# Metagenome-based identification of functional traits of the black soldier fly gut microbiome associated with larval performance

IJdema, F.; Arias-Giraldo, L.M.; Vervoort, E.; Struyf, T.; Ende, W. van den; Raaijmakers, J.M.; ... ; n

BMC Microbiology

RESEARCH

Open Access

# Metagenome-based identification of functional traits of the black soldier fly gut microbiome associated with larval performance

F. IJdema[1,2], L. M. Arias-Giraldo[3], E. Vervoort[4], T. Struyf[5], W. Van den Ende[5], J. M. Raaijmakers[3,6], B. Lievens[2] and Jeroen De Smet[1,7*]

## Abstract

**Background** The relationship between microbiomes and their hosts has been the subject of intensive study in recent years. For black soldier fly larvae (BSFL) (*Hermetia illucens* L., Diptera: Stratiomyidae), correlations between shifts in its microbial gut community composition and its health and performance suggest that the BSFL gut microbiome encodes important functions that complement the insect's own immune system and metabolism. To date, most BSFL microbiome studies have been based on 16S rRNA sequencing data. Because this approach derives a lot of information from very short sequencing reads, it was hypothesized that more insight into bacterial functionality could be generated using more extensive sequencing technologies. Here, whole genome shotgun (WGS) metagenomic sequencing was employed to investigate which microbiome-associated taxa and functions were associated with increased performance of larvae reared on a chicken feed (CF) or artificial supermarket food waste (SFW) based diet.

**Results** Taxonomic and functional profiling of the BSFL gut microbiome revealed a significant shift in response to diet, where bacterial genes encoding specific metabolic functions, such as the metabolism of sorbitol, were significantly enriched in the microbiome of larvae reared on SFW-diet. This indicates that the nutritional composition of the substrate alters the gut bacterial composition by providing competitive benefits or new niches for specific bacteria that can utilise these compounds. Moreover, specific microbial functions, such as cobalamin synthesis, appear to be correlated with larval performance. Aside from metabolic functions, biosynthetic gene cluster analysis revealed potential antimicrobial competition and protective functions among bacterial species. Improved taxonomic resolution provided by WGS led to the identification of several metagenome assembled genomes (MAGs), including a potentially novel BSFL-associated *Scrofimicrobium* species. Furthermore, there were differences in larval performance between rearing diets, and larval growth was correlated with high abundance of several MAGs.

**Conclusions** Variation in the nutritional and bacterial load of a diet can result in functional shifts in the gut microbiome of the larvae. Analysis of the BSFL metagenome identified several bacteria that are positively correlated

*Correspondence:
Jeroen De Smet
jeroen.desmet@kuleuven.be

Full list of author information is available at the end of the article

with larval performance, which could potentially provide beneficial metabolic functions for the host that should be further explored.

**Keywords** *Hermetia illucens*, Microbiome-derived functions, Metagenomics, Insect gut, Microbiota, Food waste

## Background

The relationship between microbiomes (i.e. the community of microorganisms and their genomes inhabiting a particular environment) and their hosts has been the subject of intensive study in recent years. Metagenomic analyses of human [1], animal [2, 3] and plant [4] microbiomes have identified various microbial genes that can fulfil diverse functions for their host. For example, metagenomic profiling of the gut microbiome of wood-feeding beetles has led to the identification of several microbial genes associated with metabolic functions, such as the production of lignin- and cellulose-degrading enzymes, and of detoxification enzymes [3]. Furthermore, it has been shown in many organisms, including humans, that factors such as diet significantly impact the gut microbiome, leading to beneficial or detrimental effects on an individual's health and/or performance [5].

Understanding functional shifts in the microbial gut community, and their impact on the host, can provide valuable information that can be exploited to boost either health or performance (e.g. through detoxification of harmful dietary compounds or the degradation of long-chain polymers into usable metabolites) depending on the organism being studied. Accordingly, the current study aims to generate new insights on the functions of the gut microbiota for black soldier fly larvae (BSFL) (*Hermetia illucens* L., Diptera: Stratiomyidae), a commercially produced farm insect, and the impact of these microbial functions on the larvae's performance. Due to its excellent nutritional profile [6], this insect species has obtained considerable attention over the last years as an attractive nutritional supplement to chicken- and pig feed [7], petfood [8] and feed used in the aquatic industry [9]. It also received interest for its capacity to convert a wide range of waste streams into valuable insect body mass [10, 11]. Broeckx et al. [11] demonstrated that larvae had a similar larval performance in terms of larval growth when reared on various grain-based (grain middling, beer draff) diets, fruit- and vegetable- (fruit puree, vegetable overproduction) diets and household food waste as larvae reared on a Gainesville- or chicken feed-control diet. Moreover, larvae reared on industrial food waste and chicken manure even outperformed larvae reared on regularly used diets [11].

Numerous studies suggest that the gut microbiome plays an important role in enabling BSFL to adapt to diverse nutritional environments. The microbiota supports the larvae's ability to survive and thrive on various organic waste streams [11], in part by encoding a range of functions [12–14], that are thought to complement the insect's immune system and metabolism [15]. An indication for this is that the gut microbiome community composition of BSFL is largely influenced by the rearing diet, although a "core" set of microorganisms persists across different feeding conditions [16–19]. However, limited taxonomic resolution, inconsistent definitions of the core microbiome, and a lack of functional data continue to raise questions about the true presence and role of this core community. The varying nutritional availability in the diets enables certain microbial species to flourish within the BSFL gut microbiome [20]. This microbial community flexibility could potentially lead to increased expression of certain microbial genes that provide metabolic functions which benefit the host, boosting the ability of the larvae to metabolise specific diets such as industrial food waste or chicken manure [11]. Therefore, identification of microbial functions present in larvae reared on different diets can provide valuable insights into the role of the gut microbiome on larval growth and bioconversion efficiency.

Although studies on the BSFL Gut microbiome are numerous, the identification of the BSFL Gut microbiome and its dynamics has so far mostly been studied through bacterial 16S ribosomal RNA (rRNA) gene amplicon sequencing from either dissected guts or whole larvae [19, 21]. While 16S rRNA gene amplicon sequencing offers a rapid and cost-effective means to profile entire bacterial communities, its significant Limitations hinder a deeper understanding of BSFL microbiome functionality. A key Limitation of 16S rRNA gene sequencing is that it targets only a small, specific region of the bacterial genome, offering no direct insight into the community's functional potential (unless inferred from taxonomic data) and entirely excluding other microbial kingdoms, such as fungi and viruses, which may also influence the host [22]. Additionally, taxonomic identification becomes less accurate at the species level [23], which can potentially lead to incorrect conclusions about bacterial presence and functionality. Alternatively, methods such as whole genome shotgun (WGS) sequencing can cover broad regions of both prokaryotic- and eukaryotic genomes, providing adequate information to predict genes and identify protein-coding regions [23], which makes this technique more suitable for identifying bacterial- and fungal functions. Although WGS has recently been applied in BSF research to explore specific microbial functions, such as plastic degradation [14], its broader use to comprehensively map the functional

landscape of the BSFL gut microbiome remains largely unexplored. Leveraging WGS in this way could uncover a wide array of metabolic pathways and microbial capabilities that may underpin the larvae's adaptability to diverse substrates, opening new avenues for targeted and in-depth experimental research. Therefore, the aim of the current study was to use WGS to investigate which microbial functions and microbiome-associated taxa change with a change in diet and hypothesize how these functions may affect larval performance. To achieve this, growth trials were conducted with larvae reared on two nutritionally distinct diets, chicken starter mash and an artificial supermarket food waste stream, followed by an in-depth metagenomic- and functional analysis of their gut microbiomes.

## Materials and methods

### Diet preparation

Two separate diets were prepared for this study. The first diet consisted of an artificial supermarket food waste (SFW) produced as described by Van Looveren et al. [24], by mixing various food items (bread (22.0 wt%), lettuce (13.0 wt%), apple (12.0 wt%), zucchini (12.0 wt%), cucumber (12.0 wt%), tomato (12.0 wt%), potato (9.0 wt%), kidney beans (6.0 wt%), cheese (1.5 wt%) and yoghurt (0.5 wt%)). The diet approaches the composition of supermarket food waste, which is a considerable source of global food waste that can be utilised as a nutritious rearing diet for BSFL production [24]. The SFW diet was prepared separately for each trial from fresh ingredients. The second diet consisted of a chicken feed (CF, Chicken Start Mash 259, AVEVE, Belgium) diet which is commonly used as a rearing diet for BSFL. All replicates of the CF diet originated from the same feed stock. The SFW dry matter (DM) concentration of the SFW diet was 24.4% and the DM of the CF diet at water binding capacity was 38.1%.

### Chemical analysis of diets

For both diets, the sugar and amino acid composition were determined. The SFW diet was freeze-dried for 48 h at −55 °C and 200 mbar (Büchi, Lyovapor L-200) until a DM concentration of at least 90% was obtained. Since CF already met the desired DM concentration, no additional freeze-drying step was required for this diet. The DM concentration of the freeze-dried SFW diet and the CF diet was measured gravimetrically, by weighing 3–5 g of sample material before and after it was dried at 105 °C for 17 h in a forced-air oven (Practum 224–1s, Sartorius). The SFW and CF samples were defatted by adding technical grade petroleum ether (VWR Internation BVBA, Belgium), shaking vigorously, centrifuging for 5 min at 1,000 x g and decantation of the dissolved fat. This procedure was repeated until a clear solution indicated that the samples were defatted. The final DM concentration of the defatted samples was determined gravimetrically and these samples ($n = 3$ for both substrates) were used in the amino acid- and sugar- analyses.

The amino acid content was determined using ultra-performance liquid chromatography – mass spectrometry (UPLC-MS) after acidic (HCl) and basic (LiOH) hydrolysis and derivatisation with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) as described in [25], with a small adaptation to the described UPLC elusion: the mobile phase consisted of ultrapure water + 0.1% formic acid (A) and acetonitrile + 0.1% formic acid (B) and the following gradient was used: 1% Bat 0–2 min, 1–13% Bat 2–4 min, 13–35% Bat 4–8 min, 35–95% Bat 8–9 min, 95% Bfrom 9–9.5 min, 95–1% Bat 9.5–9.6 min, 1% Buntil 12 min. To determine the sugar concentration, 50.0 mg of each sample was extracted twice in 600 µL 80% ethanol. After centrifugation (10 min at 15 000 g), the supernatant was diluted to 10% ethanol and passed over an anion exchange column (150 µL Dowex 1 × 8, $Ac^-$) and a cation exchange column (150 µL Dowex 50Wx8, $H^+$) before being analysed by HPAEC-IPAD (ICS-5000, Thermo Fisher Scientific; PA-100 and MA-1 columns, with pre-column). Quantification of the carbohydrates was performed using arabinose, fructose, galactose, glucose, maltose, maltotriose, mannose, raffinose, rhamnose, ribose, ribulose, sorbitol, sorbose, stachyose, sucrose and tagatose as reference compounds.

### Rearing experiment BSFL

*Hermetia illucens* larvae were collected from a stock colony maintained by the Sustainable Biomass and Chemistry group of the Thomas More University of Applied Sciences (Geel, Belgium). After a nursing period on CF (for details, see [26]) eight days old larvae were used to initiate the experiment. Two separate trials of the rearing experiment were done under the same experimental conditions. In each trial, the larvae were reared on a freshly prepared batch of the SFW diet, or were continued rearing on the CF diet from the nursery phase. At the beginning of each rearing trial, six groups of 500 eight days old larvae (three replicates for both the CF and SFW diet) were manually counted and their total start weight measured. Three of these groups were then placed in plastic containers with the SFW diet, while the remaining three other groups were placed in plastic containers with the CF diet). To each container 100 g DM of diet was added. Substrates were covered with aluminium foil to prevent moisture evaporation during the experiment. Container replicates were kept in a climatic chamber at 27 °C and 60% RH for eight days.

### Sample preparation and DNA sequencing

Following incubation, the larvae were separated from their substrates, rinsed with running tap water, washed with 70% ethanol and air-dried on paper towels. Next, five groups of ten randomly selected larvae from each replicate container were weighed to determine the final larval mass. Cleaned larvae were sedated on ice for several minutes until the insect reached a comatose state and moved to a sterilised laminar flow cabinet, where the larvae were slaughtered through decapitation and their complete gut system was extracted using sterilised scissors and tweezers in sterile PBS-buffer [20]. Five larval guts from each replicate container were pooled per diet and stored at −20 °C until DNA extraction, yielding three independent samples per condition. For this, the pooled larval gut samples were mechanically crushed using a sterilised metal rod after which the E.Z.N.A. ® Soil DNA Kit (Omega Bio-Tek, Norcross, GA, USA) was used, following the manufacturer's instructions. The quality and quantity of the DNA samples was checked using a Nanodrop device (ND-1000, Isogen Life Science, Utrecht, the Netherlands) and the samples were sent to Novogene Bioinformatics Technology Co., Ltd. for WGS sequencing. DNA Quantitation, integrity and purity was evaluated with the Agilent 5400 Fragment Analyzer System (Agilent Technologies, Santa Clara, CA, United States). Genomic DNA was randomly Sheered into short fragments, end repaired, A-tailed and ligated with Illumina adapters before PCR amplification, size selection and purification. Quantified Libraries were pooled and sequenced on the Illumina Novoseq 6000 platform using 150 bp paired-end sequencing, yielding 67.2 G (raw reads x sequence length) of raw data.

### Metagenome quality filtering and annotation

Adapters, PhiX control signals and reads mapping to the host genome (NCBI RefSeq assembly: GCF_905115235.1) were removed using inhouse scripts. The remaining metagenomic sequences of all samples were pooled together for a co-assembly using Megahit [27] following the SqueezeMeta workflow [28]. Short contigs (< 200 bps) were removed using prinseq [29], which was also used to perform contig statistics. Open reading frames (ORFs) were predicted using Prodigal [30]. The original reads were mapped to the contigs resulting from the assembly to estimate the abundance of the genes and contigs in each sample [28] using Bowtie2 [31]. ORFs that did not have reads in all three biological replicates of each diet were discarded. Additionally, ORFs that had a total number of reads over all samples of less than 20 were removed, resulting in 194 417 ORFs.

### Taxonomic analysis of 16S rRNA gene sequences extracted from the metagenome

RNAs were predicted using Barrnap and additionally tRNA/tmRNA sequences were predicted using Aragorn [32]. A subset of 151 ORFs annotated as 16S rRNA gene sequences were taxonomically classified with a BLAST search using the rRNA/ITS databases in GenBank against exemplar species or 'type materials' [33]. Identification was based on the highest max score, the lowest E-value and a similarity percentage of over 97%. When these criteria were not met at species level, the top 100 hits received for that sequence were evaluated and the (higher level) taxonomic annotation shared among them was selected. ORFs that could not be identified by the NCBI BLAST search on any taxonomic level were removed from the analysis.

Downstream analysis of the extracted 16S rRNA gene sequences and ORFs sequences was performed using the *DESeq2* package [34] in R version 4.3.0 [35]. Filtered reads were normalised through an estimation of size factors, estimation of dispersion and fitting of a negative binomial generalised linear model (GLM) [34] using the *DESeq* and *counts* functions. To compare the taxonomic classification of the bacteria in the metagenomes of the 16S rRNA gene sequences and the classification annotated by the ORFs sequences, Krona graphs based on the normalised bacterial reads were produced using KronaTools. Relative abundance reads were normalised across all trials and used in a permutational analysis of variance (PERMANOVA) of the beta diversity estimates based on Bray-Curtis distances, and visualised through a Principal Coordinates Analysis (PCoA) using the *vegan* and *phyloseq* [36] packages. For the next analysis the two trials were analysed separately. Reads were normalised per trial, aggregated on the family-level and the *DESeq* function was used to calculate log2fold changes between the SFW-diet-reared larval metagenome (SFW-LM) and the CF-diet-reared larval metagenome (CF-LM). Significant differences were determined using a Wald test, where families were considered significantly different if the Benjamini and Hochberg [37] adjusted p-values were lower than 0.05 and the log2fold changes were larger than 1. The results of this analysis were visualised for both trials in a dot plot using the *ggplot2* package in R.

### Taxonomic and functional analysis of the metagenome

In order to provide taxonomic and functional annotation for each ORF, the ORF sequences were compared against several taxonomic and functional databases [28] using Diamond [38]. Specifically, taxonomic annotation was assigned through similarity search against the GenBank nr database [39], whereas sequences were compared to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [40] for functional KEGG-ID annotation.

Pathway prediction for KEGG was done using MinPath [41]. ORFs sequences that did not provide a match with these databases and therefore yielded no functional or taxonomical annotation were removed from this analysis. The remaining ORFs were used to compare the differences in taxonomical and functional composition of the larval metagenomes per diet. Each analysis and normalisation of the reads was done separately per trial. First, the unique combinations of each taxonomic- and functional (KEGG)- annotation were used to create new groups. Then, the ORF reads were aggregated for each of these unique group, log2fold changes were calculated between the normalised reads per diet and significant differences were determined using a Wald test from the *DESeq* function. Taxonomic- and functional- groups were considered significantly different if the Benjamini and Hochberg adjusted p-values were lower than 1e-5 and the log2fold changes were larger than 1. The results of the tests for all samples were combined per trial and visualised in Voronoi Treemaps using the *WeightedTreemaps* package in R.

To study which functions were potentially more enriched in the CF- and SFW- reared larval metagenomes, an additional analysis was done in which the normalised reads per trial were aggregated per unique KEGG annotated function, omitting all taxonomic data. The differences between unique KEGG functions were analysed as described earlier, i.e. using the *DESeq* function to determine log2foldchanges larger than 1 and a Wald test with adjusted p-values lower than 1e-5. The results of these tests were visualised in a volcano plot and the number of significant metabolic functions per diet type were visualised in a heatmap, using the *ggplot* and *ComplexHeatmap* [42] packages, respectively. Additionally, a *DESeq* analysis was performed per substrate type to identify the differences in KEGG enrichment between the two trials. The KEGG pathways that showed the largest differences in enriched microbial functions between the two trials were visualised in a bar graph. Details of important KEGG pathways were visualised using BioRender.

#### Genomic reconstruction of the metagenome

Binning (grouping of contigs and assigning them to genomes) was done using Metabat2 [43] and CONCOCT [44], and combination of the best binning results was done using DAS Tool [45]. Bin statistics were computed using CheckM [46]. Bins were considered of good quality when completeness (presence of marker genes) was >75% and contamination (fraction of marker genes duplicated) <10%. Good quality bins or Metagenome Assembled Genomes (MAGs) were classified using the *classify_wf* function from the Genome Database Taxonomy toolkit (GTDB-Tk), which determines the classification of the genome based on its average nucleotide identity (ANI) similarities, the alignment fraction (AF) and the relative evolutionary divergence (RED) of the genome compared to the GTDB-TK reference tree [47]. Additionally, a phylogenetic tree of the MAGs was constructed based on genomic features annotated by Prokka [48], gene clustering by PIRATE [49] inferred using RAxML (v8) with the GTRGAMMA substitution model and 100 rapid bootstrap replicates [50] and visualised using iTol [51]. The total number of reads were aggregated for each MAG and differences in MAG abundance between the different trials and substrates were determined using a Wald test from the *DESeq* function, where MAG abundances were considered significantly different if the Benjamini and Hochberg adjusted p-values were lower than 1e-5 and the log2fold changes were larger than 1. After this comparison, a second generalised linear model was fitted using the *DESeq* function to identify the effects of the individual factors "Trial" and "Substrate" and their interaction effect on the abundance of the MAGs.

#### Statistical analysis of the larval final weight

The differences in larval final weight across the different diets and trials were calculated in JMP Pro 17 [52]. Normality and homogeneity of variances of the final larval weight were checked using the Shapiro-Wilk and Levene's test for equality of variances, respectively. Since the homogeneity of variances could not be assumed, the non-parametric Kruskal-Wallis rank sum test followed by a Steel-Dwass multiple comparison analysis was used. The results of the larval weight analysis were visualised using the *ggplot* package in R. Furthermore, a linear regression model was used to predict correlations between the MAGs normalised reads transformed by $\log10(x+1)$ and the mean larval weight across all diets and trials.

#### Identification of bio-synthetic gene clusters

Secondary metabolite gene clusters were identified, annotated and analysed using antiSMASH [53]. Gene clusters were compared against antiSMASH-predicted clusters and gene clusters from the Minimum Information about a Biosynthtic Gene cluster (MIBiG) [54] database. The Biosynthetic Gene Similarity Clustering and Prospecting Engine (BiG-SCAPE) software package was used to construct similarity networks of Biosynthetic Gene Clusters (BGCs) and to group these clusters into Gene Cluster Families (GCFs), using a raw distance cut-off value of 0.70 [55]. GCFs with two or less BGCs were removed. Of the remaining GCFs, the reads of the contigs included in each BGC were aggregated per diet. For each separate trial, reads were normalised and the variation in the relative abundances of the BGCs between the diets was analysed using a Wald test. BGCs were considered significantly more present in the microbiome of

**Table 1** Amino acid concentration of the two substrate types. The table shows the average concentration of 18 amino acids (left column) per substrate type. The mean concentration of each AA in gram per 100 gram dried substrate sample and the coefficient of variance in percentage is provided, where $n=2$ for the chicken feed (CF) substrate and $n=3$ for the supermarket feed waste (SFW) substrate

| | CF substrate | | SFW substrate | |
|---|---|---|---|---|
| **Amino acid** | **Mean** | **CV** | **Mean** | **CV** |
| Histidine | 1.87 | 16.90 | 0.67 | 3.90 |
| Arginine | 4.74 | 0.50 | 1.25 | 5.60 |
| Serine | 3.29 | 2.20 | 1.74 | 4.80 |
| Glycine | 3.01 | 5.60 | 1.18 | 7.90 |
| Aspartic acid | 2.44 | 28.80 | 1.74 | 7.60 |
| Glutamic acid | 5.95 | 17.90 | 6.49 | 4.40 |
| Threonine | 2.75 | 2.00 | 1.10 | 5.10 |
| Alanine | 2.15 | 8.00 | 0.99 | 5.90 |
| Proline | 2.84 | 12.30 | 2.53 | 0.70 |
| Lysine | 1.64 | 47.20 | 0.90 | 3.20 |
| Cystine | 0.77 | 13.60 | 0.24 | 10.30 |
| Tyrosine | 2.54 | 9.30 | 1.03 | 7.50 |
| Methionine | 1.10 | 0.80 | 0.12 | 36.10 |
| Valine | 2.44 | 7.10 | 1.32 | 1.60 |
| Isoleucine | 2.13 | 7.90 | 1.08 | 1.70 |
| Leucine | 3.87 | 7.20 | 2.02 | 2.40 |
| Phenylalanine | 4.33 | 13.60 | 1.71 | 5.60 |
| Tryptophan | 1.15 | 24.80 | 0.67 | 14.00 |

BSFL reared on a specific diet when the adjusted p-values were lower than 0.05 and the log2fold changes were larger than 1, and when the results of these test were the same in both trials. The results of the clustering and the statistical tests were visualised using Cytoscape [56].

## Results

### Sugar and amino acid composition of the diets

To investigate the microbial functions related to sugar and amino acid metabolism identified in the metagenome, the specific sugar and amino acid composition of both the CF and SFW diets was analysed. Since one of the three CF replicates had amino acid concentrations twice as high as the other replicates, this sample was considered an outlier and discarded from the analysis. Given that the larvae were provided with 100 g of diet on a dry matter basis, the subsequent sugar and amino acid concentrations are Likewise expressed per 100 g DM, thereby representing the quantities effectively supplemented. Results show that the CF diet contained higher levels of most amino acids compared to the SFW diet, with the exception of glutamic acid and proline, which had comparable concentrations in both diets (Table 1). Specifically, valine, leucine and isoleucine were present in higher quantities (2.44 g/100 g DM, 3.87 g/100 g DM and 2.13 g/100 g DM, respectively) in the CF diet compared to the SFW diet (1.32 g/100 g DM, 2.02 g/100 g DM and 1.08 g/100 g DM, respectively).

**Table 2** Sugar concentration of the two substrate types. The table shows the average concentration of 16 sugars (left column) per substrate type. The mean concentration of each sugar in gram per 100 gram dried substrate sample and the coefficient of variance in percentage is provided, Where $n=3$ for both the chicken feed (CF) and for the supermarket feed waste (SFW) substrates. Sugars with N.D. Where not detected in the samples

| | CF substrate | | SFW substrate | |
|---|---|---|---|---|
| **Sugar** | **Mean** | **CV** | **Mean** | **CV** |
| Sorbitol | 0.04 | 14.1 | 0.17 | 3.9 |
| Glucose | 0.11 | 9.1 | 3.90 | 11.6 |
| Fructose | 0.10 | 10.0 | 5.40 | 9.1 |
| Sucrose | 2.86 | 7.6 | N.D. | N.D. |
| Raffinose | 0.69 | 10.4 | N.D. | N.D. |
| Stachyose | 1.15 | 4.8 | N.D. | N.D. |
| Maltose | 0.10 | 14.4 | 2.16 | 7.7 |
| Maltotriose | N.D. | N.D. | 0.39 | 8.5 |
| Galactose | N.D. | N.D. | 0.02 | 6.9 |
| Arabinose | N.D. | N.D. | N.D. | N.D. |
| Ribulose | N.D. | N.D. | N.D. | N.D. |
| Ribose | N.D. | N.D. | N.D. | N.D. |
| Tagatose | N.D. | N.D. | N.D. | N.D. |
| Thamnose | N.D. | N.D. | N.D. | N.D. |
| Mannose | N.D. | N.D. | N.D. | N.D. |
| Sorbose | N.D. | N.D. | N.D. | N.D. |

Regarding sugar composition, the CF diet consisted mainly of sucrose (2.86 g/100 g DM) followed by stachyose (1.15 g/100 g DM) and raffinose (0.69 g/100 g DM). The SFW diet contained mostly fructose (5.40 g/100 g DM), followed by glucose (3.90 g/100 g DM) and maltose (2.16 g/100 g DM). Interestingly, the SFW diet contained more than four times the sorbitol concentration as compared to CF (Table 2).

### Effect of the diets on larval weight

The larvae reared on the CF diet reached a significantly higher final weight compared to the larvae reared on the SFW diet ($p<0.0001$ for trial 1, $p=0.0011$ for trial 2). Moreover, there was a difference in final larval weight between the two trials per diet, where larvae from the second trial reached a significant higher larval weight compared to larvae from the first trial for both the CF ($p=0.0123$) and SFW ($p=0.0011$) diet (Fig. 1).

### Taxonomic diversity of the BSFL gut Microbiome

The taxonomic diversity of the bacterial communities from BSFL reared on the SFW and CF diets from trial 1 and 2 was assessed using the 16S rRNA gene sequences extracted from the metagenome data set. A total of 331 601 normalised reads across 129 16S rRNA gene sequences and 180 217 normalised reads across 120 16S rRNA gene sequences were identified in trial 1 and trial

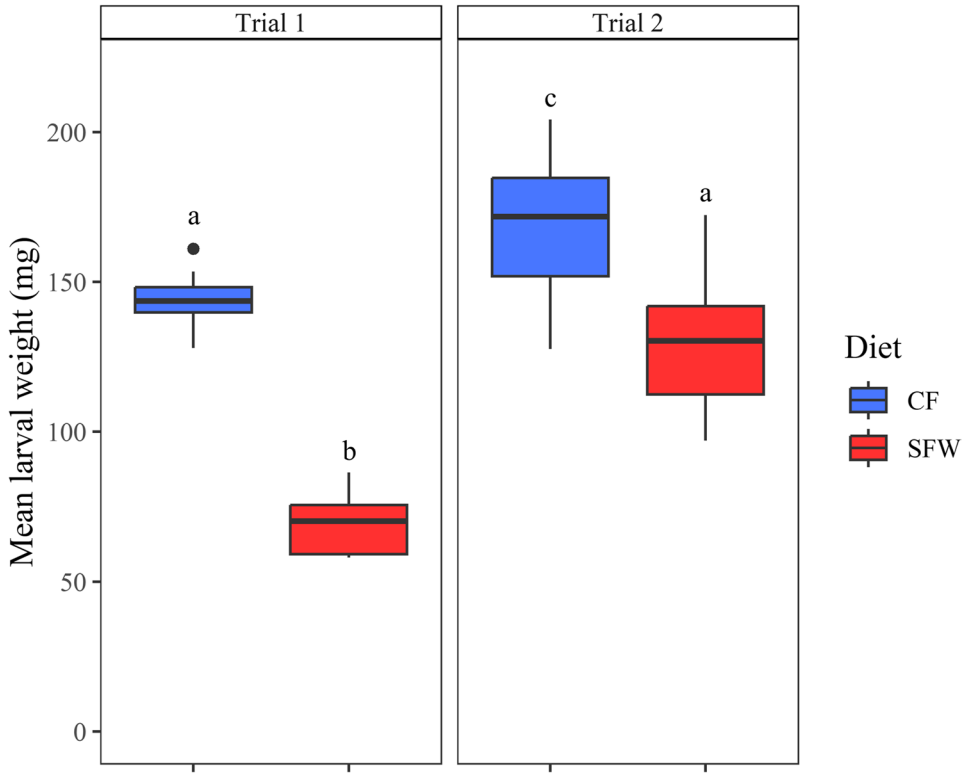


**Fig. 1** Box plot representing the larval final weight. Larval final weight (mg) was determined by randomly weighing 5 × 10 larvae from each replicate at the final day of the experiment. The average weight was determined per trial and rearing diet (chicken feed (CW). blue; supermarket food waste (SFW). red) and the differences between groups were determined using a non-parametric Kruskal-wallis rank sum test followed by a Steel-Dwass multiple comparison analysis. Different annotation letters indicate that the larval final weight was significantly ($p < 0.05$) different between treatments

2, respectively. The bacterial community composition of the larvae was more similar when the larvae were reared on the same diet as exemplified by the separation of clusters between the different diet samples on the first axis in both trials (Fig. 2A). Moreover, the second axis also showed differences between the two trials within the larval groups reared on the same diet. The PERMANOVA results revealed that both the difference in diet type ($p = 0.016$) and the variation between trials ($p < 0.001$) had a significant effect on the bacterial composition in the larval gut samples. The SFW-LM consisted largely of members of the Pseudomonadota and Bacillota phyla in near equal ratios, whereas the CF-LM was dominated by members of the Actinomycetota phylum (Fig. 2B). Specifically, members of the Actinomycetes class, including Micrococcales, Brevibacteriaceae, Corynebacteriaceae and Microbacteriaceae as well as several Bacillota members such as Bacillaceae and Staphylococcaceae, were significantly more present in CF-LM in both trials (Fig. 2C). In the SFM-LM, taxonomic groups belonging to the Pseudomonadota phylum such as Alphaproteobacteria, Enterobacterales and Enterobacteriaceae, and members of the Bacillota family Lactobacillaceae, were more enriched in both trials.

Similar as seen in the 16S rRNA gene sequences subset, taxonomical differences between the BSFL gut microbiome from different diets and trials were observed when analysing the other ORFs extracted from the metagenome. To first evaluate whether the taxonomic identification of the ORFs sequences was in Line with the taxonomy based on the 16S rRNA gene sequences, Krona graphs of the average bacterial composition of the CF-LM and SFW-LM from trial 1 and 2 were compared (Supplementary Figure S1). On the higher taxonomic levels (phyla and class), the taxonomic annotation of the bacterial composition of the larvae was similar between the 16S rRNA gene sequences and ORFs sequences. On the lower taxonomic levels, the ORFs sequences provided a better taxonomic resolution of the Actinomycetota and Bacillota phyla, whereas the 16S rRNA gene sequences provided a better resolution of the Pseudomonadota phylum. Considering the ORFs that contained both taxonomical and functional information, the BSFL metagenome of trial 1 (Fig. 3A) consisted of 87 317 792 normalised reads mapping to 53 337 unique ORFs, of which 12 711 and 15 342 were more enriched in the SFW-LM (red) and CF-LM (blue), respectively. In trial 2, 101 532 919 normalised reads were obtained that mapped to 47 480 unique ORFs, of which 4 543 and 18

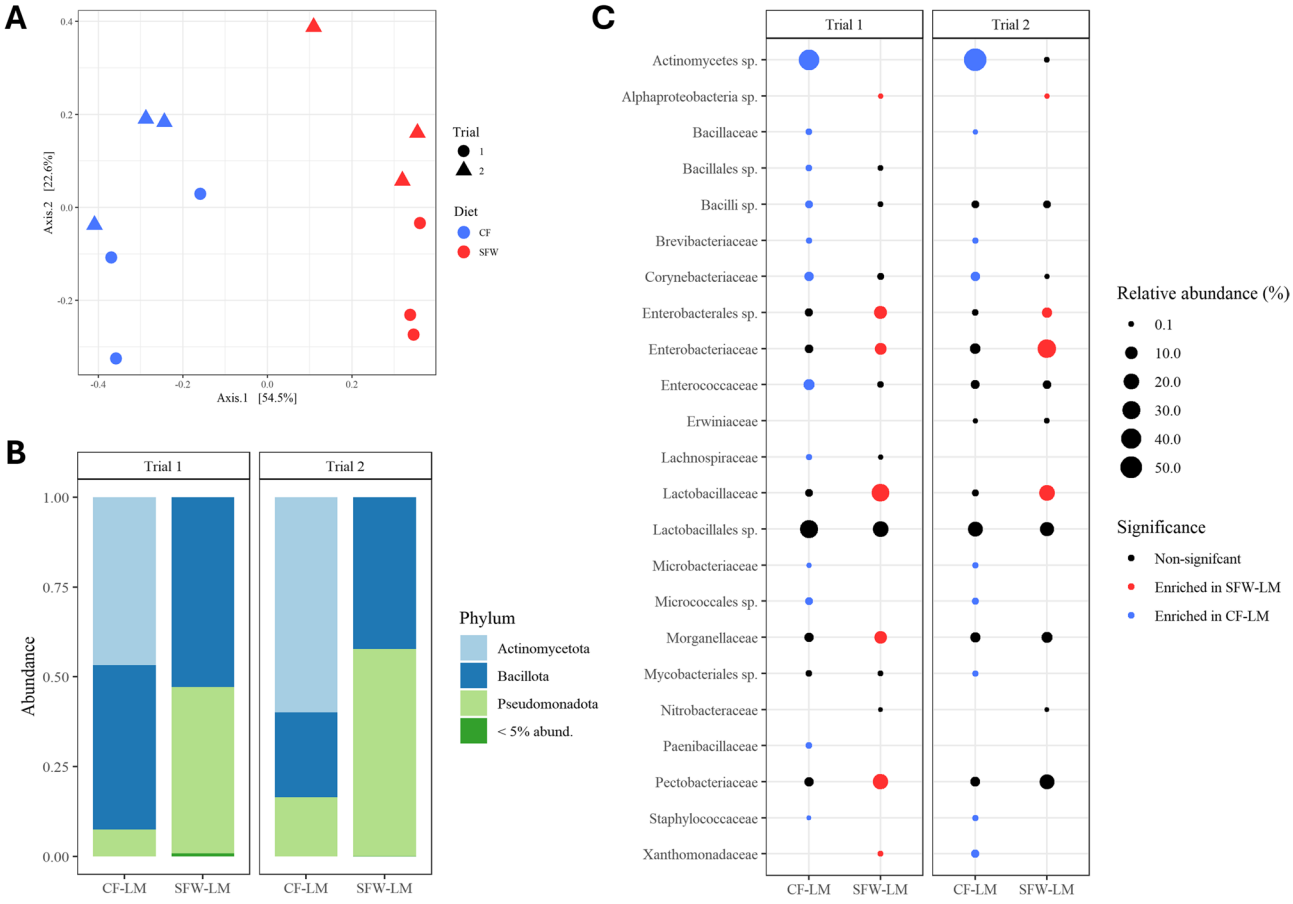


**Fig. 2** Ordination-. bar- and dot plot of 16S rRNA gene sequences. All plots are based on the normalised 16S rRNA gene sequencing reads extracted from the metagenome. Panel **A** shows the results of the Principal Component Analysis (PCoA) in an ordination plot. The plot shows the effect of the factors "Trial" (represented by the shapes) and "Diet" (represented by the colours) on the clustering of the normalised reads from the larval gut samples. where CF (blue) stands for the chicken feed substrate and SFW (red) for the supermarket food waste diet. Panel **B** shows the relative abundance of the microbiota community based on the normalised reads of the larval gut samples averaged (n = 3 for all treatments) per trial and per diet. where CF-LM stands for the larval microbiome reared on CF and SFW-LM for the larval microbiome reared on SFW. Panel **C** shows a dot plot where each dot represents the presence of the bacterial taxonomic group shown on the y-axis found in CF-LM and SFW-LM. The size of the plot represents the averaged (n = 3 for all treatments) relative abundance or normalised reads in the sample whereas the colour of the dot indicates whether this taxonomic group was significantly more enriched in CF-LM (blue) or SFW-LM (red)

293 were more enriched in the SFW-LM and CF-LM, respectively (Fig. 3B). When comparing the ORFs that were enriched in either the CF-LM or SFW-LM in both trials 1 and 2 (Supplementary Figure S2A), a total of 8 841 and 3 935 ORFs were significantly enriched in CF-LM and SFW-LM in both trials, respectively. In both trials, ORFs taxonomically delineated as Actinomycetota were predominantly present in the CF-LM (Fig. 3). Members of the phyla Bacillota and Pseudomonadota were present in the microbiomes of BSFL reared on both diets, although the abundance and prevalence of the taxonomic subgroups varied between trials. Most notably, although the relative abundance of Pseudomonadota was lower in trial 1 (25.3%) compared to trial 2 (31.6%), this phylum was largely dominated by *Providencia* (77.4% of the Pseudomonadota reads) in trial 1, while this genus was practically absent in trial 2, where *Klebsiella* was the most dominant (73.9% of the Pseudomonadota reads). Additionally, apart from bacteria, over 20% of the BSFL gut microbiome in this study consisted of fungi from the Ascomycota phylum, which was largely made up of *Pichia* spp. in the microbiome of BSFL from both diet types. Furthermore, only a small proportion of the microbiome (< 2%) consisted of phyla from other kingdoms (Fig. 3).

### Functional diversity of the BSFL gut Microbiome

To gain a better understanding which bacterial functions contribute to BSFL growth in response to the diet, the occurrence and abundance of each unique KEGG-annotated function was assessed by *DESeq* analysis. There were 311 and 365 KEGG functions significantly enriched in the SFW-LM and 413 and 548 KEGG functions significantly enriched in the CF-LM, in trial 1 and

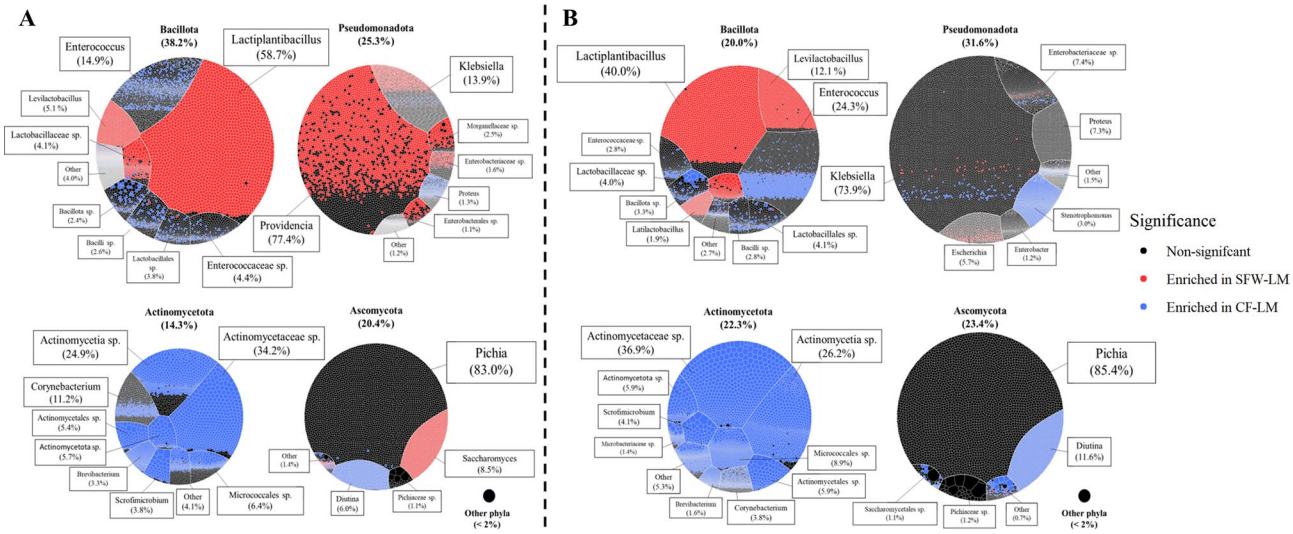


**Fig. 3** Taxonomic composition of the black soldier fly metagenome. Circles represent the relative abundances of the filtered and normalised ORFs reads taxonomically annotated per phylum for trial 1 (panel **A**) and trial 2 (panel **B**). The circles are a combination of the samples from both diets ($n = 6$ for both trials) and are further subdivided in lower taxonomic groups. Each tile represents a unique KEGG-annotated function within that taxonomic group. The size of the tile represents the average relative abundance of reads assigned to that unique group per genera across all six samples and the colour indicates whether that group was significantly ($p < 1e\text{-}5$. log2foldchange > 1) more enriched in the microbiome of either supermarket food (SFW-LM) - or chicken feed (CF-LM) fed larvae

2, respectively (Fig. 4A and B). Most of these enriched KEGG functions were associated with amino acid metabolism ($n = 140$ for trial 1, $n = 153$ for trial 2), carbohydrate metabolism ($n = 177$ for trial 1, $n = 260$ for trial 2) and energy metabolism ($n = 78$ for trial 1, $n = 90$ for trial 2) (Fig. 4C), indicating that dietary changes strongly influenced these metabolic pathways.

In both rearing diets, larvae from the second trial reached a higher final weight than larvae from the first trial. Since the diets used in trials 1 and 2 consisted of the same ingredients and were prepared in the same way, their nutritional values were considered comparable, and the rearing conditions were identical. Therefore, the remaining variable was the abundance of specific KEGG functions, which was analysed using *DESeq* to investigate potential correlations with larval weight. There were 259 and 452 KEGG functions significantly enriched in the larvae from the second trial reared on the CF and SFW diet, respectively. Figure 5A shows a selection of metabolic pathways were, on either substrate, three or more KEGG functions were enriched in the second trial compared to the first trial. Most noticeably, larvae reared on SFW in trial 2 contained a large ($n = 17$) number of enriched KEGG functions compared to larvae reared on SFW in trial 1 ($n = 5$), which were involved in the porphyrin and chlorophyll metabolism pathway. Specifically, $n = 11$ of these KEGG functions were involved in cobalamin biosynthesis (Fig. 5B). Other metabolic pathways that showed a significant difference in enriched KEGG function between the SFW- and CF fed larvae involved sorbitol- and riboflavin metabolism (Supplementary Figures S3 and S4, respectively).

### Metagenome assembled genomes (MAGs)

Binning of the BSFL metagenomic reads yielded a total of 94 bins, of which 26 met the criteria of good quality bins and were classified as MAGs. However, one of these MAGs had less than 20 reads across all samples and was therefore removed during the filtering step. The length, completeness and contamination percentages of the remaining 25 MAGs and their delineated taxonomy is shown in Table 3A and their phylogeny is visualised in Fig. 6. Three MAGs could not be reliable classified based on ANI similarities and the AF and were annotated on the genus level only. The abundances of the MAGs were influenced by the factors diet, trial or an interaction effect (Table 3B) and consequently there were significant differences between the four different experimental groups (Supplementary Figure S5). For five MAGs, the number of reads was only significantly ($p < 1e\text{-}5$) influenced by diet, while there was no significant effect of differences between trials or the interaction between trial and diet (Table 3B). Eight other MAGs were significantly influenced by the differences between trials, of which three MAGs were solely influenced by this inter-trial effect. The influences of these factors were clearly visible when the abundances of the MAGs were plotted against the mean larval weight in a regression analysis (Supplementary Figure S6). For twelve MAGs, a significant ($p < 0.05$) positive ($n = 4$) or negative ($n = 8$) correlation between MAG abundance and mean larval weight

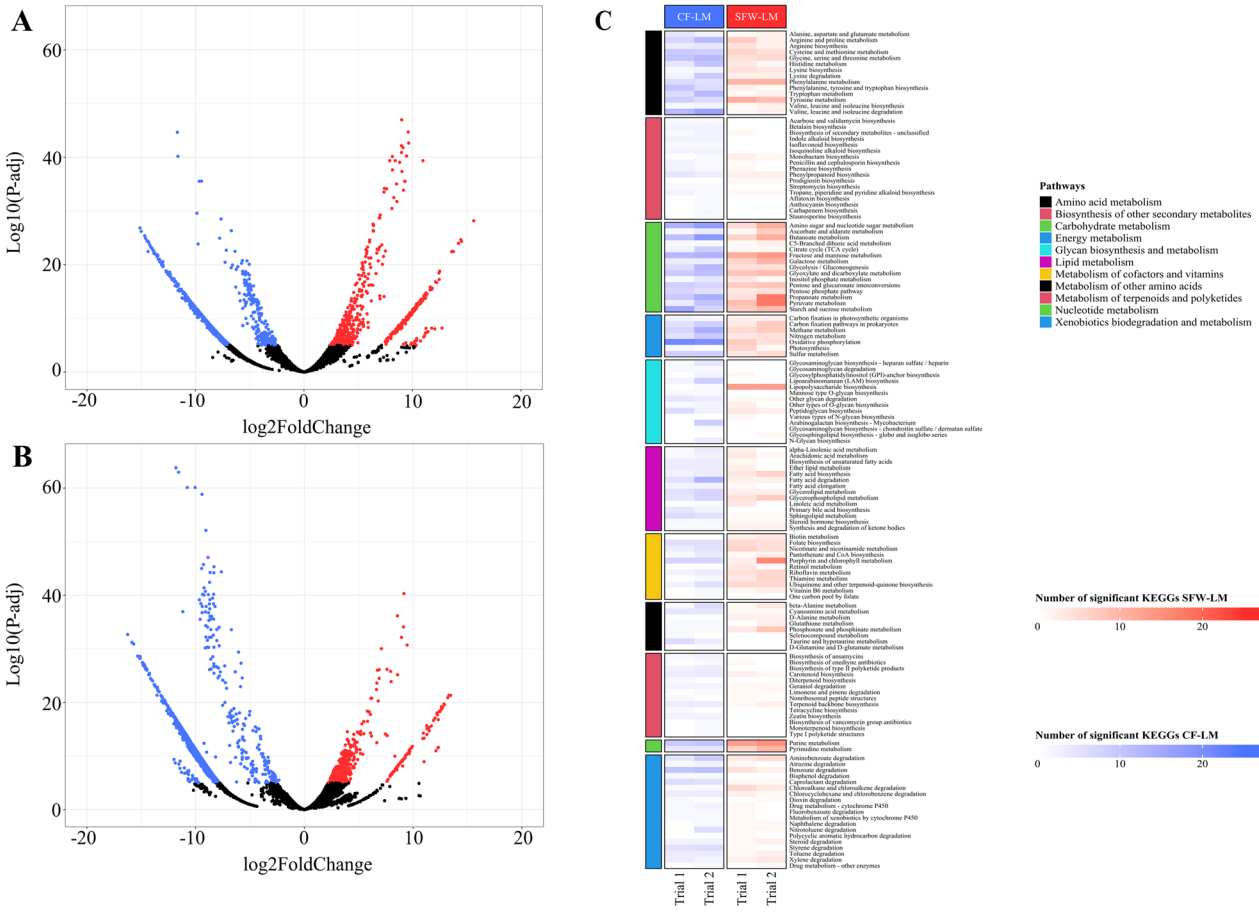


**Fig. 4** Occurrence and significance of KEGG functions from the black soldier fly metagenome. In panel **A** (results of trial 1) and panel **B** (results of trial 2). each unique metabolic KEGG function is represented as a dot in a volcano plot. where the colour indicates whether this KEGG function is significantly more present in the metagenome of BSFL reared on either chicken feed (CF-LM. blue) or supermarket food waste (SFW-LM. red) diet. or not significantly enriched in either of the metagenomes (black). In panel **C** the significant KEGGs from panel **A** and panel **B** are counted per functional pathway and visualised in a heatmap. An increasing intensity in the colouring indicates that more significant KEGG functions were found per specific pathway in the CF-LM (blue) or SFW-LM (red)

was found. The effects of the factor "diet" on MAG abundance is illustrated by the three MAGs, taxonomically delineated as *Scrofimicrobium* sp. ($p = 0.003$), *Latilatcobacillus curvatus* ($p < 0.001$) and *Levilactobacillus brevis* ($p < 0.001$), where the sample cluster heavily based on the larval diets. In contrast, the regression analyses of the MAGs taxonomically delineated as *Providencia stuartti* ($p = 0.019$), *Providencia rettgeri* ($p = 0.004$) and *Mammaliicoccus sciuri* ($p = 0.039$) reveal strong clustering based on the differences between trials.

### Biosynthetic gene clusters identified in the BSFL gut Microbiome

The results of the antiSMASH and BiGSCAPE BGC analysis of the entire metagenome dataset (Fig. 7) revealed that after filtering, a total of 39 BGCs were identified across 16 distinct gene cluster families (GCF). Two GCFs encoding non-ribosomal peptide synthetases (NRPS), one GCF encoding polyketide synthase (PKS) and two GCFs categorised as "Other", contained BGCs that were significantly enriched in the CF-LM, whereas one GCF encoding Ribosomally Synthesised and Post-translationally Modified Peptides (RiPPs) and two GCFs of the "Other" category, contained BGCs that were significantly enriched in the SFW-LM. In both the CF-LM and SFW-LM, GCFs were identified that encode for the production of antimicrobial peptides (cluster 1,2 and 3), as well as GCFs that provide protective functions for the bacteria through the production of aryl polyenes (clusters 4 and 5) or siderophores (cluster 6).

## Discussion

### Study scope, limitations, and the need for experimental validation

This study aimed to explore diet-dependent shifts in the taxonomic and functional diversity of the BSFL gut microbiome and identify microbial traits potentially linked to larval weight using WGS. It is important to

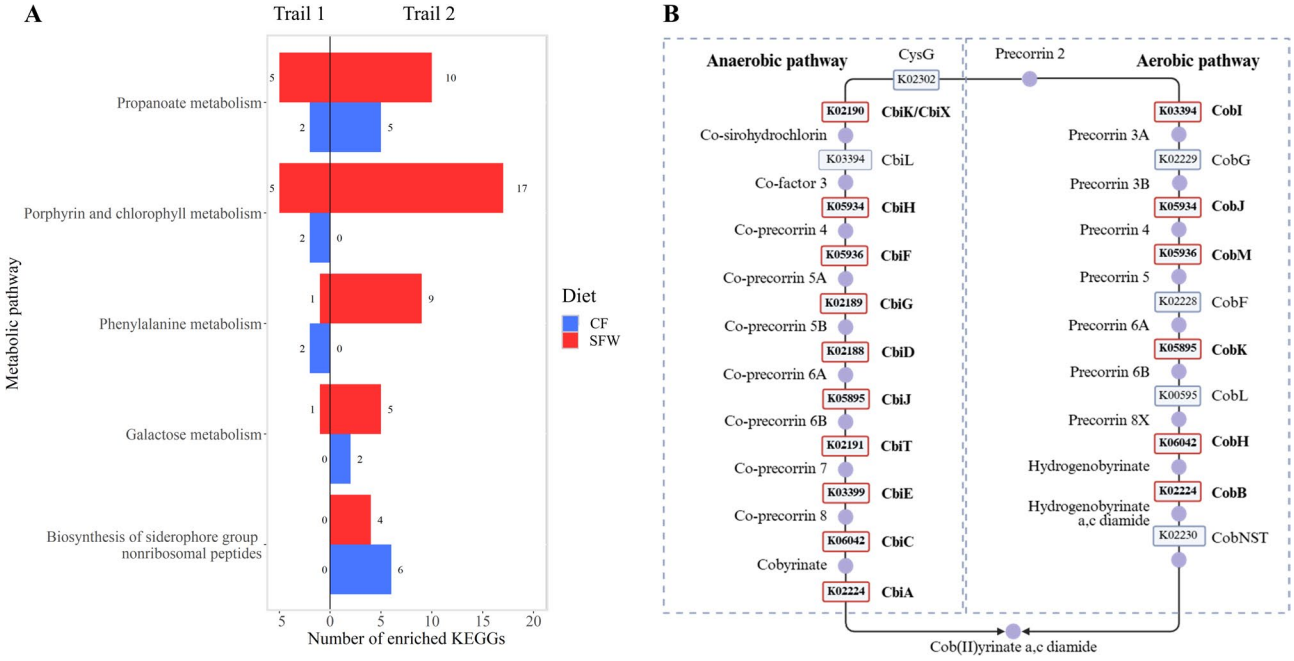


**Fig. 5** KEGG pathways with enriched bacterial functions in the second trial. Panel **A** shows the number of enriched KEGGs on either the chicken feed (CF, blue) or supermarket food waste (SFW, red) diet in the first (left) and the second (right) trial, for six pathways were the differences between the two trials was the largest. Panel **B** shows the part of the porphyrin metabolism pathway that involves aerobic- and anaerobic cobalamin synthesis, adapted from the KEGG pathway map number 00860 [40]. KEGG numbers in bold with a red border represent functions that were significantly enriched in the second trial of larvae reared on the SFW diet compared to larvae reared on a similar diet during the first trial

recognise that BSFL development is influenced by both the nutritional composition of the diet and the functional contributions of its associated microbiome. In our study, larvae reared on the SFW substrate consistently exhibited lower final body weights compared to those reared on the CF substrate. This difference is likely attributable, at least in part, to the nutritional disparity between the two different diets. For instance, amino acid profiling revealed that most amino acids were present at lower concentrations in the SFW diet. Specifically, the methionine concentration was significantly lower in SFW (0.12 g/100 g DM) compared to the CF diet (1.10 g/100 g DM) and fell below the recommended threshold for optimal BSFL development (0.23 g/100 g larval growth) [57]. This deficiency may be a key factor contributing to the reduced larval growth observed on the SFW diet.

However, nutritional composition alone may not fully account for the observed differences in larval performance. Microbial functions within the gut microbiome likely also contribute to host physiology, influencing factors such as nutrient assimilation, digestion, and immune modulation. This is particularly relevant when considering the variation in larval growth observed between trials using the same substrate, suggesting that shifts in microbial community structure or function may play a role. Indeed, our analysis revealed notable differences in the microbiome composition of larvae reared on the same diet across different trials. This discrepancy may be attributed to variations in the microbial load of the feed substrates, which could introduce competing microorganisms into the larval gut. Such interactions may disrupt the native gut microbiota, potentially leading to gut dysbiosis and altered microbial functionality, ultimately affecting larval development.

These findings highlight the complex interplay between diet, microbiota, and host development, and underscore the need for integrated analyses that consider both nutritional and microbial factors. Although the microbial functions identified through this metagenomic analysis provide a valuable starting point for hypothesis generation, targeted experimental validation is essential to confirm the actual roles of these microbial taxa and pathways in the metabolism of diverse waste streams within the BSFL gut. Such empirical studies will be critical to determine whether these functions contribute meaningfully to larval development and waste bioconversion, and to assess their potential for practical applications.

### Functional analysis of the metagenome indicates that several metabolic and protective functions are encoded by the BSFL gut microbiome and diet-dependently enriched

Taxonomic and functional profiling of the BSFL gut microbiome revealed diet-driven shifts, notably by the enrichment of Actinomycetota in larvae reared on CF. Actinomycetes, including the Actinomycetaceae family and the genera *Corynebacterium*, *Brevibacterium*,

**Table 3** Overview good quality Metagenome Assembled Genomes (MAGs) found in the black soldier fly Gut microbiome. The length, GC content, number of contigs, completeness and contamination of the 25 good quality MAGs extracted from the BSFL metagenome are shown in panel A. The MAGs were identified using average nucleotide identity similarities and the alignment fraction, however when this technique did not yield a taxonomic consensus, relative evolutionary divergence was used for topology- based classification. The taxonomic classification of the MAGs is provided on family, genus and species level. In panel B, the mean number of normalised reads of the open reading frames (ORFs) that were found in each MAG are shown per substrate type (chicken feed larval microbiome (CF-LM) and supermarket food waste larval microbiome (SFW)) and per trial. The last five columns of panel B show the effect of the factors "diet", "l" and their interaction effect on the relative abundance of the MAGs

| **A** MAG ID | Length (bp) | GC content (%) | Nr. Contigs | Completeness (%) | Contamination (%) | Average Nucleotide Identity (%) | Alignment Fraction (%) | RED | Family | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|---|
| concoct.67 | 2 335 575 | 56.66 | 122 | 100.00 | 4.20 | 79.92 | 8 | 0.95 | Actinomycetaceae | *Scrofimicrobium* | *Scrofimicrobium* sp. |
| concoct.25 | 4 450 607 | 66.45 | 158 | 99.80 | 0.48 | 97.62 | 91 | - | Xanthomonadaceae | *Stenotrophomonas* | *Stenotrophomonas maltophilia* |
| concoct.14 | 4 428 003 | 41.43 | 87 | 99.64 | 2.74 | 99.18 | 93 | - | Enterobacteriaceae | *Providencia* | *Providencia stuartii* |
| metabat2.24 | 2 709 774 | 37.66 | 80 | 99.29 | 2.43 | 98.73 | 89 | - | Enterococcaceae | *Enterococcus* | *Enterococcus faecalis* |
| concoct.59 | 2 765 894 | 32.38 | 75 | 98.90 | 2.80 | 95.77 | 89 | - | Staphylococcaceae | *Mammaliicoccus* | *Mammaliicoccus sciuri* |
| concoct.44 | 5 418 733 | 54.15 | 308 | 98.35 | 0.33 | 99.21 | 97 | - | Paenibacillaceae | *Paenibacillus* | *Paenibacillus cookii* |
| concoct.42 | 2 640 142 | 40.83 | 81 | 98.28 | 1.72 | 98.45 | 87 | - | Lactobacillaceae | *Weissella* | *Weissella paramesenteroides* |
| concoct.81 | 2 747 970 | 63.76 | 419 | 98.11 | 4.21 | 95.89 | 89 | - | Mycobacteriaceae | *Corynebacterium* | *Corynebacterium phoceense* |
| concoct.86 | 2 634 772 | 70.70 | 110 | 97.48 | 0.00 | 79.06 | 45 | 0.94 | Beutenbergiaceae | *Salana* | *Salana* sp. |
| concoct.64 | 4 830 826 | 51.00 | 219 | 96.82 | 6.85 | 96.62 | 83 | - | Enterobacteriaceae | *Escherichia* | *Escherichia coli* |
| concoct.60 | 2 006 135 | 41.82 | 98 | 96.41 | 0.08 | 98.61 | 84 | - | Lactobacillaceae | *Latilactobacillus* | *Latilactobacillus curvatus* |
| concoct.87 | 4 367 244 | 38.92 | 108 | 95.79 | 8.88 | 98.56 | 90 | - | Enterococcaceae | *Enterococcus* | *Enterococcus avium* |
| concoct.36 | 2 901 778 | 43.27 | 80 | 95.44 | 3.20 | 79.99 | 88 | - | Enterococcaceae | *Enterococcus* | *Enterococcus diestrammenae* |
| concoct.57 | 4 307 426 | 40.12 | 223 | 95.18 | 4.91 | 99.35 | 92 | - | Enterobacteriaceae | *Providencia* | *Providencia rettgeri* |
| metabat2.56 | 2 179 535 | 46.13 | 83 | 92.48 | 0.98 | 96.97 | 93 | - | Lactobacillaceae | *Levilactobacillus* | *Levilactobacillus brevis* |
| concoct.17 | 4 660 081 | 34.88 | 435 | 91.56 | 0.33 | 98.55 | 91 | - | Bacillaceae | *Heyndrickxia* | *Heyndrickxia oleronia* |
| concoct.80 | 3 906 350 | 38.50 | 172 | 91.19 | 3.38 | 99.14 | 93 | - | Enterobacteriaceae | *Proteus* | *Proteus mirabilis* |
| metabat2.61 | 2 900 304 | 41.06 | 58 | 89.47 | 7.92 | 97.65 | 84 | - | Enterococcaceae | *Enterococcus* | *Enterococcus gallinarum* |
| concoct.8 | 2 887 172 | 37.59 | 83 | 88.24 | 7.50 | - | - | 0.93 | Planococcaceae | *Sporosarcina* | *Sporosarcina* sp. |
| concoct.12 | 3 837 656 | 51.00 | 66 | 87.21 | 6.47 | 96.98 | 91 | - | Enterobacteriaceae | *Morganella* | *Morganella morganii* |
| metabat2.2 | 1 522 108 | 38.38 | 229 | 86.79 | 2.64 | 96.38 | 97 | - | Lactobacillaceae | *Lapidilactobacillus* | *Lapidilactobacillus dextrinicus* |
| metabat2.13 | 2 564 801 | 35.26 | 176 | 85.84 | 1.01 | 99.84 | 84 | - | Lactobacillaceae | *Companilactobacillus* | *C. paralimentarius* |
| concoct.66 | 4 042 039 | 52.02 | 274 | 82.81 | 1.69 | 98.81 | 93 | - | Enterobacteriaceae | *Citrobacter* | *Citrobacter werkmanii* |
| metabat2.19 | 4 170 314 | 55.97 | 271 | 81.84 | 6.00 | 93.14 | 84 | - | Enterobacteriaceae | *Enterobacter* | *E. hormaechei/quasihormaechei* |
| metabat2.44 | 3 892 066 | 56.07 | 275 | 78.05 | 0.93 | 98.58 | 93 | - | Enterobacteriaceae | *Enterobacter* | *Enterobacter bugandensis* |

| **B** | | Reads trial 1 | | Reads trial 2 | | P-values affecting factors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| MAG ID | Species | CF-LM | SFW-LM | CF-LM | SFW-LM | Effect diet in Trial 1 | Effect diet in Trial 2 | Effect trial on CF-LM | Effect trial on SFW-LM | Interaction effect |
| concoct.67 | *Scrofimicrobium* sp. | 2 746 187[a] | 2618[b] | 4 592 685[a] | 6669[b] | 0.00000 | 0.00000 | 0.39982 | 0.06716 | 0.65492 |
| concoct.25 | *Stenotrophomonas maltophilia* | 19[a] | 0[a] | 217 529[b] | 69[a] | 0.00479 | 0.00000 | 0.00000 | 0.00003 | 0.00029 |
| concoct.14 | *Providencia stuartii* | 139 698[a] | 3 419 988[b] | 0[c] | 0[c] | 0.00000 | 0.65769 | 0.00000 | 0.00000 | 0.00591 |
| metabat2.24 | *Enterococcus faecalis* | 863[a] | 245[a] | 253 798[b] | 3 625 186[c] | 0.12427 | 0.00061 | 0.00000 | 0.00000 | 0.00068 |

**Table 3** (continued)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| concoct.59 | *Mammaliicoccus sciuri* | 2[a] | 0[a] | 4529[b] | 5955[b] | 0.63197 | 0.83556 | 0.00000 | 0.00000 | 0.65492 |
| concoct.44 | *Paenibacillus cookii* | 49 826[a] | 717[b] | 7[c] | 0[d] | 0.00000 | 0.00001 | 0.00000 | 0.00000 | 0.93419 |
| concoct.42 | *Weissella paramesenteroides* | 618[a] | 815[a] | 752[a] | 3879[a] | 0.63197 | 0.00214 | 0.76574 | 0.00332 | 0.10498 |
| concoct.81 | *Corynebacterium phoceense* | 191 209[a b] | 724 893[b] | 84 809[a] | 2432[c] | 0.02482 | 0.00000 | 0.20772 | 0.00000 | 0.00000 |
| concoct.86 | *Salana* sp. | 19 109[a b] | 7198[b] | 277 350[a] | 33[c] | 0.15589 | 0.00000 | 0.00010 | 0.00000 | 0.00000 |
| concoct.64 | *Escherichia coli* | 36 154[a b] | 46 095[b] | 57 358[ac] | 900 629[c] | 0.69369 | 0.00001 | 0.51627 | 0.00000 | 0.00714 |
| concoct.60 | *Latilactobacillus curvatus* | 2083[a] | 452 115[b] | 1278[a] | 313 205[b] | 0.00000 | 0.00000 | 0.50475 | 0.53418 | 0.93419 |
| concoct.87 | *Enterococcus avium* | 105 840[a] | 368 815[a] | 13[b] | 287[b] | 0.14908 | 0.00034 | 0.00000 | 0.00000 | 0.16537 |
| concoct.36 | *Enterococcus diestrammenae* | 524[a] | 66[a] | 29 524[b] | 0[c] | 0.00320 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| concoct.57 | *Providencia rettgeri* | 171 474[a] | 23 138 717[b] | 3[c] | 64[c] | 0.00000 | 0.00048 | 0.00000 | 0.00000 | 0.17178 |
| metabat2.56 | *Levilactobacillus brevis* | 1471[a] | 2 437 167[b] | 544[a] | 895 756[b] | 0.00000 | 0.00000 | 0.06492 | 0.04751 | 0.99219 |
| concoct.17 | *Heyndrickxia oleronia* | 59 158[a] | 1120[b] | 2213[c] | 12[d] | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.12080 |
| concoct.80 | *Proteus mirabilis* | 63 860[a b] | 19 594[b] | 267 083[a c] | 2 152 390[c] | 0.11760 | 0.00306 | 0.06128 | 0.00000 | 0.00169 |
| metabat2.61 | *Enterococcus gallinarum* | 325 206[a] | 44[b] | 31 190[a] | 196[b] | 0.00000 | 0.00000 | 0.00791 | 0.07987 | 0.00237 |
| concoct.8 | *Sporosarcina* sp. | 0[a] | 30[a] | 0[a] | 0[a] | 0.06255 | 0.83556 | 0.84169 | 0.05274 | 0.16537 |
| concoct.12 | *Morganella morganii* | 549[a] | 88 001[b c] | 886[a c] | 4963[a] | 0.00000 | 0.00490 | 0.50475 | 0.00000 | 0.00026 |
| metabat2.2 | *Lapidilactobacillus dextrinicus* | 2099[a] | 410[a] | 157[b] | 1[c] | 0.00591 | 0.00000 | 0.00001 | 0.00000 | 0.00072 |
| metabat2.13 | *C. paralimentarius* | 3[a] | 1699[b] | 0[a] | 41[c] | 0.00000 | 0.00000 | 0.00659 | 0.00000 | 0.56185 |
| concoct.66 | *Citrobacter werkmanii* | 51[a] | 622[b] | 20 770[c] | 5879[b c] | 0.00000 | 0.01708 | 0.00000 | 0.00002 | 0.00000 |
| metabat2.19 | *E. hormaechei/quasihormaechei* | 5292[a] | 437[b] | 2528[c] | 6970[d] | 0.00013 | 0.11535 | 0.32347 | 0.00001 | 0.00026 |
| metabat2.44 | *Enterobacter bugandensis* | 66[a b] | 242[b] | 71[a] | 47 482[c] | 0.12427 | 0.00000 | 0.92218 | 0.00000 | 0.00002 |

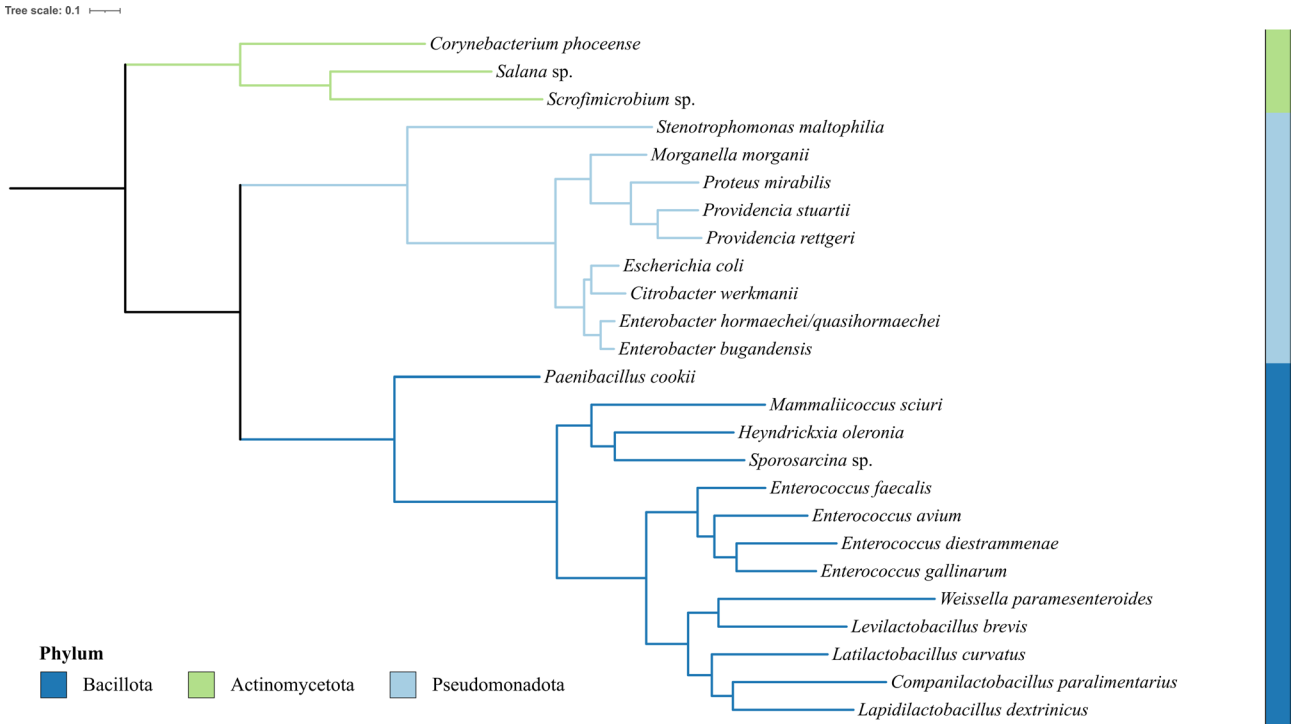


**Fig. 6** Maximum Likelihood phylogenetic tree of the 25 metagenome-assembled genomes (MAGs). Branch colours and the coloured sidebar indicate the phylum-level taxonomic classification of each MAG. Branch lengths represent the number of nucleotide substitutions per site, reflecting evolutionary distances between MAGs, with longer branches indicate greater divergence

*Micrococcales* and *Scrofimicrobium* were consistently abundant in the CF-LM in both trials. Actinomycetes can produce a wide range of useful bioactive compounds, such as secondary metabolites with antibacterial or antiviral activities, enzymes amylase and protease that help with food digestion, or the production of single-cell protein that can be used as a potential animal feed [58]. Although the taxonomic resolution of this class is limited, the ability of this class to produce bioactive compounds and their presence in the microbiome may contribute to improved larval growth. Bacillota was a dominant phylum across diets, with most reads taxonomically delineated as lactic acid bacteria (LAB) (e.g. *Lactiplantibacillus*, *Levilactobacillus*, *Enterococcus*). Most LAB were enriched in SFW-fed larvae, where the presence of *Leuconostoc* and *Latilacobacillus* aligned with previous findings on organic waste substrates [17, 59, 60]. In contrast, Enterococcaceae were more abundant in CF-fed larvae. Pseudomonadota were enriched in the SFW diets (Fig. 2), although their abundance varied between trials. Specifically, *Providencia*, a putative member of the BSFL core microbiome and therefore potentially providing important functions for the insect host [19], was more abundant in trial 1 compared to trial 2 (Fig. 3).

KEGG analysis was used to investigate diet-specific microbial functions. The most striking functional changes in the BSFL gut microbiome revolved around amino acid, carbohydrate and energy metabolism (Fig. 4C), pathways that are involved in bacterial growth and proliferation. Highlighting some of the most prominent differences in pathway enrichment between the CF and SFW-fed larvae might lead to new hypotheses regarding bacterial functionalities. A potential diet-related microbial function involves metabolism of sorbitol, where several genes involved in this metabolic pathway were significantly enriched in the SFW-LM (Supplementary Figure S3). Sugar content analysis of the diets revealed that the SFW diet had a high concentration of sorbitol (Table 2), which most likely originates from apples included in the SFW formulation. Microbial breakdown of sorbitol can lead to a more complete breakdown of the substrate, potentially generating additional metabolites that are beneficial to the host insect. Even bacteria that cannot directly metabolise sorbitol may benefit from the metabolic activity of other microbes by utilizing downstream products for their own proliferation. Such cross-feeding and metabolic cooperation among microbes have previously been shown to support larval growth under nutritional stress in *Drosophila* [61], yet these interactions have not yet been reported on in BSFL and warrant further investigation.

Riboflavin metabolism functions were also enriched in the SFW-LM. It was demonstrated by Pei et al. that certain bacteria can provide riboflavin for their BSF host,

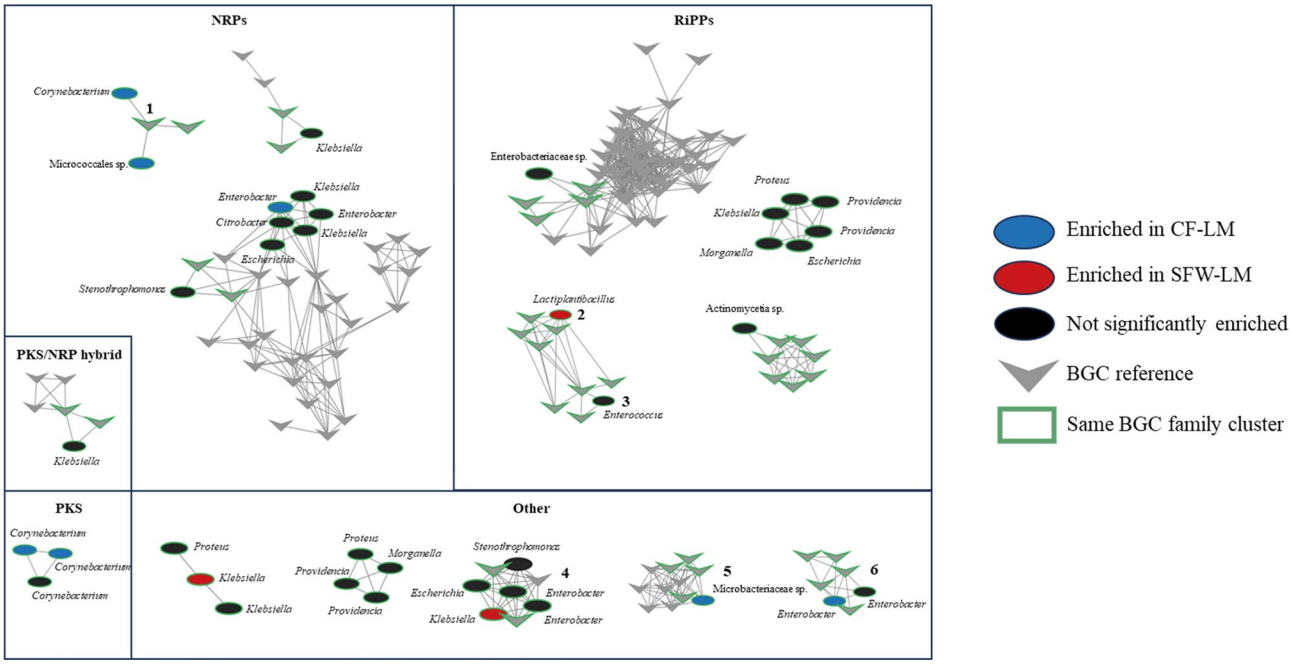


**Fig. 7** Overview of the biosynthetic gene clusters and Gene Cluster Families. The ovals in the figure represent the contigs extracted from the metagenomics dataset that were identified as secondary metabolite gene clusters (or biosynthetic gene clusters, BGCs) by antiSMASH. The taxonomic identification of each BGC based on a similarity search in GenBank is provided. Next, BGCs from the metagenomics dataset and BGCs from the MIBiG reference database (shown as grey arrowheads) with similar functionalities were clustered together in a network of gene cluster families (GCFs). A total of 16 GCFs were identified and classified per category (non-ribosomal peptide synthetase (NRPS). Ribosomally Synthesised and Post-translationally Modified Peptides (RiPPs). polyketide synthases (PKS) or hybrids between these groups) or placed in the "Other" category. BGCs that belonged to the same GCF are outlined in green. Six GCFs that were of interest in this study were numbered in the figure. The colour of the oval indicates whether the BGCs that were part of the GCF were significantly more present in the black soldier fly metagenome of larvae reared on chicken feed- (CF-LM. blue) or supermarket food waste- (SFW-LM) diet (red)

thereby enhancing larval survival [62]. Since the riboflavin concentrations of the diets were not measured, it is unclear whether microbial activity could have contributed to increased riboflavin availability for the larvae, and thereby contributed to larval performance. Moreover, SFW-fed larvae had reduced growth compared to CF-fed larvae, despite the presence of riboflavin-rich feed ingredients such as cheese and yoghurt, and potentially higher microbial riboflavin metabolic activity. Still, further research towards the isolation and characterization of riboflavin-producing microorganisms could provide valuable insights into the role of riboflavin in BSFL development, microbiome-host interactions and strategies to enhance BSFL performance through targeted microbial supplementation.

Furthermore, both aerobic and anaerobic cobalamin (vitamin B12) biosynthesis functions were more abundant in SFW-fed larvae in trial 2, even thought the ingredients of the SFW diets from both trials were identical. As insects cannot synthesise B-vitamins, microbial production may compensate for dietary deficiencies [63]. Although the specific cobalamin requirements for BSFL remain unknown, the low larval weight observed in trial 1 may have resulted from insufficient cobalamin in the SFW diet, limiting the larvae's ability to metabolise key nutrients. The enrichment of cobalamin-related KEGG functions in trial 2 suggests a greater presence of cobalamin-producing bacteria, potentially due to different microbial loads of the SFW batches used. Possibly, the cobalamin produced by these microorganisms could have been utilised by the larvae during the second rearing trial, compensating for the nutritional deficiencies of the diet and resulting in higher larval final weight. These findings highlight the potential role of microbiome-derived cobalamin in supporting BSFL development and underscore the need for further studies to quantify vitamin levels and identify key microbial producers.

### BGCs analysis reveals potential antimicrobial competition and protective functions among bacterial species

Beyond 16S rRNA profiling, BGC analysis offers deeper insights into the functional potential of the BSFL Gut microbiome. This approach revealed gene clusters encoding metabolites with antimicrobial properties. For instance, GCF number 2 (Fig. 7) consists of a contig mapped to *Lactiplantibacillus*, and reference BGCs extracted from *Lactobacillus gasseri* involved in the production of the bacteriocins gassericin E and T. Gassericin

T is reported to have strong antimicrobial effect against several Gram-positive bacteria and, interestingly, this activity appears to target other closely related lactic acid bacteria [64, 65]. Given the high abundance of *Lactiplantibacillus* in the SFW-fed larvae and the presence of this GCF, it is hypothesised that these bacteriocins may suppress competing LAB such as BSFL-native *Enterococcus* strains, which were more prevalent in CF-fed larvae.

Another notable cluster, GCF 1, is a non-ribosomal peptide synthetase (NRPS) and contains two contigs and two reference genes linked to ε-Poly-L-lysine and γ-poly-L-2,4-diaminobutyric acid BGCs, two naturally occurring homo-poly amino acids known for broad-spectrum antimicrobial activity [66]. ε-Poly-L-lysine in particular is a commercially valuable antimicrobial agent that has been found to inhibit both Gram-positive and Gram-negative bacteria [67]. The contigs, taxonomically delineated as *Corynebacterium* and Micrococcales sp., may encode antimicrobial peptides that can limit the growth of competing microorganisms in the BSFL gut microbiome. Their potential to biosynthesise ε-Poly-L-lysine is industrially significant, particularly given that some microbes can convert organic waste into this valuable antimicrobial [67], offering a sustainable route for high-value bioproduct generation.

In addition to antimicrobial compounds, three clusters were Linked to protective functions for the bacteria. Cluster 6, associated with *Enterobacter*, included five BGCs involved in the production of aerobactin and ochrobactin, which enhance microbial iron acquisition under limiting conditions [68]. Cluster 5 was enriched in CF-fed larvae, and contained a single Microbacteriaceae linked to a series of ectoine producing BGCs, which are known to help other Actinomycetia species to survive osmotic stress [68]. A similar protective function is found in cluster 4, which linked several Gram-negative Pseudomonadota members enriched in SFW-fed larvae to aryl polyenes gene clusters. These effectors may protect (pathogenic) Gram-negative bacteria against oxidative stress, for example that caused by the host immune system during infection of an eukaryotic host [69]. This protective role potentially aids the dominance of Pseudomonadota members such as *Klebsiella* in the SFW-fed larval microbiome. While the presence of BGCs alone does not confirm activity, these findings provide a foundation for future transcriptomic studies to explore the roles of secondary metabolite production in microbial competition and host protection within the BSFL gut.

### Genome reconstruction identifies potential BSFL specific, beneficial bacteria

Whole genome shotgun sequencing offers higher taxonomic resolution than 16S rRNA gene profiling, enabling species-level identification and functional analysis of microbial genomes (MAGs) that can be Linked to the phenotypic characteristics of the larvae. Larval performance varied between diets and between trials. Although the nutritional composition of the CF and SFW diets was not analysed between trials 1 and 2, consistent diet preparation across trials renders it unlikely that the observed performance differences between the two trials were due to nutritional variation. Instead, factors such as larval genetics, microbial load of the diets or microbial community compositions of the insects may have contributed to observed differences in larval performance. Therefore, MAGs that were also affected by differences in trial, diet or an interaction effect of both, might point to some important members of the BSFL microbiome.

Eight MAGs showed significant inter-trial differences in abundance, suggesting that factors beyond diet composition (such as MAG abundance) influence larval growth. Noticeably, MAGs *Providencia stuartii* and *P. rettgeri* were abundant in trial 1 in larvae reared on both diets and negatively correlated with larval weight. *Providencia rettgeri* has been linked to pathogenicity in the silkworm [70] and the Mexican fruit fly [71], though some studies suggest native *Providencia* may support BSF growth and development [72]. Other negatively correlated MAGs with inter-trial differences include *Paenibacillus cookii*, *Heyndrickxia oleronia* and *Enterococcus avium*. Some *Paenibacillus* species are known to produce insecticidal toxins [73], cause diseases in honeybee larvae [74] or impede BSF larval development and cause mortality [75]. Possibly, *P. cookii* possesses similar virulent qualities as these other species that might negatively affect BSFL development. In contrast, *Mammaliicoccus sciuri* showed a slight positive correlation with larval weight and has been reported to have a protective function in silkworm larvae due to its inhibitory effect towards parasitic fungi [76], suggesting a potential beneficial role in BSFL as well.

MAGs influenced solely by diet provide insights into diet-dependent microbiome shifts. *Latilactobacillus curvatus* and *Levilactobacillus brevis* were enriched in SFW-fed larvae but negatively correlated with mean larval weight. In contrast, *Scrofimicrobium* sp. and *Enterococcus gallinarum* were also diet-associated but positively correlated with larval weight. While these correlations suggest potential microbial contributions to larval performance, separating microbial effects from dietary influences remains challenging. Regardless, these findings support previous work suggesting that dietary shifts, such as the switch from CF to SFW after the nursery phase, can induce gut dysbiosis [20]. This results in a stochastic process where different microbial taxa may become dominant, as exemplified by the high abundance of *Providencia* in the SFW-LM treatment of trial 1 compared to its lower abundance in CF-LM. Such larval gut dysbiosis

may be more pronounced with real-world supermarket food waste, which is more variable in both nutritional content and microbial composition than the artificial SFW used here. Microbes thriving on artificial substrates due to their ability to utilise specific metabolites may not persist in real waste streams, especially those containing meat, fish, or spoiled components. Consequently, their dominance may not translate to real-world conditions.

Future research should account for the variability in both the nutritional and microbial composition of real-world food waste. It is also essential to determine whether the prevalence of *Providencia* or specific lactic acid bacteria is driven by subtle dietary differences and whether their presence impacts larval performance. These insights are critical for evaluating the suitability of supermarket food waste as a BSFL diet.

### Metagenomic assembly reveals novel Scrofimicrobium species

Several of the MAGs identified in this study, such as *Providencia*, *Enterococcus*, *Morganella* and *Scrofimicrobium*, have been consistently reported in BSFL microbiome research and may represent part of the BSFL "core" microbiome [19]. However, their functional roles remain unclear. Of particular interest is MAG *concoct.67*, which could not be identified through average nucleotide identify and alignment fraction statistics, but was placed near *Scrofimicrobium canadense* based on relative evolutionary divergence. While *Scrofimicrobium* species have previously been detected in BSFL [19, 77, 78], the identification of these bacteria has solely been based on the comparison of 16S rRNA gene amplicon sequences against a limited number of *Scrofimicrobium* genomes available in genome databases such as NCBI. This study is the first to assemble a *Scrofimicrobium* genome directly from BSFL, offering a valuable reference for future taxonomic and functional studies. Indeed, phylogenetic and BLAST analysis revealed that the *Scrofimicrobium* concoct.67 differs from the genomes *Scrofimicrobium canadense* (MN537492.1), *Scrofimicrobium appendicitides* (R652275.1) and *Scrofimicrobium* sp. (PQ736596.1) (Supplementary Figures S7.1 and S7.2), yet shows the highest similarity to BSFL-derived 16S rRNA sequences from previous studies [19, 77]. This highlights the improved taxonomic resolution that this genome can provide to BSFL related metagenome research and the potential of shotgun metagenomics to uncover novel or poorly characterised microbes in insect microbiomes. Given the frequent detection of *Scrofimicrobium* in BSFL, understanding its functional role in BSFL is important. Although the functionality of *Scrofimicrobium* has not been investigated, related Actinomycetaceae members have been implicated in the metabolism of insect deterrent toxic compounds, such as gossypol in cottonseed press cake [79]. In the same study, the ability of Actinomycetales members to degrade complex molecules such as lignocellulose is suggested. Further investigation of this novel *Scrofimicrobium* genome may reveal substrate-specific microbial functions that can support BSFL with the degradation of complex or deleterious molecules, offering more insights into the role of this specific member of the BSFL gut microbiome.

## Conclusions

The current study showed that WGS analysis of the BSFL microbiome can lead to new and valuable insights on potential metabolic- and protective functions microorganisms can provide for its insect host. A diet shift during the rearing cycle of BSFL, combined with the variation in the nutritional and bacterial load of the diet, can lead to larval gut dysbiosis, resulting in compositional shift in the microbiome of the larvae. Metagenomic reconstruction identified several bacteria, including a novel *Scrofimicrobium* species, which provided insights into the factors that affected the abundance of these MAGs and revealed potential correlations between their abundances and larval performance. These results suggest that some of these bacteria can potentially provide beneficial metabolic functions for the host, or protective functions against other microorganisms through the production of antimicrobial peptides, which should be further explored in future research.

To conclude, this study shows the added value offered by complementing the existing knowledge on BSFL Gut microbiome dynamics generated by 16S rRNA sequencing with the use of WGS. Application of this more extensive sequencing technique results in a more detailed idea at functional level of changes happening in the microbiota, as well as the ability to assemble genomic data that can be used to mine for interesting functions within these insect metagenomes.

#### Abbreviations

| | |
|---|---|
| AF | Alignment fraction |
| ANI | Average nucleotide identity |
| BCAA | Branched-chain amino acids |
| BGC | Biosynthetic Gene Clusters |
| BSF | Black soldier fly |
| BSFL | Black soldier fly larvae |
| CF | Chicken feed |
| CF-LM | CF-diet-reared larval metagenome |
| GCF | Gene Cluster Families |
| GLM | Generalised linear model |
| MAG | Metagenomic assembled genome |
| NRP | Non-ribosomal peptide synthetase |
| ORF | Open reading frame |
| PERMANOVA | Permutational analysis of variance |
| PKS | Polyketide synthases |
| RED | Relative evolutionary divergence |
| RiPPs | Ribosomally Synthesised and Post-translationally Modified Peptides |
| SFW | Supermarket food waste |
| SFW-LM | SFW-diet-reared larval metagenome |

WGS Whole genome shotgun

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-025-04327-3.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

**Acknowledgements**
The authors would like to thank dr. Nathaniel Sibinga for proofreading the manuscript.

**Author contributions**
F. IJdema – Methodology, Formal analysis, Investigation, Software, Writing – Original Draft, Writing – Review & Editing, Visualization. L.M. Arias-Giraldo – Methodology, Formal analysis, Software, Writing – Review & Editing. E. Vervoort – Investigation, Writing – Review & Editing. T. Struyf – Investigation. W. Van den Ende – Methodology, Writing – Review & Editing. J. Raaijmakers – Methodology, Writing – Review & Editing. B. Lievens – Writing – Review & Editing, Funding acquisition. J. De Smet – Conceptualization, Methodology, Supervision, Writing – Review & Editing.

**Funding**
F. IJdema was funded by the FWO (Research Foundation Flanders) under grant agreement S008519N (ENTOBIOTA) and additionally through a KU Leuven Impuls grant (IMP20028), and is currently funded by the Francqui Start-up Grant of Jeroen De Smet. J. De Smet held a postdoctoral fellowship (12V5222N) from the FWO and has now obtained a Francqui Start-up Grant. W. Van den Ende is funded through FWO project G0C4622N. E. Vervoort is funded by the C3 project CHITINERY [grant C3/22/041], funded by the KU Leuven. Research in the lab of B. Lievens is funded by FWO, Vlaio and internal KU Leuven funds.

**Data availability**
The data for this study has been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB81711.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare no competing interests.

**Author details**
[1]CLMT Research Group for Insect Production and Processing, Department of Microbial and Molecular Systems (M2S), KU Leuven, Geel Campus, Kleinhoefstraat 4, Geel 2440, Belgium
[2]CMPG Laboratory for Process Microbial Ecology and Bioinspirational Management (PME&BIM), Department of Microbial and Molecular Systems ($M^2S$), KU Leuven, Willem De Croylaan 46, Leuven 3001, Belgium
[3]Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO- KNAW), Droevendaalsesteeg 10, Wageningen 6708 PB, The Netherlands
[4]CSCE Research Group for Insect Production and Processing, Department of Microbial and Molecular Systems (M2S), KU Leuven, Geel Campus, Kleinhoefstraat 4, Geel 2440, Belgium
[5]Lab of Molecular Plant Biology, Department of Biology, KU Leuven, Kasteelpark Arenberg 31, Leuven 3001, Belgium
[6]Institute of Biology, Leiden University, Sylviusweg 72, Leiden 2333 BE, Netherlands
[7]Geel, Belgium

Received: 2 June 2025 / Accepted: 14 August 2025
Published online: 02 October 2025

**References**

1. Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. Science (1979). 2006;312(5778):1355–9.
2. Brulc JM, Antonopoulos DA, Berg Miller ME, Wilson MK, Yannarell AC, Dinsdale EA, et al. Gene-centric metagenomics of the fiber-adherent bovine rumen Microbiome reveals forage specific glycoside hydrolases. Proc Natl Acad Sci. 2009;106(6):1948–53.
3. Scully ED, Geib SM, Hoover K, Tien M, Tringe SG, Barry KW, et al. Metagenomic profiling reveals lignocellulose degrading system in a microbial community associated with a Wood-Feeding beetle. PLoS ONE. 2013;8(9):e73827.
4. Turner TR, James EK, Poole PS. The plant Microbiome. Genome Biol. 2013;14(6):209.
5. Ross FC, Patangia D, Grimaud G, Lavelle A, Dempsey EM, Ross RP, et al. The interplay between diet and the gut microbiome: implications for health and disease. Nat Rev Microbiol Nat Res. 2024.
6. Spranghers T, Ottoboni M, Klootwijk C, Ovyn A, Deboosere S, De Meulenaer B, et al. Nutritional composition of black soldier fly (*Hermetia illucens*) Prepupae reared on different organic waste substrates. J Sci Food Agric. 2017;97(8):2594–600.
7. Barragan-Fonseca KB, Dicke M, van Loon JJA. Nutritional value of the black soldier fly (*Hermetia illucens* L.) and its suitability as animal feed – a review. J Insects Food Feed. 2017;3(2):105–20.
8. Freel TA, McComb A, Koutsos EA. Digestibility and safety of dry black soldier fly larvae meal and black soldier fly larvae oil in dogs. J Anim Sci. 2021;99(3):skab047.
9. Belghit I, Liland NS, Gjesdal P, Biancarosa I, Menchetti E, Li Y, et al. Black soldier fly larvae meal can replace fish meal in diets of sea-water phase Atlantic salmon (*Salmo salar*). Aquaculture. 2019;503:609–19.
10. Lalander C, Diener S, Zurbrügg C, Vinnerås B. Effects of feedstock on larval development and process efficiency in waste treatment with black soldier fly (*Hermetia illucens*). J Clean Prod. 2019;208:211–9.
11. Broeckx L, Frooninckx L, Slegers L, Berrens S, Noyens I, Goossens S, et al. Growth of black soldier fly larvae reared on organic Side-Streams. Sustainability. 2021;13(23):12953.
12. Jiang C, Jin W, Tao X, Zhang Q, Zhu J, Feng S, et al. Black soldier fly larvae (*Hermetia illucens*) strengthen the metabolic function of food waste biodegradation by gut Microbiome. Microb Biotechnol. 2019;12(3):528–43.
13. Zhineng Y, Ying M, Bingjie T, Rouxian Z, Qiang Z. Intestinal microbiota and functional characteristics of black soldier fly larvae (*Hermetia illucens*). Ann Microbiol. 2021;71(1):13.
14. De Filippis F, Bonelli M, Bruno D, Sequino G, Montali A, Reguzzoni M, et al. Plastics shape the black soldier fly larvae gut Microbiome and select for biodegrading functions. Microbiome. 2023;11(1):205.
15. Vogel M, Shah PN, Voulgari-Kokota A, Maistrou S, Aartsma Y, Beukeboom LW, et al. Health of the black soldier fly and house fly under mass-rearing conditions: innate immunity and the role of the Microbiome. J Insects Food Feed. 2022;8(8):857–78.
16. Auger L, Deschamps MH, Vandenberg G, Derome N. Microbiota is structured by gut regions, life stage, and diet in the black soldier fly (*Hermetia illucens*). Front Microbiol. 2023;14:1221728.
17. Vandeweyer D, Bruno D, Bonelli M, IJdema F, Lievens B, Crauwels S et al. Bacterial biota composition in gut regions of black soldier fly larvae reared on industrial residual streams: revealing community dynamics along its intestinal tract. Front Microbiol. 2023;14:1276187.
18. Klammsteiner T, Walter A, Bogataj T, Heussler CD, Stres B, Steiner FM et al. The core gut Microbiome of black soldier fly (*Hermetia illucens*) larvae Raised on Low-Bioburden diets. Front Microbiol. 2020;11:993.

19. IJdema F, De Smet J, Crauwels S, Lievens B, Van Campenhout L. Meta-analysis of larvae of the black soldier fly (*Hermetia illucens*) microbiota based on 16S rRNA gene amplicon sequencing. FEMS Microbiol Ecol. 2022;98(9):fiac094.
20. Bruno D, Bonelli M, De Filippis F, Di Lelio I, Tettamanti G, Casartelli M et al. The intestinal microbiota of *Hermetia illucens* larvae is affected by diet and shows a diverse composition in the different midgut regions. Appl Environ Microbiol. 2019;85(2):e01864-18.
21. Eke M, Tougeron K, Hamidovic A, Tinkeu LSN, Hance T, Renoz F. Deciphering the functional diversity of the gut microbiota of the black soldier fly (*Hermetia illucens*): recent advances and future challenges. Volume 5. Animal Microbiome. BioMed Central Ltd; 2023.
22. Sun S, Jones RB, Fodor AA. Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. Microbiome. 2020;8(1):46.
23. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun. 2016;469(4):967–77.
24. Van Looveren N, Verbaet L, Frooninckx L, Van Miert S, Van Campenhout L, Van Der Borght M, et al. Effect of heat treatment on Microbiological safety of supermarket food waste as substrate for black soldier fly larvae (*Hermetia illucens*). Waste Manag. 2023;164:209–18.
25. Smets R, Van Der Borght M. Enhancing the specificity of Chitin determinations through glucosamine analysis via ultra-performance LC-MS. Anal Bioanal Chem. 2021;413(11):3119–30.
26. IJdema F, Lievens S, Smets R, Poma G, Van Der Borght M, Lievens B et al. Modulating the fatty acid composition of black soldier fly larvae via substrate fermentation. Animal. 2025;19(1):101383.
27. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. Bioinformatics. 2015;31(10):1674–6.
28. Tamames J, Puente-Sánchez F. SqueezeMeta A highly portable, fully automatic metagenomic analysis pipeline. Front Microbiol. 2019;9:3349.
29. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27(6):863–4.
30. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11(1):119.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–9.
32. Laslett D. ARAGORN, a program to detect tRNA genes and TmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32(1):11–6.
33. Federhen S. Type material in the NCBI taxonomy database. Nucleic Acids Res. 2015;43(D1):D1086–98.
34. Love MI, Huber W, Anders S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
35. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [Internet]. 2021; Available from: https://www.R-project.org/
36. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of Microbiome census data. PLoS ONE. 2013;8(4):e61217.
37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 1995;57(1):289–300.
38. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60.
39. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank Nucleic Acids Res. 2016;44(D1):D67–72.
40. Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
41. Ye Y, Doak TG. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol. 2009;5(8):e1000465.
42. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9.
43. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive Binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.
44. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144–6.
45. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3(7):836–43.
46. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.
47. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2020;36(6):1925–7.
48. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
49. Bayliss SC, Thorpe HA, Coyle NM, Sheppard SK, Feil EJ. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. Gigascience. 2019;8(10):giz119.
50. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
51. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78–82.
52. JMP®. Version 17.0.0. Cary, NC: SAS Institute Inc.; 2024.
53. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39(suppl2):W339–46.
54. Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Res. 2023;51(D1):D603–10.
55. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, et al. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol. 2020;16(1):60–8.
56. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
57. Spranghers T, Moradei A, Vynckier K, Boudrez M, Pinotti L, Ottoboni M. Amino acid requirements of yellow mealworm and black soldier fly larvae. J Insects Food Feed. 2024;n(6):1047–58.
58. Selim MSM, Abdelhamid SA, Mohamed SS. Secondary metabolites and biodiversity of actinomycetes. J Genetic Eng Biotechnol. 2021;19(1):72.
59. Quan J, Wang Y, Cheng X, Li C, Yuan Z. Revealing the effects of fermented food waste on the growth and intestinal microorganisms of black soldier fly (*Hermetia illucens*) larvae. Waste Manag. 2023;171:580–9.
60. Gold M, von Allmen F, Zurbrügg C, Zhang J, Mathys A. Identification of bacteria in two food waste black soldier fly larvae rearing residues. Front Microbiol. 2020;11:582867.
61. Lange C, Boyer S, Bezemer TM, Lefort MC, Dhami MK, Biggs E, et al. Impact of intraspecific variation in insect microbiomes on host phenotype and evolution. ISME J. 2023;17(11):1798–807.
62. Pei Y, Zhao S, Chen X, Zhang J, Ni H, Sun M et al. Bacillus velezensis EEAM 10B strengthens nutrient metabolic process in black soldier fly larvae (Hermetia illucens) via changing gut Microbiome and metabolic pathways. Front Nutr. 2022;9:880488.
63. Douglas AE. The B vitamin nutrition of insects: the contributions of diet, Microbiome and horizontally acquired genes. Curr Opin Insect Sci. 2017;23:65–9.
64. Pandey N, Malik RK, Kaushik JK, Singroha G. Gassericin A: a circular bacteriocin produced by lactic acid bacteria *Lactobacillus gasseri*. World J Microbiol Biotechnol. 2013;29(11):1977–87.
65. Maldonado-Barragán A, Caballero-Guerrero B, Martín V, Ruiz-Barba JL, Rodríguez JM. Purification and genetic characterization of gassericin E, a novel co-culture inducible bacteriocin from *Lactobacillus gasseri* EV1461 isolated from the vagina of a healthy woman. BMC Microbiol. 2016;16(1):37.
66. Yamanaka K, Fukumoto H, Takehara M, Hamano Y, Oikawa T. The stereocontrolled biosynthesis of Mirror-Symmetric 2,4-Diaminobutyric acid homopolymers is critically governed by adenylation activations. ACS Chem Biol. 2020;15(7):1964–73.
67. Wang L, Zhang C, Zhang J, Rao Z, Xu X, Mao Z et al. Epsilon-poly-L-lysine: recent advances in biomanufacturing and applications. Front Bioeng Biotechnol. 2021;9:748976.
68. Ofer N, Wishkautzan M, Meijler M, Wang Y, Speer A, Niederweis M, et al. Ectoine biosynthesis in *Mycobacterium smegmatis*. Appl Environ Microbiol. 2012;78(20):7483–6.

69. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell. 2014;158(2):412–21.
70. Zhang J. Isolation and identification of a pathogen, *Providencia rettgeri*, in *Bombyx Mori*. J Bacteriol Res. 2013;5(2):22–8.
71. Salas B, Conway HE, Vacek DC, Vitek C, Schuenzel EL. Pathogenicity of multiple *Providencia* species (Enterobacteriales: Morganellaceae) to the mass-reared Mexican fruit fly (Diptera: Tephritidae). J Insect Sci. 2023;23(3):4.
72. Li X, Mei C, Luo X, Wulamu D, Zhan S, Huang Y, et al. Dynamics of the intestinal bacterial community in black soldier fly larval guts and its influence on insect growth and development. Insect Sci. 2023;30(4):947–63.
73. Keller A, Brandel A, Becker MC, Balles R, Abdelmohsen UR, Ankenbrand MJ et al. Wild bees and their nests host *Paenibacillus bacteria* with functional potential of avail. Microbiome. 2018;6(1):229.
74. Genersch E. American Foulbrood in honeybees and its causative agent, *Paenibacillus larvae*. J Invertebr Pathol. 2010;103(SUPPL. 1).
75. She W, Xiao Q, Meng Y, Zhao P, Wu C, Huang F, et al. Isolated and identified pathogenic bacteria from black soldier fly larvae with soft rot reared in mass production facilities and its incidence characteristics. Waste Manag. 2023;163:85–95.
76. Zhao P, Hong S, Li Y, Chen H, Gao H, Wang C. From phyllosphere to insect cuticles: silkworms gather antifungal bacteria from mulberry leaves to battle fungal parasite attacks. Microbiome. 2024;12(1):40.
77. Van Looveren N, IJdema F, van der Heijden N, Van Der Borght M, Vandeweyer D. Microbial dynamics and vertical transmission of *Escherichia coli* across consecutive life stages of the black soldier fly (*Hermetia illucens*). Anim Microbiome. 2024;6(1):29.
78. Silvaraju S, Zhang Q, hui, Kittelmann S, Puniamoorthy N. Genetics, age, and diet influence gut bacterial communities and performance of black soldier fly larvae (*Hermetia illucens*). Anim Microbiome. 2024;6(1):56.
79. Tegtmeier D, Hurka S, Klüber P, Brinkrolf K, Heise P, Vilcinskas A. Cottonseed press cake as a potential diet for industrially farmed black soldier fly larvae triggers adaptations of their bacterial and fungal gut microbiota. Front Microbiol. 2021;12:634503.

## Publisher's note