

**Title of the dataset:**

“Data from: Substrate-dependent impact of black soldier fly larvae on bacterial community composition in substrate and larval body”

**Creators:**

Stijn J.J. Schreven, Laboratory of Entomology, Plant Sciences Group, Wageningen University & Research, P.O. Box 16, 6700 AA, Wageningen, The Netherlands. ORCID <https://orcid.org/0000-0002-5837-7025>

**Contributors:**

Hugo de Vries, Laboratory of Microbiology, Agrotechnology & Food Sciences Group, Wageningen University & Research, P.O. Box 8033, 6700 EH, Wageningen, The Netherlands. ORCID <https://orcid.org/0000-0002-8888-7202>

Gerben D.A. Hermes, Laboratory of Microbiology, Agrotechnology & Food Sciences Group, Wageningen University & Research, P.O. Box 8033, 6700 EH, Wageningen, The Netherlands. ORCID <https://orcid.org/0000-0003-4314-9553>

Giacomo Zeni, Laboratory of Entomology, Plant Sciences Group, Wageningen University & Research, P.O. Box 16, 6700 AA, Wageningen, The Netherlands; Laboratory of Microbiology, Agrotechnology & Food Sciences Group, Wageningen University & Research, P.O. Box 8033, 6700 EH, Wageningen, The Netherlands.

Hauke Smidt, Laboratory of Microbiology, Agrotechnology & Food Sciences Group, Wageningen University & Research, P.O. Box 8033, 6700 EH, Wageningen, The Netherlands. ORCID <https://orcid.org/0000-0002-6138-5026>

Marcel Dicke, Laboratory of Entomology, Plant Sciences Group, Wageningen University & Research, P.O. Box 16, 6700 AA, Wageningen, The Netherlands. ORCID <https://orcid.org/0000-0001-8565-8896>

Joop J.A. van Loon, Laboratory of Entomology, Plant Sciences Group, Wageningen University & Research, P.O. Box 16, 6700 AA, Wageningen, The Netherlands. ORCID <https://orcid.org/0000-0002-4260-0501>

**Related publication(s):**

Schreven SJJ, De Vries H, Hermes GDA, Zeni G, Smidt H, Dicke M, Van Loon JJA (in press). Black soldier fly larvae influence internal and substrate bacterial community composition depending on substrate type and larval density. *Applied and Environmental Microbiology*. DOI: <https://doi.org/10.1128/aem.00084-22>.

Schreven SJJ, De Vries H, Hermes GDA, Zeni G, Smidt H, Dicke M, Van Loon JJA (2021). Power of the crowd: substrate-dependent impact of black soldier fly larvae on bacterial community composition in substrate and larval gut (Chapter 3). In: Schreven SJJ, *Of maggots and microbes*. PhD thesis, Wageningen University and Research, Wageningen, The Netherlands. DOI: <https://doi.org/10.18174/542262>.

All sequence data have been deposited in the European Nucleotide Archive under the study accession number PRJEB40667 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB40667>).

**Description:**

We investigated how the type of feed substrate and the black soldier fly larval density (number of larvae per container) interacted to influence bacterial community composition in the substrates and larvae over time. Substrate type was the strongest driver of bacterial community composition, and the magnitude of the impact of the larvae depended on the substrate type and larval density.

**Keywords:**

16S rRNA gene, amplicon sequencing, *Hermetia illucens*, larval density, pH, manure, *Camelina sativa*, microbiota

**Files:**

Folder / file name	Explanation
Schreven_Ch3_data	Folder containing the datasets used for publication
Ch3_Dataset_1.txt	Data on larval performance parameters and substrate pH.
Ch3_Dataset_2.txt	Metadata of 16S rRNA gene amplicon sequencing dataset as deposited in the European Nucleotide Archive.
Ch3_Dataset_3.txt	File with data on pH of freshly prepared feed substrates (chicken feed and camelina) and defrosted reference samples of chicken manure batches used in the main experiment.
Ch3_Dataset_1_explanation.txt	Explanation of variables of Dataset 1.
Ch3_Dataset_2_explanation.txt	Explanation of variables of Dataset 2.
Ch3_Dataset_3_explanation.txt	Explanation of variables of Dataset 3.
Explanation_mocks.txt	Explanation of variables of dataset Mocks_composition.txt in input_data subfolder.
Schreven_Ch3_R_analysis	Folder containing the R project that includes all analyses of the publication. The .Rproj-file can be opened in Rstudio and will allow use of all R markdown files (.Rmd-files, containing the actual R scripts). Using these R markdown files, the user can produce the Tables, Figures and Supplements from the publication. The R scripts have also been exported as PDF-files to read the R script without use of R. R markdown file 3_subsetting.Rmd is essential to run prior to subsequent markdown files in order to create the required RDS-compressed formats of the phyloseq objects.
1. figures	Subfolder containing all figures produced in the R project
2. input_data	Subfolder containing the input files used in the R project:
Mocks_composition.txt	File with theoretical composition of the synthetic mock communities used as positive controls in sequencing.

Schreven_harvest_data.txt	File with data on larval performance and substrate pH. Contains same data as Ch3_Dataset_1.txt
Schreven_libAE_allotus.tre	TREE-file containing the phylogenetic relationships between ASVs of the sequences in the BIOM-file.
Schreven_libAE_mapping_file.csv	File with metadata on sequencing data. Contains same data as Ch3_Dataset_2.txt
Schreven_libAE_seqdata.biom1	BIOM-file containing the sequence data per sample.
Schreven_pHdiets_fresh.txt	File with data on pH of freshly prepared feed substrates (chicken feed and camelina) and defrosted reference samples of chicken manure batches used in the main experiment. Contains same data as Ch3_Dataset_3.txt
3. phyobjects	Subfolder containing all RDS-compressed formats of the phyloseq objects used in the R project. These files are created using R markdown file 3_subsetting.Rmd
4. renv	Subfolder required for reproducibility of the R environment (renv package).
5. tables	Subfolder containing all tables produced in the R project, and the following file:
Contaminant_OTU_by_plot.txt	File with the page numbers referring to ASVs (OTUs) that are likely contaminants, based on the PDF file Correlations_Taxa_DNA.pdf in the "figures" subfolder.

### Explanation of the variables:

Dataset\_1, Dataset\_2, and Dataset\_3 are accompanied with separate TXT-files with explanations of variables. Explanation of variables of the data files in the input\_data subfolder is also included in the folder Schreven\_Ch3\_data.

### Reproducing the figures and tables using the R markdown files:

For Chapter 3 of PhD thesis:

Figure/Table	Description	R markdown file
Figure 2	Larval performance and substrate pH	2_performance_pH.Rmd
Figure 3	Faith's phylogenetic diversity	5_alphadiversity.Rmd
Figure 4	NMDS substrate microbiota	7_betadiversity_nmds.Rmd
Figure 5	Heatmap most abundant genera	6_betadiversity_heatmaps.Rmd
Figure 6	NMDS larval and substrate microbiota	7_betadiversity_nmds.Rmd
Figure S1	Stacked bar graph bacterial phyla	6_betadiversity_heatmaps.Rmd
Figure S2	dbPRC substrate microbiota	10_betadiversity_dbPRC.Rmd
Figure S3	dbPRC larva vs substrate microbiota	10_betadiversity_dbPRC.Rmd
Figure S4	Heatmap of biofilms	4_quality_controls.Rmd
Table S1	Bacterial genera in NTCs	4_quality_controls.Rmd
Table S2	ASVs identified as contaminants	3_subsetting.Rmd
Table S3	ANOVA of phylogenetic diversity of substrate microbiota	5_alphadiversity.Rmd
Table S4	ANOVA of phylogenetic diversity of larval and substrate microbiota	5_alphadiversity.Rmd

Table S5	PERMANOVA output of dbRDA on substrate microbiota for all three diets combined	9_betadiversity_dbRDA.Rmd
Table S6	PERMANOVA output of dbRDA on substrate microbiota per diet separately	9_betadiversity_dbRDA.Rmd
Table S7	PERMANOVA output of dbRDA on larval and substrate microbiota for all three diets combined	9_betadiversity_dbRDA.Rmd
Table S8	PERMANOVA output of dbRDA on larval and substrate microbiota per diet separately	9_betadiversity_dbRDA.Rmd

For the publication in *Applied and Environmental Microbiology*:

Figure/Table	Description	R markdown file
Table 1	Larval performance and substrate pH	2_performance_pH.Rmd
Figure 2	Faith's phylogenetic diversity	5_alphadiversity.Rmd
Figure 3	NMDS larval and substrate microbiota	8_betadiversity_nmds_all.Rmd
Figure 4	Heatmap most abundant genera	6_betadiversity_heatmaps.Rmd
Table S1	Bacterial genera in NTCs	4_quality_controls.Rmd
Table S2	ASVs identified as contaminants	3_subsetting.Rmd
Table S3	PERMANOVA output of dbRDA on substrate microbiota for all three diets combined	9_betadiversity_dbRDA.Rmd
Table S4	PERMANOVA output of dbRDA on substrate microbiota per diet separately	9_betadiversity_dbRDA.Rmd
Table S5	PERMANOVA output of dbRDA on larval and substrate microbiota for all three diets combined	9_betadiversity_dbRDA.Rmd
Table S6	PERMANOVA output of dbRDA on larval and substrate microbiota per diet separately	9_betadiversity_dbRDA.Rmd
Figure S1	Shannon diversity	5_alphadiversity.Rmd
Figure S2	Stacked bar graph bacterial phyla	6_betadiversity_heatmaps.Rmd
Figure S3	dbPRC substrate microbiota	10_betadiversity_dbPRC.Rmd
Figure S4	dbPRC larval microbiota	10_betadiversity_dbPRC.Rmd
Figure S5	dbPRC larva vs substrate microbiota	10_betadiversity_dbPRC.Rmd
Figure S6	Heatmap of biofilms	4_quality_controls.Rmd

#### Methods, materials and software:

This information can be found in the publication, see: <https://doi.org/10.18174/542262> (Chapter 3 in document).

**Funding statement:**

This work was supported by the European Union's Horizon 2020 research and innovation programme (grant agreement no. 635405), as part of the COSMOS (Camelina & crambe Oil crops as Sources for Medium-chain Oils for Specialty oleochemicals) project. GDAH was supported by the project MASTER (Microbiome Applications for Sustainable food systems through Technologies and Enterprise) funded by the European Union's Horizon 2020 research and innovation programme (grant agreement no. 818368).

**Licence:** CC-BY-SA