

COST STSM

Gut microbiome of black soldier fly larvae – a meta-analysis of marker gene datasets

Thomas Klammsteiner, Blaž Stres

ML4Microbiome Final Conference

8 June 2023 - Cork, Ireland



STSM

- **When:**
April – May 2021
- **Where:**
University of Ljubljana
Biotechnical Faculty
Ljubljana, Slovenia
- **Host:**
Blaž Stres



Univerza v Ljubljani

Biotehniška fakulteta



Goals

- ✦ **Collect** all accessible 16S rRNA marker gene datasets on the **gut microbiome** of an **industrially relevant insect**
- ✦ **Merge + analyse** datasets using a **standardized workflow**
- ✦ Reduce/address **batch effects**
- ✦ Ensure **reproducible documentation**

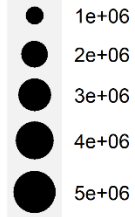
Why?

- Interdisciplinary studies at the intersection between **ecology, engineering, waste management, and animal science**
- Studies often published in journals **without microbiology focus**
 - Availability of data not mandatory
 - No specific author guidelines for microbiome data
 - Reviewers from other scientific fields
- Sequence data analysed/interpreted by **non-experts**
- Lacking **awareness** about best-practice approaches?

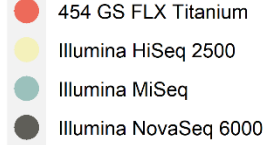
Data sets



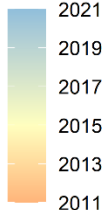
Generated reads



Instrument model



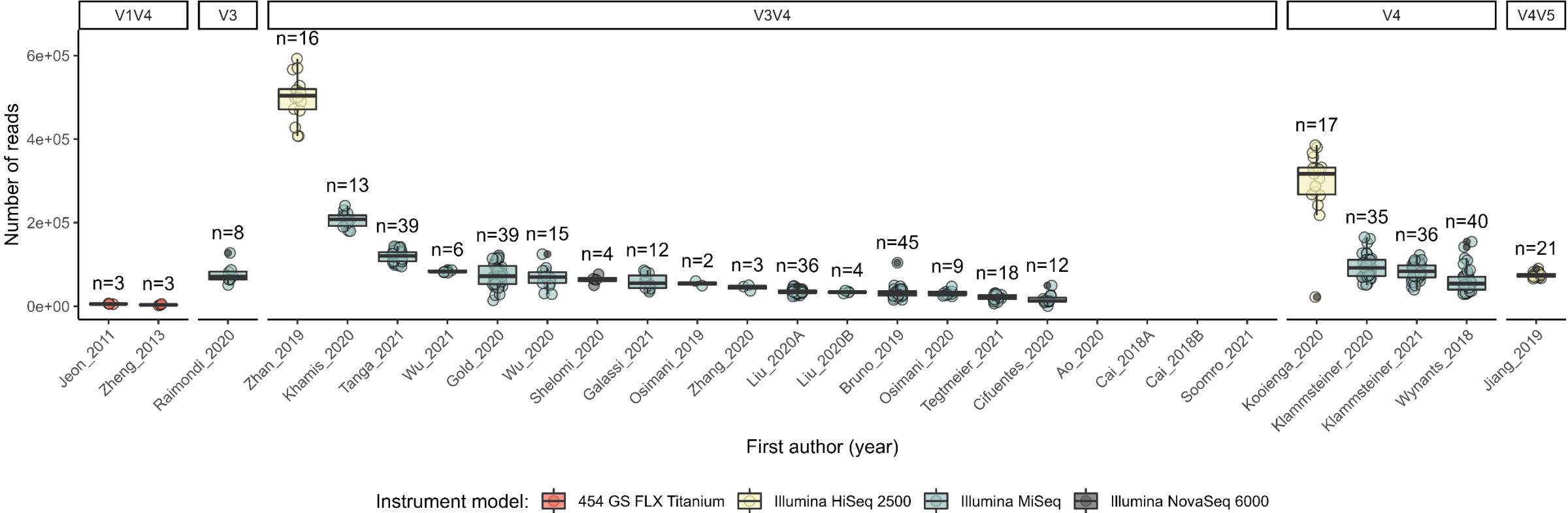
Year



Data sets

| Accession | First Author | Publ Date | Publ Lab | Publ ID | Seq T | Platform | Organism | S | Fw | Rev | Fw | Rev | ImSec | DOI | Jc |
|-----------|------------------|------------|------------------|-------------|------------|-------------|-----------|--------------------------|---------|--------|-------|--------------|---------|---|----|
| PRJEB23 | Jeon et al. | 2011-01-26 | Jeon et al. (2) | Jeon_2011 | Pyro | Roche454 | (Bacteria | V1V | 27F | 518R | GAGT | WTTA | 16741 | doi.org/10.1007/s00 | ✓ |
| PRJEB31 | Zheng et al. | 2013-05-01 | Zheng et al. (1) | Zheng_2013 | Pyro | Roche454 | (Bacteria | V1V | 28F | 519R | GAGT | GTNT | 17366 | doi.org/10.1603/mf | ✓ |
| NO DATA | Cai et al. | 2018-07-03 | Cai et al. (201) | Cai_2018B | Amplicor | Illumina Hi | (Bacteria | V3v | 341F | 806R | CCTAI | GGACTACNVG | | doi.org/10.1016/j.en | ✓ |
| NO DATA | Cai et al. | 2018-10-15 | Cai et al. (201) | Cai_2018A | Amplicor | Illumina Hi | (Bacteria | V3v | 341F | 806R | CCTAI | GGACTACNVG | | doi.org/10.1111/1462- | ✓ |
| PRJNA47 | Wynants et al. | 2018-11-14 | Wynants et al. | Wynants_20 | Amplicor | Illumina Mi | (Bacteria | V4 | 515F | 806R | GTGC | GGAC | 9038031 | doi.org/10.1007/s00 | ✓ |
| PRJNA47 | Bruno et al. | 2019-01-09 | Bruno et al. (1) | Bruno_2019 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 785R | CCTAI | GACT | 2581708 | doi.org/10.1128/AEM | ✓ |
| PRJNA40 | Jiang et al. | 2019-03-18 | Jiang et al. (2) | Jiang_2019 | Amplicor | Illumina Hi | (Bacteria | V4v | 515F | 907R | GTGC | CCGT | 4101142 | doi.org/10.1111/1751- | ✓ |
| PRJNA54 | Zhan et al. | 2019-11-25 | Zhan et al. (2) | Zhan_2019 | Amplicor | Illumina Hi | (Bacteria | V3v | 341F | 805R | CCTAI | GACT | 1.1E+07 | doi.org/10.1038/s41 | ✓ |
| PRJNA54 | Osimani et al. | 2019-12-10 | Osimani et al. | Osimani_20 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 805R | CCTAI | GACT | 549506 | doi.org/10.1371/jour | ✓ |
| PRJNA56 | Shelomi et al. | 2020-01-06 | Shelomi et al. | Shelomi_20 | Amplicor | Illumina No | (Bacteria | V3v | 341F | 805R | CCTAI | GACT | 550363 | doi.org/10.1093/eeh | ✓ |
| PRJNA70 | Wu et al. | 2020-04-01 | Wu et al. (20) | Wu_2020 | Amplicor | Illumina Mi | (Bacteria | V3v | 338F | 806R | GTAC | GTGG | 1030162 | doi.org/10.1016/j.ecv | ✓ |
| PRJEB33 | Klammstein | 2020-05-21 | Klammsteine | Klammstein | Amplicor | Illumina Mi | (Bacteria | V4 | 515F | 806R | GTGY | GGAC | 4001397 | doi.org/10.3389/fmi | ✓ |
| NO DATA | Ao et al. | 2020-05-25 | Ao et al. (202) | Ao_2020 | Amplicor | Illumina Mi | (Bacteria | V3v | 338F | 806R | ACTCI | GGACTACHVG | | doi.org/10.1111/1751- | ✓ |
| PRJNA63 | Liu et al. | 2020-06-12 | Liu et al. (202) | Liu_2020A | Amplicor | Illumina Mi | (Bacteria | V3v | 314F | 806R | CCTA' | GGAC | 1261263 | doi.org/10.1016/j.en | ✓ |
| PRJNA57 | Raimondi et al. | 2020-06-15 | Raimondi et al. | Raimondi_2 | Amplicor | Illumina Mi | (Bacteria | V3 | Probioc | Probio | CCTAI | ATTA | 1234955 | doi.org/10.3390/mic | ✓ |
| PRJNA57 | Cifuentes et al. | 2020-07-07 | Cifuentes et al. | Cifuentes_2 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 785R | CCTAI | GACT | 525869 | doi.org/10.1007/s10 | ✓ |
| PRJNA62 | Khamis et al. | 2020-07-07 | Khamis et al. | Khamis_202 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 785R | CCTAI | GACT | 2691975 | doi.org/10.3389/fmi | ✓ |
| PRJNA61 | Zhang et al. | 2020-11-03 | Zhang et al. (1) | Zhang_2020 | Amplicor | Illumina Mi | (Bacteria | V3v | 338F | 806R | ACTCI | GGAC | 447523 | doi.org/10.1007/s00 | ✓ |
| PRJNA64 | Gold et al. | 2020-11-23 | Gold et al. (2) | Gold_2020 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 806R | CCTAI | GGAC | 4550493 | doi.org/10.3389/fmi | ✓ |
| PRJNA66 | Kooienga et al. | 2020-11-24 | Kooienga et al. | Kooienga_21 | Amplicor | Illumina Hi | (Bacteria | V4 | 515F | 806R | GTGY | GGAC | 5007700 | doi.org/10.3389/fmi | ✓ |
| PRJNA64 | Liu et al. | 2020-12-15 | Liu et al. (202) | Liu_2020B | Amplicor | Illumina Mi | (Bacteria | V3v | 314F | 806R | CCTA' | GGAC | 135478 | doi.org/10.1016/j.sci | ✓ |
| PRJNA64 | Osimani et al. | 2020-12-18 | Osimani et al. | Osimani_20 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 785R | CCTAI | GACT | 914345 | doi.org/10.1016/j.foc | ✓ |
| PRJNA65 | Wu et al. | 2021-01-16 | Wu et al. (20) | Wu_2021 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 805R | CCTAI | GACT | 1009105 | doi.org/10.1016/j.ecv | ✓ |
| PRJEB39 | Klammstein | 2021-01-21 | Klammsteine | Klammstein | Amplicor | Illumina Mi | (Bacteria | V4 | 515F | 806R | GTGC | GGAC | 3798171 | doi.org/10.3389/fmi | ✓ |
| Data prov | Galassi et al. | 2021-01-22 | Galassi et al. | Galassi_202 | Amplicor | Illumina Mi | (Bacteria | V3v | Pro34' | Pro805 | CCTAI | GACT | 699513 | doi.org/10.1093/ise | ✓ |
| Data prov | Tanga et al. | 2021-02-12 | Tanga et al. (1) | Tanga_2021 | Amplicor | Illumina Mi | (Bacteria | V3v | 314F | 805R | CCTAI | GACT | 4665023 | doi.org/10.3389/fmi | ✓ |
| NO DATA | Yang et al. | 2021-02-16 | Yang et al. (2) | Yang_2021 | Transcrip | Illumina Hi | (Bacteria | | | | | | | doi.org/10.3389/fmi | ✓ |
| NO DATA | Soomro et al. | 2021-02-23 | Soomro et al. | Soomro_202 | Amplicor | Illumina Hi | (Bacteria | V3v | 341F | 806R | ACTCI | GGACTACHVG | | doi.org/10.3920/JIF | ✓ |
| NO DATA | Zhineng et al. | 2021-03-12 | Zhineng et al. | Zhineng_20 | Metagenome | | (Bacteria | | | | | | | https://doi.org/10.11 | ✓ |
| PRJNA67 | Tegtmeier et al. | 2021-03-29 | Tegtmeier et al. | Tegtmeier_2 | Amplicor | Illumina Mi | (Bacteria | V3v | 314F | 806R | CCTA' | GGAC | 711390 | https://doi.org/10.3 | ✓ |
| PRJEB40 | Schreven et al. | 2021-03-30 | Schreven et al. | Schreven_2 | Amplicor | Illumina No | (Bacteria | V5v | 784F | 1064R | FGGA | CGACRCCAT | | https://doi.org/10.10 | ✓ |
| PRJNA64 | Liu et al. | 2021-06-15 | Liu et al. (202) | Liu_2021 | Amplicor | Illumina Mi | (Bacteria | V3v | 314F | 806R | CCTAI | GGAC | 989208 | https://doi.org/10.3 | ✓ |
| PRJNA72 | Shumo et al. | 2021-09-22 | Shumo et al. | Shumo_202 | Metagen | Oxford Min | (Bacteria | | 27F | 1492R | AGAGI | GGTTACCTTGTT | | https://doi.org/10.3 | ✓ |
| PRJNA68 | Greenwood et al. | 2021-12-01 | Greenwood et al. | Greenwood_ | Metagen | Ion Torrent | (Bacteria | V2, V4, V8, V3, V6-7, V9 | | | | | | https://doi.org/10.3 | ✓ |
| PRJNA76 | Li et al. | 2021-12-06 | Li et al. (202) | Li_2021 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 805R | CCTAI | GACTACHVGE | | https://doi.org/10.3 | ✓ |
| PRJNA42 | Marasco et al. | 2022-01-05 | Marasco et al. | Marasco_20 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 785R | CCTAI | GACTACHVGE | | https://doi.org/10.11 | ✓ |
| NO DATA | Mei et al. | 2022-01-15 | Mei et al. (20) | Mei_2022 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 806R | CCTAI | GGACTACHVG | | https://doi.org/10.10 | ✓ |
| PRJNA74 | Gorrens et al. | 2022-03-27 | Gorrens et al. | Gorrens_20 | Amplicor | Illumina Mi | (Bacteria | V4 | 515F | 806R | GTGC | GGACTACHVG | | https://doi.org/10.3 | ✓ |
| NO DATA | Luo et al. | 2022-04-01 | Luo et al. (20) | Luo_2022 | Amplicor | Illumina Mi | (Bacteria | V3v | 341F | 785R | CCTAI | GACTACHVGE | | https://doi.org/10.10 | ✓ |
| PRJNA72 | Klueber et al. | 2022-05-06 | Klueber et al. | Klueber_202 | Amplicor | Illumina Mi | (Bacteria | V3v | U341F | U806R | CCTA' | GGACTACNNE | | https://doi.org/10.3 | ✓ |
| PRJNA78 | Pei et al. | 2022-05-19 | Pei et al. (202) | Pei_2022 | Amplicor | Illumina Mi | (Bacteria | V3v | 338F | 806R | ACTCI | GGACTACHVG | | https://doi.org/10.3 | ✓ |
| PRJNA75 | Li et al. | 2022-07-11 | Li et al. (2022) | Li_2022 | Amplicor | Illumina No | (Bacteria | V3v | 338F | 806R | ACTCI | GGACTACHVG | | https://doi.org/10.11 | ✓ |
| NO DATA | Zhang et al. | 2022-07-26 | Zhang et al. (1) | Zhang_2022 | Amplicor | Illumina Mi | (Bacteria | V3v | 338F | 806R | ACTCI | GGACTACHVG | | https://doi.org/10.11 | ✓ |
| PRJNA85 | Querejeta et al. | 2022-11-25 | Querejeta et al. | Querejeta_2 | Amplicor | Illumina Mi | (Bacteria | V3v | S-D-B | S-D-B | CCTAI | GACTACHVGE | | https://doi.org/10.10 | ✓ |
| PRJNA87 | Yu et al. | 2023-01-10 | Yu et al. (202) | Yu_2023 | Amplicor | Illumina Mi | (Bacteria | V3v | 338F | 806R | ACTCI | GGACTACHVG | | https://doi.org/10.3 | ✓ |
| PRJNA78 | Xu et al. | 2023-01-15 | Xu et al. (202) | Xu_2023 | Amplicor | Illumina Hi | (Bacteria | V4v | 515F | 806R | GTGC | GGACTACHVG | | https://doi.org/10.10 | ✓ |
| NO DATA | Hao et al. | 2023-03-02 | Hao et al. (20) | Hao_2023 | Amplicor | Illumina Mi | (Bacteria | V3v | 338F | 806R | GTAC | GTGGACTACH | | https://doi.org/10.3 | ✓ |
| NO DATA | She et al. | 2023-05-15 | She et al. (20) | She_2023 | Amplicor | Illumina Hi | (Bacteria | V3v | 338F | 806R | ACTCI | GGACTACHVG | | https://doi.org/10.10 | ✓ |

Data sets



What is a **healthy** larval microbiome?

High-performance computing lifts the understanding of insect-based gut microbiomes

Thomas Klammsteiner^a, Heribert Insam^b, Bostjan Murovec^b, Blaz Stres^{c,d,e}

^aDepartment of Microbiology, University of Innsbruck, Innsbruck, Austria
^bFaculty of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia
^cDepartment of Animal Science, University of Ljubljana, Ljubljana, Slovenia
^dInstitute of Sanitary Engineering, University of Ljubljana, Ljubljana, Slovenia
^eFaculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

Background

As the flagship species of an emerging insect biotechnology with the goal to convert organic wastes into a source of insect-based protein and fat, the black soldier fly (BSF; *Hermetia illucens*) has awoken popular interest in science and industry. Research showed, a complex network of microbial gut communities strongly contributes to waste degradation processes and shapes the larval environment [1, 2]. However, the lack of a best practice approach in preparing and analyzing raw sequence data affects the informative value and reproducibility between studies. Thus, a comprehensive re-analysis of published data following a standardized protocol in the context of a meta-analysis could drastically improve comparability and enable robust conclusions across studies (Fig. 1.). The volume of sequencing data and the iterative preprocessing optimization require for high computational effort, creating the need for parallelization of analysis tools on high-performance computing (HPC) platforms.

Results

In total, we collected and curated 32 studies carried out in the past 10 years, comprising more than a thousand samples from primarily larvae, substrates, and residues (Tab. 1.). As indicated by the available metadata, the study collection showed a high diversity for sequencing approaches in terms of sequencing platforms, protocols, depth, and targeted variable regions on the 16S rRNA gene. Moreover, a high variability in experimental designs including sample size, rearing conditions, and composition of fed diets was observed. Consequently, binning and processing datasets with similar sequence features was needed. The V4 region was then isolated for downstream analyses, as it had been covered by 96% of all suitable studies. After isolating the most frequently targeted gene segment, the preprocessed datasets were merged and prepared for taxonomic and functional analysis using GUMPP [3].

Conclusion

Microbiome meta-studies as means to summarize past studies by applying reproducible and well-established protocols have the potential to point out informative cross-study patterns. However, inconsistent or missing metadata and sequence data unavailability hinder such analyses. In addition, to allow for more efficient processing of sequence data on HPC platforms, options for parallelization of analysis tools need improvement.

Funding

This research was funded by a short-term scientific mission grant from EU COST action "Statistical and machine learning techniques in human microbiome studies" (CA18131).

References

- Klammsteiner, T., Walder, A., Bogataj, T., Houshali, C.D., Stres, B., Stres, F.H., Schick-Spanner, B.C., Insam, H., 2021. Impact of Processed Food (Canteen and Oil Wastes) on the Development of Black Soldier Fly (*Hermetia illucens*) Larvae and Their Gut Microbiome Functions. *Front. Microbiol.* 12. <https://doi.org/10.3389/fmicb.2021.619112>
- Klammsteiner, T., Walder, A., Bogataj, T., Houshali, C.D., Stres, B., Stres, F.H., Schick-Spanner, B.C., Artstiller, B., Insam, H., 2020. The core gut microbiome of black soldier fly (*Hermetia illucens*) larvae raised on low-bioburden diets. *Front. Microbiol.* 11, 905. <https://doi.org/10.3389/fmicb.2020.300993>
- Murovec, B., Debatich, L., Stres, B., 2021. General unified microbiome profiling pipeline (GUMPP) for large scale, streamlined and reproducible analysis of bacterial 16S rRNA data to predicted microbial metabolism, enzymatic reactions and metabolic pathways. *Metabolites*. <https://doi.org/10.3390/metab11060239>

Searching for consensus in black soldier fly microbiomes

A cross-study analysis of published datasets
 Klammsteiner, T.¹, Murovec, B.², Insam, H.¹, Stres, B.^{3,4,5}
 Poster: PS18.555

In brief

Why the black soldier fly?
 Converts decaying organics into own biomass
 High in nutrients (>40% protein, >30% fat)
 Chitin-rich residues as organic fertilizer
 Relevance of its gut microbiome
 Broad digestion capabilities of the larvae
 Larva's protection against pathogens
 Improve waste conversion process

Conclusion

Sequencing has become an easily accessible and economically feasible method also for disciplines outside of microbiology, however, special attention should be paid to reproducibility and standardization. Despite these variations, our meta-analysis pointed out that high bioburden diets may lead to an intestinal dysbiosis in black soldier fly larvae. A diet-independent core microbiome, however, might help to sustain larval health and support digestion.

Figure 1. Overview on study sizes, sequencing depths, and sequencing platforms.

Figure 2. CCA with "feed type" as a constraining variable.

Background

The larvae of the industrially relevant black soldier fly naturally live in direct contact with decomposing organic matter, which they can degrade efficiently. The high microbial load in their environment and the larvae's remarkable digestive abilities point out the great importance of microorganisms for the insect's thriving.

Study design

16S rRNA marker gene datasets generated over the past decade adding up to more than 1000 samples of the gut and diet microbiome of black soldier fly larvae were collected. The workflow presented in Fig. 3 was followed to prepare and analyze the data.

Main findings

Close to 20% of studies did not make their sequence data publicly available after publication and meta-data accompanying the sequence data was in multiple cases incomplete or inaccurate. However, despite the large variation in experimental set-ups (Fig. 1), groups such as *Enterococcus*, *Providencia*, *Morganella*, and *Actinobacteria* were found in high abundances in the majority of samples. Diets carrying a presumably higher bioburden led to an increased diversification of the microbiome compared to standard diets such as wheat bran (Fig. 2).

Workflow from data collection to analysis.

Project website
 Further information, data, and downloads available at: www.mikroclimate.eu
[klammsteiner.github.io/metaguts](https://github.com/klammsteiner)

Contact:
 Thomas Klammsteiner, PhD
 University of Innsbruck
 Department of Microbiology
 Department of Ecology
 6020 Innsbruck, Austria
 Email: thomas.klammsteiner@uibk.ac.at
 Phone: +43 512 507 51322
 Web: klammsteiner.github.io

This research was funded by a STSM grant within the COST action CA18131 (<https://www.mikroclimate.eu/>)

Cost European Cooperation in Science and Technology
universität innsbruck
 Black Soldier Fly Innsbruck
www.innsbruck.ac.at
www.black-soldier-fly.com
www.trawnsaustria.at

Benefits

- Collaboration with thesis supervisor
- Finish up work for the completion of PhD
- **Dissemination:** Opportunities to present own work
- **Skill development:** New methods & approaches
- **Personal development:** Lockdowns, COVID, comfort zone, etc.
- **Broader perspective:** escape the daily grind

Learnings

- Advocate for open data and **FAIR** principles
- Highlight importance of **reproducibility** wherever possible
- Be **fast** 😊



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

New Results

 [Follow this preprint](#)

Meta-analysis of the black soldier fly (*Hermetia illucens*) microbiota based on 16S rRNA gene amplicon sequencing

Freek Ijdema, Jeroen De Smet, Sam Crauwels, Bart Lievens, Leen Van Campenhout

doi: <https://doi.org/10.1101/2022.01.17.476578>

Acknowledgements

Collaborators

- Blaž Stres
- Bostjan Murovec
- Heribert Insam

STSM coordinator

- Tatjana Loncar-Turucalo



Univerza v Ljubljani

Biotehniška fakulteta