J. H., Baldo, L. & Clark, M. E. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Micro* **6**, 741–751 (2008).
- Mingchay, P. *et al.* *Wolbachia* supergroups A and B in natural populations of medically important filth flies (Diptera: Muscidae, Calliphoridae, and Sarcophagidae) in Thailand. *Southeast Asian J. Trop. Med. Public Health* **45**, 309–318 (2014).
- Baldo, L. *et al.* Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Applied and Environmental Microbiology* **72**, 7098–7110 (2006).
- Moreira, L. A. *et al.* A *Wolbachia* symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and *Plasmodium*. *Cell* **139**, 1268–1278 (2009).
- Martin, O. Y., Gubler, A., Wimmer, C., Germann, C. & Bernasconi, M. V. Infections with *Wolbachia* and *Spiroplasma* in the Scathophagidae and other Muscoidea. *Infection, Genetics and Evolution* **12**, 315–323 (2012).
- Firstenberg-Eden, R. & Rowley, D. B. Factors affecting inactivation of *Moraxella-Acinetobacter* cells in an irradiation process. *Applied and ...* **49**, 457–463 (1980).
- Lemaitre, B. & Miguel-Aliaga, I. The digestive tract of *Drosophila melanogaster*. *Annu. Rev. Genet.* **47**, 377–404 (2013).
- Grimaldi, D. A. & Engel, M. S. *Evolution of the Insects*. (Cambridge University Press, 2005).
- Martinson, V. G., Moy, J. & Moran, N. A. Establishment of characteristic gut bacteria during development of the honeybee worker. *Applied and Environmental Microbiology* **78**, 2830–2840 (2012).
- Wang, Y., Gilbreath, T. M., Kukutla, P., Yan, G. & Xu, J. Dynamic gut microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya. *PLoS ONE* **6**, e24767 (2011).
- Sanders, W. E. & Sanders, C. C. *Enterobacter* spp.: pathogens poised to flourish at the turn of the century. *Clinical Microbiology Reviews* **10**, 220–241 (1997).
- Matic, I. *et al.* Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science (New York, N.Y.)* **277**, 1833–1834 (1997).

34. Mende, D. R., Sunagawa, S., Zeller, G. & Bork, P. Accurate and universal delineation of prokaryotic species. *Nature Methods* **10**, 881–884 (2013).
35. Gillespie, J. J. *et al.* PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect. Immun.* **79**, 4286–4298 (2011).
36. He, J. *et al.* Complete Genome Sequence of *Bacillus thuringiensis* subsp. *chinensis* strain CT-43. *J. Bacteriol.* **193**, 3407–3408 (2011).
37. Davidsson, P. R., Kariola, T., Niemi, O. & Palva, E. T. Pathogenicity of and plant immunity to soft rot pectobacteria. *Front Plant Sci* **4**, 191 (2013).
38. Ma, B. *et al.* Host range and molecular phylogenies of the soft rot enterobacterial genera pectobacterium and dickeya. *Phytopathology* **97**, 1150–1163 (2007).
39. Reith, M. E. *et al.* The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. *BMC Genomics* **9**, 427–15 (2008).
40. Vendrell, D. *et al.* *Lactococcus garvieae* in fish: a review. *Comparative Immunology, Microbiology and Infectious Diseases* **29**, 177–198 (2006).
41. Bergogne-Bérézin, E. & Towner, K. J. *Acinetobacter* spp. as nosocomial pathogens: microbiological, clinical, and epidemiological features. *Clinical Microbiology Reviews* **9**, 148–165 (1996).
42. Brooke, C. J. & Riley, T. V. *Erysipelothrix rhusiopathiae*: bacteriology, epidemiology and clinical manifestations of an occupational pathogen. *J. Med. Microbiol.* **48**, 789–799 (1999).
43. Burall, L. S. *et al.* *Proteus mirabilis* genes that contribute to pathogenesis of urinary tract infection: identification of 25 signature-tagged mutants attenuated at least 100-fold. *Infect. Immun.* **72**, 2922–2938 (2004).
44. Erdmann, G. R. & Khalil, S. K. W. Isolation and identification of two antibacterial agents produced by a strain of *Proteus mirabilis* isolated from larvae of the screwworm (*Cochliomyia Hominivorax*) (Diptera: Calliphoridae). *J. Med. Entomol* **23**, 208–211 (1986).
45. Ma, Q. *et al.* *Proteus mirabilis* interkingdom swarming signals attract blow flies. *ISME J* **6**, 1356–1366 (2012).
46. Pietri, J. E., DeBruhl, H. & Sullivan, W. The rich somatic life of *Wolbachia*. *MicrobiologyOpen* **5**, 923–936 (2016).
47. Galac, M. R. & Lazzaro, B. P. Comparative pathology of bacteria in the genus *Providencia* to a natural host, *Drosophila melanogaster*. *Microbes Infect.* **13**, 673–683 (2011).
48. Galac, M. R. & Lazzaro, B. P. Comparative genomics of bacteria in the genus *Providencia* isolated from wild *Drosophila melanogaster*. *BMC Genomics* **13**, 612 (2012).
49. Cao, D.-M. *et al.* Comparative genomics of *H. pylori* and non-*pylori Helicobacter* species to identify new regions associated with its pathogenicity and adaptability. *BioMed Research International* 1–15, <https://doi.org/10.1155/2016/6106029> (2016).
50. Cravedi, P., Mori, G., Fischer, F. & Percudani, R. Evolution of the selenoproteome in *Helicobacter pylori* and Epsilonproteobacteria. *Genome Biology and Evolution* evv177–13, <https://doi.org/10.1093/gbe/evv177> (2015).
51. Moodley, Y. *et al.* Age of the association between *Helicobacter pylori* and man. *PLoS Pathog* **8**, e1002693–16 (2012).
52. Marshall, B. J. & Warren, J. R. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **1**, 1311–1315 (1984).
53. Grübel, P. *et al.* Vector potential of houseflies (*Musca domestica*) for *Helicobacter pylori*. *J. Clin. Microbiol.* **35**, 1300–1303 (1997).
54. Censini, S. *et al.* *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. USA* **93**, 14648–14653 (1996).
55. Palframan, S. L., Kwok, T. & Gabriel, K. Vacuolating cytotoxin A (VacA), a key toxin for *Helicobacter pylori* pathogenesis. *Front. Cell. Inf. Microbio* **2** (2012).
56. Hasan, N. A. *et al.* Microbial community profiling of human saliva using shotgun metagenomic sequencing. *PLoS ONE* **9**, e97699–10 (2014).
57. Greenberg, B. Model for destruction of bacteria in the midgut of blow fly maggots. *J. Med. Entomol.* **5**, 31–38 (1968).
58. Hall, M. & Wall, R. Myiasis of humans and domestic animals. *Advances in Parasitology (United Kingdom)* (1995).
59. Fasanella, A. *et al.* Evaluation of the house fly *Musca domestica* as a mechanical vector for an anthrax. *PLoS ONE* **5**, e12219 (2010).
60. Human Microbiome Jumpstart Reference Strains Consortium *et al.* A catalog of reference genomes from the human microbiome. *Science (New York, N.Y.)* **328**, 994–999 (2010).
61. Afshinnekoo, E., Meydan, C., Chowdhury, S. & Jaroudi, D. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Systems*, <https://doi.org/10.1016/j.cels.2015.01.001> (2015).
62. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
63. 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**, 1–16 (2014).
64. Zhao, Y., Tang, H. & Ye, Y. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28**, 125–126 (2012).
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
66. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Research* **17**, 377–386 (2007).
67. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs* **27**, 325 (1957).
68. Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423 (1948).
69. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217–11 (2013).
70. Dhariwal, A. *et al.* MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Research* **45**, W180–W188 (2017).
71. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488–6 (2008).
72. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Research* **19**, 1639–1645 (2009).

## Acknowledgements

The authors thank Rosangela Aparecida Rodrigues for technical assistance and sample processing, Gabriel Lopes Centoducatte for the help with sample collection, Tay Qi Xiang Martin and Sebastian Noel Wittekindt for 3D modelling of the fly. This project was funded by an institutional grant from Singapore Ministry of Education Academic Research Fund Tier 3 under the research grant MOE2013-T3-1-013, and the Singapore Centre for Environmental Life Sciences Engineering (SCLESE), whose research is supported by the National Research Foundation Singapore, Ministry of Education, Nanyang Technological University, and National University of Singapore, under its Research Centre of Excellence Program. Partial support was also provided by United Nations Joint FAO/IAEA missions to A.M.L.A.-E (IAEA: 14856/RO).

### Author Contributions

A.C.M.J. and S.C.S. designed and contributed equally to the study; A.C.M.J., S.C.S., A.M.L.A.E., B.L. and D.F.P. collected flies; A.C.M.J., D.F.P. and N.E.G. extracted DNA; A.C.M.J., D.I.D.-M. and S.C.S. performed library construction and sequencing; A.C.M.J., A.R., E.A., B.N.V.P., P.I.C., P.B., B.L., R.W.P. and S.C.S. analysed raw data and conducted bioinformatics analyses. B.L. performed colony growth experiments; A.R. developed the script for dbAssign; P.S., N.H. and R.C.C. performed the CosmosID metagenomics analysis; D.A.B. and A.R. helped to interpret results. A.C.M.J. and S.C.S. wrote the manuscript with input from D.A.B., R.R.C., P.I.C., P.B. and A.R.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-16353-x>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017