Investigating genetic diversity and microRNA of *Hermetia illucens* (the black soldier fly) to breed for mass production of a novel sustainable protein.

SARAH JOANNE DE RAEDT

A thesis submitted in fulfilment of the requirements for the degree of Doctor Philosophy in Bioinformatics at the South African National Bioinformatics Institute, University of the Western Cape.

UNIVERSITY of the
WESTERN CAPE

2023

Supervisor: Prof Alan Christoffels

http://etd.uwc.ac.za

Investigating genetic diversity and microRNA of *Hermetia illucens* (the black soldier fly) to breed for mass production of a novel sustainable protein.

by S. J. De Raedt

Keywords

Hermetia illucens

Black soldier fly

Sustainable protein

Mass rearing

microRNA expression

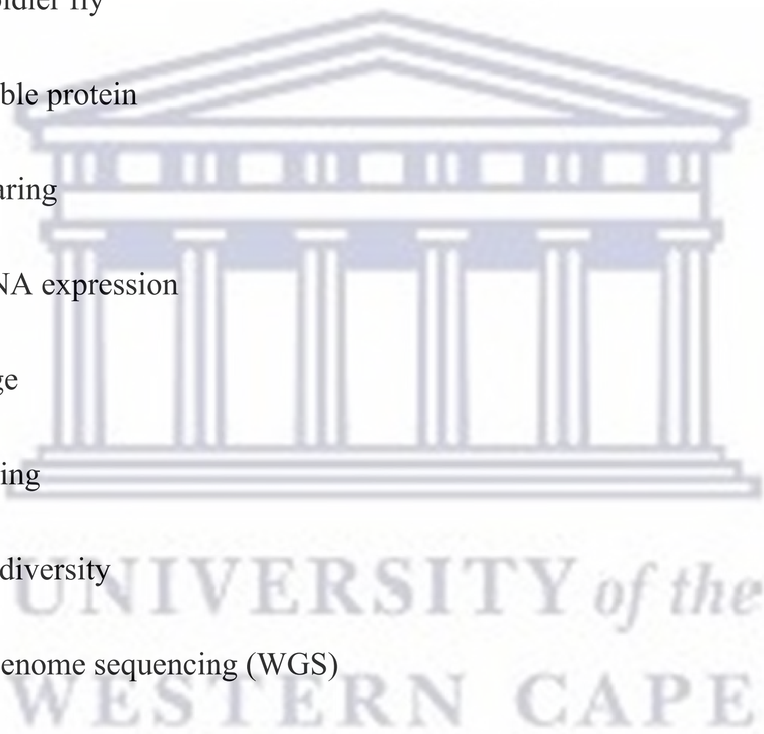
Life stage

Sequencing

Genetic diversity

Whole genome sequencing (WGS)

Variant calling


UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

Abstract

Investigating genetic diversity and microRNA of *Hermetia illucens* (the black soldier fly) to breed for mass production of a novel sustainable protein.

by S. J. De Raedt (PhD Thesis, South African National Bioinformatics Institute, University of the Western Cape).

Introduction: A new sustainable source of protein is needed to meet the demands of the growing global population. Insect meal is a suitable replacement, and the black soldier fly (BSF) is the most used insect in industrial rearing. The black soldier fly larvae (BSFL) are not only high in sought-after nutrients (protein, fat, and chitin/source of fiber), but they also reduce organic waste that would go into landfills by consuming the waste and leaving behind a beneficial residue, which is used in fertilizers. However, little has been published on the genetics of BSF which are crucial to optimizing mass breeding programs necessary to meet the population demands. The aim of this study was to further the base of knowledge beneficial to mass rearing protocols by describing the genetic diversity of 3 populations, under differing scales of rearing, and the microRNA expression profile across 5 life stages, along with the first report of the novel microRNA of BSF.

Methods: Three African populations of BSF reared at large-, medium- and small-scale were selected for whole genome sequencing (WGS) of male and female flies. Chromosomal regions of interest were discovered through genome nucleotide diversity and Tajima's D statistical analysis. Fifteen microRNA libraries were sequenced covering the egg, larva, pupa, adult unmated female, and adult unmated

http://etd.uwc.ac.za

male flies stages of the life cycle. Differential expression was ascertained and categorized by the log fold change (LFC) between life stages highlighting microRNA with significant changes in expression across and specific to a life stage.

Results: Many high-quality single nucleotide polymorphisms (SNPs) were found across all populations with small-scale reared flies producing the most unique SNPs and highest nucleotide diversity. Large- and medium-scale reared flies were shown to be indistinguishable from each other genetically, but medium-scale reared did retain more unique SNPs and higher genetic diversity than large-scale reared flies. The large- and medium-scale reared populations had six regions of interest on four chromosomes that were identified as potential areas undergoing selection pressure. Of the 192 novel microRNAs found across the five life stages, 24 were unique to BSF and 168 were orthologous to other arthropod species. A search of experimentally validated *Drosophila* target genes resulted in 56 genes from 22 conserved microRNAs as predicted target genes for BSF.

Conclusions: This study finds the loss of genomic diversity occurs strongly between small- and medium-scale reared populations and identifies chromosomal regions undergoing potential selection pressures. Additionally, novel BSF microRNAs and potential target genes were reported along with a life stage and sex expression profile. The findings of this study expand on the rapidly growing genetic landscape needed to mass rear the economically and environmentally important black soldier fly.

http://etd.uwc.ac.za

Declaration

I declare that *Investigating genetic diversity and microRNA of Hermetia illucens (the black soldier fly) to breed for mass production of a novel sustainable protein* is my own work (except where acknowledgements indicate otherwise), that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full name: Sarah Joanne De Raedt August 2023

Signed: ………………………………

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

Acknowledgements

This work was financially supported by the South African Research Chairs Initiative of the Department of Science and Technology, the National Research Foundation of South Africa, and AgriProtein Technologies (PTY) Ltd.

There are so many people I would like to thank for all the support, motivation, help, and time that has been given to me throughout this process. I would first like to thank my supervisor, Prof Alan Christoffels of the South African National Bioinformatics Institute (SANBI) at the University of the Western Cape. He took a chance on an unknown Medical Entomologist looking for a supervisor in bioinformatics and gave me the support I needed when the project seemed insurmountable. His patience in teaching me that the way to succeed with coding is to try, try, try until it works was invaluable. I thank you so much!

Second, I would like to thank Dr. Cameron Richards and AgriProtein for having the vision to fund a project benefiting the future of the black soldier fly industry. I learned so many life lessons for which I am grateful. Many of the people working at AgriProtein helped to make this work possible, but especially the friendship and discussions I had with Olivia Erispe, Bridget Ristow, Noxolo Sibhozo, and Anandi Bierman. Rozane Badenhorst and Riaan Rispel, a value cannot be put on our nights discussing buckets!

There is no doubt that I would not have been able to complete this project without the help and support of the faculty, staff, and fellow students of SANBI. I am so grateful for the hours of discussions, problem-solving, understanding, and learning that I had with everyone. I would like to especially thank Dominique Anderson for

http://etd.uwc.ac.za

her friendship, support, and general awesomeness. Also, thank you to Tracey Calvert-Joshua and Linley Abdoll for our study group and conversations. To the guru of coding, Peter van Heudsen, I am beyond grateful for the patience and insight you showed me as I fumbled through learning to code. Thank you to Catherine Rossouw for your friendship and editing!

I am so lucky to have found the support of so many wonderfully intelligent, strong, and understanding friends from all around the world. I would like to thank Derica Lambrechts and Nuša Tucik, for the many video calls and never-failing words of wisdom. Kezia Spicer, Peter Winskill, Sophie Grant, Allie Robinson, Vanessa Chen-Hussey, Andrew Riley, Angela Rory Kay, James Logan, and Nicki Deidrick, you all have made this journey better and gave me the confidence to finish with your time, encouragement, and the many drinks along the way! I also thank my two oldest friends, Stephanie Nocella and Bekah Bradish, for listening to me talk about my thesis when I am sure it was boring, always finding a way to make me laugh and making me know you had my back no matter what!

I want to thank my parents for all they have provided me throughout my life and teaching me I could do anything I put my mind to. I also want to thank my mother- and father-in-law for all of their support and encouragement.

I have been lucky enough to have two sisters that I call my best friends. To Mary, who talked me through anxieties, reminded me that I was not insane, and constantly shined a light for me when all I could see was a tunnel. I am eternally grateful to have you in my life. And to Kate, who always cut through the crazy, pulled me back to reality and lent me her strength when I thought I had none left. I do not know

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

where I would be without you and luckily, I never had to find out. The three of us are never alone, because we hear and have each other inside our heads (whether we want to in the moment or not)!

And finally, to my husband, Ian. I sometimes wonder how I functioned before I found the missing parts of me that you are. You brought joy when I couldn't, vision when I struggled to see, and understanding when all I wanted to do was quit. You make my life better than I could have ever imagined, and I can't wait to be the Drs. Banks together!

http://etd.uwc.ac.za

List of abbreviations

AVG	Average of 3 biological replicates
BCE	Before the common (or current) era
BSF	Black soldier fly (*Hermetia illucens*)
BSFL	Black soldier fly larvae
BSQR	Recalibration of base quality scores
CHR	Chromosome
CE	Common (or current) era
CRISPR/Cas9	clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9
°C	Degrees Celsius
ddRADSeq	Double digestion Double digest restriction-site associated DNA sequencing
DNA	Deoxyribonucleic acid
E	Egg
EU	European Union
F	Adult unmated female
$F_{(IS)}$	Inbreeding coefficient

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za

FS	Fisher Strand (Phred-scaled probability that there is a strand bias at the site)
GATK	Genome analysis toolkit (Broad Institute)
GBS	Genotype by sequencing
GVCF	Genomic variant call format
H. illucens	*Hermetia illucens*
IPIFF	International Platform of Insects for Food and Feed
InDels	Insertion/Deletions
kb	Kilobase
kg	kilogram
L	Larva
LFC	Log fold change
LS	Large-scale
M	Male
MS	Medium-scale
mRNA	Messenger RNA
MQ	Root mean square (mapping quality over all the reads at the site)
MQRS	Mapping quality rank sum test (U-based z-approximation from the rank sum test for mapping qualities)

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za

ND, π	Nucleotide diversity
P	Pupa
PC	Principal component
PCA	Principal component analysis
PQV	Phred quality score
PCR	Polymerase chain reaction
QD	Quality by depth (variant confidence divided by unfiltered depth of non-homozygous reference samples)
RNA	Ribonucleic acid
RPRS	Read position rank sum test (U-based z-approximation from the rank sum test for site position within reads)
SNP	Single nucleotide polymorphism
SOR	Strand odds ratio (strand bias that considers the ratios of reads that cover both alleles)
SS	Small-scale
TD	Tajima's D
Ts/Tv	Transition vs transversion ratio
VCF	Variant call format
WG	Whole genome
WGS	Whole genome sequencing

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za

±	Plus, or minus

http://etd.uwc.ac.za

List of figures

Figure 1: *Hermetia illucens* (black soldier fly) life cycle and potential mass rearing locations. Each life stage and average length of time for the stage to develop is listed in the space following the image. Images are of each life stage and the substrate/containers in which they are mass reared. 36

Figure 2: Schematic representation of insect production in a circular food production system. Insects can transform organic waste into high-quality animal protein for food and feed. Here, we discuss the possible use of insect residual streams as soil amendments to stimulate beneficial microbiota and improve soil and crop health. (© Trends in Plant Science 2022) 39

Figure 3. Domestication process and insect domestication level. Numbers 0–5 refer to the domestication levels [12, 18]. Characteristics of each domestication level are provided on the left. Lines and points near the insect species names show the range of domestication degrees observed among populations of the species. (© Animal Domestication 2018) 41

Figure 4: Steps of microRNA binding to the 3′ UTR region of the target mRNA (adapted from © Vasulu, 2015) microRNA binds to mRNA causing down-regulation of the mRNA target gene. (mRNA = messenger RNA) 45

Figure 5: Principal Component Analysis (PCA) of three populations of *Hermetia illucens* using 14 individual's single nucleotide polymorphisms. Populations were large-scale (n=2), medium-scale (n=6), and small-scale (n=6) reared BSF. Only the small-scale population displayed distinct clusters, while large- and medium-scale populations were indistinguishable. Percent variation explained by Principal

http://etd.uwc.ac.za

Component 1 (PC1) and Principal Component 2 (PC2) in parentheses. Populations are colour coded with large--scale as red, medium-scale as green, and small-scale as blue. Large-scale is hard to see as it clustered with medium- and small-scale clusters in the centre. 55

Figure 6: Venn diagram of unique & shared SNPs between three populations of *Hermetia illucens*. The large-scale population (n=2) is represented in yellow; medium-scale population (n=6) in green; and small-scale population (n=6) in blue. Shared SNPs are in overlapping sections of circles with medium- and small-scale populations sharing the most SNPs, and SNPs shared by all three populations in the centre overlapped by all three circles. Numbers are in the millions. 56

Figure 7: Whole genome sliding window box plot of nucleotide diversity (π) for three populations of *Hermetia illucens.* Sliding windows of 20kb nucleotides were used to cover whole genomes for each population. The small-scale population contained the highest π of the three populations. The large-scale population (n=2) is represented in yellow; medium-scale population (n=6) in blue; and small-scale population (n=6) in green. 57

Figure 8: Chromosomal sliding window box plot of nucleotide diversity (π) for three populations of *Hermetia illucens*. Sliding windows of 20kb nucleotides were used to cover chromosomal genomes for each population. The large-scale population (n=2) is represented in yellow; medium-scale population (n=6) in blue; and small-scale population (n=6) in green. (Chrom # = chromosomal number).. 58

Figure 9: Chromosomal sliding window box plot of nucleotide diversity (π) of three populations of male and female *Hermetia illucens*. Sliding windows of 20kb nucleotides were used to cover chromosomal genomes for each population. The

http://etd.uwc.ac.za

large-scale male population (n=1) is represented in light yellow and female (n=1) in dark yellow; medium-scale male population (n=3) in light blue and female (n=3) in dark blue; and small-scale male population (n=3) in light green and female (n=3) in dark green. (Chrom # = chromosomal number)... 59

Figure 10: Mean Tajima's D sliding window chromosomal values across three *Hermetia illucens* populations. Sliding windows of 20kb nucleotides were used to span chromosomal regions. The large-scale population is represented in yellow, medium-scale in green, and small-scale in blue. (Chr# = chromosome number; WG = whole genome).. 60

Figure 11: Percent positive and negative Tajima's D of three populations of *Hermetia illucens* by chromosome. Green themed brackets are positive, red themed brackets are negative. Neutral (0) Tajima's D were found only in the LS population and were too small in scale to visualize therefore, a black line indicates neutral. 62

Figure 12: Nucleotide diversity scatter plots with regions of interest along the *Hermetia illucens* genome. (A) Chromosome 1 and 2 of the LS population regions with Tajima's D highly positive (> 2) values highlighted in red. (B) Chromosome 1 and 5 of the MS regions with Tajima's D highly negative (< -2) values highlighted in red. 64

Figure 13: Heatmap of microRNA expression of *Hermetia illucens* with a log fold change (LFC) of ±10 (p-adjusted value <0.05) across life stages. Expression is based on an average read count of triplicate samples for each life stage............. 78

Figure 14: Heatmap of unique microRNA expression of *Hermetia illucens* across life stages. Expression is based on an average read count of triplicate samples for each life stage... 78

http://etd.uwc.ac.za

List of tables

Table 1: Whole Population SNP Summary by genome and chromosome of *Hermetia illucens*. Abbreviations: Ts/Tv = Transition / Transversion; F = inbreeding coefficient.. 54

Table 2: Summary of *Hermetia illucens* reads sequenced and mapped. Abbreviations: E =egg; L= larva; P= pupa; F=female; M=male; 1-3=replicate; avg=average of 3 replicates.. 75

Table 3: microRNA with statistically significant regulation between life stages of novel *Hermetia illucens* microRNA with the same seed as known arthropod microRNA. 77

Table 4: Summary of *Hermetia illucens* microRNAs differential expression log fold change (LFC). 79

Table 5: Novel *Hermetia illucens* microRNA with orthologous microRNA and their predicted target genes in Drosophila. Abbreviations: LFC = Log Fold Change between life stages listed... 80

Table S1: Tajima's D values by chromosome for three populations of *Hermetia illucens*. Chr# = chromosome #; LS = Large-Scale population (n=2); MS = Medium-Scale population (n=6); SS = Small-Scale population (n=6). 103

Table S2: Conserved *Hermetia illucens* microRNA identified in other arthropod species. microRNA provisionally named using hil for species and number taken from microRNA identified in other arthropod species. Identical locations collapsed and those with multiple locations included below the bold. If no other location is

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za

listed, all microRNA were found on the same location. *Life stage abbreviations: E=Egg; L=Larva; P=Pupa; F=Female; M=Male 104

Table S3: Unique *Hermetia illucens* microRNA not identified in any other species. microRNA provisionally named; identical locations collapsed and those with multiple locations included below the bold. If no other location is listed, all microRNA were found on the same location. *Life stage abbreviations: E=Egg; L=Larva; P=Pupa; F=Female; M=Male... 109

Table S4: Unique *Hermetia illucens* microRNA log fold change (LFC) and P-Adjusted Value between life stages. microRNA provisionally named; Green filled boxes represent statistically significant up regulated LFC; Red filled boxes represent statistically significant down regulated LFC...................................... 110

Table S5: Conserved *Hermetia illucens* microRNA log fold change (LFC) and DESeq2 P-Adjusted Value between life stages. microRNA provisionally named; Green filled boxes represent statistically significant up regulated LFC; Red filled boxes represent statistically significant down regulated LFC............................ 111

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

Table of contents

Keywords ... ii

Abstract ... iii

Declaration ... v

Acknowledgements ... vi

List of abbreviations ... ix

List of figures ... xiii

List of tables ... xvi

Table of contents ... xviii

Publications from this thesis ... xxiv

Chapter 1: Introduction to thesis and research rationale ... 25

1.1 Summary of the literature ... 25

1.2 Research rationale ... 26

1.3 Aims and objectives of thesis research project ... 28

1.4 Thesis overview ... 29

Chapter 2: Literature review ... 31

2.1 Food sustainability ... 31

2.1.1 Increase in protein demand ... 32

2.2 Insects for food and feed ... 32

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za

2.2.1 Entomophagy ... 32

2.2.2 Insects for feed in agriculture ... 34

2.3 *H. illucens*/Black soldier fly (BSF) ... 35

2.3.1 Life cycle ... 35

2.3.2 High value products and sustainability of BSF ... 37

2.3.3 Animal feed ingredients ... 37

2.3.4 Fertilizer ... 38

2.4 How to avoid unsustainable issues of mass rearing ... 39

2.4.1 Population genetics ... 42

2.4.2 microRNA ... 43

2.5 Current knowledge of BSF genetics ... 46

2.6 Conclusions ... 46

Chapter 3: The impact of scaled rearing on the genetic diversity among three African populations of *Hermetia illucens* (black soldier fly) ... 48

3.1 Introduction ... 48

3.2 Methods ... 49

3.2.1 Sample collection ... 49

3.2.2 DNA extraction ... 50

3.2.3 Sequencing ... 50

3.2.4 Read processing and analysis ... 51

3.3 Results ... 53

UNIVERSITY *of the* WESTERN CAPE

http://etd.uwc.ac.za

3.3.1 Population diversity 53

3.3.2 Principal component analysis (PCA) highlight population similarities 54

3.3.3 Unique and shared SNPs between BSF populations 55

3.3.4 SS rearing produces nucleotide diversity 56

3.3.5 Tajima's D (TD) more extreme in MS and LS rearing 60

3.3.6 Chromosomal regions of interest under potential selection pressure 63

3.4 Discussion 65

3.4.1 Limitations and future work 68

3.4.2 Conclusion 68

Chapter 4: microRNA profile of *Hermetia illucens* (black soldier fly) and its implications on mass rearing 70

4.1 Introduction 70

4.2 Methods 71

4.2.1 Sample collection 71

4.2.2 microRNA extraction 72

4.2.3 Sequencing 73

4.2.4 Read processing and analysis 73

4.2.5 Differential expression 73

4.3 Results 74

http://etd.uwc.ac.za

4.3.1 Sequencing and mapping ... 74

4.3.2 Identification of novel microRNAs in *Hermetia illucens* ... 75

4.3.3 Differential expression ... 76

4.3.4 Predicted target genes ... 79

4.4 Discussion ... 81

4.4.1 Stage specific microRNA expression ... 82

4.4.2 microRNA target prediction ... 83

4.4.3 Future work ... 84

4.4.4 Conclusions ... 85

Chapter 5: Conclusions and future work ... 86

5.1 Key findings ... 86

5.2 Limitations and future work ... 88

5.3 Concluding remarks ... 89

References ... 91

Appendix A: Supplementary tables ... 103

A.1 Supplementary Table 1 ... 103

A.2 Supplementary Table 2 ... 104

A.3 Supplementary Table 3 ... 109

A.4 Supplementary Table 4 ... 110

A.5 Supplementary Table 5 ... 111

Appendix B: Variant calling pipeline ... 120

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za

B.1 QC and trimming sample whole genomes: 120

B.1.1 Fastqc 120

B.1.2 Trimmomatic and concatenating.. 120

B.2 Reference genome preparation:.. 120

B.2.1 BWA indexing of reference genome.. 120

B.2.2 Create GATK sequence dictionary 120

B.2.3 Create Samtools reference genome index 120

B.3 Sample alignment, sorting, marking duplicates, and read group verification: 121

B.3.1 Align, index, and sort.. 121

B.3.2 Mark duplicates.. 121

B.3.3 Read group confirmation 122

B.4 GATK haplotype caller: 122

B.4.1 Create GVCF... 122

B.4.2 Create GenomicsDBImport database... 123

B.4.3 Create VCF, Genotype GVCFs in GenomicsDBImport database 124

B.5 VCF filtering: 124

B.5.1 Select SNPs / remove InDels 124

B.5.2 Density plots for hard filtering parameters 124

B.5.3 Hard filtering.. 125

B.6 VCF Analysis: 125

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za

B.6.1 PCA ... 125

B.6.2 Nucleotide Diversity ... 125

B.6.3 Tajima's D ... 126

B.6.4 Heterozygosity/Inbreeding coefficient ... 126

B.6.5 Unique & shared SNPs ... 126

Appendix C: microRNA discovery pipeline ... 128

C.1 QC and trimming sample sequences: ... 128

C.2 Reference genome and conserved microRNA prep: ... 128

C.2.1 Concatenate reference genome files ... 128

C.2.2 Bowtie index reference genome ... 128

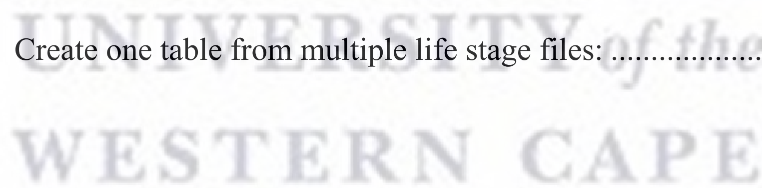
C.2.3 Select appropriate conserved microRNA ... 128

C.3 Map sample microRNA to reference genome: ... 129

C.3.1 Mapper.pl from MiRDeep2 software ... 129

C.4 Call microRNA using miRDeep2: ... 130

C.5 Create one table from multiple life stage files: ... 130



http://etd.uwc.ac.za

Publications from this thesis

1. DeRaedt S, Bierman A, Van Heusden P, Richards C, Christoffels A. microRNA profile of *Hermetia illucens* (black soldier fly) and its implications on mass rearing. PLOS ONE. 2022;17(3):e0265492.

http://etd.uwc.ac.za

Chapter 1: Introduction to thesis and research rationale

1.1 Summary of the literature

The Earth's population is set to reach 9.7 billion by the year 2050, with projections indicating food production demands will increase by 50% of current levels by the same year (1, 2). Unsustainable farming methods and production practices, along with climate change have led to a predicted 25% reduction of food in that time (1). The growing population is located mainly in low- and middle-income economies, and growth in these groups is linked to higher demands in protein for food (1-4). Agriculture utilizes large amounts of land (40% of all land), water, and fertilizer in the production of food, with 28% of that land being devoted to producing feed for the protein sector (1, 4-6). As precious as food is, 21% of the food produced is either wasted or lost (7, 8). Insects have been a part of the food chain throughout history as food, medicine, and agriculture (9-12). They are an excellent source of nutrients, require less land and water usage than traditional products, and valorise organic waste streams into high value products (13-15). A major challenge facing the industry is scalability in order to meet demand (10, 16). *Hermetia illucens* (the black soldier fly; BSF) is the "crown jewel" of the industry because of the beneficial biology of the fly, including a short life span, omnivorous and fast digestion, non-nuisance, and non-vector status (17-21). Coupled with the numerous high value products like protein and oil for feed ingredients (22-24), chitin (25), biodiesel (26), and fertilizer (27), BSF are uniquely fitted to being a large-scale sustainable protein source. However, large-scale rearing demands an understanding of the genetic landscape of the BSF so that optimal traits are not selected against and bred out of a population due to short generational times and environmental conditions in

http://etd.uwc.ac.za

facilities (28-31). Population genetic tools applied to whole genome sequences can highlight areas of diversity and adaptation across populations by finding regions on the genome potentially undergoing genetic selection in different scales of rearing (32-35). Quantifying microRNA levels, which are gene expression regulators, has been able to elucidate behavioural and developmental traits through the life stages and sexes of various agricultural and insect species (36, 37) and can identify the regulation of genes of interest for optimizing breeding and creating more efficient environments for rearing (38-42).

1.2 Research rationale

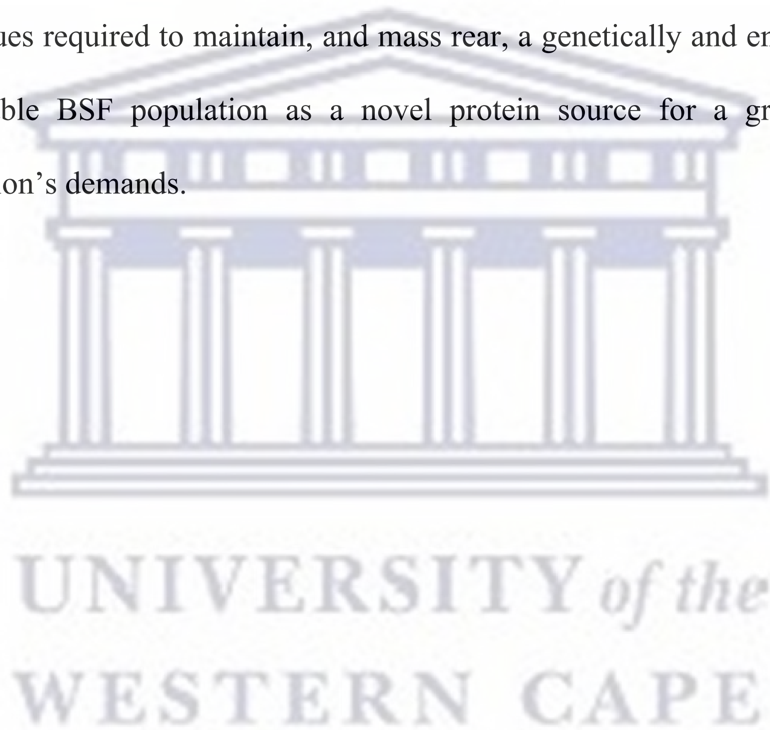
The growing demand for protein by the year 2050 requires sustainable and scalable alternatives. BSF fulfils both aspects and provides multiple high value products while reducing waste. However, the majority of BSF research has been completed on the fly products and gut microbiome only, but there is a growing body of research investigating the genetics of BSF. In order to develop BSF into an alternative protein source, the flies must be reared on a large-scale. Viable mass rearing requires an understanding of its genetic profile to avoid rapidly selecting genetic traits that may not be beneficial for the industry.

Extensive selective breeding may have a cost-benefit to BSF in the long run, especially considering that BSF are being reared on a mass scale. Therefore, understanding the genetic implications of this process is crucial to ensure a genetically diverse and healthy population. Investigating the population diversity and structure across different levels of scaled rearing will bring insights into how the species adapts to different environmental conditions and selective pressures. Combined analyses at the individual life stage and population levels characterize

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

the genetic resources available to create sustainable breeding protocols for a species. BSF undergo multiple life stages that require very different environmental and resource needs. Understanding how microRNA are regulating gene expression throughout these life stages will highlight potential environmental and developmental impacts leading to healthier and more efficient fly rearing. The genetic diversity and microRNA expression profiles of BSF can help to better understand and identify the relatively unknown physiological and behavioural phenotypes needed to streamline the process for developing the most appropriate breeding protocols. This PhD intends to develop and expand the molecular techniques required to maintain, and mass rear, a genetically and environmentally sustainable BSF population as a novel protein source for a growing human population's demands.


UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

1.3 Aims and objectives of thesis research project

The aims of this PhD were to:

1. Determine genetic diversity and potential genomic regions undergoing selection pressures across three African BSF populations being reared at varying scale and,
2. Sequence baseline microRNA to understand genetic regulation for 5 different life stages and 2 sexes of an African BSF population.

These aims were accomplished through:

1. Obtaining 15 BSF samples from five different life stages: egg, larva, pupa, adult unmated female, and adult unmated male, in triplicate.

1.1 Sequencing the extracted small RNA from the 15 samples.

1.2 Establishing a pipeline of code and programs (Appendix B) in order to quality check, identify and filter novel and conserved microRNA from millions of sequences.

1.3 Run statistical analysis to establish expression levels across life stages and sexes.

1.4 Search for known microRNA that has been shown to regulate specific genes in laboratory validated experiments of the closest relative of BSF.

2. The second aim was completed by obtaining 14 BSF adult male and female samples that had been reared in small-, medium-, and large-scale facilities.

2.1 Extracted genomic DNA was sequenced for whole genomes from the 14 samples.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

2.2 A pipeline was created (Appendix C) to quality check, map to the reference BSF genome, and call high quality filtered variants/SNPs using the GATK established variant calling pipeline.

2.3 Analysis of the SNPs called was completed using various genetic programs designed for large datasets.

1.4 Thesis overview

Literature review

A review of the literature on food security and sustainability, with a focus on the BSF as an alternative protein source and the challenges of rearing scalable amounts to meet the demand.

The impact of scaled rearing on the genetic diversity among three African populations of *Hermetia illucens* (black soldier fly)

This study describes the genetic diversity and reports chromosomal regions of interest due to selection pressures of three populations of BSF reared at large-scale (LS), medium-scale (MS), and small-scale (SS).

microRNA profile of *Hermetia illucens* (black soldier fly) and its implications on mass rearing

This study creates a database of novel microRNAs and their expression levels across five life stages and two sexes in BSF and provides candidate microRNAs with potential impact on BSF mass rearing.

Conclusions and future work

Conclusions based on the data of both studies were drawn about the genetic landscape of BSF and how it can be utilized for mass rearing of the fly, with a focus

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za

on chromosomal selection pressures and life stage specific expression changes. Future work could provide a deeper understanding of highlighted regions of BSF genetics.

http://etd.uwc.ac.za

Chapter 2: Literature review

2.1 Food sustainability

The globe is facing an insecure food, land, and water future based on a number of old and new problems. The world population reached 8 billion in November of 2022 and is projected to reach 9.7 billion by the year 2050, peaking at 10.4 billion by 2080, but level off by the year 2100 (2). Eight countries are projected to account for more than half of the population increase up to 2050, with more than half of those countries belonging to sub-Saharan Africa (2). The rise in population means food production needs to increase by 50% (1). However, food production is expected to decrease by as much as 25% by 2050 due to climate change and unsustainable traditional methods of manufacturing food (1). Around 40% of all land is used for agriculture (5) and an estimated 70% of all agricultural land is utilized to supply feed to livestock (1, 4). The clearing of land for agriculture is responsible for a loss of biodiversity through deforestation, and a loss of approximately 70% of human water consumption through irrigated crops (4, 6). Fertilizer use has also increased by a staggering 700% over the past 40 years to help the growing agricultural land (5). Food loss and waste is another major contributor to inefficient food production systems. Food loss occurs from the harvest to retail portion of food production, and accounts for a loss of 14% of food produced (8). While food waste occurs on the retail and consumer level of the food system and represents a further 17% loss in global food production (7). In order to combat these challenges, new methods of food production and more sustainable usage of the resources involved, must be found (43).

http://etd.uwc.ac.za

2.1.1 Increase in protein demand

The demand for protein is increasing with the growth of low- and middle-income economies, as economic growth is linked to an increased consumption of meat, dairy, and aquatic products (3, 8). The 8 countries projected to make up more than half of the population increase until 2050, mainly belong to sub-Saharan Africa and parts of Asia (2). These same regions, along with parts of Latin America and the Caribbean, are also experiencing a rise in the number of working age people (4). The consumption of aquatic animals increased globally at a rate of 3% for the years between 1961 and 2019, while the consumption per person remained the same. This rate increase is double the global population increase rate (1.6% annually) for the same time and indicates how the growing population's need for protein is expanding (44). The expanding protein demand is explained by the increase of higher earnings in populations that previously were priced out of protein in their diets. However, protein producers are one of the highest consumers of natural resources such as land and water. Due to the increasing demands for protein, it is important to explore sustainable, alternative protein sources which are environmentally friendly.

UNIVERSITY of the
WESTERN CAPE

2.2 Insects for food and feed

2.2.1 Entomophagy

Insects have been a part of the human diet since before we developed tools to hunt and collect food, and there are still an estimated 2,000 species being consumed mainly in low- and middle-income countries to this day (12). The discovery of insect remains in fossilized faeces from ancient humans provided proof that insects were being consumed (10, 45). A wide range of insects were found to have been

http://etd.uwc.ac.za

eaten in the Americas, such as ants, ticks, lice, beetles, and mites (14, 46), and 3200 years ago, the Chinese were also consuming insects (47). Aristotle gave tips in 384-322 BCE, that cicada females were better tasting once they were full of mature eggs (48). In 77 CE, Pliny the Elder wrote about how popular the larvae of *Cerambix cerdo* (the great capricorn beetle), and many texts from the Christian, Islamic, and Judaism religions mention entomophagy (9).

Throughout history, humans have partaken in entomophagy as part of celebrations (9), everyday diet, medicine, and agriculture (10). Maggot therapy and vaccine development are two ways insects have been utilized for medicine (11, 49). They have also been reared for food colourant, biological control efforts, and research studies (50-52). Insect applications for sustainable practices have also been found with *Tenebreo molitor* (mealworm) and their ability to digest polystyrene (53). The short life cycle, small rearing space, and high nutritional value of insects has made them a part of agriculture as well. Insect flours have been found dated as far back as 1550 in Northern Africa (9), and rearing *Apis mellifera* (honeybee) for honey and pollination or *Bombyx mori* (silkworm) for silk (10) in Asian cultures.

Entomophagy, while still widespread in Eastern tropical countries, has not endured in Western countries throughout the years, even developing a stigma around the practice. The maintained tradition in tropical countries is most likely due to the opportunistic nature of entomophagy, and that insects are able to grow faster and larger in year-round heat and humidity (10). However, the use of insects for feed in agriculture has started to develop globally despite any stigma surrounding insect use.

http://etd.uwc.ac.za

2.2.2 Insects for feed in agriculture

Numerous reasons exist for why insects are a good alternative resource for feed of livestock, especially in the context of growing climate changes and increasing demands for protein. Rearing insects for feed requires less land use (13), up to 50% less water usage (14), and the conversion of waste into valuable protein and other products (15). However, introducing a new industry and product to the market can hit multiple barriers, such as legality and consumer acceptance.

While more studies need to be conducted, a review of 8 studies on consumer opinion showed that replacing traditional meal with insect meal for fish and poultry is widely accepted (54), potentially due to fish and poultry already eating insects in nature. (22, 55). This acceptance only grows when the consumer is educated about the environmental benefits of using insect meal. Studies have also shown consumer acceptance is not hindered by any negative sensory or taste profiles of meat-fed insect protein in their diet (54).

The legal barriers are more complex and require cooperation and acceptance by governments. The regulations regarding the use of insect meal are varied across the globe and can restrict the benefits of insect meal. For instance, the EU allows the use of insect meal for animal feed, but the insect diet must not contain animal by-products (56). These regulations limiting food sources for rearing insects cause an increase in the environmental impact of insect rearing (56, 57). Limitations on what insects can be reared on can also affect an obstacle that the International Platform of Insects for Food and Feed (IPIFF) has identified for this new industry: scalability (16).

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za

Since the demand for protein is growing at an enormous rate around the world (1), the IPIFF has stated that the industry needs to be able to increase the supply of insect meal for competitive pricing and to produce a stable, reliable product (16). Currently, mass production of insect meal is still in its developing stage. There are few companies that can produce tonnes of larvae a day. The goal of replacing unsustainable feed ingredients, such as soybean and fish meal, demands production to be much higher than is currently available. For example, to feed several small chicken farms (30,000 hens/farm) would require at least 10 tonnes of larvae a day (10), which is what is produced from the largest BSF company.

There are multiple edible insects being utilized for making feed ingredients, like *Tenebreo molitor* (meal worms), *Gryllus bimaculatus* (Mediterranean field crickets), *Bombyx mori* (silkworms), *Musca domestica* (common houseflies), *Lucilia sericata* (common green bottle flies), *Cetonia aurata* (green rose chafers), *Pachnoda marginata* (sun beetles), and *H. illucens* (black soldier flies; BSF) (10). However, the most commonly used insect grown for feed ingredients is *H. illucens*, based on widespread utilization in the nascent industry (57). The reasons for BSF's popularity are the fly's short life cycle, non-nuisance or vector behaviour, and the high-level products made from the larvae.

2.3 *H. illucens/*Black soldier fly (BSF)

2.3.1 Life cycle

BSF is a dipteran insect species found worldwide but originated in South America (19, 20, 28). They are uniquely adapted to being used as an alternative protein and feed ingredient due to multiple reasons. The BSF have a short life cycle (Figure 1) (8-9 generations in a year) that involves the laying of 300-600 eggs (18, 28),

http://etd.uwc.ac.za

followed by six larval instar stages that consume organic matter of a wide range (58).

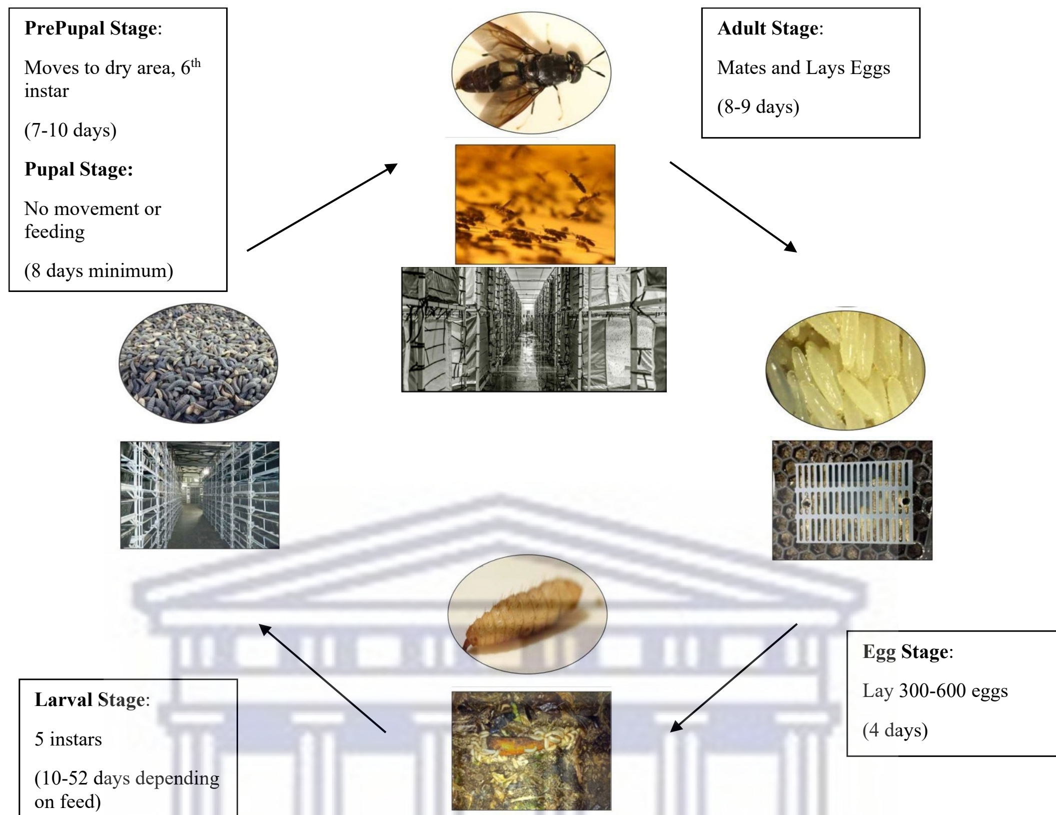


Figure 1: ***Hermetia illucens*** **(black soldier fly) life cycle and potential mass rearing locations. Each life stage and average length of time for the stage to develop is listed in the space following the image. Images are of each life stage and the substrate/containers in which they are mass reared.**

UNIVERSITY of the
WESTERN CAPE

During this larval stage, BSF are at their highest nutrient density at the sixth and final larval instar before pre-pupation and can be harvested for feed ingredients. Once pupation ends and adult flies emerge, they spend the next 8-10 days mating and laying eggs (59, 60). The adult BSF feeds mainly on nectars, honeydew, and pollen, and relies on the stores built up during the larval stage (61).

The lack of feeding as an adult, along with egg laying adjacent to a food source, makes the BSF a non-nuisance and non-mechanical vector, unlike other flies, such

http://etd.uwc.ac.za

as *Musca domestica* (common housefly) (21, 62). Also, the black soldier fly larvae (BSFL) can contain a range of 40-60% protein, up to 47% lipid, and 3-8% chitin in prepupae, depending on the organic substrate they are reared on and the extraction process (15, 25, 63-66). Their feed conversion is the most efficient of the mainstream insects being used with only 1.4 kg of feed required to produce 1 kg of larvae compared to crickets (2.3 kg of feed) and meal worm (3.8 kg of feed) (13). This makes them not only a good protein source, but shows potential for a variety of derived products, and has minimal environmental impacts.

2.3.2 High value products and sustainability of BSF

The uses of the BSF are varied from protein and oils for animal feed ingredients, biodiesel (26), food stabilizing agents, fertilizer (27) and waste reduction (67-69). The larvae are known to reduce municipal organic waste (68), faecal waste (69) and animal waste (67, 70) with varying levels of impact on BSFL nutrition and have been shown to reduce harmful pathogens in the food source (71).

2.3.3 Animal feed ingredients

A total of 17% of animal protein and 7% of all proteins come from consuming fish (72). This crucial food source currently uses fishmeal as the main feed for fish farming, which comes from dredging oceans and using the fish considered too small as the protein source. Considering that roughly 90% of the world's fish have been fully fished or overfished, a more sustainable method for feeding farmed fish is necessary (44). Studies have shown that BSFL as a fishmeal replacement, or partial replacement, not only do not have a negative effect on the fish, but also have beneficial anti-microbial and anti-bacterial properties (23, 73-75). The growing list of studies for fish that are shown to benefit from a BSFL diet includes many

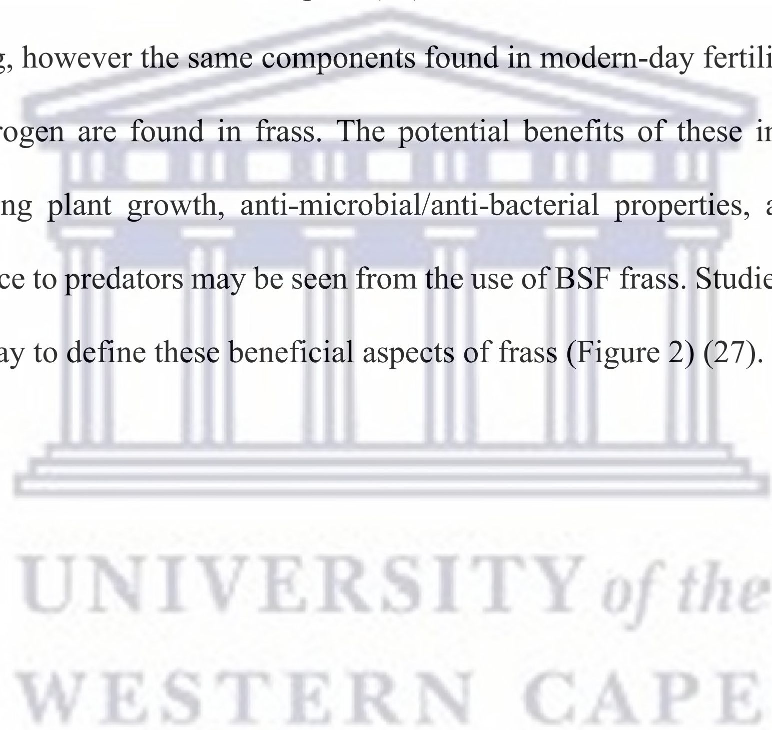
UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

varieties such as turbot (76), seabass (77-79), carp (80, 81), catfish (82, 83), tilapia (84, 85), trout (86), salmon (87), and sturgeon (88). The BSFL have also been used in feeding livestock of poultry (89, 90), and have added to the environmental enrichment and behavioural welfare of the poultry (22, 24). Using BSFL as a feed replacement for swine produces pigs of the same quality protein and even enhances desirable traits, such as growth performance and yield (22, 91-93).

2.3.4 Fertilizer

After consuming organic wastes, BSFL leave behind a substrate, or frass (insect faeces), that is similar to compost (94). The frass for fertilizer industry is still growing, however the same components found in modern-day fertilizer, like chitin and nitrogen are found in frass. The potential benefits of these ingredients like promoting plant growth, anti-microbial/anti-bacterial properties, and increasing resistance to predators may be seen from the use of BSF frass. Studies are currently underway to define these beneficial aspects of frass (Figure 2) (27).


UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

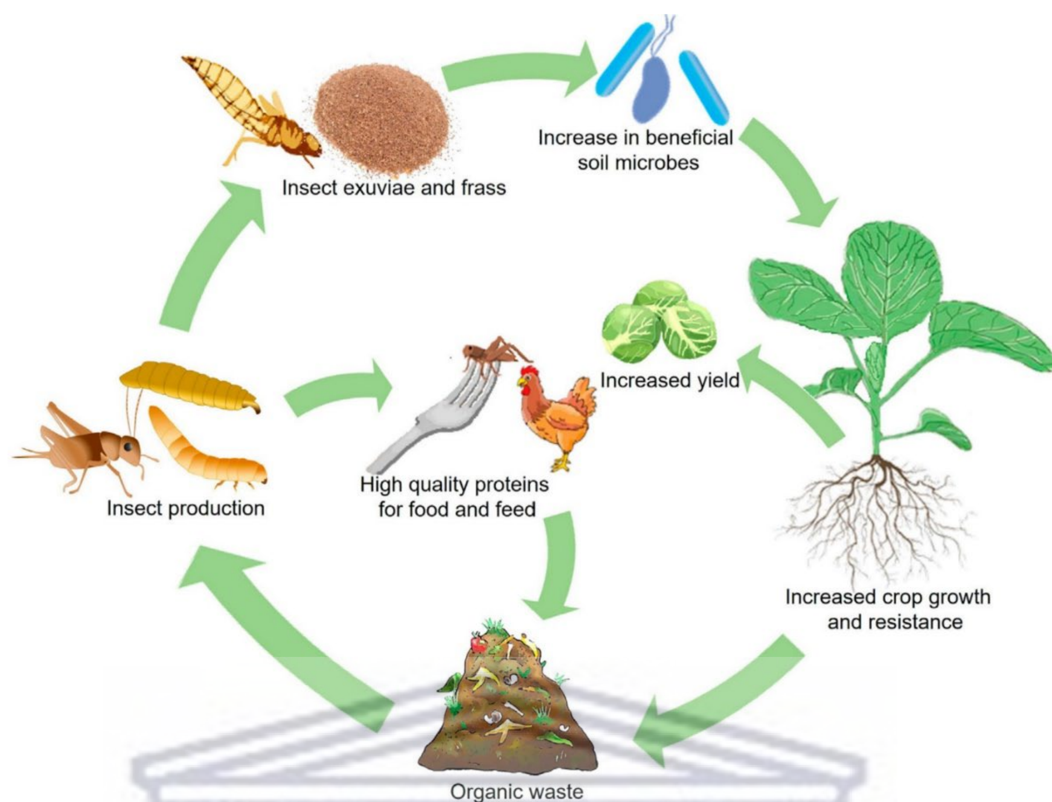


Figure 2: Schematic representation of insect production in a circular food production system. Insects can transform organic waste into high-quality animal protein for food and feed. Here, we discuss the possible use of insect residual streams as soil amendments to stimulate beneficial microbiota and improve soil and crop health. (© Trends in Plant Science 2022)

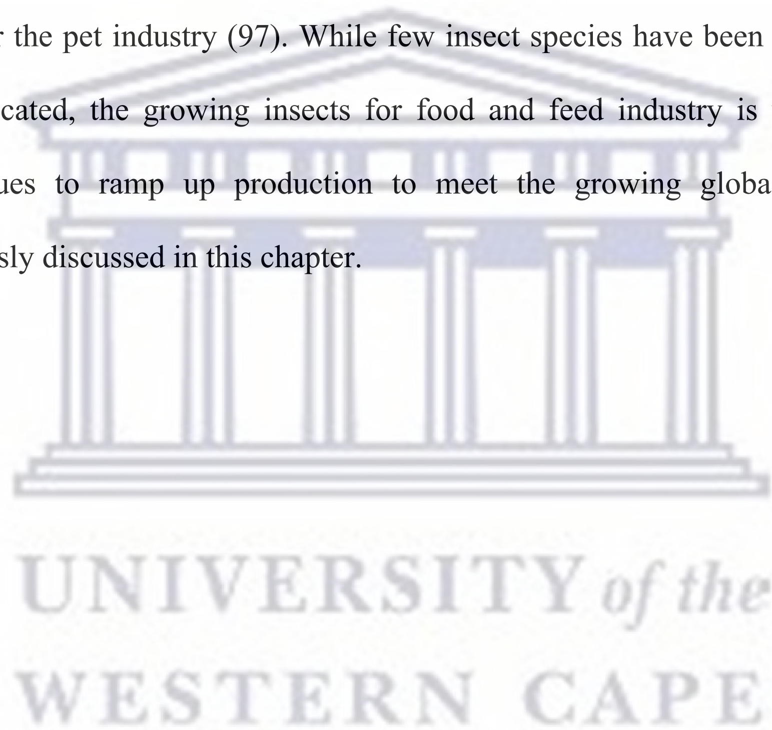
The mass production of BSFL requires minimal land and utilizes organic waste streams that would normally be dumped in a landfill. The food waste is recycled by feeding it to the BSFL. Therefore, the use of BSFL as a protein source for a feed ingredient not only results in multiple products, but also has minimal environmental impact making them an ideal candidate for large-scale rearing for food sustainability and waste reduction.

2.4 How to avoid unsustainable issues of mass rearing

Domestication and breeding programs selecting beneficial and desired phenotypic traits have been around for thousands of years and were the start of agriculture and

http://etd.uwc.ac.za

livestock rearing (95-97). The level of domestication can be broken into five sections (Figure 3) and are dependent on the amount of gene flow from wild type species and the control over the population by humans (97). Though insects have been documented for many uses for millennia (10, 12, 14, 45), few insect populations have been domesticated to the highest level which is entirely dependent on human control with no gene flow from the wild. Only *Bombyx mori* and *Apis mellifera* have been domesticated to this extent (Figure 3)(97). The domesticated house cricket *Acheta domesticus* has been bred at level 4 (Figure 3) for agriculture and feed; along with the *Gromphadorhina portentosa* (Madagascar hissing beetle) bred for the pet industry (97). While few insect species have been categorized as domesticated, the growing insects for food and feed industry is utilizing these techniques to ramp up production to meet the growing global demands as previously discussed in this chapter.


UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

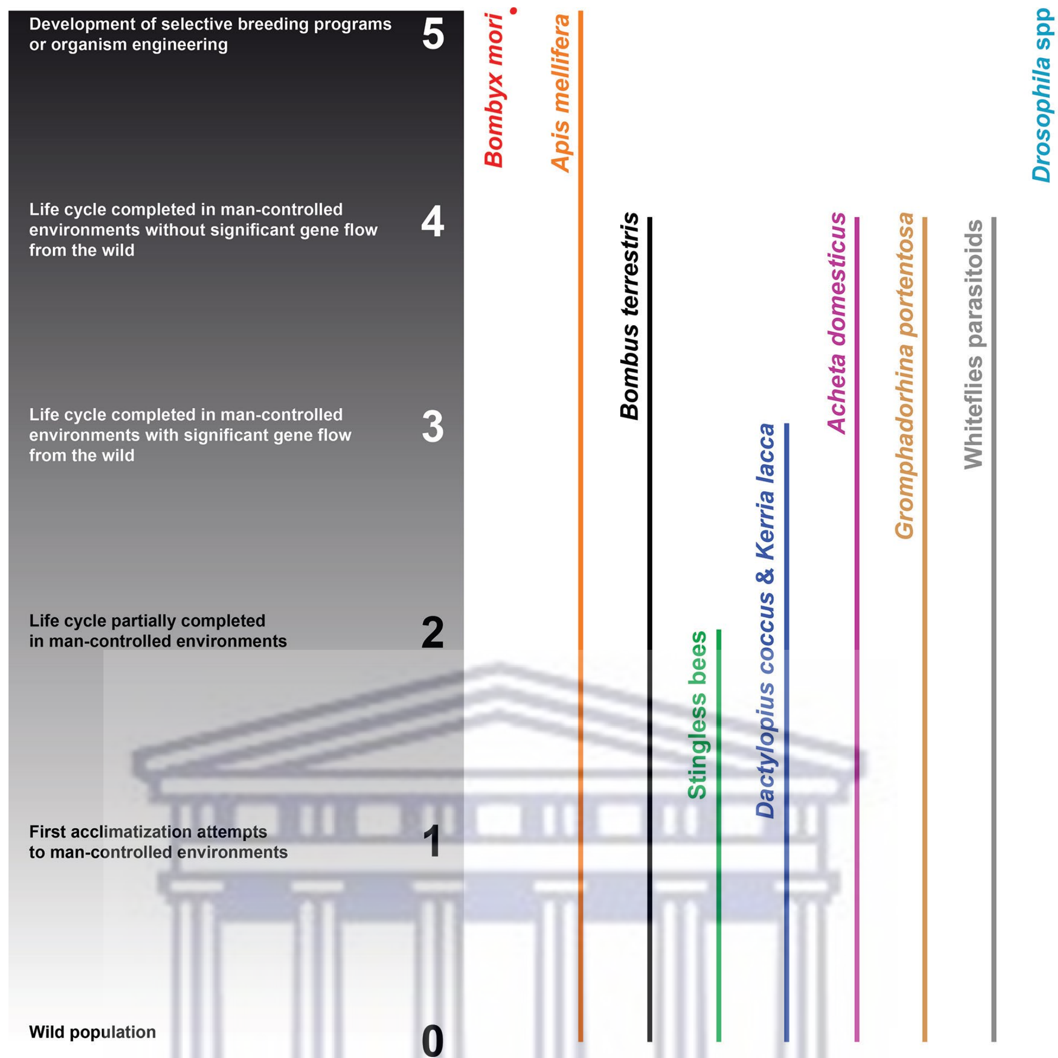


Figure 3. Domestication process and insect domestication level. Numbers 0–5 refer to the domestication levels [12, 18]. Characteristics of each domestication level are provided on the left. Lines and points near the insect species names show the range of domestication degrees observed among populations of the species. (© Animal Domestication 2018)

However, these domestication strategies are known to result in consequences to the species being reared, such as phenotypic alterations and sustained genetic variation (97). As the science progressed, the genotypes causing desirable phenotypes started to be more understood and became the driving force behind selective breeding operations (96, 98, 99). Large-scale rearing in controlled environments over generations can have negative effects on the population being reared and is why

http://etd.uwc.ac.za

large-scale breeding operations consider not only the beneficial phenotypic traits, but also the genetic health and welfare of the species (29, 100). When focusing on phenotypic traits advantageous to mass production, traits like growth rate, generational length, and muscle size, without the understanding of genetic trade-offs for selecting specific traits can lead to the overall decline of a breeding line of animals (98, 101). This was seen in chicken, cattle, sheep, and pig rearing, where they selectively bred out too many traits without understanding the population genetics and the animals became sick (29, 30). Specific to insects, mass rearing techniques can lead to degradation of a colony's genetic diversity, beneficial traits, and robustness of the population. Lab colonies of insects show high loss of genetic diversity, often ending up not being representative of wild type species. (31, 102, 103). Also, the domestication of *Bombyx mori* has led to the loss of flight and camouflage (104). Therefore, selective breeding programs should consider the entire genetic landscape of a species to better provide a sustainable and healthy population.

2.4.1 Population genetics

Understanding the population structure and genetic diversity of a species is pivotal to identifying and incriminating causal genetic variations of phenotypes needed for large-scale breeding protocols (32-34). The use of whole genome sequencing (WGS), as opposed to marker/ddRADSeq/Genotyping by Sequencing (GBS), gives a more precise picture of the variation in a species. By looking at the whole genome, "ascertainment bias" of using markers and SNP chips developed to find previously identified regions of diversity in a population is reduced (105, 106). In non-model species, where relatively, little is known about the genome, these regions may give

http://etd.uwc.ac.za

a slanted view of genetic diversity, either less or more variation due to the smaller sampling of the genome (32, 107).

Classifying a species population structure and diversity can be achieved through cataloguing both neutral and adaptive variations in the genome. Neutral variation takes a summary statistics approach to classify a genome, highlighting the heterozygosity and relatedness of the individual and populations (32-35). Identifying the type of variant, transition versus transversion, can highlight potential amino acid changes (108). Inbreeding coefficients highlight relatedness of populations by measuring the probability of two individuals having identical genes, either on the individual or population level (109). More statistics include nucleotide diversity (π) across genomes, which has the base assumption that alleles are randomly distributed and measures the number of substitutions and polymorphisms between populations (110, 111). Higher nucleotide diversity has also been linked to functional segments of the genome (112), whereas adaptive variation examines selective pressures a species has experienced based on environment, geographic locations, and resources. Tajima's D (TD) is a statistical test describing allelic frequency based on a model of neutrality across the genome in a population, meaning no selection pressures are occurring and mutation is random (113, 114). By examining these variations through the lens of adaptation to selection pressures, regions of variation due to these pressures between populations can be located.

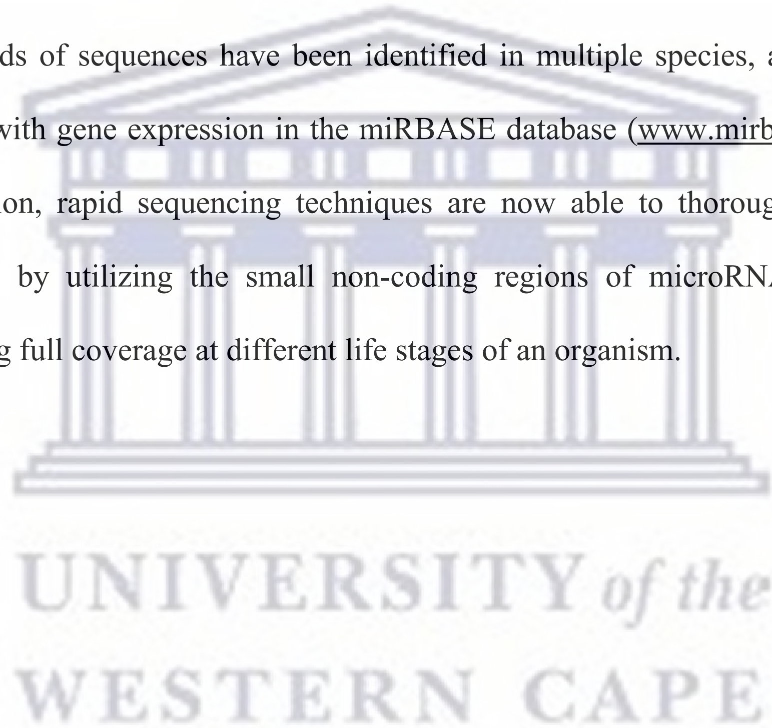
2.4.2 microRNA

The development of new rapid sequencing techniques has opened up the possibility of in-depth mapping of gene expression in populations. Understanding the gene expression of BSF will provide better information as to the most efficient and robust

http://etd.uwc.ac.za

methods of mass rearing the flies as an alternative protein source. Gene expression studies, used to develop mass rearing techniques for agriculture, are well accepted throughout the industry (36, 37). In the context of insects, gene expression studies have shown that microRNAs affect biological traits such as a fecundity and larval development through gene regulation (38-40). In locusts, microRNA regulation allows for phenotypic plasticity due to environmental changes (41).

microRNA is a small, highly conserved regulator of gene expression (Figure 4) (38). microRNA functions by aligning with messenger RNA (mRNA) to post-transcriptionally regulate gene expression (42). Since the discovery of microRNA, thousands of sequences have been identified in multiple species, and have been linked with gene expression in the miRBASE database (www.mirbase.org). Next generation, rapid sequencing techniques are now able to thoroughly cover the genome by utilizing the small non-coding regions of microRNA (115, 116), allowing full coverage at different life stages of an organism.



http://etd.uwc.ac.za

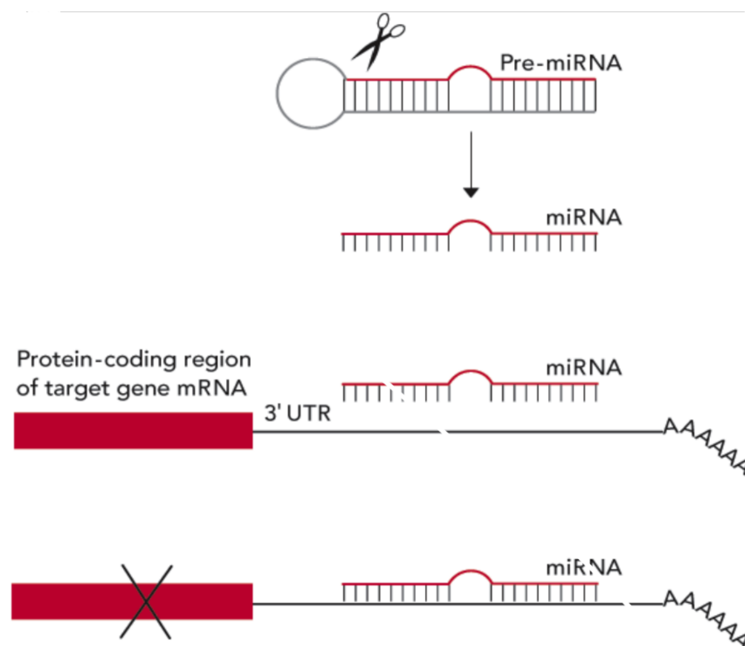


Figure 4: Steps of microRNA binding to the 3′ UTR region of the target mRNA (adapted from © Vasulu, 2015) microRNA binds to mRNA causing down-regulation of the mRNA target gene. (mRNA = messenger RNA)

Multiple insects have already been mapped for their microRNA, showing how biological and behavioural functions are regulated by microRNA at different life stages (38, 41, 42, 117, 118). Understanding the biological and behavioural functions of BSF will provide data to improve environmental rearing conditions that are suited to BSF genetic predisposition. The use of microRNA expression maps to establish selective breeding protocols has not yet been completed. However, this novel technique would allow for rapid selection processes for breeding, and innovative methods of environmental control, to increase production efficiency of an alternative sustainable protein source that is greatly needed.

http://etd.uwc.ac.za

2.5 Current knowledge of BSF genetics

Studies being conducted on the genetics of the BSF are greatly increasing in number but are still relatively few. Two high quality genomes have been reported now (119, 120), but for years before that only a draft genome was known (121). These advances have allowed for further investigation into the genetic landscape of this important agronomic fly. Until more recently, most genetic work on BSF has focused on the gut biome (122-125). In 2021, two studies were published describing population structure between BSF that built a complex, geographically distinct and resilient BSF genetic history (28, 126). The Kaya *et al.* study described 16 distinct genetic clusters of BSF from both wild and captive populations from 150 global populations, and a species originating in South America (28). Global expansion of the species is believed to have stemmed from shipping trades and the relative historically recent BSF farming industry (28, 126). One study has looked at the genetic link to phenotypic traits in 4 genetically distinct populations and found a direct "strain-by-diet" connection highlighting the importance of understanding the genetic background of BSF for optimised rearing (127).

2.6 Conclusions

A growing global population and changing climate have made food sustainability and security a world-wide concern (1, 2). The main challenges in meeting the growing demand are increasing food production and reducing waste using sustainable methods (1). The growing demands on protein producers and the dwindling resources available has made it clear that novel sources of protein need to be found (3, 8). Insects have had a long history of being part of the human food chain (12), and are an excellent source of protein and other nutrients in feed for

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

animals (13). Specifically, BSF is uniquely suited to the task of meeting scalability challenges, has a wide range of valuable products, and digests organic waste leaving a beneficial frass for fertilizer, therefore meeting protein demand while reducing waste (22, 27, 67, 68). When scaling up production of BSF to meet protein demands, there are potential pitfalls of which to be aware. Mass rearing causes loss of genetic diversity which could be reducing the efficiency of feed conversion and health of the fly (31, 102, 103). While studies involving BSF have increased exponentially (17), the genetics of the fly are still not fully discovered (28). Investigating the genetic diversity and microRNA expression of BSF may help to inform optimised breeding protocols and environmental conditions for mass rearing.

http://etd.uwc.ac.za

Chapter 3: The impact of scaled rearing on the genetic diversity among three African populations of *Hermetia illucens* (black soldier fly)

3.1 Introduction

By the year 2050, world demand for food production is predicted to increase by 50% due to a growing population (1, 128), but the supply of world food production is expected to drop by 25% during that same time (1). Climate change and unsustainable methods of manufacturing are key factors in this decreased production (1). The management of the coinciding organic waste associated with this food production increase provides another problem, as one third of food production becomes waste (129). The need to find both sustainable and scalable resolutions to these problems has led to the exploration of insects as one possible alternative resource, due to their relatively short life cycle and small environmental impact (120). Out of the 2,000 species already being utilised as a food source (12), BSF has become popular for its nutrient content and omnivorous diet. BSF are found worldwide and have been approved to provide a sustainable source of protein for fish and animal feeds (19, 20, 75, 78, 130). They contain high protein, crude fats, and chitin during their different developmental life stages (15, 25, 63-66). BSF produces beneficial products such as feed, biofuel, and fertilizer by feeding on waste streams that would have previously gone to landfills (67-69). These benefits, along with the fact that BSF are non-nuisance and not mechanical vectors (21, 25), make them an ideal sustainable and scalable solution to the growing demands on food production and waste management.

http://etd.uwc.ac.za

The genetics of BSF are just now starting to be reported, with two genomes recently published to high quality resolution (119, 120). Zhan recently published the midgut transcriptome and the use of CRISPR/Cas-9 to map genetic pathways to phenotypes in industrial scale reared BSF, as well (119). A baseline study has identified sex and developmental microRNAs of interest, along with another study characterising some population structure in one mass reared population (120, 131). Identifying and describing the population structure and genetic variation of a species is important in determining sustainable and genetically robust breeding protocols for LS rearing facilities (32, 52, 132). Understanding the genetic landscape of BSF is necessary to develop LS breeding protocols that can avoid colony collapse (133) and loss of genetic diversity through colonization found in other arthropods (25, 31, 124).

This study aimed to describe the genetic diversity and report chromosomal regions of interest due to selection pressures of three populations of BSF: LS, MS, and SS reared. Investigating the genetic diversity of BSF provides a genetic understanding of the environmental impacts of scalable breeding protocols and how BSF adapts to its environment.

3.2 Methods

3.2.1 Sample collection

BSF samples were collected from three different populations: large-scale (LS), medium-scale (MS), and small-scale (SS) reared. These scaled labels were determined by the level of artificial rearing conditions and the number of flies reared per day. LS rearing involves mass numbers (> 10 tonnes/day) of BSF being reared under controlled but artificial environments. The LS flies were reared under

http://etd.uwc.ac.za

large-scale factory breeding conditions (by AgriProtein Technologies Ltd in Philippi, South Africa), fed a standardized proprietary diet created around commercial ingredients of layer hen feed, and allowed to consume *ad libitum*. Flies collected in Nairobi, Kenya, were reared under standard but artificial conditions with a population under 10 tonnes/day and were considered MS reared. The third population was collected from Durban, South Africa in a SS, "free range" non-artificial environment, with under 1 tonne produced per day. Both the LS and MS flies were kept at 28°C (±2°C), 80% relative humidity, under 12-hour day and night cycles. The SS population did not control environmental conditions. Due to budgetary constraints, three adult females and three adult males were collected from each population, except from the LS population, of which only 2 samples, one male and one female, were used.

3.2.2 DNA extraction

Samples were collected and stored in 100% ethanol at room temperature until DNA extraction (134). The flies were then frozen with liquid nitrogen and ground into a powder using a mortar and pestle. A DNeasy Blood & Tissue Kit from Qiagen (135) was used to extract all DNA following the standard protocols of the kit. A Qubit dsDNA Broad-Range (BR) Assay (Invitrogen™) was performed, according to manufacturer's instructions, for quality control and to determine the concentration of DNA extracted for sequencing. After extraction, all samples were stored at -80°C. Samples were then shipped for sequencing.

3.2.3 Sequencing

Extracted samples were shipped to Macrogen Inc. (Seoul, South Korea) for WGS. Paired-end library construction of 14 samples was completed using the Illumina

http://etd.uwc.ac.za

TruSeq DNA PCR-free kit (136). Read lengths were 151 base pairs. Only two samples, one male and one female, were used from the LS population, but were sequenced at 30X coverage. The lower sample number was due to a supplier change in what sequencing was offered, causing the cost to be prohibitive for 6 samples. At the time of sequencing in 2018, no reference genomes were available (first reference genome NCBI: assembly iHerIll2.2.curated20191125 published in 2020)(119), therefore sequencing at a higher depth (30X) of two samples for comparison was thought to be best practice. Three males and three females were sequenced for both the MS and SS populations at 10X coverage. Sequences were quality checked using FastQC (v0.11.7) (137).

3.2.4 Read processing and analysis

Reads were trimmed using Trimmomatic (v0.39)(138, 139) software and the provided Illumina adapter sequences (140). The reference genome of *H. illucens* was used to map the reads (NCBI: assembly iHerIll2.2.curated20191125). After indexing, reads were aligned to the reference genome using BWA (v0.7.17) (141). Aligned reads were then sorted, and duplicates marked using the Genome Analysis Toolkit (GATK4) (v4.2.0.0) (142). The GATK4 Broad Institute best practices for germline short variant discovery was followed for variant discovery(143, 144), except the recalibration of base quality scores (BSQR) step, as there were no known single nucleotide polymorphism (SNPs) databases available at the time of processing to run BSQR. Variant calling was completed using the HaplotypeCaller in GVCF mode. Samples were then jointly called to provide unfiltered variant files for each sample. Variant files were then filtered to remove Insertion/Deletions (InDels) because of coverage being under 100X and filtered according to GATK

http://etd.uwc.ac.za

hard filtering recommendations (143, 144). Density plots of all filters were completed to ensure that the parameters were appropriate for sample filtration (QD min 2, FS max 60, SOR max 3, MQ min 40, MQRS -12.5 to 12.5, RPRS -8 to 8). Further filtration was run on sites missing more than 15% data. BCF stats from BCFtools (v0.7.17) (145) were used to calculate transition vs transversion ratios , VCFtools (v0.1.16) was used to calculate inbreeding coefficients (146). Calculating SNPs that were unique or shared between populations was done using BCFtools ISEC command (145). PLINK (v1.90b4) calculated the principal component analysis (PCA) of the three populations (147). Nucleotide diversity (ND, π) and Tajima's D (TD) were run on these filtered data using VCFtools (v.0.1.16) over 20 kb sliding windows (146). The diversity of the genome was calculated over total populations and represented in box and scatter plots to compare the statistical values of π between sexes and chromosomes. Higher values for π are indicative of more diverse genomes, and is a descriptive statistic used in conjunction with TD and PCA to locate areas of genetic differentiation between populations. As this was a whole genome scan, TD values were considered significant if they varied ±2 from the majority of TD values found across the sliding windows and were described using box and scatter plots of the statistical values. Overall positive or negative trends in TD values were also included to describe selection pressure the genomes may be experiencing, where positive TD values are linked to less rare alleles being identified and population balancing selection pressures. Negative TD values are indicative of rare alleles being present in large numbers across genomes that have potentially experienced recent selective pressures, such as population growth after bottlenecks. Microsoft® Excel and R studio were used for analysis and

http://etd.uwc.ac.za

visualization of data in scatter plots for regions of interest (148, 149). See Appendix B for analysis pipeline.

3.3 Results

3.3.1 Population diversity

The total number of variants found after running HaplotypeCaller, and before filtering for all samples, was 896 million. Selecting only SNPs and filtering for high quality scoring, as described in the GATK hard filtering, 127 million SNPs were considered for further analysis. After filtering for high quality biallelic SNPs, the LS population had a mean number of SNPs for the whole genome at 22 million. The number of SNPs per chromosome followed according to the size of each chromosome, with chromosome 1, the largest in size having the largest number and chromosome 7 having the lowest number of SNPs reported. This pattern was followed by all BSF populations. The MS population had the next highest number across the whole genome with 39 million, and the SS population identified 47 million. The Transition vs Transversion (Ts/Tv) mutation rate across all three populations was similar, with LS having 1.757 Ts/Tv, MS equaling 1.759 Ts/Tv, and SS reporting 1.757 as well. The inbreeding coefficient (F_{IS}) for the LS population was lower than for the MS and SS populations. LS showed a 0.083 F_{IS}, while MS was 0.281 F_{IS}, and the SS population was 0.327 F_{IS}. All populations

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

showed significant drops in mean SNPs and Ts/Tv when examining chromosome 7, with their inbreeding coefficient rising instead (Table 1).

Table 1: Whole Population SNP Summary by genome and chromosome of *Hermetia illucens*. Abbreviations: Ts/Tv = Transition / Transversion; F = inbreeding coefficient

	Large Scale (n=2)			Medium Scale (n=6)			Small Scale (n=6)		
	Mean Biallelic SNPs	Ts/Tv	F	Mean Biallelic SNPs	Ts/Tv	F	Mean Biallelic SNPs	Ts/Tv	F
Whole Genome	21,712,568	1.757	0.083	39,170,830	1.759	0.281	46,611,091	1.757	0.327
Chr 1	4,804,470	1.745	0.120	8,211,962	1.748	0.174	10,032,836	1.748	0.303
Chr 2	3,721,833	1.754	0.029	8,057,305	1.756	0.448	8,947,633	1.755	0.333
Chr 3	3,466,328	1.756	0.000	7,187,492	1.757	0.239	7,950,174	1.757	0.295
Chr 4	4,002,006	1.764	0.112	7,136,217	1.765	0.294	8,164,077	1.764	0.332
Chr 5	2,812,014	1.738	0.130	4,266,360	1.742	0.262	5,859,050	1.741	0.347
Chr 6	2,687,823	1.790	0.095	3,925,424	1.793	0.210	5,155,450	1.779	0.359
Chr 7	218,096	1.806	0.235	386,072	1.820	0.486	501,872	1.824	0.523

3.3.2 Principal component analysis (PCA) highlight population similarities

The PCA was performed on 14 individuals in three populations and their SNP profiles. PCA predicts the three populations are similar, with over 30% of the variation explained in the first two principal components (PC). PC1 explains almost 20% of variance, and PC2 following with 11%, but the following PCs 3 through 14 all explained below 10% each of the variance (Figure 5). The SS population displayed the only separation from the other samples. The PCA predicts that the LS and MS populations are more similar in variation than the SS population.

http://etd.uwc.ac.za

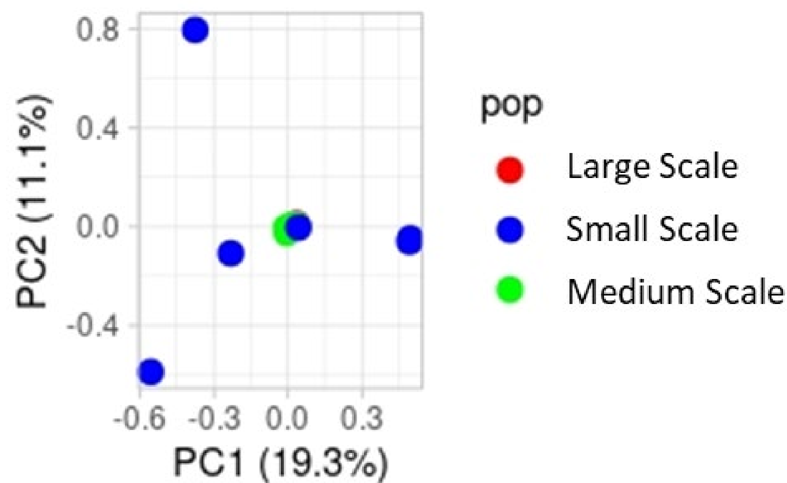


Figure 5: Principal Component Analysis (PCA) of three populations of *Hermetia illucens* using 14 individual's single nucleotide polymorphisms. Populations were large-scale (n=2), medium-scale (n=6), and small-scale (n=6) reared BSF. Only the small-scale population displayed distinct clusters, while large- and medium-scale populations were indistinguishable. Percent variation explained by Principal Component 1 (PC1) and Principal Component 2 (PC2) in parentheses. Populations are colour coded with large--scale as red, medium-scale as green, and small-scale as blue. Large-scale is hard to see as it clustered with medium- and small-scale clusters in the centre.

3.3.3 Unique and shared SNPs between BSF populations

The number of unique SNPs found for the LS population was 2,879,596, representing 2.3% of the SNPs found. The MS population had 9,272,861 (7.3%) unique SNPs, and the SS population had the largest number of unique SNPs with 21,654,251 (17.1%) (Figure 6). A total of 18,002,677 shared SNPs, comprising 14.2% of SNPs, were found in all three populations. MS and SS populations shared the most SNPs at 31,078,615 (24.5%), followed by the MS and LS populations sharing 23,544,872 (18.5%), and SS and LS sharing the least at 20,571,080 (16.2%) SNPs (Figure 6).

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

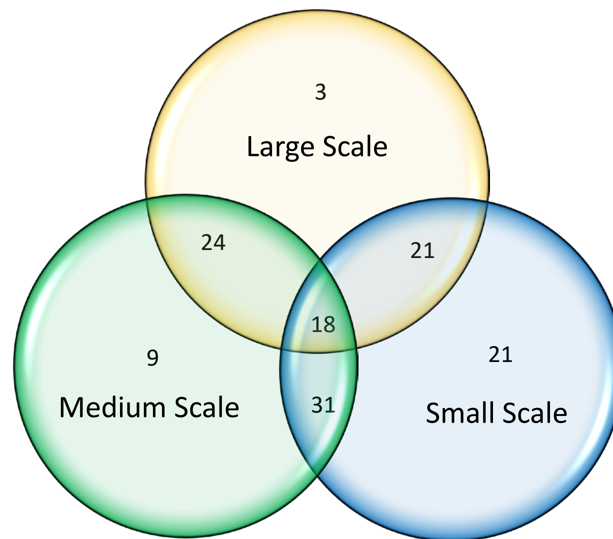


Figure 6: Venn diagram of unique & shared SNPs between three populations of *Hermetia illucens*. The large-scale population (n=2) is represented in yellow; medium-scale population (n=6) in green; and small-scale population (n=6) in blue. Shared SNPs are in overlapping sections of circles with medium- and small-scale populations sharing the most SNPs, and SNPs shared by all three populations in the centre overlapped by all three circles. Numbers are in the millions.

3.3.4 SS rearing produces nucleotide diversity

The nucleotide diversity (π) was calculated using 20 kb sliding windows across the whole genome and each chromosome. The LS population had the lowest whole genome nucleotide diversity at 0.014 π, MS population in the middle with slightly higher 0.015 π, and the SS displaying the highest diversity with 0.018 π (Figure 7).

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

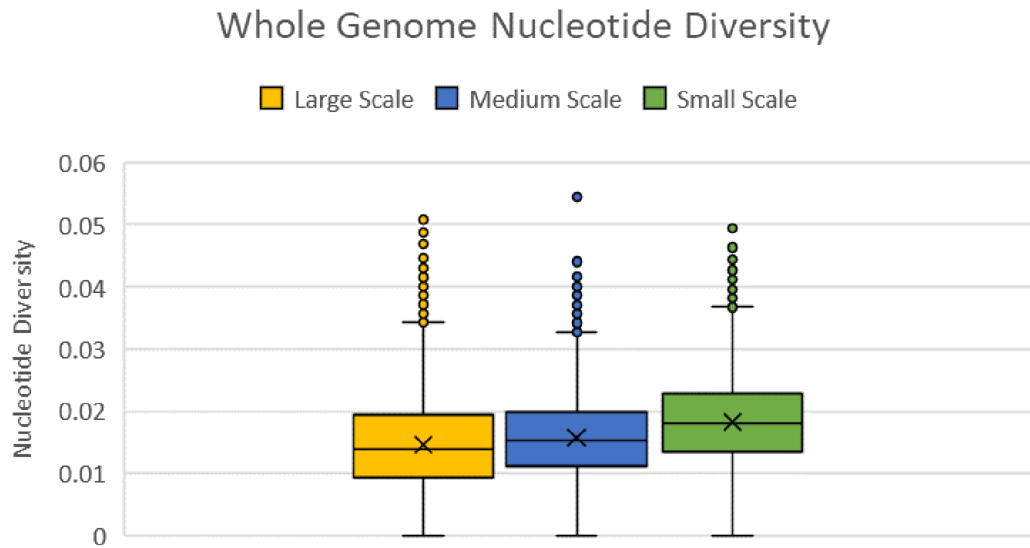


Figure 7: Whole genome sliding window box plot of nucleotide diversity (π) for three populations of *Hermetia illucens.* Sliding windows of 20kb nucleotides were used to cover whole genomes for each population. The small-scale population contained the highest π of the three populations. The large-scale population (n=2) is represented in yellow; medium-scale population (n=6) in blue; and small-scale population (n=6) in green.

A similar pattern holds across the first three chromosomes for each population, where the LS population has the lowest, MS in the middle, and the SS population with the highest nucleotide diversity. Chromosome 4 appeared equally diverse across the populations. The LS population topped the MS population mean diversity on chromosome 5, but still followed the SS. Chromosome 6 displayed the only time that the LS population had the highest mean nucleotide diversity of the three populations. Chromosome 7 expressed lower diversity than the other six chromosomes, LS and MS populations equal to 0.009 π and the SS population at 0.012 π (Figure 8).

http://etd.uwc.ac.za

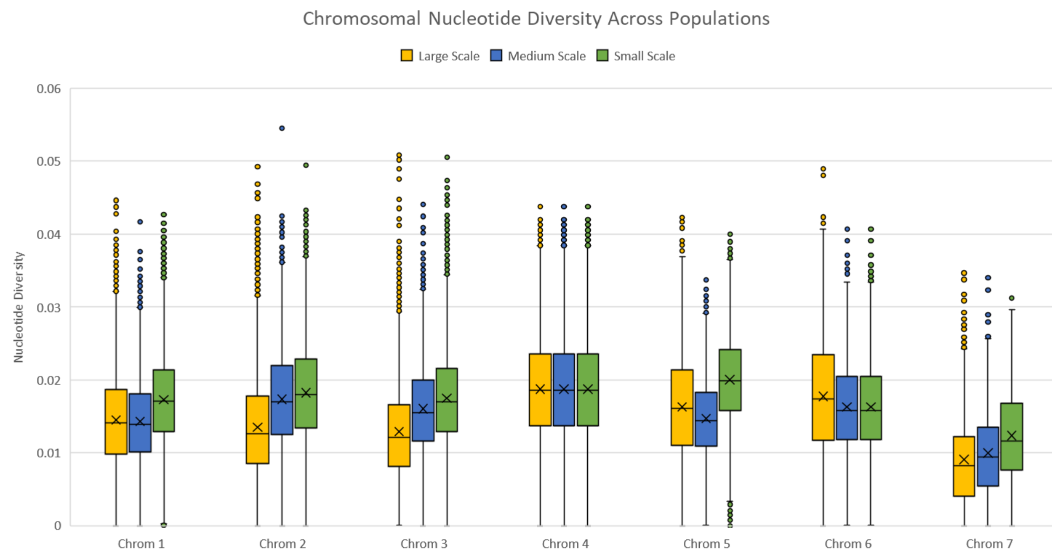


Figure 8: Chromosomal sliding window box plot of nucleotide diversity (π) for three populations of *Hermetia illucens*. Sliding windows of 20kb nucleotides were used to cover chromosomal genomes for each population. The large-scale population (n=2) is represented in yellow; medium-scale population (n=6) in blue; and small-scale population (n=6) in green. (Chrom # = chromosomal number)

The 14 individuals were split into male and female for further analysis of nucleotide diversity. Chromosomes 1 and 2 both displayed male diversity higher than the females in the LS population, then switching for the females to be higher on chromosomes 3, 4, 6, and 7. Chromosome 5 was almost identical between sexes at 0.015 π for each. The MS males were higher in diversity on chromosomes 1, 2, 4, and 5. Chromosome 5 was distinctively lower diversity for females with 0.012 for females and 0.016 for males. Females had higher nucleotide diversity on chromosomes 3, 6, and 7. Interestingly, the SS population males had higher nucleotide diversity across all chromosomes except chromosome 7 (Figure 9).

http://etd.uwc.ac.za

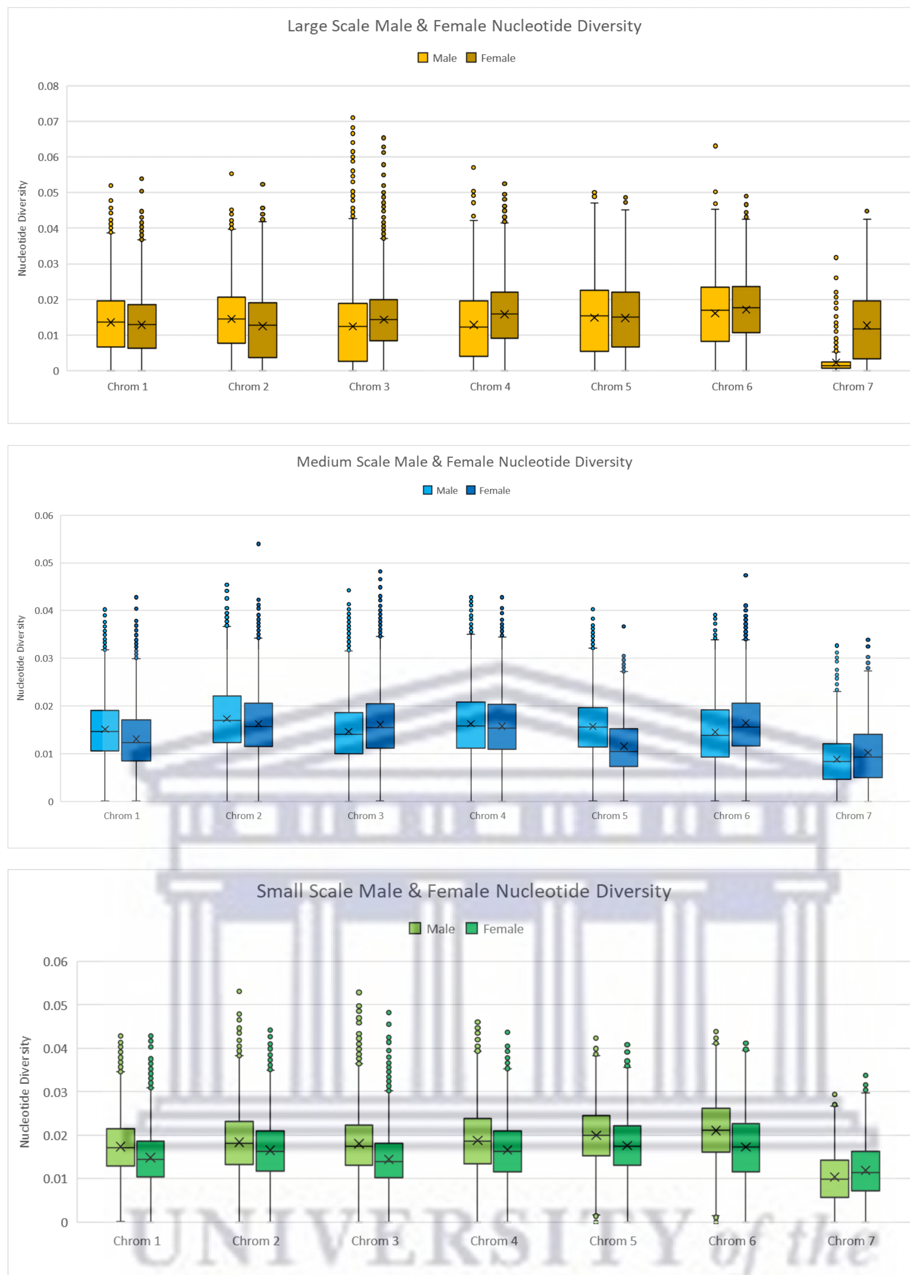


Figure 9: Chromosomal sliding window box plot of nucleotide diversity (π) of three populations of male and female *Hermetia illucens*. Sliding windows of 20kb nucleotides were used to cover chromosomal genomes for each population. The large-scale male population (n=1) is represented in light yellow and female (n=1) in dark yellow; medium-scale male population (n=3) in light blue and female (n=3) in dark blue; and small-scale male population (n=3) in light green and female (n=3) in dark green. (Chrom # = chromosomal number)

http://etd.uwc.ac.za

3.3.5 Tajima's D (TD) more extreme in MS and LS rearing

The mean TD across the whole genome for the LS population was the lowest (0.171) of the three populations, with MS (0.263) and SS (0.289) populations being very similar (Figure 6). Across the chromosomes, at least 50% of TD values in the LS population fell between -0.5 to 0.5. However, chromosomes 3 and 7 had more TD values in the negative than positive with 52% and 57% respectively. The range of TD values were tighter in the LS population (2.33 to -0.87) compared to MS (2.76 to -2.29) and SS (2.6 to -2.04). In general, the LS population mean TD values were lower than the MS and SS populations, except on chromosomes 1 and 2 (Figure 10). The MS population trended along the same values as the SS population, except for chromosome 2 where it was higher, and chromosome 5 where it was lower (Figure 10). The SS population remained steadily above 0.2 TD across all chromosomes except 1 and 2 where the values were closer to neutral (0).

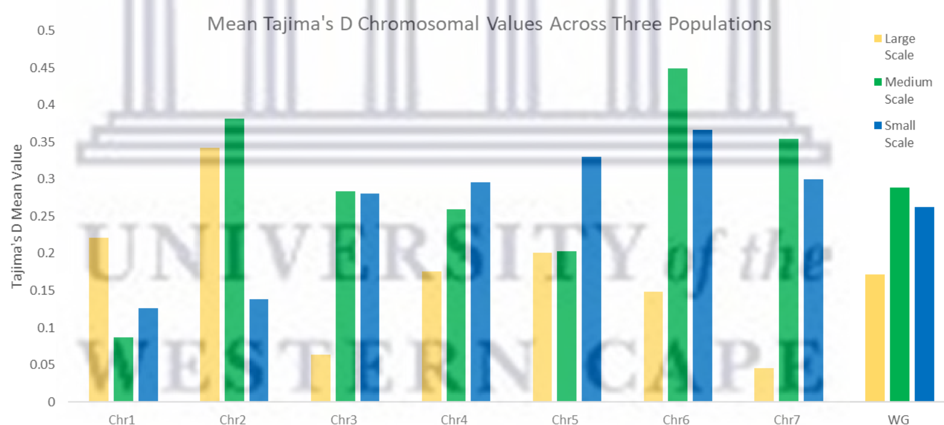


Figure 10: Mean Tajima's D sliding window chromosomal values across three *Hermetia illucens* populations. Sliding windows of 20kb nucleotides were used to span chromosomal regions. The large-scale population is represented in yellow, medium-scale in green, and small-scale in blue. (Chr# = chromosome number; WG = whole genome)

http://etd.uwc.ac.za

Less than 10% of TD values in the LS population were between +1.5 to 2.5, except chromosome 2 and 7 which had 16.8% and 11.3%, respectively, as shown in Supplementary Table S1 (Appendix A) and Figure 11.

http://etd.uwc.ac.za

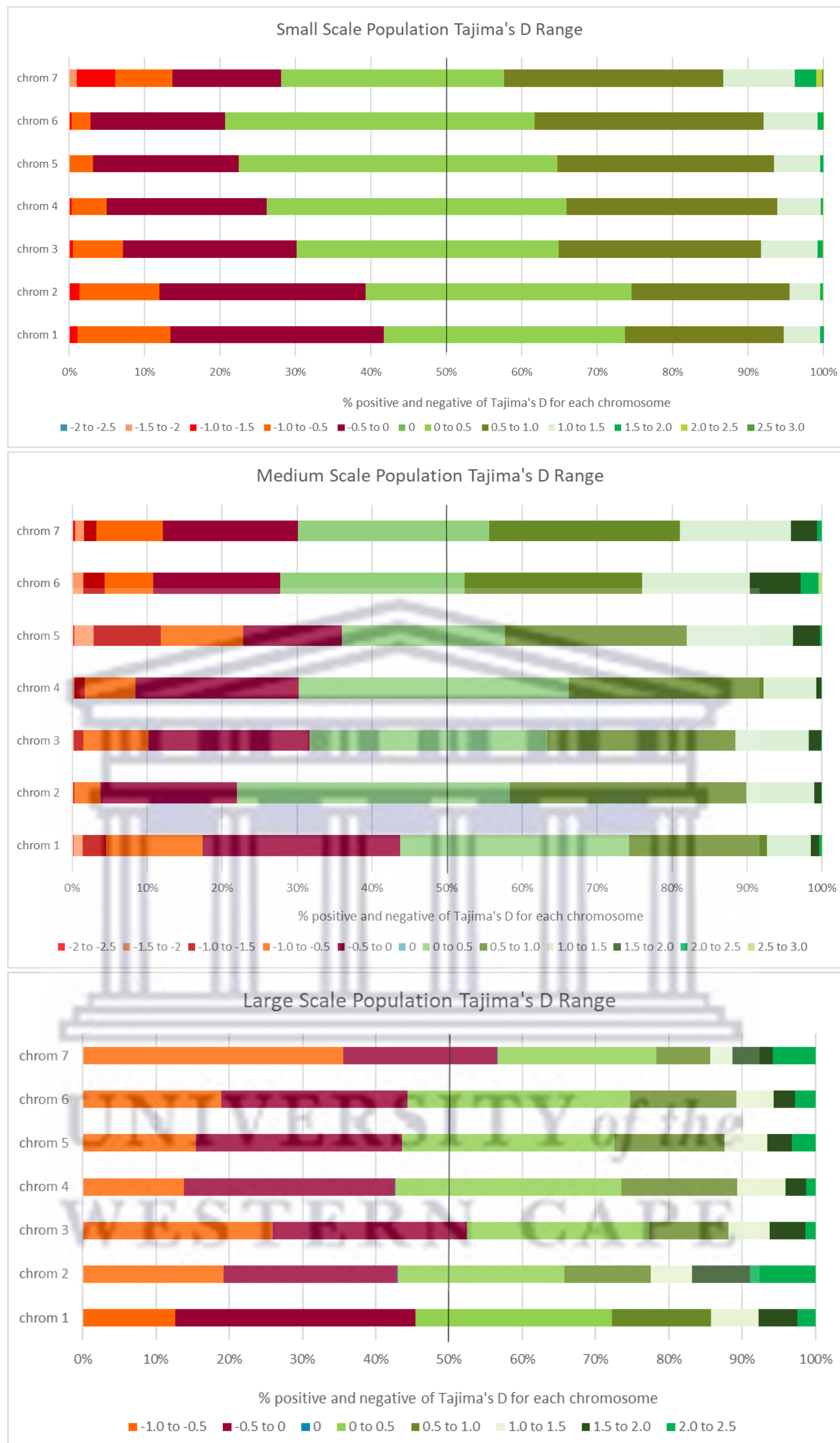


Figure 11: Percent positive and negative Tajima's D of three populations of *Hermetia illucens* by chromosome. Green themed brackets are positive, red themed brackets are negative. Neutral (0) Tajima's D were found only in the LS population and were too small in scale to visualize therefore, a black line indicates neutral.

http://etd.uwc.ac.za

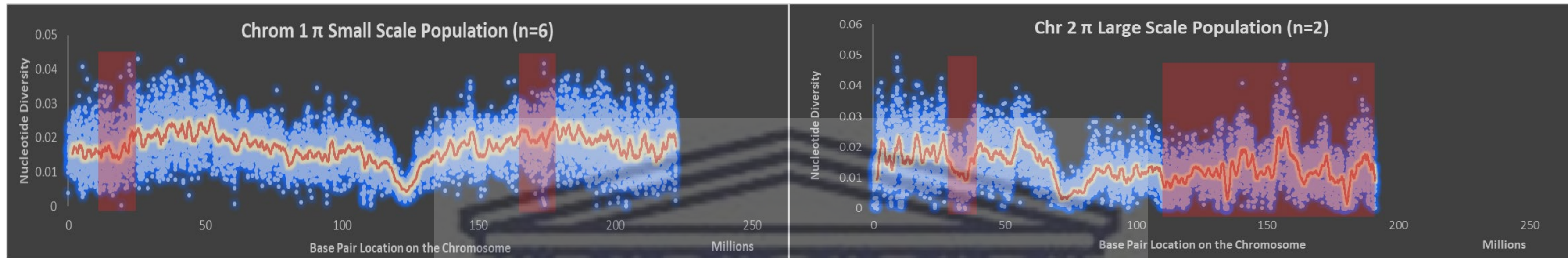
Chromosome 2 in the LS population had the highest percentage of TD values greater than + 2 comprising 9% of the values, with all other chromosomes only having 3% or less of their values hitting above 2 TD. The MS population was all majority positive TD, with chromosome 1 coming the closest to majority negative with 44% negative TDs. The MS population had the most TD values hitting above 2 (240) or below -2 (50) (Supplementary Table Table S1 (Appendix A); Figure 11). The SS population displayed a bell-like curve with the majority of values in the -0.5 to 0.5 range and only a few on the far ends of the values (Supplementary Table Table S1 (Appendix A); Figure 11).

3.3.6 Chromosomal regions of interest under potential selection pressure

Regions of interest were defined as areas on the chromosomes that displayed both low nucleotide diversity and high TD (±2). The SS population produced no TD values above or below ±2 across all chromosomes. Chromosomes 1 and 2 of the LS population displayed highly positive TD values (> 2) that were grouped together in large runs (Figure 11, A). These areas coincided with drops in the nucleotide diversity on the chromosome as well. However, chromosome 2 showed a large region, almost 100 kb in length, with varying levels of nucleotide diversity (Figure 11, A). The MS population is the only one that displayed TD values lower than -2 (Figure 11, B). Chromosome 1 had a roughly 20 million base pair region of highly negative TD values, and chromosome 5 had a smaller 10 million base pair region.

http://etd.uwc.ac.za

A



B

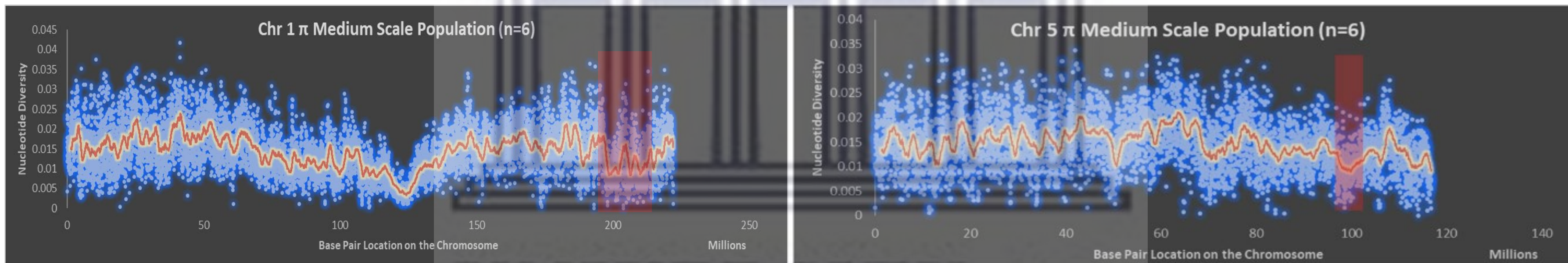


Figure 12: Nucleotide diversity scatter plots with regions of interest along the *Hermetia illucens* genome. (A) Chromosome 1 and 2 of the LS population regions with Tajima's D highly positive (> 2) values highlighted in red. (B) Chromosome 1 and 5 of the MS regions with Tajima's D highly negative (< -2) values highlighted in red.

UNIVERSITY of the
WESTERN CAPE
http://etd.uwc.ac.za

3.4 Discussion

The value of BSF as a scalable resource for both feed and waste management continues to increase over the years, especially considering current global climate change. This study investigated genetic diversity and established SNPs across three populations at varying levels of scaled rearing. Understanding the population structure and genetic variation between populations and the impact of scaled rearing on *H. illucens* will help to create sustainable breeding protocols for large-scale operations.

A high number of SNPs were found on all samples, even after quality filtering. The average 1.75 Ts/Tv ratios were only slightly lower than the expected non-human 1.8 ratio, however this is not unusual, particularly in insects (150). The inbreeding coefficients were all positive but low values, indicating that there were fewer sites of heterozygosity than expected, however, no values that would indicate poor mapping. Chromosome 7 has not yet been defined as a sex chromosome, but the data from this study corroborates the data in other studies that suggest chromosome 7 is a sex chromosome (120, 121).

The results of the PCA showed all three populations very closely grouped together. The SS were the only ones to have outlying samples of the three populations. Even though the three populations were collected from separate geographic locations and rearing conditions, the parental origins of each population are unknown. Considering that most artificially reared BSF populations are propagated from similar breed stocks purchased from limited breeder sources in the United States or China (28, 119, 151), the similarities in the population's genetic profiles are not unexpected. The separation of the SS population may represent the higher genetic

http://etd.uwc.ac.za

diversity found in that population's genomes. The split of the SS population in the PCA highlights the potential impact of scaling up rearing on the genetic diversity of BSF, as it has been shown that BSF undergo intense founder effect selective pressures in the first generation that is artificially reared (133). The number of unique and shared variants found between the three populations also corresponds with the grouping found in the PCA (Figure 5). The SS population produced the highest number of unique SNPs between the three populations. In addition, the large unique genomic diversity found in the SS population was not found in the other two populations. The MS group had higher numbers of unique variants than the LS group, however both were a steep decline from the SS population. The MS population shared the most SNPs with the other populations, potentially due to selection pressures of medium scale rearing helping to maintain genetic diversity more than current large scale rearing practices. The unique SNPs found in each population may offer insight into what selection pressures scaled rearing practices generate. The patterns of the LS and MS populations separating from the SS group in the PCA and displayed in the lack of unique SNPs, corresponds to similar changes in lab-reared colonies of insects (31, 102, 103).

The mean nucleotide diversity (π) of all three populations was low (1.4% to 1.8%) but is similar to the other reported π in another BSF population of 1.5% (120). Nucleotide diversity remained similar across all chromosomes, except for chromosome 7, which had below 1% for all populations. This may be due to it being an unreported sex chromosome. The MS population's chromosome 5 reported the lowest π, outside of chromosome 7, with 1.2%. Interestingly, the TD mean values described in these populations were substantially lower than the only other reported

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za

TD value of BSF, which may be due to differences between the genetically distinct populations (127). While there are no mean TDs of negative value, a closer look at their breakdown (Figure 11) shows a more equal spread between negative and positive values for chromosomes across all three populations. Only the LS population had two chromosomes (3 & 7) with the majority being negative TD values. Both low nucleotide diversity and negative TD are indicators of higher numbers of rare alleles being found and could highlight areas undergoing a selective sweep in the LS genome (111-113). The low mean TDs highlighted specific regions across the genomes that had significant increases or decreases in TD values. When defining significance using TD changes on WGS data, any shift from the "norm" or average value can be considered significant due to the large amount of data from which to calculate the values (152). As this study is a first pass through reporting WGS genetic variation data of BSF, only TD values that went above 2 or below -2 were examined, to focus on potential regions of interest that may be undergoing selection pressure. The SS population did not show any extreme ranges in TD, which corresponds to a more diverse genome undergoing less selection pressures. The LS population had two chromosomal regions (chromosomes 1 and 2) with highly positive TD values (> 2) that may indicate a balancing selection pressure on a population that has undergone a bottleneck (Figure, A & B). The MS population's uniquely negative (< -2) TD regions may indicate areas that are experiencing a selection pressure in those regions and could be the focus of a future study. Chromosome 5 was also found to be of interest in another study of BSF, with a large run of homozygosity found (120). Overall, the LS population appears to be under the highest level of selection pressure, with chromosomes 3 and 7 displaying

http://etd.uwc.ac.za

mainly negative potentially selective sweeps. Along with the negative TDs, the LS population had the greatest number of TD values outside of the majority (above or below -0.5 to 0.5). While it did not have the highest values found, which was in the MS population, it appeared to have the most values differing from the majority.

3.4.1 Limitations and future work

The genetic variation data established in this study were all completed as a baseline study for a relatively small group of BSF. The WGS data provides an excellent resource for narrowing the search for regions of interest on the BSF genome that have not yet been described. However, future work should involve a larger sample size, as this study was unable to make direct comparisons between sexes due to the change in samples sequenced in the LS population. Therefore, future studies focusing on larger sample sizes, with the same low coverage will help to increase the certainty of the genetic diversity of BSF. A study with deeper sequencing of chromosomal regions of interest and corresponding phenotypic data aimed at the unique SNPs found in the populations could emphasize pressures of different scaled rearing protocols on BSF. Also, focusing on the areas that may be under selection pressure will highlight potential candidate genes for breeding protocols.

3.4.2 Conclusion

The description of the genetic diversity between three populations of BSF highlights the effect of scaled rearing environments on the species genome. While the differences are not enough to distinguish the populations between LS and MS, the SS population expressed more diversity and less selective pressure on its genome. The need to fully understand the genome of BSF and how it adapts to LS artificial rearing environments will only continue to grow in importance as the

http://etd.uwc.ac.za

demand for sustainable feed and waste management resources grows in the coming years. This study therefore helps to create a wider base of knowledge and understanding surrounding the BSF populations being mass produced to meet this need.

http://etd.uwc.ac.za

Chapter 4: microRNA profile of *Hermetia illucens* (black soldier fly) and its implications on mass rearing

4.1 Introduction

The world population is projected to reach 9.8 billion by the year 2050 (128), meaning food production needs to increase by 50% (1). However, food production is expected to decrease by as much as 25% by 2050 due to climate change and unsustainable traditional methods of manufacturing food (1). To combat this decline, new methods of attaining the three major resources involved in food production including food, water, and energy, must be found (43). Due to the increasing demands for protein, it is important to explore sustainable, alternative, large-scale protein sources that are environmentally friendly.

Insects are one alternative source of protein, with an estimated 2,000 species already being consumed mainly in low- and middle-income countries (12, 130). The black soldier fly (BSF) is a species found worldwide and can be used as a sustainable animal and fish feed (19, 20, 75, 78). Depending on the organic substrate the black soldier fly larvae (BSFL) are fed and the extraction process, they can contain a range of 40-60% protein, up to 47% lipid, and 3-8% chitin in prepupae (15, 25, 63-66). This makes them not only a good protein source, but shows potential for a variety of derived products, and has minimal environmental impacts. Due to these facts, and that they are not a nuisance species, or mechanical disease vector (21, 102), they are an ideal candidate for LS rearing for food sustainability and waste reduction (70, 153).

http://etd.uwc.ac.za

The potential uses for BSF are numerous; however, until recently little has been done to fully understand their genetics and the implications of mass rearing. The majority of studies have only explored the BSF midgut microbiota (122-125). In 2015, a draft genome was published (121), and in 2020 the full genome was sequenced at a depth of 300x coverage (119). Another study still in preprint, has covered the genome using long read PacBio sequencing (120). The midgut transcriptome, and the application of the gene modification tool CRISPR/Cas-9 have also been recently reported for adapting BSF to industrialization (119). Multiple insect genomics studies (barring BSF) have demonstrated the role of microRNA in regulating biological and behavioural functions at different life stages (38, 41, 42, 117). As arthropod genetic diversity is known to diminish through colonization (102, 154), looking directly at gene regulation via microRNA expression explores how the BSF adapts to the mass rearing environment. Mapping the baseline microRNA expression data will provide crucial information on the large-scale development of BSF into an economically sustainable protein source. This study creates a database of novel microRNAs and their expression levels across five life stages and two sexes in BSF and provides candidate microRNAs with potential impact on BSF mass rearing.

UNIVERSITY of the
WESTERN CAPE

4.2 Methods

4.2.1 Sample collection

Hermetia illucens (black soldier fly, BSF) were reared under factory breeding conditions by AgriProtein Technologies Ltd. in Philippi, South Africa. The flies were fed *ad libitum* on a standardized proprietary diet based on the commercial

http://etd.uwc.ac.za

composition of layer hen feed. All life stages were kept at 28°C (±2°C), 80% relative humidity, under 12-hour day and night cycles.

Fifteen specimens were selected across five life stages: egg, larva, pupa, adult unmated female, and adult unmated male. The 15 specimens consisted of triplicates (biological replicates) for each of the five life stages. In order to acquire enough material to extract microRNAs according to methodology(155, 156), an egg batch laid from a single female was used for each egg replicate. The egg batches were laid one day prior to collection. Larvae were collected during the L5 instar, which was defined as day 21 (±2) from egg. The L5 instar was selected due to its size and ease of identification, along with the life stage's importance in protein production. The pupal stage was defined as day 28 (±2) from egg. Adult females and males were collected before being allowed to mate.

4.2.2 microRNA extraction

All specimens were harvested on the same day and transported to the lab in sterile falcon tubes (Corning, USA). The flies were left to acclimate to the lab for 60 minutes after transport to allow any stress from travel to reduce. The samples were then flash frozen with liquid nitrogen and ground into a powder. The lysing agent QIAzol (157) was used, and the microRNA extraction kit miRNeasy Mini Kit (155)(MD, USA) from Qiagen following standard manufacturer's protocols. A Qubit RNA Assay (Invitrogen™) was performed for quality control and to ensure enough microRNA was extracted for sequencing. After extraction, all samples were stored at -80°C. To ship samples for sequencing, the microRNA was stabilized at room temperature using RNAstable® (Biomatrica®).

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

4.2.3 Sequencing

Extracted samples were shipped to Macrogen Inc. (Seoul, South Korea) for small RNA sequencing. Illumina TruSeq Small RNA Library construction of the 15 samples was completed. Sequencing was done on an Illumina HiSeq 2500 machine with 8 million reads per sample.

4.2.4 Read processing and analysis

Sequences from the 15 different libraries were quality checked using FastQC (v0.11.7)(137) Reads were filtered for any non-canonical letters, and the 3' adapters were trimmed, followed by the removal of any reads shorter than 18 nucleotides long using the MiRDeep2 (v.2.0.1.2) program (158). The genome of *H. illucens* was used to map the reads (NCBI: assembly iHerIll2.2.curated20191125). The genome was indexed for mapping using Bowtie (v1.1.1)(159).

After mapping the reads, MiRDeep2 was used to identify any exact matches to published microRNAs (miRBase, v22.1) (160-167) found in the sample reads. Initially, all known microRNAs from miRBase were used, then filtered for arthropod species only for more closely related relevance of the microRNAs. The microRNAs found were given read counts using the Quantifier.pl script from MiRDeep2.

4.2.5 Differential expression

Differential expression was calculated using the DESeq2 (168) package in the R programming language. Expression was calculated based on read counts that were normalized using DeSeq2 statistical program. DeSeq2 normalizes read counts for sequencing depth and RNA composition. Five different life stages were compared:

http://etd.uwc.ac.za

Egg to Larva, Larva to Pupa, Pupa to Female, and Pupa to Male. DESeq2 utilized a negative binomial generalized linear model to test for statistical significance in expression between the life stages. Raw read counts are normalized using a median of ratios method which accounts for sequencing depth and RNA composition. Therefore, read counts were normalized for more than reads per million to account for differences in sequencing quality and length of microRNAs. All analyses were completed using R (169-171). See Appendix C for analysis pipeline.

4.3 Results

4.3.1 Sequencing and mapping

The microRNA from five different life stages, (egg, larva, pupa, adult unmated female, and adult unmated male) of *H. illucens* were extracted and sequenced to complete a baseline assessment of the species. Illumina TruSeq Small Library platform was used to create microRNA libraries. The total raw read count for the life stages ranged from 10.6-16.4 million reads with an average sequence length of 51 nucleotides (Table 2). More than 98% of the reads had Phred quality values (PQV) of 20 and 96% with PQV of 30. After the reads were filtered for lengths smaller than 18 nucleotides and any non-canonical values, the average reads were 10.9, 9.1, 11.4, 12.1, and 10.5 million for egg, larva, pupa, adult unmated female, and adult unmated male, respectively.

The percentages of total filtered reads that mapped to the BSF genome were: 3.0% of egg, 0.5% larva, 1.4% pupa, 8.2% female, and 6.4% male libraries. The second larval replicate only had 5.4 million reads after filtering, which lowered the overall larval library mapping (Table 2).

http://etd.uwc.ac.za

Table 2: Summary of *Hermetia illucens* reads sequenced and mapped. Abbreviations: E =egg; L= larva; P= pupa; F=female; M=male; 1-3=replicate; avg=average of 3 replicates.

	Total Raw Reads	Total Filtered Reads	Mapped Reads	Unmapped Reads	Mapped (%)	Unmapped (%)
1E	10,634,038	10,485,727	202,996	10,282,731	1.9%	98.1%
2E	11,391,345	1,058,752	312,488	10,746,264	2.8%	97.2%
3E	11,630,268	11,143,061	471,985	10,671,076	4.2%	95.8%
E avg	11,218,550	10,895,847	329,156	10,566,690	3.0%	97.0%
1L	13,716,268	11,013,749	46,243	10,967,506	0.4%	99.6%
2L	16,363,975	5,374,681	18,318	5,356,363	0.3%	99.7%
3L	14,753,374	1,077,735	74,337	10,703,020	0.7%	99.3%
L avg	14,944,539	5,822,055	46,299	9,008,963	0.5%	99.5%
1P	11,717,837	10,156,757	157,402	9,999,355	1.5%	98.5%
2P	14,142,734	11,927,696	265,560	11,662,136	2.2%	97.8%
3P	12,674,729	11,982,861	89,089	11,893,772	0.7%	99.3%
P avg	12,845,100	11,355,771	170,684	11,185,088	1.5%	98.5%
1M	10,958,506	9,495,353	955,239	8,540,114	10.1%	89.9%
2M	12,428,650	11,973,265	474,068	11,499,197	4.0%	96.0%
3M	10,665,444	10,037,438	527,529	9,509,909	5.3%	94.7%
M avg	11,350,867	10,502,019	652,279	9,849,740	6.5%	93.5%
1F	13,126,901	12,795,661	585,428	12,210,233	4.6%	95.4%
2F	14,284,210	13,860,109	579,082	13,281,027	4.2%	95.8%
3F	11,691,808	9,695,009	1,560,275	8,134,734	16.1%	83.9%
F avg	13,034,306	12,116,926	908,262	11,208,665	8.3%	91.7%

4.3.2 Identification of novel microRNAs in *Hermetia illucens*

Analysis of the 15 BSF libraries was completed using MiRDeep2 software (v.2.0.1.2), and alignment of all known arthropod microRNAs from miRbase (v22.1) was used to discover the baseline microRNAs of BSF. As BSF is a novel species with no known/reported microRNA, all microRNA found were considered novel. Any microRNA identified to be conserved across other known arthropods are referred to as conserved, and the microRNA that have yet to be identified in any other species are defined as unique microRNA. In order to be considered unique,

http://etd.uwc.ac.za

both the mature and star sequence needed to be present and found in at least 2 of the 15 libraries.

A total of 192 novel microRNAs were found across the 15 BSF libraries. Of these, 168 were found to be orthologous to known arthropod microRNAs (Supplementary Table S2, Appendix A), and 24 were found to be unique microRNA of BSF (Supplementary Table S3, Appendix A), not identified in any other species.

4.3.3 Differential expression

microRNA differential expression was calculated with log fold changes (LFC) of 0, 2, 5, and 10 in expression between life stages of egg to larva, larva to pupa, pupa to adult female, pupa to adult male.

Twenty-six of the 168 microRNAs conserved across arthropods had a statistically significant (p-adjusted value < 0.05) differential expression between egg to larval stages. Nearly half (14) of these 26 microRNA were downregulated and 12 were upregulated. Development from larva to pupa was characterized by 16 statistically significant (p-adjusted value < 0.05) differentially expressed microRNA with equal amount up- and downregulated transcripts. Differential expression of seven microRNA were detected as statistically significant (p-adjusted value < 0.05) between pupa to adult female (3 microRNA upregulated and 4 microRNAs downregulated). microRNAs from pupa to adult male had 9 statistically significant expression changes (4 microRNAs upregulated and 5 downregulated (Table 3).

http://etd.uwc.ac.za

Table 3: microRNA with statistically significant regulation between life stages of novel *Hermetia illucens* microRNA with the same seed as known arthropod microRNA.

Life Stage	Egg to Larva	Larva to Pupa	Pupa to Female	Pupa to Male
Up Regulated	12	8	3	4
Down Regulated	14	8	4	5
Total microRNA w/ p-adj < 0.05	26	16	7	9
Total microRNA	168	168	168	168

A total of 44 microRNAs (23% of the total microRNAs) showed differential expression between at least two stages with an LFC of ±2. These included 31 microRNAs conserved across all arthropods and 13 unique to BSF. A total of 40 microRNAs (21% of the total microRNAs) showed differential expression between at least two life stages with an LFC of ±5 (30 conserved across all arthropods and 10 unique to BSF). Finally, 21 microRNA were identified with an LFC of ±10 (11% of total microRNA). These included 18 microRNAs conserved across arthropods and 3 microRNAs unique to BSF (Figure 13 and Table 4).

Seven of the unique 24 microRNA were detected exclusively in single life stages. The egg life stage expressed five microRNAs not seen in any other life stages (*hil-miR-m*, *hil-miR-p, hil-miR-r, hil-miR-s,* and *hil-miR-u*; Supplementary Table S3, Appendix A). The female adult and pupa life stages expressed one microRNA each (*hil-miR-h* and *hil-miR-ac* respectively; Supplementary Table S3, Appendix A). Both male and female adult life stages expressed 3 shared microRNA (*hil-miR-a, hil-miR-b,* and *hil-miR-y;* Supplementary Table S3, Appendix A). There were no unique microRNAs found only in the larva stage.

http://etd.uwc.ac.za

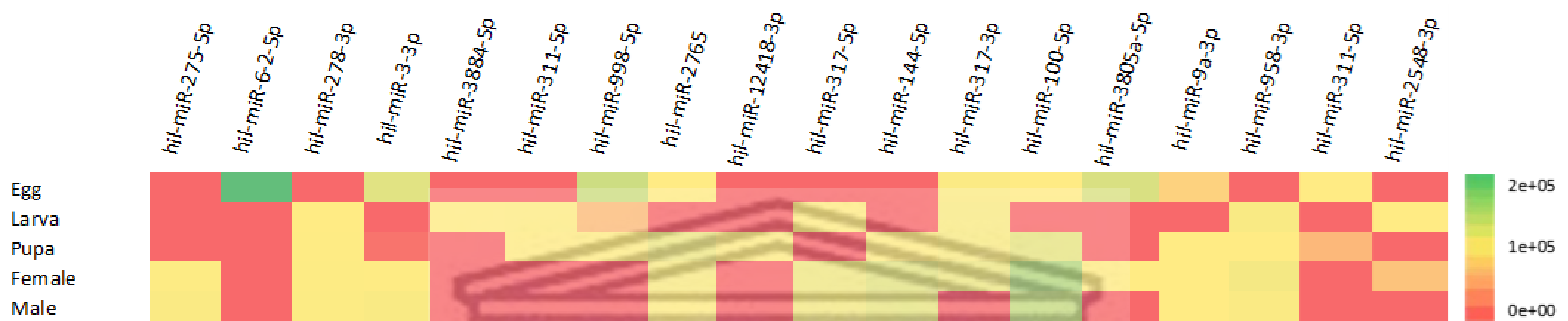


Figure 13: Heatmap of microRNA expression of *Hermetia illucens* with a log fold change (LFC) of ±10 (p-adjusted value <0.05) across life stages. Expression is based on an average read count of triplicate samples for each life stage.

UNIVERSITY of the
WESTERN CAPE

Figure 14: Heatmap of unique microRNA expression of *Hermetia illucens* across life stages. Expression is based on an average read count of triplicate samples for each life stage.

http://etd.uwc.ac.za

Table 4: Summary of *Hermetia illucens* microRNAs differential expression log fold change (LFC).

Log Fold Change	Total microRNA	LFC ± 2	LFC ± 5	LFC ± 10
Known arthropod microRNA	168	31	30	18
***H. illucens* unique**	24	13	10	3
Total microRNA	192	44	40	21

4.3.4 Predicted target genes

BSF microRNAs were searched against experimentally validated target genes in *Drosophila Melanogaster* (*D. melanogaster)* using Flybase.org (FB2020_06) (172) and MirTarBase (http://mirtarbase.cuhk.edu.cn/php/index.php) (173) to extrapolate BSF genes targeted for microRNA regulation. *D. melanogaster* was chosen as the closest relative to BSF with experimentally validated target genes. These databases represent curated published, and experimentally validated microRNA-target gene pathways. Fifty-six genes were found to be targets of 22 BSF microRNAs. The target genes covered a range of developmental functions in Drosophila (Table 5). In the absence of experimentally verified BSF microRNA targets, the closely related insect species (*D. melanogaster)* provided greater understanding of the development of the fly during rearing. The predicted target genes are characterized by 21 known gene families, the majority identified as C2H2 Zinc Finger Transcription Factors gene family. The second highest representative family was the Basic Helix-Loop-Helix Transcription Factors (negative regulators of notch signaling pathway), followed by RHG Proteins family. The Bearded Gene Family and ABCG ATP-Binding Cassette Transporter Subfamily matched two genes each. The other 16 gene families were each represented by single genes. Each of these genes were targeted by multiple microRNAs.

http://etd.uwc.ac.za

Table 5: Novel *Hermetia illucens* microRNA with orthologous microRNA and their predicted target genes in Drosophila. Abbreviations: LFC = Log Fold Change between life stages listed

Novel *H. illucens* microRNA with same seed	Known Predicted Target Gene in *Drosophila*	Egg to Larva LFC	Larva to Pupa LFC	Pupa to Female LFC	Pupa to Male LFC
hil-bantam-3p	W	2.563	-0.926	-0.78	-1.172
hil-let-7-5p	ab	2.074	8.84	0.832	0.192
hil-miR-124-3p	ana, gli	-4.198	5.269	1.485	0.65
hil-miR-1-3p	Amyotrophic lateral sclerosis 2, CG11377, CG17065, CG18542, CG31121, Chd64, crim, DI, Jafrac2, Msr-110, Nedd4, sinu, tutl	0.403	0.48	-1.166	-1.493
hil-miR-14-3p	EcR, Ice, IP3K2, sug	1.84	0.078	-0.675	-0.828
hil-miR-263a-5p	W	-0.335	2.089	1.466	1.184
hil-miR-276a-3p	DopR	1.816	2.127	0.417	0.008
hil-miR-278-3p	ex, tup	10.304	-0.466	2.175	1.362
hil-miR-279-3p	esg, nerfin-1, os, SP555	-3.171	0.839	0.388	0.604
hil-miR-283-5p	cos, smo	1.562	-2.857	-0.412	-0.267
hil-miR-2a-3p	grim, reaper, skl, hid, malpha, HLHmdelta	-3.466	1.563	0.821	0.693
hil-miR-315-5p	Axn, Notum	-1.824	1.066	0.843	0.772
hil-miR-316-5p	IA-2	7.816	0.174	2.418	1.813
hil-miR-317-3p	yellow-c	1.164	-1.118	1.486	14.041
hil-miR-34-5p	Eip74EF, Su(z)12	6.163	-6.769	-6.126	-5.636
hil-miR-5-5p	smo	2.074	0	-4.82	-6.715
hil-miR-7-5p	e, fng, h, HLHm5, I(1)MZ4, iHog, ttk	-9.097	3.818	-0.892	0.181
hil-miR-8-3p	CG13060, CG32767, CG8420, Cpr56F, ena, Gug, pan, ush, wls	2.353	-0.304	-0.043	0.173
hil-miR-92a-3p	sha	-7.694	2.342	0.993	2.052
hil-miR-9a-5p	Bx, dg, sens	-2.725	-0.306	1.621	1.746
hil-miR-iab-4-5p	abd-A, Ubx	-6.179	4.019	-0.978	1.564
hil-miR-iab-8-5p	abd-A, Abd-B, Ubx	2.074	0	0	-7.464

http://etd.uwc.ac.za

4.4 Discussion

This study established a baseline survey of novel microRNAs, both conserved and unique, in the BSF and their expression levels across 5 different life stages and identified likely target genes for these microRNAs. Like most species, the BSF genome retains a highly conserved microRNA library as evidenced by the large numbers of orthologous microRNAs found from a wide range of arthropod species in the miRBase database. Focusing our microRNA discovery pipeline on known arthropod species microRNA only, allowed for a stronger comparative analysis of potential microRNA function in BSF. Finding orthologs of microRNAs across species results in a higher power of discovery for *de novo* microRNAs (174). The high number of orthologs found for each microRNA family, along with the strict identification criterion utilized by the miRDeep2 program for mature, star, and precursor sequences, provide high confidence in the microRNAs found in the BSF genome. These new BSF microRNA add another species that follow the highly conserved nature of microRNA across all species.

Overall, 192 novel microRNAs were found in BSF, with the vast majority conserved (87%). The 24 unique microRNAs, that passed the criteria to be included, came from a group of 91 potentially unique sequences found. The criteria used ensured that reporting would only include highly likely candidates. However, the many sequences not considered candidates represent a potential larger network of microRNAs for future study. The limited number of libraries per life stage (3) could mean some candidate sequences were missed. BSF has already been seen to have a divergent and large genomic landscape (119), and the number of microRNAs follows suit.

UNIVERSITY of the WESTERN CAPE

http://etd.uwc.ac.za

4.4.1 Stage specific microRNA expression

Understanding the basic expression levels across the life stages has led to a better understanding of stage specific microRNA regulation. Seventy-four conserved microRNAs were found to be specific to a single life stage, while only 27 were found across all life stages (Supplementary Table Table S2, Appendix A). These may be important for gene regulation linked to life stage specific development. The unique microRNAs identified had a number of life stage specific microRNAs: (Egg) *hil-miR-m*, *hil-miR-p, hil-miR-r, hil-miR-s, hil-miR-u,* (Pupa) *hil-miR-ac*, (Female) *hil-miR-h*, (Male and Female) *hil-miR-a, hil-miR-b,* and *hil-miR-y.* These 10 represent nearly half (42%) of the unique microRNAs found. This life stage specificity may be due to too low frequencies of the fragments to be detected through sequencing in the other life stages. However, these 10 unique microRNAs provide insight into the egg, larva, pupa, and adult life stages of BSF, and can be used to compare the developmental regulatory differences of microRNAs to those of other insect species (175). The inclusion of separate male and female life stages allows insight into the sex-linked microRNAs of BSF, and their potential functions in development in sex differentiation (176, 177). For instance, *hil-miR-h, hil-miR-317-3p, hil-miR-3805a-5p,* and *hil-miR-2548-3p* displayed different expression levels between female and male samples. The unique microRNA of *hil-miR-af* was found in the larval and male stages, but not the female life stage. These stage specific microRNAs provide valuable developmental information unique to BSF showing their genetic adaptability and divergence from other insect species (118, 119, 175).

http://etd.uwc.ac.za

Eighteen conserved microRNAs were found with substantial (LFC 10) expression regulation between at least two life stages. While these do not have any known experimentally validated predicted target genes, the level of differential expression makes them microRNA of interest for future study.

4.4.2 microRNA target prediction

Looking at the functions of possible target genes highlights how important the regulation of microRNA is in the development of insects. While most of the predicted target genes deal with cell development, differentiation, and death, the target genes discussed below were selected for having the greatest potential impacts on the mass rearing of BSF (Table 5).

bantam-3p has been well established in *D. melanogaster* and involved in the cell differentiation, apoptosis, neural development, and germ line maintenance (178, 179). The microRNA is expressed during the 3rd larval instar stage, regulating optical disc and photo receptor differentiation (179). *bantam* is linked to mediating the circadian rhythm proteins (*clk*) in *D. melanogaster*. Overexpression of *bantam* causes a lengthening of the circadian period (180). BSF *bantam* expression follows a path consistent with the developmental role it has been identified to play with higher expression in larva, followed by a decreasing trend from larva to pupa, and both adult stages where it was heavily downregulated. Both factors may be exploited for optimal mass rearing as BSF have been shown to have a direct relationship between length of light exposure and egg production (181).

The *let-7* microRNA is one of the first to be identified in *C. elegans* and is known to be highly conserved across most species (182). Controlling expression of *let-7* is

http://etd.uwc.ac.za

essential to prevent deleterious phenotypes such as wing, fertility, motility, and flight deficiencies from unusual abdominal musculature maturation (39, 182-185). *let*-7 has been shown to regulate the ecdysis pathways during molting stages in ticks (186) and silkworms (187). The microRNA has been seen to increase in expression over the third larval instar, with highest expression levels in the pupa stage, in *D. melanogaster* (182). BSF *let*-7 follows the same expression pattern, peaking in expression during the pupal stage.

Overexpression of *mir-14* leads to lean *D. melanogaster* (188). *Sugarbabe* is one of the target genes for *mir-14* and has been shown to be controlled by both diet nutrient levels and *mir-14* targeting in the face of nutrient deprivation (189). The non-nutrient dependent nature of *mir-14* regulation has led to flies being obese during starvation (189). As BSFL fat content is a known beneficial part of the role of BSF as sustainable feed additive (190), being able to potentially increase fat content would be a useful tool in the mass rearing process.

4.4.3 Future work

The baseline of conserved and unique novel microRNAs in the BSF genome were all identified by utilizing established computational methods. As BSF is a novel species with a relatively new genome, the baseline data identified here should be confirmed through laboratory validation. These microRNA sequences should be used to confirm the predicted target genes and regulatory pathways. Once validated, the microRNA can be harnessed in breeding protocols to improve mass rearing of BSF by understanding how they adapt to different environments (38-40) and are being used in the grape and olive industries (36, 191).

http://etd.uwc.ac.za

4.4.4 Conclusions

The conserved and unique microRNA described in this study form an essential understanding of gene regulation in the economically and environmentally important species of *H. illucens* and will provide potential targets for genetic manipulation of this species in order to improve its use as an alternative protein source.

http://etd.uwc.ac.za

Chapter 5: Conclusions and future work

5.1 Key findings

The global challenges facing the future of food security include being able to meet the protein demands of growing populations of low- and middle-income countries. Not only does protein production need to increase, sustainable and resource preserving methods also need to be found (1, 2, 4-6). The use of insect meal as a feed ingredient is one solution, and BSF are considered the "crown jewel" of the industry (17). The biology of the BSF lends to a shortened production cycle due to their shorter life cycle (67), and reduction in organic waste that would normally be sent to landfill (67-70). However, the challenges facing the BSF industry are threefold: public acceptance, legalisation, and scalability (16). This study focused on the scalability issues involved with mass rearing and finding genetic data to support selective breeding programs ensuring efficient and healthy fly colonies.

Genomic diversity of a population is an indicator of the general health of a species, as well as adaptability and selection pressures faced (30). Lower genetic diversity has been found in captive populations of BSF, regardless of geography, highlighting the importance of understanding the trade-offs involved with artificial rearing of populations (28). By comparing WGS data of three populations reared at different scales, the effect of scale on the genetic profile of BSF was explored. The inbreeding coefficients and Ts/Tv ratios showed nothing abnormal in the sequence data or between populations. PCA and identification of unique and shared SNPs across the three populations indicated that the SS population did not group together with the MS and LS populations. The SS population also had the most unique SNPs found between the three, followed by MS and LS. Examining areas of the genome

http://etd.uwc.ac.za

that had low nucleotide diversity and extreme Tajima's D values (±2) revealed regions at a chromosomal level that may be undergoing selection pressures. Six regions of interest on four chromosomes were defined in the LS and MS populations. The SS population did not have any extreme TD values. Interestingly, the LS population displayed the most negative TD values, while not all were extreme, variation of TD values in WGS data can be considered important based on deviation from the majority values. This indicates that the LS population was undergoing the most pressures of the three populations. Captive BSF have been shown to lose genetic diversity and display little distinct genetic footprints without suffering from high inbreeding coefficients (28). Breeding programs in Asia that currently utilize outcrossing with local wild populations did not see meaningful divergence from other non-outcrossed captive populations (28), therefore discovering the causes for such uniformity in reared populations is necessary. While it is expected that LS, MS, and SS populations would have large, medium, and small effects, identifying the specific corresponding genetic data in this study provides a platform to understanding the potentially negative or positive rearing conditions for BSF.

UNIVERSITY of the
WESTERN CAPE

Along with describing genetic diversity and regions undergoing potential selection pressures, a baseline survey of novel microRNA, both unique to BSF and conserved across arthropod species, was completed. Of the 192 BSF microRNAs found, the majority were conserved across other insects, with 24 being unique to BSF. In other studies, mapping of insect microRNA has helped the understanding of biological and behavioural functions being regulated at different developmental life stages (38, 41, 42, 117). The expression profile reported in this study was able to define

http://etd.uwc.ac.za

74 conserved microRNAs specific to a single life stage and 27 conserved found in all life stages (Supplementary Table S2, Appendix A). Almost half of the unique microRNAs found were life stage specific. Separating out male and female adult microRNAs identified multiple with differing expression levels between the sexes. As this a baseline study, stringent parameters were adhered to for distinguishing microRNAs and their expression levels. Differential expression analysis was able to specify changes in log fold values of 2, 5, and 10 between life stages, with highlighting 18 microRNAs a LFC of 10. The highly conserved nature of microRNA provides the ability to compare BSF profiles to other closely related insects who have already had their functions experimentally validated (192). Some microRNAs have known to have target genes they are regulating through experimental validation (38, 178, 182, 186, 189). Three validated microRNA (*bantam, let-7, mir*-14) target genes were discussed as potential targets based on their relevance to mass rearing. *bantam* is linked to circadian periods (180). *let-7* has been shown to influence negative phenotypes and moulting stages (185-187). Finally, controlling the expression of *mir-14* has manifested in non-nutrient dependent obesity in flies (189).

5.2 Limitations and future work

The novel work completed in this study required some limitations on conclusions that could be drawn. During the course of investigation, two high quality genomes were published. While these publications greatly advanced the work being done, experimental design for sequencing had already been decided and completed. The genomes for the populations of the genetic diversity study were sequenced at different depths. Due to cost restraints, a higher depth of coverage could only be

http://etd.uwc.ac.za

completed on two genomes, which were going to be used as a "reference" genome, with the other two populations being mapped to it. However, that was no longer necessary, and in the end all three populations were mapped to the published reference genome. Had this reference genome been available before sequencing, more samples would have been sequenced at lower coverage to obtain a higher sample size for stronger power in experimental design.

The baseline nature of the data described contributes towards the growing body of knowledge on the genetics of BSF and lays groundwork for future clarification of the interactions between the optimal genotypes and phenotypes of mass reared BSF. Future work would be able to zero in on the areas of interest discovered in this study, allowing for a deeper coverage of these potentially pivotal chromosomal regions. Additionally, more samples being added to a database of SNPs of mass reared BSF will benefit validation and implications for linking beneficial phenotypes. Further validation of the novel microRNAs found through established computational methods would include experimental validation and potentially knock out regulatory studies linked to phenotype.

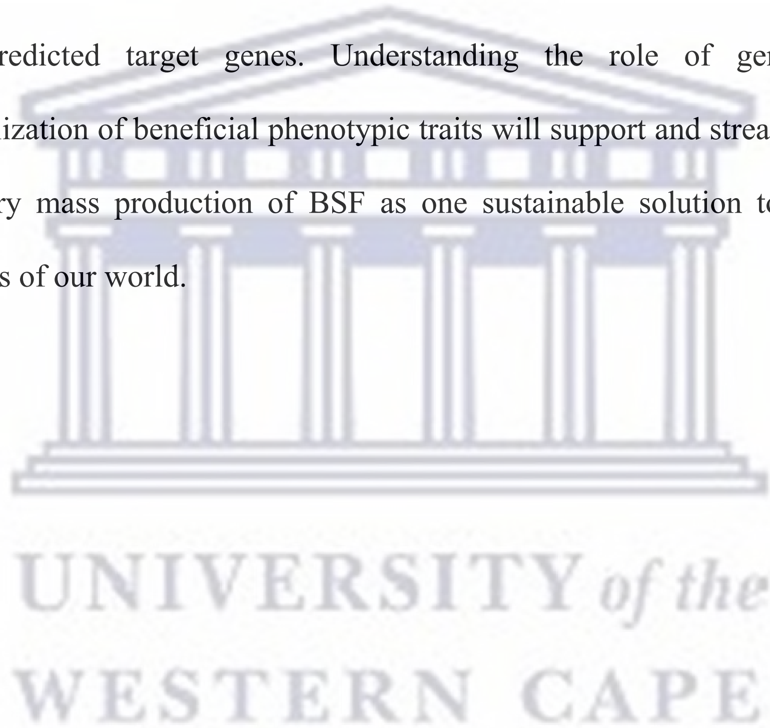
5.3 Concluding remarks

The current concerns for food security are centred around a growing population and the consequences of unsustainable food production methods. The problem calls for solutions that are not only able to meet protein demand but be able to be environmentally sustainable by not relying on overuse of precious natural resources like land and water (1, 2, 6). Along with resource draining methods, the system currently includes inefficient waste management protocols by losing 21% of the

http://etd.uwc.ac.za

food produced (44, 193). The BSF industry involves upcycling and valorising waste into much needed products like protein and fertilizer.

There are limited data available on the genetics of BSF, which have a divergent and complex genomic landscape (28, 119, 120, 127, 131). Understanding the genetic pathways to beneficial phenotypes used to make in-demand products is crucial for this nascent industry. This study adds to the current published genetics by describing the implications to the genetic diversity and regions of chromosomal selection pressures under varying levels of scaled rearing. Additionally, this project was the first to report a survey and life stage expression profiling of microRNA and their predicted target genes. Understanding the role of genetics in the materialization of beneficial phenotypic traits will support and streamline the very necessary mass production of BSF as one sustainable solution to the growing demands of our world.


UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

References

1. Nations FaAOotU. The State of Food and Agriculture: Climate Change, Agriculture and Food Security. Rome: United Nations; 2016. Report No.: 0081-4539.
2. United Nations Department of Economic and Social Affairs PD. World population prospects 2022: Summary of results. UN DESA/POP/2022/TR/NO3. 2022.
3. Salter AM. Improving the sustainability of global meat and milk production. Proc Nutr Soc. 2017;76(1):22-7.
4. IPBES. IPBES (2019): Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn, Germany: IPBES secretariat; 2019.
5. Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, et al. Global consequences of land use. Science. 2005;309.
6. Garnett T. Food sustainability: problems, perspectives and solutions. Proc Nutr Soc. 2013;72(1):29-39.
7. Programme UNE. Food waste index report 2021. Nairobi. 2021.
8. FAO, IFAD, UNICEF, WFP, WHO. The State of Food Security and Nutrition in the World 2022: Repurposing food and agricultural policies to make healthy diets more affordable. Rome: FAO; 2022.
9. Van Huis A, Van Itterbeeck J, Klunder H, Mertens E, Halloran A, Muir G, et al. Edible insects future propspects of food and feed security. FAO Forestry Paper 171 ed. UR FaW, editor: Food and Agriculture Organization of the United Nations; 2013 May 2013.
10. Cortes Ortiz JA, Ruiz AT, Morales-Ramos JA, Thomas M, Rojas MG, Tomberlin JK, et al. Insect Mass Production Technologies. Insects as Sustainable Food Ingredients. 2016. p. 153-201.
11. Madhan S, Prabakaran M, Kwang J. Baculovirus as vaccine vectors. Current Gene Therapy. 2010;10:201-13.
12. Ramos-Elorduy J, Gonzalez EA, Hernandez AR, Pino JM. Use of *Tenebrio molitor* (Coleoptera: Tenebrionidae) to Recycle Organic Wastes and as Feed for Broiler Chickens. Veterinary Entomology. 2002;95(1):214-20.
13. Oonincx DG, de Boer IJ. Environmental impact of the production of mealworms as a protein source for humans - a life cycle assessment. PLoS One. 2012;7(12):e51145.
14. Govorushko S. Global status of insects as food and feed source: A review. Trends in Food Science & Technology. 2019;91:436-45.
15. Meneguz M, Schiavone A, Gai F, Dama A, Lussiana C, Renna M, et al. Effect of rearing substrate on growth performance, waste reduction efficiency and chemical composition of black soldier fly (*Hermetia illucens*) larvae. J Sci Food Agric. 2018;98(15):5776-84.
16. (IPIFF) IPoIfFF. IPIFF position paper on the use of insect larvae as feed for food producing animals. 2021.

http://etd.uwc.ac.za

17. Tomberlin JK, van Huis A. Black soldier fly from pest to 'crown jewel' of the insects as feed industry: an historical perspective. Journal of Insects as Food and Feed. 2020;6(1):1-4.
18. Tomberlin JK, Sheppard DC. Factors influencing mating and oviposition of black soldier flies (Diptera: Stratiomyidae) in a colony. Journal of Entomological Science. 2002;37:345-52.
19. Bondari K, Sheppard DC. Soldier Fly Larvae as Feed in Commercial Fish Production. Aquaculture. 1981;24:103-9.
20. Hale OM. Dried *Hermetia illucens* larvae (Diptera: Stratiomyidae) as a feed additive for poultry. Journal of the Georgia Entomological Society. 2013;8:16-20.
21. Furman DP, Young RD, Catts EP. *Hermetia illucens* (Linnaeus) as a Factor in the Natural Control of *Musca domestica* Linnaeus. Journal of Economic Entomology. 1959;52(5):917-21.
22. Biasato I, Renna M, Gai F, Dabbou S, Meneguz M, Perona G, et al. Partially defatted black soldier fly larva meal inclusion in piglet diets: effects on the growth performance, nutrient digestibility, blood profile, gut morphology and histological features. J Anim Sci Biotechnol. 2019;10:12.
23. Foysal MJ, Fotedar R, Tay C-Y, Gupta SK. Dietary supplementation of black soldier fly (*Hermetica illucens*) meal modulates gut microbiota, innate immune response and health status of marron (*Cherax cainii*, Austin 2002) fed poultry-by-product and fishmeal based diets. PeerJ. 2019;7.
24. Veldkamp T, Vernooij AG. Use of insect products in pig diets. Journal of Insects as Food and Feed. 2021;7(5):781-93.
25. Diener S, Zurbrugg C, Tockner K. Conversion of organic material by black soldier fly larvae: establishing optimal feeding rates. Waste Manag Res. 2009;27(6):603-10.
26. Surendra KC, Tomberlin JK, van Huis A, Cammack JA, Heckmann LL, Khanal SK. Rethinking organic wastes bioconversion: Evaluating the potential of the black soldier fly (*Hermetia illucens* (L.)) (Diptera: Stratiomyidae) (BSF). Waste Manag. 2020;117:58-80.
27. Barragan-Fonseca KY, Nurfikari A, van de Zande EM, Wantulla M, van Loon JJA, de Boer W, et al. Insect frass and exuviae to promote plant growth and health. Trends Plant Sci. 2022;27(7):646-54.
28. Kaya C, Generalovic TN, Stahls G, Hauser M, Samayoa AC, Nunes-Silva CG, et al. Global population genetic structure and demographic trajectories of the black soldier fly, *Hermetia illucens*. BMC Biol. 2021;19(1):94.
29. Petracci M, Cavani C. Muscle growth and poultry meat quality issues. Nutrients. 2012;4(1):1-12.
30. van Marle-Koster E, Visser C. Unintended consequences of selection for increased production on the health and welfare of livestock. Arch Anim Breed. 2021;64(1):177-85.
31. Berlocher SH, Friedman S. Loss of Genetic Variation in Laboratory Colonies of *Phormia regina*. *Entomologia Experimentalis et Applicata*. 1981;30(3):205-8.
32. Benjelloun B, Alberto FJ, Streeter I, Boyer F, Coissac E, Stucki S, et al. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. Front Genet. 2015;6:107.

http://etd.uwc.ac.za

33. Marsden CD, Lee Y, Kreppel K, Weakley A, Cornel A, Ferguson HM, et al. Diversity, Differentiation, and Linkage Disequilibrium: Prospects for Association Mapping in the Malaria Vector *Anopheles arabiensis*. G3 Genes|Genomes|Genetics. 2014;4(1):121-31.
34. Giri L, Jugran AK, Bahukhandi A, Dhyani P, Bhatt ID, Rawal RS, et al. Population Genetic Structure and Marker Trait Associations Using Morphological, Phytochemical and Molecular Parameters in *Habenaria edgeworthii*—a Threatened Medicinal Orchid of West Himalaya, India. Applied Biochemistry and Biotechnology. 2017;181(1):267-82.
35. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, et al. Reconstructing Native American population history. Nature. 2012;488(7411):370-4.
36. Khadari B, El Aabidine AZ, Grout C, Ben Sadok I, Doligez A, Moutier N, et al. A Genetic Linkage Map of Olive Based on Amplified Fragment Length Polymorphism, Intersimple Sequence Repeat and Simple Sequence Repeat Markers. Journal of the American Society for Horticultural Science. 2010;135(6):548-55.
37. Borneman AR, Schmidt SA, Pretorius IS. At the cutting-edge of grape and wine biotechnology. Trends Genet. 2013;29(4):263-71.
38. Asgari S. MicroRNA functions in insects. Insect Biochem Mol Biol. 2013;43(4):388-97.
39. Caygill EE, Johnston LA. Temporal regulation of metamorphic processes in Drosophila by the *let-7* and *miR-125* heterochronic microRNAs. Curr Biol. 2008;18(13):943-50.
40. Sokol NS, Ambros V. Mesodermally expressed Drosophila *microRNA-1* is regulated by Twist and is required in muscles during larval growth. Genes Dev. 2005;19(19):2343-54.
41. Wei Y, Chen S, Yang P, Ma Z, Kang L. Characterization and comparative profiling of the small RNA transcriptomes in two phases of locust. Genome Biol. 2009;10(1):R6.
42. Behura SK. Insect microRNAs: Structure, function and evolution. Insect Biochem Mol Biol. 2007;37(1):3-9.
43. Trines E, Hohne N, Jung M, Skutsch M, Petsonk A, Silva-Chavez G, et al. Climante Change Scientific Assessment and Policy Analysis. Netherlands: Netherlands Environmental Assessment Agency; 2006. Contract No.: 500102002.
44. FAO. The State of World Fisheries and Aquaculture 2022: Towards blue transformation. Rome: FAO; 2022.
45. Hardy K, Radini A, Buckley S, Blasco R, Copeland L, Burjachs F, et al. Diet and environment 1.2 million years ago revealed through analysis of dental calculus from Europe's oldest hominin at Sima del Elefante, Spain. Naturwissenschaften. 2017;104(1-2):2.
46. Mitsuhashi J, Capinera JL. Entomophagy: Human consumption of insects. Heidelberg: Springer Science + Business Media B. V.; 2008.
47. Yi C, He Q, Wang L, Kuang R. The utilization of insect-resources in Chinese rural area. Journal of agricultural science. 2010;2.
48. Akhtar Y, Isman MB. Insects as an Alternative Protein Source. Proteins in Food Processing. 2018. p. 263-88.

http://etd.uwc.ac.za

49. Costa-Neto EM. Entomotherapy, or the Medicinal Use of Insects. Journal of Ethnobiology. 2005;25(1):93-114.
50. Roque-Rodriguez FJ. Controlled Mass Rearing of Cochineal Insect (Hemiptera: Dactylopiidae) Using Two Laboratory-Scale Production Systems in Peru. J Insect Sci. 2022;22(1).
51. Zimmermann H, H. E, McFayden R. Annotated list of some cactus-feeding insects of South America Acta Zoologica Lilloana. 1979;33:101-12.
52. Yoon JS. Drosophilidae II: Drosophila species other than *D. melanogaster*. *Handbook of Insect Rearing 2.* 1985. p. 85-91.
53. Yang Y, Yang J, Wu WM, Zhao J, Song Y, Gao L, et al. Biodegradation and Mineralization of Polystyrene by Plastic-Eating Mealworms: Part 2. Role of Gut Microorganisms. Environ Sci Technol. 2015;49(20):12087-93.
54. Sogari G, Amato M, Biasato I, Chiesa S, Gasco L. The Potential Role of Insects as Feed: A Multi-Perspective Review. Animals. 2019;9(4):119.
55. Sealey WM, Gaylord TG, Barrows FT, Tomberlin JK, McGuire MA, Ross C, et al. Sensory analysis of Rainbow Trout, *Oncorhynchus mykiss*, fed enriched black soldier fly prepupae, *Hermetia illucens*. Journal of the World Aquaculture Society. 2011;42.
56. Lalander C, Vinnerås B. Actions needed before insects can contribute to a real closed-loop circular economy in the EU. Journal of Insects as Food and Feed. 2022;8(4):337-42.
57. Smetana S, Palanisamy M, Mathys A, Heinz V. Sustainability of insect use for feed and food: Life Cycle Assessment perspective. Journal of Cleaner Production. 2016;137:741-51.
58. Barros-Cordeiro KB, Bao SN, Pujol-Luz JR. Intra-puparial development of the black soldier-fly, *Hermetia illucens*. Journal of Inset Science. 2014;14.
59. 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). Journal of Cleaner Production. 2019;208:211-9.
60. Harnden LM, Tomberlin JK. Effects of temperature and diet on black soldier fly, Hermetia illucens (L.) (Diptera: Stratiomyidae), development. Forensic Sci Int. 2016;266:109-16.
61. Bertinetti C, Samayoa AC, Hwang SY. Effects of Feeding Adults of Hermetia illucens (Diptera: Stratiomyidae) on Longevity, Oviposition, and Egg Hatchability: Insights Into Optimizing Egg Production. J Insect Sci. 2019;19(1).
62. Copello A. Biology of *H. illucens*, the fly of Argentine Beehives. . Revista De La SEA. 1926;1:23-6.
63. Franco A, Scieuzo C, Salvia R, Petrone AM, Tafi E, Moretta A, et al. Lipids from *Hermetia illucens*, an Innovative and Sustainable Source. Sustainability. 2021;13(18):10198.
64. Booram CVJ, Newton GL, Hale OM, Barker RW. Manure residue as a substrate for protein production via *Hermetia illucens* larvae (for swine rations). . Cornell Agricultural Waste Management Conference; Ann Arbor, Michigan. 1977.
65. Smets R, Verbinnen B, Van De Voorde I, Aerts G, Claes J, Van Der Borght M. Sequential Extraction and Characterisation of Lipids, Proteins, and Chitin from Black Soldier Fly (*Hermetia illucens*) Larvae, Prepupae, and Pupae. Waste and Biomass Valorization. 2020;11(12):6455-66.

http://etd.uwc.ac.za

66. Shumo M, Osuga IM, Khamis FM, Tanga CM, Fiaboe KKM, Subramanian S, et al. The nutritive value of black soldier fly larvae reared on common organic waste streams in Kenya. Scientific Reports. 2019;9(1).
67. Sheppard DC, Newton GL, Thompson SA, Savage S. A value added manure management system using the black soldier fly. Bioresource Technology. 1994;50:275-9.
68. Diener S, Zurbrugg C, Gutierrez FR, Nguyen DH, Morel A, Koottatep T, et al., editors. Black soldier fly larvae for organic waste treatment-prospects and constraints. Proceedings of the WasteSafe 2011 - 2nd International Conference on Solid Waste Management in the Developing Countries; 2011 13-15 February, 2011; Khulna, Bangladesh.
69. Banks IJ, Gibson WT, Cameron MM. Growth rates of black soldier fly larvae fed on fresh human faeces and their implication for improving sanitation. Trop Med Int Health. 2014;19(1):14-22.
70. Miranda CD, Cammack JA, Tomberlin JK. Mass Production of the Black Soldier Fly, *Hermetia illucens* (L.), (Diptera: Stratiomyidae) Reared on Three Manure Types. Animals. 2020;10(7):1243.
71. Erickson MC, Islam M, Sheppard C, Liao J, Doyle MP. Reduction of *Escherichia coli* O157:H7 and *Salmonella enterica* Serovar Enteritidis in chicken manure by larvae of the black soldier fly. Journal of Food Protection. 2004;67:685-90.
72. Obiero K, Meulenbroek P, Drexler S, Dagne A, Akoll P, Odong R, et al. The Contribution of Fish to Food and Nutrition Security in Eastern Africa: Emerging Trends and Future Outlooks. Sustainability. 2019;11(6).
73. Park S-I, Chang BS, Yoe SM. Detection of antimicrobial substances from larvae of the black soldier fly,Hermetia illucens(Diptera: Stratiomyidae). Entomological Research. 2014;44(2):58-64.
74. Elhag O, Zhou D, Song Q, Soomro AA, Cai M, Zheng L, et al. Screening, Expression, Purification and Functional Characterization of Novel Antimicrobial Peptide Genes from *Hermetia illucens* (L.). PLOS ONE. 2017;12(1):e0169582.
75. Abdel-Latif HMR, Abdel-Tawwab M, Khalil RH, Metwally AA, Shakweer MS, Ghetas HA, et al. Black soldier fly (*Hermetia illucens*) larvae meal in diets of European seabass: Effects on antioxidative capacity, non-specific immunity, transcriptomic responses, and resistance to the challenge with *Vibrio alginolyticus*. Fish Shellfish Immunol. 2021;111:111-8.
76. Kroeckel S, Harjes AGE, Roth I, Katz H, Wuertz S, Susenbeth A, et al. When a turbot catches a fly: Evaluation of a pre-pupae meal of the Black Soldier Fly (*Hermetia illucens*) as fish meal substitute — Growth performance and chitin degradation in juvenile turbot (*Psetta maxima*). Aquaculture. 2012;364-365:345-52.
77. Magalhães R, Sánchez-López A, Leal RS, Martínez-Llorens S, Oliva-Teles A, Peres H. Black soldier fly (*Hermetia illucens*) pre-pupae meal as a fish meal replacement in diets for European seabass (*Dicentrarchus labrax*). Aquaculture. 2017;476:79-85.
78. Abdel-Tawwab M, Khalil RH, Metwally AA, Shakweer MS, Khallaf MA, Abdel-Latif HMR. Effects of black soldier fly *(Hermetia illucens* L.) larvae meal on growth performance, organs-somatic indices, body composition, and hemato-

http://etd.uwc.ac.za

biochemical variables of European sea bass, *Dicentrarchus labrax*. Aquaculture. 2020;522.
79. Wang G, Peng K, Hu J, Yi C, Chen X, Wu H, et al. Evaluation of defatted black soldier fly (*Hermetia illucens* L.) larvae meal as an alternative protein ingredient for juvenile Japanese seabass (*Lateolabrax japonicus*) diets. Aquaculture. 2019;507:144-54.
80. Li S, Ji H, Zhang B, Zhou J, Yu H. Defatted black soldier fly (Hermetia illucens) larvae meal in diets for juvenile Jian carp (*Cyprinus carpio* var. Jian): Growth performance, antioxidant enzyme activities, digestive enzyme activities, intestine and hepatopancreas histological structure. Aquaculture. 2017;477:62-70.
81. Zhou JS, Liu SS, Ji H, Yu HB. Effect of replacing dietary fish meal with black soldier fly larvae meal on growth and fatty acid composition of Jian carp (*Cyprinus carpio* var. Jian). Aquaculture Nutrition. 2018;24(1):424-33.
82. Xiao X, Jin P, Zheng L, Cai M, Yu Z, Yu J, et al. Effects of black soldier fly (*Hermetia illucens*) larvae meal protein as a fishmeal replacement on the growth and immune index of yellow catfish (*Pelteobagrus fulvidraco*). Aquaculture Research. 2018;49(4):1569-77.
83. Fawole FJ, Adeoye AA, Tiamiyu LO, Ajala KI, Obadara SO, Ganiyu IO. Substituting fishmeal with Hermetia illucens in the diets of African catfish (*Clarias gariepinus*): Effects on growth, nutrient utilization, haemato-physiological response, and oxidative stress biomarker. Aquaculture. 2020;518.
84. Limbu SM, Shoko AP, Ulotu EE, Luvanga SA, Munyi FM, John JO, et al. Black soldier fly (*Hermetia illucens*, L.) larvae meal improves growth performance, feed efficiency and economic returns of Nile tilapia (*Oreochromis niloticus*, L .) fry. Aquaculture, Fish and Fisheries. 2022;2(3):167-78.
85. Devic E, Leschen W, Murray F, Little DC. Growth performance, feed utilization and body composition of advanced nursing Nile tilapia (*Oreochromis niloticus*) fed diets containing Black Soldier Fly (*Hermetia illucens*) larvae meal. Aquaculture Nutrition. 2018;24(1):416-23.
86. Huyben D, Vidaković A, Werner Hallgren S, Langeland M. High-throughput sequencing of gut microbiota in rainbow trout (*Oncorhynchus mykiss*) fed larval and pre-pupae stages of black soldier fly (*Hermetia illucens*). Aquaculture. 2019;500:485-91.
87. 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.
88. Caimi C, Renna M, Lussiana C, Bonaldo A, Gariglio M, Meneguz M, et al. First insights on Black Soldier Fly (*Hermetia illucens* L.) larvae meal dietary administration in Siberian sturgeon (*Acipenser baerii* Brandt) juveniles. Aquaculture. 2020;515.
89. Star L, Arsiwalla T, Molist F, Leushuis R, Dalim M, Paul A. Gradual Provision of Live Black Soldier Fly (*Hermetia illucens*) Larvae to Older Laying Hens: Effect on Production Performance, Egg Quality, Feather Condition and Behavior. Animals (Basel). 2020;10(2).
90. Tahamtani FM, Ivarsson E, Wiklicky V, Lalander C, Wall H, Rodenburg TB, et al. Feeding live Black Soldier Fly larvae (*Hermetia illucens*) to laying hens: effects on feed consumption, hen health, hen behavior, and egg quality. Poult Sci. 2021;100(10):101400.

http://etd.uwc.ac.za

91. Chia SY, Tanga CM, Osuga IM, Alaru AO, Mwangi DM, Githinji M, et al. Effect of Dietary Replacement of Fishmeal by Insect Meal on Growth Performance, Blood Profiles and Economics of Growing Pigs in Kenya. Animals (Basel). 2019;9(10).
92. Chia SY, Tanga CM, Osuga IM, Alaru AO, Mwangi DM, Githinji M, et al. Black soldier fly larval meal in feed enhances growth performance, carcass yield and meat quality of finishing pigs. Journal of Insects as Food and Feed. 2021;7(4):433-47.
93. Newton GL, Booram CV, Barker RW, Hale OM. Dried *Hermetia illucens* larvae meal as a supplement for swine. Journal of Animal Science. 1977;44.
94. Gold M, Tomberlin JK, Diener S, Zurbrugg C, Mathys A. Decomposition of biowaste macronutrients, microbes, and chemicals in black soldier fly larval treatment: A review. Waste Manag. 2018;82:302-18.
95. Jonas E, de Koning DJ. Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Front Genet. 2015;6:49.
96. Hill WG. Applications of population genetics to animal breeding, from wright, fisher and lush to genomic prediction. Genetics. 2014;196(1):1-16.
97. Lecocq T. Insects: The Disregarded Domestication Histories. Animal Domestication. 2019.
98. Tixier-Boichard M, Leenstra F, Flock DK, Hocking PM, Weigend S. A century of poultry genetics. World's Poultry Science Journal. 2019;68(2):307-21.
99. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. The Genetics Society of America. 2001;157:1819-29.
100. Nebel RL, Jobst SM. Evaluation of systematic breeding programs for lactating dairy cows: a review. Journal of Dairy Science. 1998;81:1169-74.
101. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in dairy cattle: progress and challenges. J Dairy Sci. 2009;92(2):433-43.
102. Lainhart W, Conn JE, Rios CT, Vinetz JM, Bickersmith SA, Moreno M. Changes in Genetic Diversity from Field to Laboratory During Colonization of *Anopheles darlingi* Root (Diptera: Culicidae). The American Journal of Tropical Medicine and Hygiene. 2015;93(5):998-1001.
103. Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Mol Ecol. 2008;17(24):5177-88.
104. Yu HS, Shen YH, Yuan GX, Hu YG, Xu HE, Xiang ZH, et al. Evidence of selection at melanin synthesis pathway loci during silkworm domestication. Mol Biol Evol. 2011;28(6):1785-99.
105. Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 2005;15(11):1496-502.
106. Neilson ME, Stepien CA. Historic speciation and recent colonization of Eurasian monkey gobies (*Neogobius fluviatilis* and *N. pallasi*) revealed by DNA sequences, microsatellites, and morphology. Diversity and Distributions. 2011;17(4):688-702.

http://etd.uwc.ac.za

107. Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet. 2011;43(10):956-63.
108. Watanabe K, Asai K, Oshima T, Kuchino Y. Chemical structure and thermal properties of initiator tRNA from *Euphausia sperba* in comparison with those of other eucaryotic initiator tRNAs. The Journal of Biochemistry. 1981;90:1259-66.
109. Rousset F. Inbreeding and relatedness coefficients: what do they measure? Heredity (Edinb). 2002;88(5):371-80.
110. Nei M, Wen-Hsiung L. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 1979;76:5269-73.
111. Nei M. Analysis of Gene Diversity in Subdivided Populations. Proceedings of the National Academy of Sciences. 1973;70(12):3321-3.
112. Tatarinova TV, Chekalin E, Nikolsky Y, Bruskin S, Chebotarov D, McNally KL, et al. Nucleotide diversity analysis highlights functionally important genomic regions. Scientific Reports. 2016;6(1):35730.
113. Schmidt D, Pool J. The effect of population history on the distribution of the Tajima's D statistic. Population English Edition. 2002:1-8.
114. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics Society of America. 1989;123:585-95.
115. Huang Y, Liu Z, Rong YS. Genome Editing: From Drosophila to Non-Model Insects and Beyond. J Genet Genomics. 2016;43(5):263-72.
116. Etebari K, Asgari S. Accuracy of microRNA discovery pipelines in non-model organisms using closely related species genomes. PLoS One. 2014;9(1):e84747.
117. Lucas KJ, Myles KM, Raikhel AS. Small RNAs: a new frontier in mosquito biology. Trends Parasitol. 2013;29(6):295-303.
118. Lucas KJ, Zhao B, Liu S, Raikhel AS. Regulation of physiological processes by microRNAs in insects. Current Opinion in Insect Science. 2015;11:1-7.
119. Zhan S, Fang G, Cai M, Kou Z, Xu J, Cao Y, et al. Genomic landscape and genetic manipulation of the black soldier fly *Hermetia illucens*, a natural waste recycler. Cell Res. 2020;30(1):50-60.
120. Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. A high-quality, chromosome-level genome assembly of the Black Soldier Fly (*Hermetia illucens* L.). G3 Genes|Genomes|Genetics. 2021;11(5).
121. Vicoso B, Bachtrog D. Numerous transitions of sex chromosomes in Diptera. PLoS Biol. 2015;13(4):e1002078.
122. De Smet J, Wynants E, Cos P, Campenhout LV. Microbial Community Dynamics during Rearing of Black Soldier Fly Larvae (*Hermetia illucens)* and Impact on Exploitation Potential. Applied and Environmental Microbiology. 2018;84(9).
123. Jiang CL, Jin WZ, Tao XH, Zhang Q, Zhu J, Feng SY, 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.
124. Bonelli M, Bruno D, Brilli M, Gianfranceschi N, Tian L, Tettamanti G, et al. Black Soldier Fly Larvae Adapt to Different Food Substrates through Morphological and Functional Responses of the Midgut. Int J Mol Sci. 2020;21(14).

http://etd.uwc.ac.za

125. Bruno D, Bonelli M, Cadamuro AG, Reguzzoni M, Grimaldi A, Casartelli M, et al. The digestive system of the adult *Hermetia illucens* (Diptera: Stratiomyidae): morphological features and functional properties. Cell Tissue Res. 2019;378(2):221-38.
126. Guilliet J, Baudouin G, Pollet N, Filée J. The natural history of the black soldier fly, *Hermetia illucens*: insights from complete mitochondrial genome sequences. bioRxiv preprint 2021.
127. Sandrock C, Leupi S, Wohlfahrt J, Kaya C, Heuel M, Terranova M, et al. Genotype-by-Diet Interactions for Larval Performance and Body Composition Traits in the Black Soldier Fly, *Hermetia illucens*. Insects. 2022;13(5).
128. United Nations Department of Economic and Social Affairs PD. World Population Prospects The 2017 Revision: Key Findings and Advance Tables. United Nations; 2017. Report No.: ESA/P/WP/248.
129. Gustavsson J, Cederberg C, Sonesson U. Global food losses and food waste. Save Food at Interpack; Dusseldorf, Germany: Food and Agriculture Organization of the United Nations; 2011.
130. Shafique L, Abdel-Latif HMR, Hassan F-U, Alagawany M, Naiel MAE, Dawood MAO, et al. The Feasibility of Using Yellow Mealworms (*Tenebrio molitor*): Towards a Sustainable Aquafeed Industry. Animals. 2021;11(3):811.
131. Deraedt S, Bierman A, Van Heusden P, Richards C, Christoffels A. microRNA profile of *Hermetia illucens* (black soldier fly) and its implications on mass rearing. PLOS ONE. 2022;17(3):e0265492.
132. Zhou B, Ho SS, Zhang X, Pattni R, Haraksingh RR, Urban AE. Whole-genome sequencing analysis of copy number variation (CNV) using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. bioRxiv preprint. 2018.
133. Hoffmann L, Hull KL, Bierman A, Badenhorst R, Bester-van der Merwe AE, Rhode C. Patterns of Genetic Diversity and Mating Systems in a Mass-Reared Black Soldier Fly Colony. Insects. 2021;12(6).
134. Marquina D, Buczek M, Ronquist F, Lukasik P. The effect of ethanol concentration on the morphological and molecular preservation of insects for biodiversity studies. PeerJ. 2021;9:e10799.
135. Qiagen. DY14 Purification of total DNA from insects using the DNeasy Blood Tissue Kit. Qiagen; 2006.
136. TruSeq DNA PCR-free workflow reference In: illumina, editor. 1000000039279v00 ed: illumina; 2017.
137. Andrews S. FastQC: A quality control tool for high throughput sequence data [Online]. 2015.
138. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
139. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
140. Illumina adapter sequences. In: Illumina, editor. 1000000002694 v10 ed: Illumina; 2019.
141. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997v1 [q-bioGN]. 2013.

http://etd.uwc.ac.za

142. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.
143. Depristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 2011;43(5):491-8.
144. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1110):11 0 1- 0 33.
145. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2).
146. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8.
147. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
148. Team R. RStudio: Integrated development for R. In: RStudio, editor. Boston, MA: PBC; 2020.
149. Corporation M. Microsoft Excel. 2018.
150. Keller I, Bensasson D, Nichols RA. Transition-Transversion Bias Is Not Universal: A Counter Example from Grasshopper Pseudogenes. PLoS Genetics. 2007;3(2):e22.
151. Sheppard C, Tomberlin JK, Joyce JA, Kiser BC, Sumner SM. Rearing methods for the black soldier fly (Diptera: stratiomyidae). Journal of Medical Entomology. 2002;39:695-8.
152. Huang S, Deng L, Guan M, Li J, Lu K, Wang H, et al. Identification of genome-wide single nucleotide polymorphisms in allopolyploid crop Brassica napus. BMC Genomics. 2013;14(1):717.
153. Scala A, Cammack JA, Salvia R, Scieuzo C, Franco A, Bufo SA, et al. Rearing substrate impacts growth and macronutrient composition of *Hermetia illucens* (L.) (Diptera: Stratiomyidae) larvae produced at an industrial scale. Scientific Reports. 2020;10(1).
154. Godinho DP, Cruz MA, Charlery De La Masselière M, Teodoro-Paulo J, Eira C, Fragata I, et al. Creating outbred and inbred populations in haplodiploids to measure adaptive responses in the laboratory. Ecology and Evolution. 2020;10(14):7291-305.
155. Qiagen. miRNeasy Mini Handbook Qiagen; 2014.
156. Illumina. TruSeq small RNA library prep reference guide. In: illumina, editor. 15004197 v02 ed; 2016.
157. Qiagen. QIAzol Handbook for efficient lysis of fatty tissues and all other types of tissue before RNA purification. In: Qiagen, editor.: Qiagen; 2021.
158. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 2012;40(1):37-52.
159. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

http://etd.uwc.ac.za

160. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019;47(D1):D155-D62.
161. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68-73.
162. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39(Database issue):D152-7.
163. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, et al. A uniform system for microRNA annotation. RNA. 2003;9(3):277-9.
164. Griffiths-Jones S. The microRNA Registry. Nucleic Acids Res. 2004;32(Database issue):D109-11.
165. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. Nucleic Acids Res. 2003;31(1):439-41.
166. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34(Database issue):D140-4.
167. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008;36(Database issue):D154-8.
168. 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.
169. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
170. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2009.
171. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019;4(43).
172. Larkin A, Marygold SJ, Antonazzo G, Attrill H, Dos Santos G, Garapati PV, et al. FlyBase: updates to the *Drosophila melanogaster* knowledge base. Nucleic Acids Res. 2021;49(D1):D899-D907.
173. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res. 2020;48(D1):D148-D54.
174. Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ, et al. Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. Genome Res. 2007;17(12):1865-79.
175. Ylla G, Fromm B, Piulachs M-D, Belles X. The microRNA toolkit of insects. Scientific Reports. 2016;6(1):37736.
176. Peng W, Tariq K, Xie J, Zhang H. Identification and Characterization of Sex-Biased MicroRNAs in *Bactrocera dorsalis* (Hendel). PLoS One. 2016;11(7):e0159591.
177. Zhang X, Zheng Y, Jagadeeswaran G, Ren R, Sunkar R, Jiang H. Identification and developmental profiling of conserved and novel microRNAs in *Manduca sexta*. Insect Biochemistry and Molecular Biology. 2012;42(6):381-95.
178. Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. *bantam* Encodes a Developmentally Regulated microRNA that Controls Cell Proliferatino and regulates the Proapoptotic Gene *hid* in Drosophila. Cell. 2003;113(1):25-36.

http://etd.uwc.ac.za

179. Li Y, Padgett RW. *bantam* is required for optic lobe development and glial cell proliferation. PLoS One. 2012;7(3):e32910.
180. Kadener S, Menet JS, Sugino K, Horwich MD, Weissbein U, Nawathean P, et al. A role for microRNAs in the Drosophila circadian clock. Genes Dev. 2009;23(18):2179-91.
181. Hoc B, Noel G, Carpentier J, Francis F, Caparros Megido R. Optimization of black soldier fly (*Hermetia illucens*) artificial reproduction. PLoS One. 2019;14(4):e0216160.
182. Pasquinelli A, Reinhart B, Slack F, Martindale M, Kuroda M, Maller B, et al. Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. Nature. 2000;408:86-9.
183. Lee H, Han S, Kwon CS, Lee D. Biogenesis and regulation of the *let-7* miRNAs and their functional implications. Protein Cell. 2016;7(2):100-13.
184. Sokol NS, Xu P, Jan YN, Ambros V. Drosophila *let-7* microRNA is required for remodeling of the neuromusculature during metamorphosis. Genes Dev. 2008;22(12):1591-6.
185. Hertel J, Bartschat S, Wintsche A, Otto C, Students of the Bioinformatics Computer L, Stadler PF. Evolution of the *let-7* microRNA family. RNA Biol. 2012;9(3):231-41.
186. Wu F, Luo J, Chen Z, Ren Q, Xiao R, Liu W, et al. MicroRNA *let-7* regulates the expression of ecdysteroid receptor (ECR) in *Hyalomma asiaticum* (Acari: Ixodidae) ticks. Parasit Vectors. 2019;12(1):235.
187. Ling L, Ge X, Li Z, Zeng B, Xu J, Aslam AF, et al. MicroRNA *Let-7* regulates molting and metamorphosis in the silkworm, *Bombyx mori*. Insect Biochem Mol Biol. 2014;53:13-21.
188. Xu P, Vernooy SY, Guo M, Hay BA. The Drosophila MicroRNA *Mir-14* Suppresses Cell Death and Is Required for Normal Fat Metabolism. Current Biology. 2003;13(9):790-5.
189. Varghese J, Lim SF, Cohen SM. Drosophila *miR-14* regulates insulin production and metabolism through its target, sugarbabe. Genes Dev. 2010;24(24):2748-53.
190. Kim YB, Kim DH, Jeong SB, Lee JW, Kim TH, Lee HG, et al. Black soldier fly larvae oil as an alternative fat source in broiler nutrition. Poult Sci. 2020;99(6):3133-43.
191. Pulvirenti A, Giugno R, Distefano R, Pigola G, Mongiovi M, Giudice G, et al. A knowledge base for *Vitis vinifera* functional analysis. BMC Systems Biology. 2015;9(Suppl 3):S5.
192. Ylla G, Piulachs M-D, Belles X. Comparative analysis of miRNA expression during the development of insects of different metamorphosis modes and germ-band types. BMC Genomics. 2017;18(1).
193. Kronberg SL, Provenza FD, van Vliet S, Young SN. Review: Closing nutrient cycles for animal production - Current and future agroecological and socio-economic issues. Animal. 2021;15 Suppl 1:100285.

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

Appendix A: Supplementary tables

A.1 Supplementary Table 1

Table S1: Tajima's D values by chromosome for three populations of *Hermetia illucens*. Chr# = chromosome #; LS = Large-Scale population (n=2); MS = Medium-Scale population (n=6); SS = Small-Scale population (n=6).

		Tajima's D Range												Tajima's D Summary			
		2.5 to 3.0	2.0 to 2.5	1.5 to 2.0	1.0 to 1.5	0.5 to 1.0	0 to 0.5	0	-0.5 to 0	-1.0 to -0.5	-1.0 to -1.5	-1.5 to -2	-2 to -2.5	Positive	Negative	> 2	> -2
Chr 1	LS	0	276	584	713	1500	2971	4	3621	1398	0	0	0	6048	5019	276	0
	MS	6	33	129	646	2040	3394	0	2917	1432	342	131	26	6248	4848	39	26
	SS	1	1	41	534	2337	3554	0	3141	1367	125	0	0	6468	4633	2	0
Chr 2	LS	0	856	750	538	1127	2161	5	2261	1835	0	0	0	5437	4096	856	0
	MS	0	10	89	866	3012	3480	0	1740	326	26	6	0	7457	2098	10	0
	SS	0	4	39	384	2004	3369	0	2612	1015	123	5	0	5800	3755	4	0
Chr 3	LS	0	122	442	509	975	2231	6	2385	2333	0	0	0	4285	4718	122	0
	MS	0	14	145	883	2261	2856	0	1936	781	122	16	0	6159	2855	14	0
	SS	0	2	66	677	2419	3130	0	2077	599	43	2	0	6294	2721	2	0
Chr 4	LS	0	115	236	579	1362	2666	8	2489	1201	0	0	0	4966	3690	115	0
	MS	0	6	62	609	2243	3122	0	1882	585	117	23	2	6042	2609	6	2
	SS	0	3	23	502	2416	3440	0	1837	399	31	0	0	6384	2267	3	0
Chr 5	LS	0	187	195	346	873	1694	2	1641	903	0	0	0	3297	2544	187	0
	MS	1	18	205	829	1418	1272	0	764	646	521	152	15	3743	2098	19	15
	SS	0	0	22	359	1680	2466	0	1130	178	6	1	0	4527	1315	0	0
Chr 6	LS	0	143	151	265	745	1573	2	1312	976	0	0	0	2879	2288	143	0
	MS	27	120	351	742	1225	1274	0	874	337	148	71	4	3739	1434	147	4
	SS	0	1	37	373	1570	2123	0	923	128	18	0	0	4104	1069	1	0
Chr 7	LS	0	44	42	23	56	164	1	159	270	0	0	0	330	429	44	0
	MS	0	5	27	114	196	196	0	139	68	13	9	3	538	232	5	3
	SS	1	6	22	73	224	228	0	111	59	39	7	1	554	217	7	1

http://etd.uwc.ac.za

A.2 Supplementary Table 2

Table S2: Conserved *Hermetia illucens* microRNA identified in other arthropod species. microRNA provisionally named using hil for species and number taken from microRNA identified in other arthropod species. Identical locations collapsed and those with multiple locations included below the bold. If no other location is listed, all microRNA were found on the same location. *Life stage abbreviations: E=Egg; L=Larva; P=Pupa; F=Female; M=Male

microRNA *H. illucens*	Life Stage*	Length (nt)	Mature Sequence	Chromosomal (CHR) Location			
hil-miR-11927	**F**	**18**	**aaaaagaucgucaucuga**	**5**	100529647	100529731	-
hil-miR-10361b-3p	**L; P**	**18**	**aaaauggucgguacaaug**	**3**	100641948	100642016	-
hil-miR-5-5p	**F; M**	**23**	**aaaggaaccaucaaugugagaug**	**3**	74335108	74335168	-
hil-miR-283-5p	**E;L;P,F,M**	**23**	**aaauaucaguugguaauucuggg**	**5**	78708512	78708576	-
hil-miR-929-5p	**P;F;M**	**21**	**aaauugacucuaguagggagu**	**3**	43096789	43096847	+
hil-miR-100-5p	**E;P;F;M**	**22**	**aacccguagauccgaacuugug**	**4**	72410015	7241007	-
hil-miR-6055-5p	**E**	**23**	**aagacgaagggcgcgccggacgg**	**4**	55779648	55779721	-
hil-miR-281-2-5p	**E;L;P;F;M**	**22**	**aagagagcuauccgucgacagu**	**1**	39007412	39007470	+
hil-miR-3227	**L;M**	**21**	**aagcaauugauugucggcauc**	**6**	99897234	99897278	-
hil-miR-6042-3p	**E;L;P;F;M**	**22**	**aaggaacgaucguugugauauc**	**4**	124627187	124627247	+
hil-miR-2a-3p	**L;F;M**	**19**	**aaucacaggaguauacugu**	**4**	113107875	113107929	+
hil-miR-3805a-5p	**E;F**	**22**	**aaucacaucagcucuuuucuau**	**4**	124626532	124626604	+
hil-miR-279d-5p	**P**	**18**	**aaugaguggggugaaguc**	**4**	5324583	5324640	-
hil-miR-263a-5p	**E;L;P;F;M**	**24**	**aauggcacuggaagaauucacggg**	**1**	138404014	138404086	+
hil-miR-10361b-5p	**F**	**20**	**acacgauauagggacugcca**	**4**	136417895	136417946	+
hil-miR-3859-5p	**P;M**	**18**	**accccacaccugacacca**	**6**	91542431	91542490	-
hil-miR-10-5p	**E;L;P;F;M**	**22**	**acccuguagauccgaauuuguu**	**3**	2075187	2075245	+
hil-miR-iab-4-5p	**E;P;F;M**	**22**	**acguauacugaauguauccuga**	**3**	389420	389478	-
hil-miR-2a-5p	**E**	**23**	**acucucaaagugguugugaaaug**	**6**	28456131	28456191	+
hil-miR-190-5p	**E;L;P;F;M**	**24**	**agauauguuugauauucuugguug**	**2**	147421568	147421629	+
hil-miR-6011-5p	**E;P;F**	**25**	**agaugaugcaacuuugaauggugag**	**3;7**			
			agaugaugcaacuuugaauggugag	3	90462080	90462124	+
			agaugaugcaacuuugaauggugag	3	90462101	90462150	+
			agaugaugcaacuuugaauggugag	7	7194175	7194224	+
hil-miR-11903b	**F**	**25**	**agccauauuaucguucgagcugaca**	**1**	97288015	97288101	+
hil-miR-311-5p	**P**	**18**	**aggacggugucguagacc**	**2**	8609521	8609570	-
hil-miR-iab8-3p	**E**	**23**	**aggauacauucaguauacguaca**	**3**	389421	389481	+
hil-miR-973-5p	**P;M**	**18**	**aggguuggcucugaagau**	**5**	13228815	13228862	+
hil-miR-2775b	**P**	**18**	**aguaggacauugcuaggc**	**1;2**			
			aguaggacauugcuaggc	1	168226734	168226771	-
			aguaggacauugcuaggc	2	54877373	54877411	-
hil-miR-223-3p	**F;M**	**22**	**agucaguuguucuuuucuucaa**	**3**	60062104	60062165	-
hil-miR-9378-5p	**E;F**	**19**	**aguggaguugucacuuuuc**	**2**	46865742	46865784	+
hil-miR-4-3p	**E;L;P;F;M**	**23**	**auaaagcuagauuaccaaagcau**	**1;4**			
			auaaagcuagauuaccaaagcau	1	158777862	158777923	+
			auaaagcuagaucaccaaugca	4	124627018	124627078	+
hil-miR-314-5p	**F**	**18**	**auaacuuggcgauuggac**	**2**	138688743	138688804	-
hil-miR-9893-5p	**E**	**25**	**auaugaucgagucaagcagugccgg**	**3**	109202652	109202717	-

http://etd.uwc.ac.za

microRNA *H. illucens*	Life Stage*	Length (nt)	Mature Sequence	Chromosomal (CHR) Location			
hil-3381-3p	**M**	**25**	**auauucauaugaccgucguccgggu**	**3**	152519090	152519158	+
hil-miR-1000-5p	**E;P;F;M**	**22**	**auauuguccugucacagcagua**	**6**	1819714	1819771	+
hil-miR-181b-3p	**P**	**18**	**aucacugaaagucggccu**	**5**	75841530	75841579	+
hil-miR-3752-3p	**L**	**19**	**auggauaaccgcugaaggc**	**3**	141380225	141380273	-
hil-miR-305-5p	**E;L;P;F;M**	**24**	**auuguacuucaucaggugcucugg**	**1**	190924436	190924495	-
hil-miR-2498b-5p	**P**	**20**	**caaagucugaugugaaaucc**	**2**	185673737	185673826	+
hil-miR-304-5p	**E;L;P;F;M**	**23**	**caaucucaaauuguuauuguggu**	**5**	78702696	78702759	-
hil-miR-3213-5p	**P**	**23**	**cacuacuugaucgcgacacucgg**	**3**	100784348	100784398	+
hil-miR-10373-3p	**F**	**23**	**cagagaaagaugucaggcugguc**	**1**	22883734	22883821	-
hil-miR-3228	**P**	**19**	**cagaucaacuguaguuaag**	**2**	14166820	14166910	-
hil-miR-34-3p	**E**	**23**	**cagccacuauccucccugccgcc**	**3**	44268799	44268861	-
hil-miR-311-5p	**L;P**	**21**	**cagcugggagagcaucuggcu**	**4**	66232631	66232688	+
hil-miR-2780a-5p	**F**	**18**	**cagggugyuguugugggc**	**3**	151344004	151344062	+
hil-miR-2a-3p	**L;P;F;M**	**22**	**caucacagucugaguucuugcu**	**6**	6765948	6766012	-
hil-miR-490-5p	**L**	**24**	**ccauggauucaguacguagcggac**	**5**	39325217	39325308	-
hil-miR-2856-3p	**M**	**20**	**ccauucgaggacagucuggc**	**2**	92165123	92165191	+
hil-miR-988-3p	**P;F;M**	**22**	**ccccuuguugcaaaccccacgu**	**1**	102645320	102645378	-
hil-miR-278-5p	**E**	**23**	**ccggacgagggucucagcgacc**	**4**	137680214	137680273	+
hil-miR-977-3p	**F;M**	**22**	**cgagauauucaaaucguccagu**	**6**	37625503	37625569	-
hil-miR-3338-5p	**M**	**22**	**cgcaaacagggcggauuaggca**	**7**	12253510	12253560	-
hil-miR-275-5p	**F;M**	**23**	**cgcgcuaagcaggcgccagggcu**	**1**	190924617	190924673	-
hil-miR-3375-5p	**P**	**20**	**cgcuaucugggguucagaacg**	**3**	107907703	107907752	-
hil-miR-311-5p	**E**	**22**	**cggacgguuguugguugcagc**	**2**	20179181	20179240	+
hil-miR-14-5p	**E;F**	**18**	**cggagcgaggcuuggcca**	**4**	59552727	59552773	+
hil-miR-6-2-5p	**E**	**23**	**cgggaacucuugcugugacugcu**	**6**	6765948	6766012	-
hil-miR-1-371-5p	**M**	**22**	**cguuugauugauucccugaguc**	**4**	120681598	120681663	+
hil-miR-252-5p	**E;L;P;F;M**	**22**	**cuaaguacuagugccgcaggag**	**5**	52093567	52093622	-
hil-miR-6098	**P;F;M**	**22**	**cuagugaucagucgacaucgau**	**1**	57033124	57033184	+
hil-miR-3862-3p	**P**	**18**	**cucaggcgguuagagcgc**	**5**	32965086	32965130	+
hil-miR-929-3p	**M**	**21**	**cucccuaacggagucagauug**	**3**	43096789	43096847	+
hil-miR-3338	**F**	**23**	**cuguacuugaucgcggcacucgg**	**5**	110263129	110263194	-
hil-miR-961-5p	**E;F**	**25**	**cuugaucaguaggcuuucgguuuuu**	**1**	183158507	183158596	-
hil-miR-263b-5p	**E;P;F;M**	**23**	**cuuggcacugggagaauucacag**	**2**	145800145	145800204	+
hil-miR-2548-3p	**F**	**19**	**cuugguguaaagguagacg**	**2**	172424800	172424867	+
hil-miR-210-3p	**P;F;M**	**21**	**cuugugcgugugacagcggcu**	**5**	70458922	70458979	+
hil-miR-993-3p	**E;P;F;M**	**23**	**gaagcucgucucuacagguaucu**	**3**	29290170	29290254	-
hil-miR-3323	**E**	**25**	**gaauacgauccuacggcggagcagc**	**3**	96780080	96780148	-
hil-miR-3238	**L**	**20**	**gaaugaauaucgugcaggau**	**4**	8579804	8579865	+
hil-miR-11913	**P**	**19**	**gauguuggugauccugugu**	**4**	128395253	128395306	+
hil-miR-2280-3p	**F**	**20**	**gauuuacucugaacccgcga**	**4**	140003993	140004069	+
hil-miR-10357-5p	**P**	**18**	**gcagaaccuuaccuggcc**	**1**	180215809	180215844	+

http://etd.uwc.ac.za

microRNA *H. illucens*	Life Stage*	Length (nt)	Mature Sequence	Chromosomal (CHR) Location			
hil-miR-3884-5p	**L**	**18**	**gcugaagccggucccaag**	**2**	70251684	70251742	-
hil-miR-3777-3p	**P**	**19**	**gggaaacagaaccugaaac**	**4**	43209498	43209567	-
hil-miR-9878-3p	**M**	**19**	**ggggauguagcucaguggu**	**3**	102480797	102480859	+
hil-miR-309b-5p	**F**	**23**	**ggucaaaccccguucaguuggua**	**4**	124627494	124627553	+
hil-miR-12418-3p	**P**	**19**	**gguucgauuccgguacucg**	**2**	100712477	100712526	+
hil-miR-2796	**E;P;F;M**	**23**	**guaggccggcggaaacuacuugc**	**6**	54741226	54741285	+
hil-miR-3212	**L**	**22**	**gucacugugucgucgucgggga**	**6**	90271124	90271178	-
hil-miR-2774a	**P**	**18**	**gucggugguuuuagcggu**	**6**	87657009	87657073	-
hil-miR-2999	**E**	**18**	**gugcgacggcuucuuuuu**	**1**	44402081	44402133	-
hil-miR-988-5p	**E;M**	**22**	**gugugauuuguggcaaugggau**	**1**	102645320	102645378	-
hil-miR-9a-3p	**E;P;F;M**	**22**	**uaaagcuagauuaccaaagcau**	**1**	158777862	158777923	+
hil-miR-277-3p	**E;L;P;F;M**	**23**	**uaaaugcacuaucugguacgaca**	**3**	44278531	44278596	-
hil-miR-965-3p	**E;L;P;F;M**	**22**	**uaagcguauagcuuuuccccuu**	**1**	137738280	137738342	-
hil-miR-124-3p	**E;P;F;M**	**21**	**uaaggcacgcggugaaugcca**	**3**	124681385	124681442	+
hil-miR-2767	**E**	**25**	**uaaguaaacucacggcgcagcgagc**	**1**	221172208	221172261	+
hil-miR-8-3p	**E;L;P;F;M**	**23**	**uaauacugucagguaaagauguc**	**4**	136264207	136264269	-
hil-miR-2788-3p	**P**	**22**	**uaaugcccuuggaaauuccaaa**	**2**	41199313	41199393	+
hil-miR-993-5p	**P;F**	**23**	**uacccuguaguuccgggcuuuug**	**3**	29290170	29290253	-
hil-miR-193-3p	**P;F;M**	**22**	**uacuggccuacuaagucccaac**	**2**	41199119	41199207	+
hil-miR-285-3p	**E;L;P;F;M**	**22**	**uagcaccauucgaaaucagugc**	**2;3;6**			
			uagcaccauucgaaaucagugc	2	2410532	2410598	+
			uagcaccacaugauucggcuu	3	9615069	9615127	+
			uagcaccaugagauucagcuc	6	6758949	6759006	-
hil-miR-282-5p	**L;P;F;M**	**22**	**uagccucuacuaggcuuugucu**	**2**	113258996	113259079	+
hil-miR-980-3p	**E;L;P;F;M**	**21**	**uagcugccuagugaagggcuu**	**4**	9910182	9910247	-
hil-miR-276a-3p	**E;L;P;F;M**	**22**	**uaggaacuucauaccgugcucu**	**2**	96176203	96176262	-
hil-miR-140-5p	**P**	**18**	**uagugguuggaggacacu**	**1**	168301245	168301290	+
hil-miR-2a-3p	**E;L;P;F;M**	**23**	**uaucacagccagcuuugaggagc**	**6**			
			uaucacagccagcuuugaggag	6	28447811	28447874	+
			uaucacagccagcuuugaugagcu	6	28456131	28456191	+
			uaucacagccauuuugacgaguu	6	28447603	28447667	+
hil-miR-3276	**L;P**	**18**	**uaugauuggggugaaguc**	**3**	123593599	123593651	+
hil-miR-2838	**P**	**25**	**uauucagcacgcgagcggacauuga**	**1**	136368508	136368575	+
hil-miR-92a-3p	**E;L;P;F;M**	**22**	**uauugcacauuucccggccugg**	**6**			
			uauugcacauuucccggccugg	6	438169	438233	-
			uauugcacuugucccggccu	6	5670166	5670226	+
hil-miR-137-3p	**F**	**22**	**uauugcuugagaauacacguag**	**1**	19827478	19827539	+
hil-miR-932-5p	**E;P;F;M**	**23**	**ucaauuccguagugcauugcagu**	**3**	144505295	144505354	-
hil-miR-307a-3p	**E;L;P;F;M**	**21**	**ucacaaccuccuugagugagc**	**4**	138390625	138390686	+
hil-miR-2a	**E;L;P;F;M**	**21**	**ucacagccagcuuugaugagc**	**6**			
			ucacagccagcuuugaugagc	6	28447036	28447100	+
			ucacagccagcuuugaugagc	6	28456131	28456190	+
hil-miR-3264	**E;F**	**25**	**ucacaggaugauuggguggcuuggg**	**4**	51937379	51937468	+
hil-miR-3-3p	**E;P;F;M**	**19**	**ucacugggccgaguuuguc**	**3;4**			
			ucacugggcaaaguuuaucuc	3	74334977	74335035	+
			ucacugggaaagguuugucaga	4	124627352	124627412	-
			ucacugggccgaguuuguccu	4	124627494	124627553	+

http://etd.uwc.ac.za

microRNA *H. illucens*	Life Stage*	Length (nt)	Mature Sequence	Chromosomal (CHR) Location			
hil-miR-750	**F**	**25**	**ucagaucuuguaugagacgcggccu**	**5**	9660923	9661002	-
hil-miR-275-3p	**E;L;P;F;M**	**23**	**ucagguaccugaaguagcgcgcg**	**1**			
			ucagguaccugaaguagcgcgcg	1	190924615	190924673	-
			ucagguacuuagugacucuca	1	158769156	158769218	+
hil-miR-14-3p	**E;L;P;F;M**	**22**	**ucagucuuuuucucucuccuau**	**1**	6613491	6613552	-
hil-miR-9385-5p	**L**	**18**	**ucaguguaccgaagaacc**	**1**	65538956	65538995	-
hil-miR-970-3p	**E;L;P;F;M**	**21**	**ucauaagacacacgcggcuau**	**5**	90556233	90556296	+
hil-miR-4986-3p	**F**	**21**	**ucaucggacgcugcuucaccu**	**1**	131419780	131419829	-
hil-miR-125-5p	**P;F;M**	**22**	**ucccugagacccuaacuuguga**	**4**	72383122	72383181	-
hil-miR-278-3p	**L;P;F;M**	**22**	**ucggugggacuuucguccguuu**	**4**	137680214	137680272	+
hil-miR-2a-5p	**E;L**	**23**	**ucucucaaagugguugugaaaug**	**6**	28447811	28447876	+
hil-miR-998-5p	**E;L;P**	**22**	**ucugaacucucaugggccugca**	**6**	6758949	6759006	-
hil-miR-3265	**E**	**23**	**ucugagaggccagcuugggcgag**	**4**	63355567	63355613	-
hil-miR-38390-3p	**E**	**24**	**ucuuguuggaaauguacuuggagc**	**1;4**			
			ucuuguuggaaauguacuuggagc	1	6048991	6049061	+
			ucuuguuggaaauguacuuggagc	4	80857861	80857945	+
hil-miR-9a-5p	**E;L;P;F;M**	**22**	**ucuuugguaucuagcuguaga**	**1;2**			
			ucuuugguauucuagcuguaga	1	158760631	158760693	+
			ucuuuggugauuaagcucugugg	1	158778316	158778379	+
			ucuuugguuaucuagcuguauga	2	61885951	61886011	-
hil-miR-3279	**M**	**22**	**ucuuuucaacaacugaccuuca**	**5**	37098245	37098303	-
hil-miR-957-3p	**E;P;F;M**	**22**	**ugaaaccguccaaaacugaggc**	**2**	277009	277075	+
hil-miR-317-3p	**E;L;P;F**	**25**	**ugaacacagcugguugguaucucagu**	**3**	44334763	44334824	-
hil-miR-279-3p	**E;L;P;F;M**	**22**	**ugacuagauuugcacucauccg**	**3;4**			
			ugacuagauuugcacucauccg	3	162862403	162862464	+
			ugacuagauccacacucauuaa	3	71661045	71661110	+
			ugacuagauaucaugcucgucu	3	71685196	71685263	+
			ugacuagaccgaacacucgucagc	4	124626873	12462693	+
hil-bantam-3p	**E;L;P;F;M**	**23**	**ugagaucauuuugaaagcugauu**	**2**	172657695	172657755	-
hil-miR-958-3p	**L;P;F;M**	**21**	**ugagauucuucuauucuacuu**	**7**	9454370	9454430	-
hil-let-7-5p	**P;F;M**	**21**	**ugagguaguagguuguauagu**	**4**	72384272	72384330	-
hil-miR-12-5p	**E;L;P;F;M**	**23**	**ugaguauuauauuagguacuggu**	**5**	78702132	78702198	-
hil-miR-144-5p	**P;F;M**	**23**	**ugauaucauuuuguaagcugauu**	**1**	80122407	80122459	-
hil-miR-3213-3p	**E**	**23**	**ugauaugaaauuugggagaaagg**	**4**	150288034	150288086	-
hil-miR-4982-3p	**E**	**24**	**ugcuggaacugcaggucgguugac**	**1**	97399443	97399519	-
hil-miR-7-5p	**E;L;P;F;M**	**24**	**uggaagacuagugauuuuguugu**	**6**	461937	461998	+
hil-miR-1-3p	**E;L;P;F;M**	**22**	**uggaauguaaagaaguauggag**	**3**	176783749	176783810	-
hil-miR-184-3p	**E;L;P;F;M**	**22**	**uggacggagaacugauaagggc**	**1**	43471172	43471232	+
hil-miR-219-3p	**F**	**23**	**uggacugugacucaacugacgcc**	**7**	769423	769477	+
hil-miR-31b-5p	**E;L;P;F;M**	**23**	**uggcaagaugucggcauagcuga**	**1**	47357143	47357205	+
hil-miR-34-5p	**E;L;P;F;M**	**24**	**uggcagugugguuagcugguugug**	**3**	44268799	44268861	-
hil-miR-317-5p	**L;F;M**	**22**	**ugggauacacucugugcucgcu**	**3**	44334763	44334824	-
hil-miR-2765	**P;F;M**	**23**	**ugguaacuccaccaccguuggcg**	**6**	8129158	8129216	+
hil-miR-316-5p	**E;L;P;F;M**	**22**	**ugucuuuuuccgcuuacugccg**	**2**	30659317	30659377	-
hil-miR-308-5p	**P**	**20**	**uguggguauccgagcuggacu**	**4**	91412364	91412417	+
hil-miR-263a-3p	**M**	**22**	**uguguucuuuguacuauggacu**	**3**	133966825	133966911	-

UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za

microRNA *H. illucens*	Life Stage*	Length (nt)	Mature Sequence	Chromosomal (CHR) Location			
hil-miR-999-3p	**E;L;P;F;M**	**22**	**uguuaacuguaagacugugucu**	**3**	22030039	22030098	-
hil-miR-3851o-2-5	**E**	**23**	**uguuggcguugauuugauguggu**	**7**	10614134	10614188	+
hil-miR-193-5p	**L**	**18**	**uguugggaaguuggugau**	**4**	73469962	73470002	-
hil-miR-1002-5p	**E;L;F;M**	**23**	**uuaaguaguaaauaaguagguga**	**3**			
			uuaaguaguaaauaaguaggug	3	180202834	180202894	+
			uuaaguaguaaauaaguagguga	3	180203029	180203089	+
hil-miR-iab-8-5p	**M**	**22**	**uuacguauacugaagguauacc**	**3**	389421	389481	+
hil-miR-2807a	**M**	**22**	**uuccagcuaaugaccucugaug**	**3**	164949999	164950080	+
hil-miR-146c-3p	**F**	**24**	**uuccaugguaaacucuugaggaga**	**7**	1516677	1516765	-
hil-miR-2940-3p	**F**	**23**	**uucgacaggaauucugauggacu**	**4**	148574138	148574203	-
hil-miR-3724-3p	**F**	**21**	**uucgugguauuuguuggcugc**	**1**	104888228	104888287	+
hil-miR-981-3p	**E;P;F;M**	**22**	**uucguugucgacgaaaccugca**	**5**	58835647	58835708	-
hil-miR-263b-3p	**P**	**19**	**uugaauuugauacugggug**	**1**	50551838	50551912	-
hil-miR-87-3p	**P;F;M**	**23**	**uugagcaaaauuucaggugugug**	**1**	162813456	162813518	+
hil-miR-33-5p	**E;L;P;F;M**	**21**	**uugcauuguagucgcauugua**	**5**	54162908	54162968	+
hil-miR-2499-3p	**E;L;M**	**22**	**uugcggaucuauggacgucucu**	**1**	161572119	161572180	+
hil-miR-2801	**E**	**22**	**uuggaaucgauugguaacuggc**	**2**	98304661	98304709	+
hil-miR-3861-5p	**P**	**23**	**uuggaauuucucugagacauuuu**	**2**	41199313	41199393	+
hil-miR-928-5p	**F**	**25**	**uuggcuguguugcuggagcuggcgu**	**4**	35515572	35515665	+
hil-miR-2492-5p	**F**	**23**	**uugggaucaggccauuuugcugu**	**4**	76411949	76412008	-
hil-miR-133-3p	**E;L;P;F;M**	**22**	**uugguccccuucaaccagcugu**	**3**	176248355	176248417	-
hil-miR-971-3p	**P;F;M**	**21**	**uuggguguuauaucuuacagug**	**4**	39613879	39613936	+
hil-miR-11894a	**F**	**22**	**uuuaccauuaucguugcggccc**	**6**	26336958	26337006	+
hil-miR-927-5p	**L;P;F;M**	**22**	**uuuagaauuccuacgcuuuacc**	**5**	3001374	3001432	+
hil-miR-956-3p	**E;P**	**20**	**uuucgagaccacuacaaacu**	**2**	55097617	55097700	-
hil-miR-2744	**P**	**18**	**uuugagaguuugauccug**	**2**	3655097	3655150	-
hil-miR-976-5p	**L**	**19**	**uuuggugauuauggcggag**	**3**	75118931	75118996	-
hil-miR-2548-3p	**L**	**19**	**uuugguguuguaggacugc**	**2**	41964183	41964252	+
hil-miR-274-5p	**P;F;M**	**22**	**uuugugaccgacacuaacgggu**	**2**	43581222	43581305	-
hil-miR-3844-3p	**E**	**23**	**uuuguguggguuguggaaugagg**	**2**	42800748	42800832	-
hil-miR-375-3p	**E;L;P;F;M**	**22**	**uuuguucguuuggcuuaagugc**	**3**	112525535	112525597	+
hil-miR-3296-5p	**M**	**23**	**uuuugaguugcagagucgcuggc**	**1**	30541241	30541312	+
hil-miR-315-5p	**E;L;P;F;M**	**23**	**uuuugauuguugcucagaaagcc**	**2**	141335340	141335401	+

http://etd.uwc.ac.za

A.3 Supplementary Table 3

Table S3: Unique *Hermetia illucens* microRNA not identified in any other species. microRNA provisionally named; identical locations collapsed and those with multiple locations included below the bold. If no other location is listed, all microRNA were found on the same location. *Life stage abbreviations: E=Egg; L=Larva; P=Pupa; F=Female; M=Male

microRNA *H. illucens*	Life Stage*	Length (nt)	Mature Sequence	Chromosomal (CHR) Location			
hil-miR-a	**F; M**	**22**	**aacgauuugcucccuucuccag**	**5**	97170454	97170508	-
hil-miR-b	**F; M**	**22**	**aaugagcccgguauucugaaca**	**1**	105897629	105897684	-
hil-miR-f	**E;L;P;F;M**	**22**	**agcaguguuuaugaucaggcuu**	**6**	9852651	9852712	+
hil-miR-g	**P;F**	**22**	**aggauugugacuggacaucgug**	**2**	174068470	174068531	-
hil-miR-h	**F**		**caccuaagcugucgaacaagc**	**1**	221194834	221194889	+
hil-miR-j	**F;P**	**21**	**ccagaacugagcgauuuaaau**	**2;6**			
			ccagaacugagcgauuuaaau	6	78783078	78783136	+
			ccagaacugagcgauuuaaau	2	73070312	73070370	+
hil-miR-l	**F;E**	**23**	**cccuccuugaucgcggcacucgg**	**2;3;4**			
			cccuccuuaaucgcggcacucgg	2	188148995	188149056	+
			cccuccuugaucgcggcgcucgg	4	51883729	51883807	-
			cccuccuugaucgcggcgcucgg	3	178862038	178862116	+
			cccuccuugaucgcugcacucgg	2	58573817	58573878	+
			cccugcuugaucgcggcacucgg	4	50750836	50750897	-
			cccugcuugaucgcggcacucgg	4	50757250	50757311	-
hil-miR-m	**E**	**22**	**cuauacgaucugcuagcuuccc**	4	72384272	72384330	-
hil-miR-n	**E;P;F;M**	**23**	**cucucucugccuuggcauucucc**	**1**	6471787	6471846	+
hil-miR-o	**E;L;P;F;M**	**22**	**gagagugcugucguaccugacc**	**1**	158769156	158769218	+
hil-miR-p	**E**	**22**	**gcagaggugaggcagcgucuga**	**7**	10424208	10424283	-
hil-miR-r	**E**	**23**	**ggcgaacauggaucuagugcacg**	**3**	71685196	71685262	+
hil-miR-s	**E**	**23**	**gguguccacugcaggccuguaug**	**3**	124681384	124681443	+
hil-miR-t	**F;M**	**22**	**uacuguaaugaacuuggcggaa**	**4**	149752585	149752641	-
hil-miR-u	**E**	**21**	**ucacuggauaaguuugucgua**	**4**	124626721	124626783	+
hil-miR-v	**P;F;M**	**22**	**ucagacuuugcgaucgaccuuu**	**3**	141644304	141644363	+
hil-miR-w	**E;L**	**22**	**ucguuaauuuggucgugacgug**	**6**	28447603	28447667	+
hil-miR-x	**E;F**	**23**	**ucuuuuucauuaggauccuucug**	**6**	21724849	21724909	+
hil-miR-y	**F;M**	**22**	**ugagccgguugaaucgcacgga**	**1**	75691513	75691575	+
hil-miR-aa	**P;F**	**24**	**ugcaguguauuccuguguuuuugg**	**4**	113107872	113107929	+
hil-miR-ac	**P**	**23**	**uucuuuucaacaacugaccuuca**	**5**	37098245	37098303	-
hil-miR-ad	**E;F**	**22**	**uuggucgguaguggugaggagc**	**2;7**			
			uuggucgguaguggugaggagc	7	1560975	1561018	+
			uuggucgguaguggugaggagc	2	19009895	19009959	+
hil-miR-ae	**P;F;M**	**22**	**uuguggaacucgaugaugcagc**	**2**	18803235	18803293	+
hil-miR-af	**L;M**	**25**	**uuuucggaauguuguaauaaauaua**	**7**			
			uuuucggaauguuguaauaaauaua	7	738404	738469	-
			uuuucggaauguuguaauaaauaua	7	763802	763867	+

http://etd.uwc.ac.za

A.4 Supplementary Table 4

Table S4: Unique *Hermetia illucens* microRNA log fold change (LFC) and P-Adjusted Value between life stages. microRNA provisionally named; Green filled boxes represent statistically significant up regulated LFC; Red filled boxes represent statistically significant down regulated LFC.

miRNA	Egg v Larva		Larva v Pupa		Pupa v Female		Pupa v Male		Female v Male	
	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-a	0.539	0.795	-0.352	0.843	-0.451	0.791	-2.129	0.232	-1.678	0.359
hil-miR-aa	0.533	0.807	3.470	0.328	-1.996	0.687	3.988	0.155	5.983	0.020
hil-miR-ac	0.513	0.795	3.367	0.010	5.053	4.58E-05	3.869	0.002	-1.184	0.629
hil-miR-ad	-2.474	0.097	-0.348	0.843	-0.571	0.767	0.161	0.959	0.732	0.703
hil-miR-ae	0.514	0.807	2.690	0.594	-3.815	0.169	-1.178	0.959	2.638	0.377
hil-miR-af	10.111	6.88E-06	-9.900	7.71E-05	1.442	0.734	-13.122	1.61E-08	-14.563	9.58E-11
hil-miR-b	0.530	0.795	-0.351	0.843	-2.895	0.115	-3.079	0.064	-0.184	0.868
hil-miR-f	7.893	1.46E-10	2.777	0.010	-0.790	0.687	-1.579	0.172	-0.790	0.629
hil-miR-g	0.523	NA	0.761	0.843	0.942	0.707	1.269	0.746	0.327	0.863
hil-miR-h	0.539	0.795	-0.355	0.843	-2.147	0.250	0.171	0.959	2.318	0.176
hil-miR-j	0.526	NA	0.343	0.843	0.647	0.734	0.854	0.959	0.207	0.868
hil-miR-l	-8.254	4.03E-07	-0.349	0.843	-7.268	4.58E-05	0.166	0.959	7.434	1.88E-05
hil-miR-m	-1.835	NA	-0.349	0.843	1.713	0.668	0.154	0.959	-1.559	0.581
hil-miR-n	-3.090	0.073	1.738	0.763	-0.716	0.734	0.068	0.959	0.784	0.688
hil-miR-o	-5.743	5.12E-05	2.491	0.315	0.006	0.997	-0.579	0.959	-0.584	0.761
hil-miR-p	-6.565	3.43E-05	-0.350	0.843	1.254	0.730	0.167	0.959	-1.087	0.688
hil-miR-r	-8.903	3.49E-09	-0.350	0.843	1.353	0.708	0.168	0.959	-1.186	0.668
hil-miR-s	-2.784	0.018	-0.357	0.843	1.378	0.687	0.148	0.959	-1.230	0.629
hil-miR-t	0.538	0.795	-0.352	0.843	-0.351	0.827	-2.834	0.102	-2.484	0.104
hil-miR-u	-11.735	4.64E-24	-0.348	0.843	1.872	0.596	0.168	0.959	-1.704	0.516
hil-miR-v	0.533	0.795	1.272	0.763	-0.386	0.767	-2.354	0.044	-1.968	0.080
hil-miR-w	-10.012	2.04E-17	-2.989	0.209	1.514	0.687	0.167	0.959	-1.347	0.629
hil-miR-x	-6.230	3.41E-05	-0.350	0.843	-1.028	0.730	0.166	0.959	1.194	0.629
hil-miR-y	0.521	0.795	-0.354	0.843	-3.553	0.007	-4.602	1.70E-04	-1.049	0.359

miRNA	Egg v Pupa		Egg v Female		Egg v Male		Larva v Female		Larva v Male	
	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-a	-0.187	0.926	-0.639	0.707	-2.317	0.084	-0.099	0.979	-1.777	0.368
hil-miR-aa	-4.003	0.089	-5.998	0.005	-0.015	0.994	-5.466	0.020	0.518	0.809
hil-miR-ac	-3.880	0.001	1.173	0.507	-0.011	0.994	1.686	0.310	0.502	0.803
hil-miR-ad	2.822	0.063	2.251	0.068	2.983	0.036	-0.223	0.964	0.508	0.803
hil-miR-ae	-3.204	0.180	-7.019	0.001	-4.381	0.041	-6.505	0.005	-3.867	0.151
hil-miR-af	-0.211	0.926	1.230	0.707	-13.333	1.70E-09	11.342	1.58E-06	-3.222	0.264
hil-miR-b	-0.179	0.926	-3.074	0.026	-3.258	0.018	-2.544	0.119	-2.728	0.114
hil-miR-f	-10.670	9.62E-19	-11.460	1.61E-21	-12.249	2.60E-24	-3.567	1.48E-04	-4.356	0.000
hil-miR-g	-1.284	0.431	-0.342	0.816	-0.015	0.994	0.181	0.964	0.508	0.803
hil-miR-h	-0.184	0.926	-2.331	0.069	-0.013	0.994	-1.792	0.239	0.526	0.803
hil-miR-j	-0.869	0.668	-0.222	0.858	-0.015	0.994	0.304	0.964	0.511	0.803
hil-miR-l	8.603	1.30E-07	1.335	0.432	8.768	5.85E-08	-6.919	9.36E-05	0.514	0.803
hil-miR-m	2.184	0.117	3.897	0.002	2.338	0.068	2.062	0.247	0.503	0.803
hil-miR-n	1.352	0.431	0.637	0.707	1.421	0.346	-2.453	0.233	-1.669	0.596
hil-miR-o	3.253	0.039	3.258	0.030	2.674	0.068	-2.485	0.177	-3.069	0.114
hil-miR-p	6.915	1.51E-05	8.169	2.98E-07	7.082	7.70E-06	1.604	0.468	0.517	0.803
hil-miR-r	9.253	8.35E-10	10.606	7.67E-13	9.420	3.94E-10	1.703	0.427	0.517	0.803
hil-miR-s	3.141	0.008	4.519	1.57E-04	3.289	0.004	1.735	0.310	0.505	0.803
hil-miR-t	-0.187	0.926	-0.537	0.749	-3.021	0.032	0.001	0.999	-2.482	0.151
hil-miR-u	12.083	1.94E-25	13.955	1.67E-35	12.251	3.89E-26	2.220	0.233	0.516	0.803
hil-miR-v	-1.806	0.208	-2.192	0.080	-4.160	4.42E-04	-1.658	0.247	-3.627	0.009
hil-miR-w	13.000	3.90E-20	14.514	1.16E-25	13.167	7.53E-21	4.503	0.006	3.156	0.114
hil-miR-x	6.579	1.43E-05	5.552	2.19E-05	6.745	6.92E-06	-0.678	0.817	0.516	0.803
hil-miR-y	-0.167	0.926	-3.719	0.001	-4.769	1.80E-05	-3.198	0.010	-4.247	0.001

http://etd.uwc.ac.za

A.5 Supplementary Table 5

Table S5: Conserved *Hermetia illucens* microRNA log fold change (LFC) and DESeq2 P-Adjusted Value between life stages. microRNA provisionally named; Green filled boxes represent statistically significant up regulated LFC; Red filled boxes represent statistically significant down regulated LFC.

	Egg v Female		Egg v Male		Egg v Pupa		Larva v Female	
microRNA	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-11894a	1.284	NA	-0.038	NA	-0.015	NA	1.186	0.831
hil-miR-11903b	1.668	NA	-0.039	NA	-0.016	NA	1.571	0.620
hil-miR-11913	-0.059	NA	-0.039	NA	1.428	NA	-0.155	0.930
hil-miR-11927	1.818	NA	-0.039	NA	-0.016	NA	1.721	0.555
hil-miR-263a-3p	-0.058	NA	1.306	NA	-0.016	NA	-0.155	0.930
hil-miR-278-5p	-2.223	NA	-2.203	NA	-2.180	NA	-0.155	0.930
hil-miR-2940-3p	1.668	NA	-0.039	NA	-0.016	NA	1.571	0.620
hil-miR-2a-5p	-12.213	2.05E-21	-12.192	2.27E-21	-12.171	3.81E-21	-1.735	0.445
hil-miR-309b-5p	2.020	NA	-0.037	NA	-0.015	0.995	1.922	0.470
hil-miR-10357-5p	-0.059	0.977	-0.039	0.983	2.886	0.121	-0.155	0.930
hil-miR-10361b-3p	-0.058	0.977	-0.039	0.983	1.761	0.420	-1.783	0.497
hil-miR-10361b-5p	1.949	NA	-0.039	NA	-0.016	0.995	1.852	0.497
hil-miR-10371-5p	-0.058	NA	1.028	NA	-0.016	NA	-0.155	NA
hil-miR-10373-3p	2.230	0.230	-0.037	0.983	-0.015	0.995	2.133	0.399
hil-miR-12418-3p	-0.057	0.977	-0.036	0.983	6.269	7.01E-05	-0.155	0.930
hil-miR-998-5p	-12.360	8.57E-15	-12.342	8.41E-15	-4.573	0.001	-2.482	0.312
hil-miR-2788-3p	-0.059	NA	-0.039	NA	1.428	NA	-0.155	0.930
hil-miR-3724-3p	1.147	NA	-0.039	NA	-0.016	NA	1.050	NA
hil-miR-3752-3p	-0.058	0.977	-0.037	0.983	-0.015	0.995	-2.444	0.303
hil-miR-3777-3p	-0.058	NA	-0.038	NA	1.718	NA	-0.155	0.930
hil-miR-6042-3p	-6.645	3.22E-11	-15.068	1.47E-40	-14.911	1.59E-40	4.900	3.28E-06
hil-miR-6055-5p	-4.273	0.007	-4.254	0.007	-4.232	0.010	-0.155	0.930
hil-miR-928-5p	1.397	NA	-0.038	NA	-0.015	NA	1.300	0.763
hil-miR-9878-3p	-0.057	0.977	6.916	9.32E-06	-0.014	0.995	-0.155	0.930
hil-miR-9893-5p	-2.006	NA	-1.987	NA	-1.964	0.376	-0.155	0.930
hil-miR-263b-3p	-0.058	NA	-0.039	NA	1.811	NA	-0.155	0.930
hil-miR-2744	-0.059	0.977	-0.039	0.983	2.348	0.243	-0.155	0.930
hil-miR-2774a	-0.058	0.977	-0.040	0.983	7.056	4.46E-07	-0.155	0.930
hil-miR-2775b	-0.057	0.977	-0.039	0.983	7.060	2.45E-08	-0.155	0.930
hil-miR-2780a-5p	1.491	NA	-0.039	NA	-0.016	NA	1.393	0.702
hil-miR-279d-5p	-0.059	0.977	-0.039	0.983	3.869	0.020	-0.155	0.930
hil-miR-2801	-6.941	3.30E-06	-6.921	3.28E-06	-6.898	4.91E-06	-0.155	0.930
hil-miR-2807a	-0.058	NA	1.511	NA	-0.015	NA	-0.155	0.930
hil-miR-2838	-0.058	NA	-0.037	NA	1.936	0.385	-0.155	0.930
hil-miR-2856-3p	-0.058	NA	1.688	NA	-0.015	NA	-0.155	0.930
hil-miR-2999	-6.066	3.66E-05	-6.047	3.80E-05	-6.025	5.31E-05	-0.155	0.930
hil-miR-308-5p	-0.057	0.977	-0.040	0.983	5.409	1.45E-04	-0.155	0.930
hil-miR-3212	-0.058	NA	-0.039	NA	-0.016	NA	-1.213	NA
hil-miR-3213-3p	-1.800	NA	-1.780	NA	-1.757	NA	-0.155	0.930
hil-miR-3213-5p	-0.059	0.977	-0.039	0.983	2.471	0.205	-0.155	0.930
hil-miR-3227	-0.058	0.977	2.523	0.172	-0.015	0.995	-4.445	0.012
hil-miR-3228	-0.059	NA	-0.039	NA	1.707	NA	-0.155	0.930
hil-miR-3238	-0.058	0.977	-0.037	0.983	-0.015	0.995	-4.505	0.003
hil-miR-3264	-0.866	0.779	-3.057	0.097	-3.034	0.122	2.056	0.452
hil-miR-3265	-1.395	NA	-1.376	NA	-1.353	NA	-0.155	0.930
hil-miR-3276	-0.058	0.977	-0.038	0.983	6.422	2.31E-05	-6.572	1.55E-05
hil-miR-3279	-0.058	0.977	2.239	0.232	-0.015	0.995	-0.155	0.930
hil-miR-3296-5p	-0.058	0.977	3.307	0.050	-0.015	0.995	-0.155	0.930
hil-miR-3323	-1.509	NA	-1.489	NA	-1.467	NA	-0.155	0.930
hil-miR-3338-5p	-0.058	NA	0.851	NA	-0.016	NA	-0.155	NA
hil-miR-3375-3p	-0.058	NA	-0.038	NA	1.438	NA	-0.155	0.930
hil-miR-3381-3p	-0.058	0.977	3.066	0.074	-0.015	0.995	-0.155	0.930
hil-miR-140-5p	-0.058	0.977	-0.039	0.983	4.419	0.004	-0.155	0.930

http://etd.uwc.ac.za

microRNA	Egg v Female LFC	Egg v Female P-Adjusted Value	Egg v Male LFC	Egg v Male P-Adjusted Value	Egg v Pupa LFC	Egg v Pupa P-Adjusted Value	Larva v Female LFC	Larva v Female P-Adjusted Value
hil-miR-144-5p	11.997	2.43E-13	11.672	9.14E-13	6.226	2.73E-04	11.901	7.64E-13
hil-miR-146c-3p	1.284	NA	-0.038	NA	-0.015	NA	1.186	0.831
hil-miR-181b-3p	-0.058	0.977	-0.039	0.983	7.803	3.73E-07	-0.155	0.930
hil-miR-223-3p	0.889	NA	1.082	NA	-0.015	NA	0.792	0.930
hil-miR-490-5p	-0.058	0.977	-0.037	NA	-0.015	0.995	-2.290	0.351
hil-bantam-3p	5.532	1.98E-12	5.373	7.47E-12	2.063	0.017	4.686	4.98E-09
hil-let-7-5p	10.850	4.09E-18	10.753	5.56E-18	8.807	3.78E-12	10.753	1.93E-17
hil-miR-1000-5p	4.809	3.02E-07	4.794	3.08E-07	2.565	0.011	8.307	5.81E-11
hil-miR-1002-5p	-5.672	7.94E-11	-6.687	1.64E-14	-13.032	1.92E-26	4.270	9.32E-06
hil-miR-100-5p	7.973	3.45E-22	7.622	1.41E-20	6.524	3.45E-15	13.019	7.88E-28
hil-miR-10-5p	0.315	0.971	-1.925	0.021	-4.587	7.14E-09	3.509	1.27E-05
hil-miR-124-3p	1.706	0.273	2.229	0.145	0.901	0.806	6.315	1.59E-04
hil-miR-125-5p	9.790	1.73E-17	9.566	6.99E-17	8.720	7.55E-14	9.694	8.13E-17
hil-miR-12-5p	3.974	6.47E-05	2.771	0.007	-0.827	0.634	5.942	1.95E-09
hil-miR-133-3p	8.875	2.59E-13	9.085	5.65E-14	5.523	1.11E-05	6.900	1.10E-08
hil-miR-137-3p	2.793	0.088	-0.038	0.983	-0.015	0.995	2.696	0.156
hil-miR-1-3p	5.300	4.24E-09	4.930	4.98E-08	1.541	0.157	6.589	5.42E-13
hil-miR-14-3p	5.525	1.48E-09	5.044	3.69E-08	2.202	0.030	5.776	4.37E-10
hil-miR-14-5p	0.021	0.986	-4.075	0.008	-4.054	0.011	3.961	0.015
hil-miR-184-3p	-0.199	0.977	-0.100	0.983	-1.630	0.099	4.560	1.15E-07
hil-miR-190-5p	-1.500	0.087	-1.791	0.035	-3.604	1.00E-05	3.376	8.19E-05
hil-miR-193-3p	7.732	6.68E-06	6.133	4.15E-04	6.602	1.68E-04	7.634	1.27E-05
hil-miR-210-3p	11.585	5.37E-21	11.442	1.42E-20	7.189	1.41E-08	11.488	3.02E-20
hil-miR-2280-3p	1.583	NA	-0.039	NA	-0.016	NA	1.486	0.656
hil-miR-2492-5p	1.276	NA	-0.039	NA	-0.016	NA	1.178	NA
hil-miR-2499-3p	-1.536	0.443	1.155	0.587	-1.499	0.545	-1.440	0.654
hil-miR-252-5p	8.664	5.17E-13	8.250	5.62E-12	6.854	2.38E-08	5.416	3.34E-06
hil-miR-263a-5p	3.250	1.88E-04	2.715	0.002	2.189	0.018	5.546	1.13E-10
hil-miR-263b-5p	4.362	2.50E-05	3.709	3.77E-04	2.536	0.025	6.131	2.27E-06
hil-miR-274-5p	10.383	8.72E-18	10.214	1.91E-17	6.620	9.30E-08	10.286	3.87E-17
hil-miR-275-3p	0.969	0.843	-0.117	0.983	2.088	0.416	1.789	0.616
hil-miR-275-5p	7.711	1.14E-04	9.863	4.90E-07	-0.013	0.995	7.612	2.08E-04
hil-miR-276a-3p	6.500	1.05E-10	6.449	1.42E-10	4.491	1.63E-05	6.454	2.77E-10
hil-miR-277-3p	8.367	1.73E-17	8.402	9.55E-18	0.948	0.552	9.505	3.36E-21
hil-miR-278-3p	8.415	3.43E-11	8.534	1.59E-11	7.893	1.08E-09	2.478	0.018
hil-miR-279-3p	0.283	0.977	-0.547	0.815	-1.950	0.049	5.561	7.19E-10
hil-miR-281-2-5p	11.757	1.58E-28	11.819	4.88E-29	6.958	1.49E-10	4.312	1.00E-04
hil-miR-282-5p	6.600	1.83E-05	6.256	4.95E-05	7.425	1.50E-06	1.405	0.497
hil-miR-283-5p	2.080	0.013	1.299	0.153	-0.938	0.390	2.481	0.004
hil-miR-285-3p	-4.037	1.88E-04	-4.244	7.93E-05	-6.107	1.41E-08	2.992	0.011
hil-miR-2a-3p	-0.046	0.977	-0.346	0.983	-1.700	0.171	5.466	2.00E-07
hil-miR-304-5p	3.100	1.94E-04	1.678	0.061	-1.612	0.099	6.363	4.33E-13
hil-miR-305-5p	4.104	4.57E-05	3.225	0.002	3.074	0.004	2.620	0.017
hil-miR-307a-3p	4.979	5.15E-05	4.907	6.28E-05	4.444	3.83E-04	7.098	1.19E-07
hil-miR-311-5p	-3.382	0.066	-3.365	0.067	-0.708	0.995	-0.155	0.930
hil-miR-314-5p	1.668	NA	-0.039	NA	-0.016	NA	1.571	0.620
hil-miR-315-5p	1.252	0.364	0.628	0.887	-0.288	0.995	5.435	3.83E-06
hil-miR-316-5p	8.351	9.89E-16	7.932	2.10E-14	7.616	5.11E-13	2.984	0.003
hil-miR-317-3p	0.431	0.977	-9.952	5.61E-08	-0.453	0.995	2.058	0.436
hil-miR-317-5p	6.072	0.001	11.392	1.48E-10	-0.032	0.995	3.530	0.071
hil-miR-31b-5p	7.225	4.89E-17	7.525	2.23E-18	4.853	3.71E-08	4.452	5.59E-07
hil-miR-33-5p	2.155	0.020	1.405	0.159	-0.815	0.561	4.403	7.80E-07
hil-miR-3-3p	-4.610	6.55E-05	-2.392	0.054	-11.385	6.52E-18	7.335	5.59E-07
hil-miR-34-3p	-2.557	0.163	-2.537	0.160	-2.513	0.195	-0.155	0.930
hil-miR-34-5p	8.151	5.90E-08	6.775	8.30E-06	0.114	0.995	4.392	0.006
hil-miR-375-3p	7.067	2.05E-21	7.087	1.49E-21	-0.010	0.995	1.957	0.018
hil-miR-4-3p	-1.982	0.230	-5.303	3.13E-04	-5.169	0.001	5.695	3.09E-04
hil-miR-4982-3p	-2.555	0.163	-2.536	0.160	-2.514	0.195	-0.155	0.930
hil-miR-4986-3p	1.885	NA	-0.039	NA	-0.016	NA	1.788	0.518
hil-miR-5-5p	5.189	6.19E-05	6.476	3.29E-07	-0.016	0.995	5.093	1.28E-04
hil-miR-6-2-5p	-14.288	1.89E-32	-14.267	1.09E-32	-14.244	1.55E-32	-0.155	0.930

http://etd.uwc.ac.za

	Egg v Female		Egg v Male		Egg v Pupa		Larva v Female	
microRNA	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-7-5p	-1.301	0.182	-2.900	0.001	-4.779	3.41E-08	8.363	2.44E-16
hil-miR-8-3p	5.139	5.03E-06	4.271	1.66E-04	2.507	0.045	4.311	2.15E-04
hil-miR-87-3p	4.859	1.88E-04	4.540	4.99E-04	1.941	0.249	4.762	3.77E-04
hil-miR-927-5p	10.013	4.47E-17	10.110	1.73E-17	6.380	1.90E-07	8.569	8.77E-18
hil-miR-929-3p	-0.059	0.977	2.387	0.192	-0.015	0.995	-0.155	0.930
hil-miR-929-5p	4.962	0.003	5.001	0.003	2.256	0.293	4.864	0.006
hil-miR-92a-3p	-3.290	0.002	-4.976	2.07E-06	-4.670	0.000	6.275	4.86E-09
hil-miR-932-5p	3.346	0.006	3.920	0.001	3.287	0.010	6.889	3.43E-06
hil-miR-9378-5p	-4.130	0.006	-6.329	1.39E-04	-6.307	1.78E-04	2.066	0.452
hil-miR-9385-5p	-0.058	NA	-0.039	NA	-0.016	NA	-1.705	0.568
hil-miR-956-3p	-2.566	0.163	-2.546	0.160	5.103	2.32E-04	-0.155	0.930
hil-miR-957-3p	8.177	0.000	8.636	4.14E-18	4.234	5.59E-05	9.231	2.02E-14
hil-miR-958-3p	10.310	5.24E-13	9.415	4.32E-11	5.638	1.68E-04	0.626	0.930
hil-miR-961-5p	-3.380	0.006	-5.406	1.30E-04	-5.384	1.68E-04	1.887	0.412
hil-miR-965-3p	0.941	0.664	0.352	0.983	-0.981	0.731	6.633	5.75E-06
hil-miR-970-3p	4.801	2.75E-06	4.412	1.71E-05	1.069	0.471	6.993	1.73E-11
hil-miR-971-3p	4.046	0.005	4.451	0.002	1.367	0.563	3.948	0.009
hil-miR-976-5p	-0.058	NA	-0.039	NA	-0.015	NA	-1.502	0.656
hil-miR-977-3p	1.607	NA	1.288	NA	-0.016	NA	1.510	0.656
hil-miR-980-3p	5.567	1.34E-06	5.233	5.64E-06	3.216	0.009	6.024	3.13E-07
hil-miR-981-3p	5.449	7.54E-06	5.023	4.09E-05	4.795	1.18E-04	7.582	2.01E-07
hil-miR-988-3p	6.979	1.64E-06	5.313	3.30E-04	5.699	1.41E-04	6.881	3.43E-06
hil-miR-988-5p	-7.871	1.16E-08	-4.983	9.82E-06	-7.827	2.38E-08	-0.155	0.930
hil-miR-993-3p	-4.587	6.19E-05	-4.097	3.49E-04	-7.472	1.09E-09	4.495	0.003
hil-miR-993-5p	2.367	0.211	-0.038	0.983	2.950	0.123	2.270	0.363
hil-miR-999-3p	7.117	2.34E-12	7.095	2.53E-12	4.488	2.00E-05	6.306	9.53E-10
hil-miR-9a-3p	4.049	0.021	3.013	0.100	2.679	0.184	6.529	0.001
hil-miR-9a-5p	-1.669	0.092	-2.243	0.017	-2.907	0.002	2.984	0.002
hil-miR-iab-4-5p	1.442	0.273	-1.543	0.253	-1.981	0.151	7.592	2.00E-07
hil-miR-iab-8-5p	-0.058	0.977	7.158	7.78E-09	-0.017	0.995	-0.155	0.930
hil-miR-193-5p	-0.058	NA	-0.038	NA	-0.015	NA	-1.807	0.518
hil-miR-2548-3p	1.759	0.337	-0.038	0.983	-0.015	0.995	-1.963	0.303
hil-miR-311-5p	-0.058	0.977	-0.038	0.983	6.424	6.24E-05	-8.258	1.86E-07
hil-miR-iab-8-3p	-1.790	NA	-1.770	NA	-1.746	NA	-0.155	0.930
hil-miR-973-5p	-0.060	0.977	6.256	4.31E-05	1.381	0.622	-0.155	0.930
hil-miR-2498b-5p	-0.058	0.977	-0.037	0.983	3.706	0.018	-0.155	0.930
hil-miR-2765	4.998	5.47E-06	2.117	0.079	5.772	1.61E-07	9.897	1.29E-12
hil-miR-2767	-1.611	NA	-1.591	NA	-1.569	NA	-0.155	0.930
hil-miR-2796	3.390	0.003	3.890	0.001	1.268	0.416	6.429	5.92E-06
hil-miR-2a	-2.264	0.030	-2.309	0.026	-1.338	0.279	1.781	0.159
hil-miR-3338	2.484	0.175	-0.037	0.983	-0.015	0.995	2.387	0.312
hil-miR-6098	6.659	6.68E-06	5.988	5.54E-05	4.151	0.009	6.562	1.27E-05
hil-miR-750	2.762	0.102	-0.038	0.983	-0.015	0.995	2.665	0.184
hil-miR-219-3p	2.083	0.272	-0.039	NA	-0.015	0.995	1.986	0.453
hil-miR-3805a-5p	-7.278	1.90E-11	-12.416	1.88E-19	-12.393	3.94E-19	5.001	0.001
hil-miR-3839-3p	-7.311	1.15E-07	-7.291	1.15E-07	-7.265	1.87E-07	-0.155	0.930
hil-miR-3844-3p	-1.867	NA	-1.847	NA	-1.824	NA	-0.155	0.930
hil-miR-3851o-2-5p	-1.612	NA	-1.592	NA	-1.568	NA	-0.155	0.930
hil-miR-3859-5p	-0.058	0.977	4.348	0.004	3.996	0.011	-0.155	0.930
hil-miR-3861-5p	-0.058	NA	-0.038	NA	1.324	NA	-0.155	0.930
hil-miR-3862-3p	-0.057	0.977	-0.039	0.983	5.925	6.81E-06	-0.155	0.930
hil-miR-3884-5p	-0.059	0.977	-0.039	0.983	-0.015	0.995	-3.662	0.037
hil-miR-6011-5p	0.236	0.977	-6.291	5.82E-05	-4.125	0.004	6.391	6.74E-05

http://etd.uwc.ac.za

microRNA	Egg v Larva		Larva v Pupa		Pupa v Female		Pupa v Male	
	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-11894a	-0.098	NA	0.112	NA	1.298	NA	-0.023	NA
hil-miR-11903b	-0.097	NA	0.114	NA	1.684	NA	-0.023	NA
hil-miR-11913	-0.097	NA	-1.331	NA	-1.487	NA	-1.467	NA
hil-miR-11927	-0.097	NA	0.114	NA	1.834	NA	-0.023	NA
hil-miR-263a-3p	-0.097	NA	0.113	NA	-0.042	NA	1.322	NA
hil-miR-278-5p	2.068	NA	0.113	NA	-0.043	NA	-0.023	NA
hil-miR-2940-3p	-0.097	NA	0.114	NA	1.684	NA	-0.023	NA
hil-miR-2a-5p	10.479	1.82E-22	1.692	0.488	-0.043	0.981	-0.022	0.990
hil-miR-309b-5p	-0.098	0.962	0.112	0.957	2.035	NA	-0.023	NA
hil-miR-10357-5p	-0.097	0.962	-2.789	0.189	-2.945	0.108	-2.925	0.129
hil-miR-10361b-3p	-1.725	0.518	-0.036	0.979	-1.818	0.396	-1.800	0.474
hil-miR-10361b-5p	-0.097	0.962	0.114	NA	1.966	NA	-0.023	NA
hil-miR-10371-5p	-0.097	NA	0.113	NA	-0.042	NA	1.043	NA
hil-miR-10373-3p	-0.098	0.962	0.112	0.957	2.245	NA	-0.023	NA
hil-miR-12418-3p	-0.098	0.962	-6.171	1.80E-04	-6.326	7.75E-05	-6.305	7.80E-05
hil-miR-998-5p	9.878	2.88E-12	-5.306	0.001	-7.787	2.98E-06	-7.769	2.98E-06
hil-miR-2788-3p	-0.097	NA	-1.331	NA	-1.487	NA	-1.467	NA
hil-miR-3724-3p	-0.097	NA	0.113	NA	1.163	NA	-0.023	NA
hil-miR-3752-3p	-2.387	0.278	2.401	0.310	-0.043	0.981	-0.023	NA
hil-miR-3777-3p	-0.098	NA	-1.620	NA	-1.776	NA	-1.756	NA
hil-miR-6042-3p	11.544	1.93E-30	3.367	0.010	8.267	1.23E-12	-0.157	0.990
hil-miR-6055-5p	4.117	0.015	0.114	0.957	-0.041	0.981	-0.022	0.990
hil-miR-928-5p	-0.098	NA	0.112	NA	1.412	NA	-0.023	NA
hil-miR-9878-3p	-0.098	0.962	0.112	0.957	-0.043	0.981	6.930	1.54E-05
hil-miR-9893-5p	1.851	0.492	0.114	NA	-0.041	NA	-0.023	NA
hil-miR-263b-3p	-0.097	NA	-1.714	NA	-1.869	NA	-1.850	NA
hil-miR-2744	-0.097	0.962	-2.251	0.369	-2.407	NA	-2.387	NA
hil-miR-2774a	-0.098	0.962	-6.958	1.64E-06	-7.114	5.81E-07	-7.096	5.93E-07
hil-miR-2775b	-0.098	0.962	-6.962	1.08E-07	-7.118	2.79E-08	-7.099	3.21E-08
hil-miR-2780a-5p	-0.097	NA	0.114	NA	1.507	NA	-0.023	NA
hil-miR-279d-5p	-0.097	0.962	-3.772	0.035	-3.928	0.020	-3.907	0.024
hil-miR-2801	6.786	8.47E-06	0.113	0.957	-0.043	0.981	-0.023	0.990
hil-miR-2807a	-0.098	NA	0.112	NA	-0.043	NA	1.526	NA
hil-miR-2838	-0.098	0.962	-1.839	NA	-1.994	NA	-1.974	NA
hil-miR-2856-3p	-0.098	NA	0.112	NA	-0.043	NA	1.703	NA
hil-miR-2999	5.911	9.03E-05	0.114	0.957	-0.041	0.981	-0.022	0.990
hil-miR-308-5p	-0.098	0.962	-5.311	3.64E-04	-5.466	1.47E-04	-5.449	1.58E-04
hil-miR-3212	-1.155	NA	1.170	NA	-0.042	NA	-0.023	NA
hil-miR-3213-3p	1.645	NA	0.112	NA	-0.043	NA	-0.023	NA
hil-miR-3213-5p	-0.097	0.962	-2.375	0.313	-2.530	0.192	-2.510	NA
hil-miR-3227	-4.387	0.014	4.402	0.016	-0.043	0.981	2.538	0.228
hil-miR-3228	-0.097	NA	-1.610	NA	-1.765	NA	-1.745	NA
hil-miR-3238	-4.448	0.004	4.462	0.005	-0.043	0.981	-0.023	0.990
hil-miR-3264	2.922	0.170	0.112	0.957	2.168	0.315	-0.023	0.990
hil-miR-3265	1.240	NA	0.113	NA	-0.042	NA	-0.023	NA
hil-miR-3276	-6.514	2.10E-05	0.092	0.957	-6.480	2.60E-05	-6.460	2.54E-05
hil-miR-3279	-0.098	0.962	0.112	0.957	-0.043	NA	2.254	NA
hil-miR-3296-5p	-0.098	0.962	0.112	0.957	-0.043	0.981	3.322	0.068
hil-miR-3323	1.353	NA	0.113	NA	-0.042	NA	-0.023	NA
hil-miR-3338-5p	-0.097	NA	0.113	NA	-0.042	NA	0.867	NA
hil-miR-3375-3p	-0.098	NA	-1.341	NA	-1.496	NA	-1.476	NA
hil-miR-3381-3p	-0.098	0.962	0.112	0.957	-0.043	0.981	3.081	0.102
hil-miR-140-5p	-0.097	0.962	-4.322	0.007	-4.477	0.003	-4.458	0.004
hil-miR-144-5p	-0.096	0.962	-6.130	0.001	5.771	8.20E-05	5.446	2.12E-04
hil-miR-146c-3p	-0.098	NA	0.112	NA	1.298	NA	-0.023	NA
hil-miR-181b-3p	-0.097	0.962	-7.705	1.44E-06	-7.861	5.01E-07	-7.841	5.31E-07
hil-miR-223-3p	-0.097	NA	0.113	NA	0.905	NA	1.097	NA
hil-miR-490-5p	-2.233	0.332	2.248	0.369	-0.043	NA	-0.023	NA
hil-bantam-3p	-0.846	0.536	-1.217	0.306	3.469	2.71E-05	3.311	6.33E-05
hil-let-7-5p	-0.097	0.962	-8.710	1.69E-11	2.043	0.049	1.946	0.070
hil-miR-1000-5p	3.499	0.015	-6.064	4.70E-06	2.243	0.026	2.228	0.031

http://etd.uwc.ac.za

	Egg v Larva		Larva v Pupa		Pupa v Female		Pupa v Male	
microRNA	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-1002-5p	9.942	2.95E-27	3.090	0.042	7.360	7.09E-09	6.345	7.12E-07
hil-miR-100-5p	5.046	6.35E-05	-11.570	7.01E-22	1.449	0.137	1.098	0.344
hil-miR-10-5p	3.194	9.24E-05	1.393	0.189	4.902	0.000	2.662	0.002
hil-miR-124-3p	4.609	0.010	-5.510	0.002	0.805	0.868	1.328	0.596
hil-miR-125-5p	-0.096	0.962	-8.624	5.44E-13	1.070	0.268	0.846	0.507
hil-miR-12-5p	1.967	0.099	-1.140	0.528	4.802	2.01E-06	3.598	0.001
hil-miR-133-3p	-1.975	0.237	-3.548	0.008	3.352	0.009	3.562	0.006
hil-miR-137-3p	-0.097	0.962	0.112	0.957	2.809	NA	-0.023	NA
hil-miR-1-3p	1.289	0.316	-2.830	0.005	3.759	7.24E-05	3.389	3.90E-04
hil-miR-14-3p	0.251	0.962	-2.453	0.020	3.323	0.001	2.842	0.005
hil-miR-14-5p	3.940	0.016	0.114	0.957	4.075	0.011	-0.020	0.990
hil-miR-184-3p	4.759	2.72E-08	-3.129	0.001	1.430	0.163	1.530	0.142
hil-miR-190-5p	4.877	4.30E-09	-1.272	0.306	2.104	0.017	1.814	0.053
hil-miR-193-3p	-0.098	0.962	-6.504	3.97E-04	1.129	0.700	-0.470	0.990
hil-miR-210-3p	-0.097	0.962	-7.092	5.85E-08	4.396	1.72E-06	4.253	3.44E-06
hil-miR-2280-3p	-0.097	NA	0.114	NA	1.600	NA	-0.023	NA
hil-miR-2492-5p	-0.097	NA	0.113	NA	1.292	NA	-0.023	NA
hil-miR-2499-3p	0.097	0.962	1.402	0.812	-0.038	0.981	2.654	0.157
hil-miR-252-5p	-3.248	0.017	-3.605	0.004	1.810	0.189	1.397	0.407
hil-miR-263a-5p	2.296	0.016	-4.485	5.67E-07	1.061	0.343	0.526	0.990
hil-miR-263b-5p	1.769	0.359	-4.305	0.002	1.826	0.092	1.173	0.396
hil-miR-274-5p	-0.097	0.962	-6.523	4.10E-07	3.763	2.42E-05	3.594	5.62E-05
hil-miR-275-3p	0.821	0.962	-2.908	0.281	-1.119	0.866	-2.205	0.438
hil-miR-275-5p	-0.099	0.962	0.113	0.959	7.724	1.63E-04	9.876	0.000
hil-miR-276a-3p	-0.046	0.962	-4.446	3.57E-05	2.009	0.090	1.958	0.114
hil-miR-277-3p	1.138	0.496	-2.086	0.099	7.419	2.34E-13	7.454	1.22E-13
hil-miR-278-3p	-5.937	7.63E-06	-1.956	0.097	0.522	0.903	0.641	0.914
hil-miR-279-3p	5.278	5.62E-09	-3.328	0.001	2.233	0.022	1.402	0.217
hil-miR-281-2-5p	-7.444	5.17E-12	0.486	0.957	4.798	1.47E-05	4.861	1.02E-05
hil-miR-282-5p	-5.196	0.001	-2.230	0.189	-0.825	0.807	-1.169	0.640
hil-miR-283-5p	0.402	0.962	0.537	0.957	3.018	3.18E-04	2.237	0.012
hil-miR-285-3p	7.029	5.72E-11	-0.922	0.841	2.070	0.095	1.863	0.159
hil-miR-2a-3p	5.512	1.56E-07	-3.812	0.001	1.654	0.189	1.353	0.356
hil-miR-304-5p	3.263	4.17E-04	-1.651	0.165	4.712	2.79E-08	3.290	1.56E-04
hil-miR-305-5p	-1.484	0.274	-1.589	0.267	1.030	0.478	0.151	0.990
hil-miR-307a-3p	2.119	0.246	-6.563	2.09E-06	0.534	0.981	0.463	0.990
hil-miR-311-5p	3.227	0.116	-2.519	0.310	-2.675	0.192	-2.658	0.219
hil-miR-314-5p	-0.097	NA	0.114	NA	1.684	NA	-0.023	NA
hil-miR-315-5p	4.183	0.001	-3.895	0.002	1.540	0.286	0.916	0.783
hil-miR-316-5p	-5.367	5.95E-07	-2.249	0.042	0.735	0.693	0.317	0.990
hil-miR-317-3p	1.627	0.594	-1.174	0.957	0.884	0.914	-9.499	5.31E-07
hil-miR-317-5p	-2.543	0.332	2.575	0.369	6.105	0.001	11.424	4.52E-10
hil-miR-31b-5p	-2.773	0.003	-2.080	0.043	2.372	0.013	2.672	0.005
hil-miR-33-5p	2.247	0.022	-1.433	0.250	2.970	0.001	2.219	0.025
hil-miR-3-3p	11.946	4.53E-17	-0.561	0.957	6.775	8.09E-07	8.993	3.33E-11
hil-miR-34-3p	2.402	0.274	0.111	0.957	-0.044	0.981	-0.024	NA
hil-miR-34-5p	-3.759	0.027	3.645	0.041	8.037	2.00E-07	6.661	1.86E-05
hil-miR-375-3p	-5.110	1.10E-11	5.120	2.50E-11	7.077	6.51E-21	7.097	4.61E-21
hil-miR-4-3p	7.677	5.24E-07	-2.508	0.244	3.188	0.051	-0.134	0.990
hil-miR-4982-3p	2.400	0.274	0.114	0.957	-0.041	0.981	-0.022	NA
hil-miR-4986-3p	-0.097	NA	0.114	NA	1.902	NA	-0.023	NA
hil-miR-5-5p	-0.097	0.962	0.113	0.957	5.205	8.99E-05	6.492	6.14E-07
hil-miR-6-2-5p	14.133	1.03E-31	0.111	0.957	-0.044	0.981	-0.023	0.990
hil-miR-7-5p	9.664	1.04E-21	-4.885	5.77E-06	3.478	9.98E-05	1.880	0.063
hil-miR-8-3p	-0.827	0.863	-1.680	0.309	2.631	0.035	1.764	0.219
hil-miR-87-3p	-0.097	0.962	-1.844	0.369	2.918	0.009	2.599	0.026
hil-miR-927-5p	-1.444	0.519	-4.936	2.55E-06	3.633	2.81E-05	3.730	1.59E-05
hil-miR-929-3p	-0.097	0.962	0.111	0.957	-0.044	0.981	2.401	NA
hil-miR-929-5p	-0.097	0.962	-2.158	0.428	2.706	0.107	2.745	0.115
hil-miR-92a-3p	9.565	9.47E-20	-4.895	1.21E-05	1.380	0.302	-0.306	0.990
hil-miR-932-5p	3.543	0.034	-6.830	6.91E-06	0.059	0.981	0.632	0.990

http://etd.uwc.ac.za

	Egg v Larva		Larva v Pupa		Pupa v Female		Pupa v Male	
microRNA	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-9378-5p	6.196	3.06E-04	0.110	0.957	2.176	0.315	-0.023	0.990
hil-miR-9385-5p	-1.647	NA	1.663	NA	-0.042	NA	-0.023	NA
hil-miR-956-3p	2.411	0.274	-7.514	4.66E-06	-7.669	1.96E-06	-7.649	1.91E-06
hil-miR-957-3p	1.054	0.815	-5.289	3.57E-05	3.942	4.57E-06	4.402	2.84E-07
hil-miR-958-3p	-9.684	2.18E-11	4.046	0.001	4.672	9.98E-05	3.777	0.002
hil-miR-961-5p	5.267	3.09E-04	0.118	0.957	2.005	0.274	-0.021	0.990
hil-miR-965-3p	5.692	1.33E-04	-4.711	0.003	1.922	0.241	1.332	0.590
hil-miR-970-3p	2.193	0.076	-3.261	0.005	3.732	4.62E-04	3.343	0.002
hil-miR-971-3p	-0.097	0.962	-1.269	0.841	2.679	0.050	3.085	0.023
hil-miR-976-5p	-1.443	NA	1.458	NA	-0.044	NA	-0.024	NA
hil-miR-977-3p	-0.097	NA	0.113	NA	1.623	NA	1.303	NA
hil-miR-980-3p	0.457	0.962	-3.673	0.004	2.351	0.066	2.017	0.142
hil-miR-981-3p	2.133	0.308	-6.927	4.52E-06	0.654	0.888	0.229	0.990
hil-miR-988-3p	-0.097	0.962	-5.602	3.53E-04	1.280	0.427	-0.386	0.990
hil-miR-988-5p	7.716	3.98E-08	0.112	0.957	-0.044	0.981	2.844	0.099
hil-miR-993-3p	9.082	1.80E-10	-1.610	0.607	2.885	0.036	3.376	0.014
hil-miR-993-5p	-0.097	0.962	-2.853	0.195	-0.583	0.981	-2.988	0.135
hil-miR-999-3p	-0.811	0.815	-3.678	0.001	2.629	0.020	2.606	0.025
hil-miR-9a-3p	2.480	0.358	-5.160	0.013	1.369	0.637	0.334	0.990
hil-miR-9a-5p	4.653	2.43E-07	-1.746	0.125	1.237	0.268	0.664	0.843
hil-miR-iab-4-5p	6.150	3.87E-05	-4.168	0.011	3.423	0.006	0.439	0.990
hil-miR-iab-8-5p	-0.097	0.962	0.114	0.957	-0.041	0.981	7.176	2.00E-08
hil-miR-193-5p	-1.749	NA	1.764	NA	-0.043	NA	-0.023	NA
hil-miR-2548-3p	-3.722	0.025	3.737	0.030	1.775	0.396	-0.023	0.990
hil-miR-311-5p	-8.199	2.20E-07	1.775	0.369	-6.482	6.88E-05	-6.463	6.70E-05
hil-miR-iab-8-3p	1.634	NA	0.112	NA	-0.044	NA	-0.024	NA
hil-miR-973-5p	-0.096	0.962	-1.285	0.909	-1.440	0.596	4.875	0.001
hil-miR-2498b-5p	-0.098	0.962	-3.608	0.033	-3.763	0.018	-3.743	0.023
hil-miR-2765	4.899	0.001	-10.671	7.49E-14	-0.774	0.755	-3.655	0.002
hil-miR-2767	1.455	NA	0.114	NA	-0.042	NA	-0.023	NA
hil-miR-2796	3.039	0.069	-4.307	0.005	2.122	0.093	2.622	0.034
hil-miR-2a	4.044	6.81E-05	-2.706	0.015	-0.926	0.550	-0.971	0.595
hil-miR-3338	-0.098	0.962	0.112	0.957	2.499	0.198	-0.023	NA
hil-miR-6098	-0.097	0.962	-4.054	0.016	2.507	0.065	1.837	0.231
hil-miR-750	-0.097	0.962	0.112	0.957	2.777	NA	-0.023	NA
hil-miR-219-3p	-0.097	0.962	0.111	0.957	2.097	NA	-0.024	NA
hil-miR-3805a-5p	12.279	6.27E-19	0.114	0.957	5.115	0.001	-0.023	0.990
hil-miR-3839-3p	7.156	3.32E-07	0.110	0.957	-0.046	0.981	-0.025	0.990
hil-miR-3844-3p	1.712	NA	0.112	NA	-0.044	NA	-0.024	NA
hil-miR-3851o-2-5p	1.457	NA	0.112	NA	-0.044	NA	-0.024	NA
hil-miR-3859-5p	-0.097	0.962	-3.899	0.020	-4.054	0.011	0.352	0.990
hil-miR-3861-5p	-0.098	NA	-1.227	NA	-1.382	NA	-1.362	NA
hil-miR-3862-3p	-0.098	0.962	-5.827	1.71E-05	-5.982	7.30E-06	-5.963	7.43E-06
hil-miR-3884-5p	-3.603	0.040	3.618	0.045	-0.044	0.981	-0.024	0.990
hil-miR-6011-5p	6.155	1.33E-04	-2.030	0.439	4.361	0.002	-2.166	0.338

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

	Larva v Male		Female v Male	
microRNA	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-11894a	-0.135	NA	-1.321	NA
hil-miR-11903b	-0.136	NA	-1.707	NA
hil-miR-11913	-0.135	NA	0.020	NA
hil-miR-11927	-0.136	NA	-1.857	NA
hil-miR-263a-3p	1.209	NA	1.364	NA
hil-miR-278-5p	-0.136	NA	0.020	NA
hil-miR-2940-3p	-0.136	NA	-1.707	NA
hil-miR-2a-5p	-1.714	0.411	0.021	0.992
hil-miR-309b-5p	-0.135	NA	-2.057	NA
hil-miR-10357-5p	-0.135	0.939	0.020	0.992
hil-miR-10361b-3p	-1.764	0.496	0.019	0.992
hil-miR-10361b-5p	-0.136	NA	-1.988	NA
hil-miR-10371-5p	0.930	NA	1.086	NA
hil-miR-10373-3p	-0.135	NA	-2.268	NA
hil-miR-12418-3p	-0.134	0.939	0.021	0.992
hil-miR-998-5p	-2.463	0.252	0.018	0.992
hil-miR-2788-3p	-0.135	NA	0.020	NA
hil-miR-3724-3p	-0.136	NA	-1.186	NA
hil-miR-3752-3p	-2.424	NA	0.020	NA
hil-miR-3777-3p	-0.135	NA	0.020	NA
hil-miR-6042-3p	-3.524	0.004	-8.424	7.34E-13
hil-miR-6055-5p	-0.137	0.939	0.019	0.992
hil-miR-928-5p	-0.135	NA	-1.435	NA
hil-miR-9878-3p	6.818	1.45E-05	6.973	2.19E-05
hil-miR-9893-5p	-0.136	NA	0.019	NA
hil-miR-263b-3p	-0.136	NA	0.019	NA
hil-miR-2744	-0.135	NA	0.020	NA
hil-miR-2774a	-0.137	0.939	0.018	0.992
hil-miR-2775b	-0.137	0.939	0.018	0.992
hil-miR-2780a-5p	-0.136	NA	-1.530	NA
hil-miR-279d-5p	-0.135	0.939	0.020	0.992
hil-miR-2801	-0.135	0.939	0.020	0.992
hil-miR-2807a	1.414	NA	1.569	NA
hil-miR-2838	-0.135	NA	0.020	NA
hil-miR-2856-3p	1.591	NA	1.746	NA
hil-miR-2999	-0.137	0.939	0.019	0.992
hil-miR-308-5p	-0.137	0.939	0.018	0.992
hil-miR-3212	-1.193	NA	0.019	NA
hil-miR-3213-3p	-0.135	NA	0.020	NA
hil-miR-3213-5p	-0.135	NA	0.020	NA
hil-miR-3227	-1.864	0.332	2.581	0.295
hil-miR-3228	-0.135	NA	0.020	NA
hil-miR-3238	-4.485	0.003	0.020	0.992
hil-miR-3264	-0.135	0.939	-2.191	0.477
hil-miR-3265	-0.136	NA	0.019	NA
hil-miR-3276	-6.552	1.34E-05	0.020	0.992
hil-miR-3279	2.142	NA	2.297	NA
hil-miR-3296-5p	3.210	0.070	3.365	0.093
hil-miR-3323	-0.136	NA	0.019	NA
hil-miR-3338-5p	0.754	NA	0.909	NA
hil-miR-3375-3p	-0.135	NA	0.020	NA
hil-miR-3381-3p	2.968	0.104	3.124	0.136
hil-miR-140-5p	-0.137	0.939	0.019	0.992
hil-miR-144-5p	11.576	2.99E-12	-0.325	0.992
hil-miR-146c-3p	-0.135	NA	-1.321	NA
hil-miR-181b-3p	-0.136	0.939	0.019	0.992
hil-miR-223-3p	0.984	NA	0.192	NA
hil-miR-490-5p	-2.270	NA	0.020	NA
hil-bantam-3p	4.527	1.43E-08	-0.159	0.992
hil-let-7-5p	10.656	2.76E-17	-0.096	0.992
hil-miR-1000-5p	8.292	4.40E-11	-0.015	0.992
hil-miR-1002-5p	3.254	0.001	-1.016	0.613
hil-miR-100-5p	12.669	1.83E-26	-0.351	0.992

http://etd.uwc.ac.za

	Larva v Male		Female v Male	
microRNA	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-10-5p	1.269	0.194	-2.240	0.014
hil-miR-124-3p	6.838	2.64E-05	0.523	0.992
hil-miR-125-5p	9.470	3.35E-16	-0.224	0.992
hil-miR-12-5p	4.739	1.76E-06	-1.203	0.539
hil-miR-133-3p	7.110	3.18E-09	0.210	0.992
hil-miR-137-3p	-0.135	NA	-2.832	NA
hil-miR-1-3p	6.219	8.35E-12	-0.370	0.992
hil-miR-14-3p	5.295	1.01E-08	-0.481	0.992
hil-miR-14-5p	-0.135	0.939	-4.095	0.017
hil-miR-184-3p	4.659	4.05E-08	0.099	0.992
hil-miR-190-5p	3.086	2.82E-04	-0.290	0.992
hil-miR-193-3p	6.034	0.001	-1.599	0.659
hil-miR-210-3p	11.345	6.78E-20	-0.143	0.992
hil-miR-2280-3p	-0.136	NA	-1.622	NA
hil-miR-2492-5p	-0.136	NA	-1.314	NA
hil-miR-2499-3p	1.252	0.686	2.691	0.214
hil-miR-252-5p	5.002	1.45E-05	-0.414	0.992
hil-miR-263a-5p	5.011	5.08E-09	-0.535	0.992
hil-miR-263b-5p	5.478	2.15E-05	-0.652	0.992
hil-miR-274-5p	10.117	9.86E-17	-0.169	0.992
hil-miR-275-3p	0.704	0.939	-1.085	0.992
hil-miR-275-5p	9.764	7.37E-07	2.152	0.539
hil-miR-276a-3p	6.404	3.12E-10	-0.051	0.992
hil-miR-277-3p	9.540	1.60E-21	0.035	0.992
hil-miR-278-3p	2.597	0.010	0.119	0.992
hil-miR-279-3p	4.731	1.55E-07	-0.830	0.827
hil-miR-281-2-5p	4.375	5.52E-05	0.062	0.992
hil-miR-282-5p	1.060	0.757	-0.344	0.992
hil-miR-283-5p	1.701	0.061	-0.781	0.793
hil-miR-285-3p	2.785	0.016	-0.207	0.992
hil-miR-2a-3p	5.166	7.00E-07	-0.300	0.992
hil-miR-304-5p	4.942	1.78E-08	-1.421	0.226
hil-miR-305-5p	1.741	0.144	-0.879	0.889
hil-miR-307a-3p	7.026	1.15E-07	-0.072	0.992
hil-miR-311-5p	-0.138	0.939	0.017	0.992
hil-miR-314-5p	-0.136	NA	-1.707	NA
hil-miR-315-5p	4.811	3.90E-05	-0.624	0.992
hil-miR-316-5p	2.565	0.011	-0.419	0.992
hil-miR-317-3p	-8.325	8.35E-06	-10.383	3.08E-08
hil-miR-317-5p	8.849	6.57E-08	5.319	0.003
hil-miR-31b-5p	4.752	6.10E-08	0.300	0.992
hil-miR-33-5p	3.652	4.02E-05	-0.751	0.917
hil-miR-3-3p	9.553	2.97E-11	2.218	0.152
hil-miR-34-3p	-0.135	0.939	0.020	0.992
hil-miR-34-5p	3.016	0.073	-1.376	0.827
hil-miR-375-3p	1.977	0.013	0.020	0.992
hil-miR-4-3p	2.374	0.217	-3.322	0.070
hil-miR-4982-3p	-0.136	0.939	0.019	0.992
hil-miR-4986-3p	-0.136	NA	-1.924	NA
hil-miR-5-5p	6.380	6.23E-07	1.287	0.413
hil-miR-6-2-5p	-0.134	0.939	0.021	0.992
hil-miR-7-5p	6.764	3.80E-11	-1.599	0.178
hil-miR-8-3p	3.443	0.003	-0.868	0.992
hil-miR-87-3p	4.443	0.001	-0.319	0.992
hil-miR-927-5p	8.666	2.36E-18	0.096	0.992
hil-miR-929-3p	2.290	NA	2.445	NA
hil-miR-929-5p	4.903	0.004	0.039	0.992
hil-miR-92a-3p	4.589	2.20E-05	-1.686	0.292
hil-miR-932-5p	7.463	3.07E-07	0.573	0.992
hil-miR-9378-5p	-0.133	0.939	-2.199	0.477
hil-miR-9385-5p	-1.686	NA	0.019	NA
hil-miR-956-3p	-0.135	0.939	0.020	0.992
hil-miR-957-3p	9.690	5.26E-16	0.459	0.992

http://etd.uwc.ac.za

	Larva v Male		Female v Male	
microRNA	LFC	P-Adjusted Value	LFC	P-Adjusted Value
hil-miR-958-3p	-0.269	0.939	-0.896	0.992
hil-miR-961-5p	-0.139	0.939	-2.026	0.413
hil-miR-965-3p	6.043	3.09E-05	-0.589	0.992
hil-miR-970-3p	6.604	1.79E-10	-0.389	0.992
hil-miR-971-3p	4.354	0.003	0.406	0.992
hil-miR-976-5p	-1.482	NA	0.020	NA
hil-miR-977-3p	1.191	NA	-0.319	NA
hil-miR-980-3p	5.689	1.04E-06	-0.335	0.992
hil-miR-981-3p	7.156	7.35E-07	-0.426	0.992
hil-miR-988-3p	5.216	0.001	-1.666	0.381
hil-miR-988-5p	2.733	0.102	2.888	0.131
hil-miR-993-3p	4.986	0.001	0.491	0.992
hil-miR-993-5p	-0.135	0.939	-2.405	0.364
hil-miR-999-3p	6.284	9.69E-10	-0.022	0.992
hil-miR-9a-3p	5.494	0.005	-1.035	0.992
hil-miR-9a-5p	2.410	0.011	-0.573	0.992
hil-miR-iab-4-5p	4.607	0.002	-2.985	0.029
hil-miR-iab-8-5p	7.061	1.74E-08	7.216	1.36E-08
hil-miR-193-5p	-1.786	NA	0.020	NA
hil-miR-2548-3p	-3.760	0.019	-1.797	0.588
hil-miR-311-5p	-8.238	1.39E-07	0.020	0.992
hil-miR-iab-8-3p	-0.135	NA	0.020	NA
hil-miR-973-5p	6.160	6.06E-05	6.315	9.61E-05
hil-miR-2498b-5p	-0.135	0.939	0.020	0.992
hil-miR-2765	7.016	5.97E-07	-2.881	0.024
hil-miR-2767	-0.136	NA	0.019	NA
hil-miR-2796	6.929	6.65E-07	0.500	0.992
hil-miR-2a	1.736	0.138	-0.045	0.992
hil-miR-3338	-0.135	0.939	-2.522	0.295
hil-miR-6098	5.892	8.11E-05	-0.670	0.992
hil-miR-750	-0.135	NA	-2.800	NA
hil-miR-219-3p	-0.135	NA	-2.121	NA
hil-miR-3805a-5p	-0.136	0.939	-5.138	0.001
hil-miR-3839-3p	-0.135	0.939	0.021	0.992
hil-miR-3844-3p	-0.135	NA	0.020	NA
hil-miR-3851o-2-5p	-0.135	NA	0.020	NA
hil-miR-3859-5p	4.251	0.006	4.406	0.008
hil-miR-3861-5p	-0.135	NA	0.020	NA
hil-miR-3862-3p	-0.137	0.939	0.018	0.992
hil-miR-3884-5p	-3.642	0.031	0.020	0.992
hil-miR-6011-5p	-0.136	0.939	-6.527	8.27E-05

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

Appendix B: Variant calling pipeline

B.1 QC and trimming sample whole genomes:

B.1.1 Fastqc

```
singularity exec /tools/containers/fastqc/fastqc-0.11.7.simg fastqc ./ABF3_1.fastq.gz ./-
ABF3_2.fastq.gz ./ABM1_1.fastq.gz ./ABM1_2.fastq.gz ./DF1_1.fastq.gz ./DF1_2.fastq.gz ./-
DF2_1.fastq.gz ./DF2_2.fastq.gz ./DF3_1.fastq.gz ./DF3_2.fastq.gz ./DM1_1.fastq.gz ./-
DM1_2.fastq.gz ./DM2_1.fastq.gz ./DM2_2.fastq.gz ./DM3_1.fastq.gz ./DM3_2.fastq.gz ./-
KF1_1.fastq.gz ./KF1_2.fastq.gz ./KF2_1.fastq.gz ./KF2_2.fastq.gz ./KF3_1.fastq.gz ./-
KF3_2.fastq.gz ./KM1_1.fastq.gz ./KM1_2.fastq.gz ./KM2_1.fastq.gz ./KM2_2.fastq.gz ./-
KM3_1.fastq.gz ./KM3_2.fastq.gz
```

B.1.2 Trimmomatic and concatenating

```
(py3) user@Lenovo-E51-80:~/Desktop/genome_study/genomes$ trimmomatic PE KM1_1.fastq.gz
KM1_2.fastq.gz KM1_1_trim_p.fastq.gz KM1_1_trim_up.fastq.gz KM1_2_trim_p.fastq.gz
KM1_2_trim_up.fastq.gz ILLUMINACLIP:TruSeq3-2PE.fa:2:30:10:2:keepBothReads LEADING:3
TRAILING:3 MINLEN:36

(py3) user@Lenovo-E51-80:~/Desktop/genome_study/genomes$ cat DF1_1_trim_p.fastq.gz
DF1_1_trim_up.fastq.gz > ~/Desktop/genome_study/genomes/clean_trim_genomes/-
DF1_1_cleaned.fastq.gz
```

B.2 Reference genome preparation:

B.2.1 BWA indexing of reference genome

Make sure to remove all commas and spaces from reference genome header before indexing and alignment.

```
#ran locally to remove commas and spaces in reference genome:
# sed 's/,/_/g' hil_ref.fa > hil_ref.fa

singularity exec /tools/containers/bwa/bwa-0.7.17--h84994c4_5.simg/ bwa index -p hil_ref -
a is /usr/people/sderaedt/genome_study/bwa_align/hil_ref.fa
```

UNIVERSITY of the WESTERN CAPE

B.2.2 Create GATK sequence dictionary

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"
$GATK CreateSequenceDictionary R=hil_ref.fa O=hil_ref.dict
```

B.2.3 Create Samtools reference genome index

```
SAMTOOLS="singularity exec /tools/containers/samtools/samtools-1.12--h9aed4be_1.simg
samtools"
$SAMTOOLS faidx  hil_ref.fa
```


http://etd.uwc.ac.za

B.3 Sample alignment, sorting, marking duplicates, and read group verification:

B.3.1 Align, index, and sort

```
SAMTOOLS="singularity exec /tools/containers/samtools/samtools-1.12--h9aed4be_1.simg
samtools"
BWA="singularity exec /tools/containers/bwa/bwa-0.7.17--h84994c4_5.simg bwa"
$BWA mem  -R '@RG\tID:durbanfemale\tSM:DF1\tPL:illumina\tLB:lib1\tPU:H2T22CCX2.1' -M -t $
{SLURM_CPUS_PER_TASK:-1} \
        /usr/people/sderaedt/genome_study/bwa_align/hil_ref \
        DF1_1_cleaned.fastq.gz DF1_2_cleaned.fastq.gz | $SAMTOOLS view -b  | $SAMTOOLS
sort > DF1_aligned.bam

$BWA mem  -R '@RG\tID:durbanfemale\tSM:DF2\tPL:illumina\tLB:lib2\tPU:H2T22CCX2.1' -M -t $
{SLURM_CPUS_PER_TASK:-1} \
        /usr/people/sderaedt/genome_study/bwa_align/hil_ref \
        DF2_1_cleaned.fastq.gz DF2_2_cleaned.fastq.gz | $SAMTOOLS view -b  | $SAMTOOLS
sort > DF2_aligned.bam

$BWA mem  -R '@RG\tID:durbanfemale\tSM:DF3\tPL:illumina\tLB:lib3\tPU:H2T22CCX2.1' -M -t $
{SLURM_CPUS_PER_TASK:-1} \
        /usr/people/sderaedt/genome_study/bwa_align/hil_ref \
        DF3_1_cleaned.fastq.gz DF3_2_cleaned.fastq.gz | $SAMTOOLS view -b  | $SAMTOOLS
sort > DF3_aligned.bam

$BWA mem  -R '@RG\tID:durbanmale\tSM:DM1\tPL:illumina\tLB:lib4\tPU:H2T22CCX2.1' -M -t $
{SLURM_CPUS_PER_TASK:-1} \
        /usr/people/sderaedt/genome_study/bwa_align/hil_ref \
        DM1_1_cleaned.fastq.gz DM1_2_cleaned.fastq.gz | $SAMTOOLS view -b  | $SAMTOOLS
sort > DM1_aligned.bam
```

B.3.2 Mark duplicates

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"

$GATK MarkDuplicates -I DF2_aligned.reheadered.bam -O DF2_marked_dup.bam -M
DF2_metrics.log

$GATK MarkDuplicates -I DF3_aligned.reheadered.bam -O DF3_marked_dup.bam -M
DF3_metrics.log

$GATK MarkDuplicates -I DM1_aligned.reheadered.bam -O DM1_marked_dup.bam -M
DM1_metrics.log

$GATK MarkDuplicates -I DM2_aligned.reheadered.bam -O DM2_marked_dup.bam -M
DM2_metrics.log

$GATK MarkDuplicates -I DM3_aligned.reheadered.bam -O DM3_marked_dup.bam -M
DM3_metrics.log

$GATK MarkDuplicates -I KF1_aligned.reheadered.bam -O KF1_marked_dup.bam -M
KF1_metrics.log

$GATK MarkDuplicates -I KF2_aligned.reheadered.bam -O KF2_marked_dup.bam -M
KF2_metrics.log

$GATK MarkDuplicates -I KF3_aligned.reheadered.bam -O KF3_marked_dup.bam -M
KF3_metrics.log

$GATK MarkDuplicates -I KM1_aligned.reheadered.bam -O KM1_marked_dup.bam -M
KM1_metrics.log

$GATK MarkDuplicates -I KM2_aligned.reheadered.bam -O KM2_marked_dup.bam -M
KM2_metrics.log

$GATK MarkDuplicates -I KM3_aligned.reheadered.bam -O KM3_marked_dup.bam -M
KM3_metrics.log

$GATK MarkDuplicates -I ABF3_aligned.reheadered.bam -O ABF3_marked_dup.bam -M
ABF3_metrics.log

$GATK MarkDuplicates -I ABM1_aligned.reheadered.bam -O ABM1_marked_dup.bam -M
ABM1_metrics.log
```


http://etd.uwc.ac.za

B.3.3 Read group confirmation

Make sure to check that your read groups are correct before haplotype calling.

```
SAMTOOLS="singularity exec /tools/containers/samtools/samtools-1.12--h9aed4be_1.simg
samtools"
$SAMTOOLS view -H KF1_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H KF2_md_RG.bam | grep '^@RG'

$SAMTOOLS view -H KF3_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H KM1_md_RG.bam | grep '^@RG'

$SAMTOOLS view -H KM2_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H KM3_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H DF1_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H DF2_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H DF3_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H DM1_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H DM2_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H DM3_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H ABF3_marked_dup.bam | grep '^@RG'

$SAMTOOLS view -H ABM1_marked_dup.bam | grep '^@RG'
```

To change any mislabelled ReadGroups.

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"

$GATK AddOrReplaceReadGroups \
      I=KF2_marked_dup.bam \
      O=KF2_md_RG.bam \
      RGID=kenyafemale \
      RGSM=KF2 \
      RGPL=illumina \
      RGPU=H2T22CCX2.1 \
      RGLB=lib8
```

B.4 GATK haplotype caller:

B.4.1 Create GVCF

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"

$GATK HaplotypeCaller -R hil_ref.fa -I KF1_marked_dup.bam -O KF1.g.vcf.gz -ERC GVCF

$GATK HaplotypeCaller -R hil_ref.fa -I KF2_marked_dup.bam -O KF2.g.vcf.gz -ERC GVCF

$GATK HaplotypeCaller -R hil_ref.fa -I KF3_marked_dup.bam -O KF3.g.vcf.gz -ERC GVCF

$GATK HaplotypeCaller -R hil_ref.fa -I KM1_marked_dup.bam -O KM1.g.vcf.gz -ERC GVCF

$GATK HaplotypeCaller -R hil_ref.fa -I KM2_marked_dup.bam -O KM2.g.vcf.gz -ERC GVCF

$GATK HaplotypeCaller -R hil_ref.fa -I KM3_marked_dup.bam -O KM3.g.vcf.gz -ERC GVCF
```


UNIVERSITY of the WESTERN CAPE
http://etd.uwc.ac.za

B.4.2 Create GenomicsDBImport database

B.4.2.1 All populations together

Combine all populations together to ensure the most robust variant calling. Needed to be delineated by chromosome for software.

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"

$GATK GenomicsDBImport \
        -V ABF3.g.vcf.gz \
        -V ABM1.g.vcf.gz \
        -V KF1.g.vcf.gz \
        -V KF2.g.vcf.gz \
        -V KF3.g.vcf.gz \
        -V KM1.g.vcf.gz \
        -V KM2.g.vcf.gz \
        -V KM3.g.vcf.gz \
        -V DF1.g.vcf.gz \
        -V DF2.g.vcf.gz \
        -V DF3.g.vcf.gz \
        -V DM1.g.vcf.gz \
        -V DM2.g.vcf.gz \
        -V DM3.g.vcf.gz \
        --genomicsdb-workspace-path all1_test_database \
        -L NC_051849.1_Hermetia_illucens_chromosome_1__iHerIll2.2.curated.-
20191125__whole_genome_shotgun_sequence
```

B.4.2.2 Run separate populations

Select for each population to have their own database for unique and shared SNP analysis. Still required to delineate by chromosome.

```
$GATK GenomicsDBImport \
        -V ABF3.g.vcf.gz \
        -V ABM1.g.vcf.gz \
        --genomicsdb-workspace-path Agri_chr1_database \
        -L NC_051849.1_Hermetia_illucens_chromosome_1__iHerIll2.2.curated.-
20191125__whole_genome_shotgun_sequence

$GATK GenomicsDBImport \
        -V KF1.g.vcf.gz \
        -V KF2.g.vcf.gz \
        -V KF3.g.vcf.gz \
        -V KM1.g.vcf.gz \
        -V KM2.g.vcf.gz \
        -V KM3.g.vcf.gz \
        --genomicsdb-workspace-path Kenya_chr1_database \
        -L NC_051849.1_Hermetia_illucens_chromosome_1__iHerIll2.2.curated.-
20191125__whole_genome_shotgun_sequence

$GATK GenomicsDBImport \
        -V DF1.g.vcf.gz \
        -V DF2.g.vcf.gz \
        -V DF3.g.vcf.gz \
        -V DM1.g.vcf.gz \
        -V DM2.g.vcf.gz \
        -V DM3.g.vcf.gz \
        --genomicsdb-workspace-path Durban_chr1_database \
        -L NC_051849.1_Hermetia_illucens_chromosome_1__iHerIll2.2.curated.-
20191125__whole_genome_shotgun_sequence
```


http://etd.uwc.ac.za

B.4.3 Create VCF, Genotype GVCFs in GenomicsDBImport database

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"

$GATK GenotypeGVCFs \
        -R hil_ref.fa \
        -V gendb://all1_test_database \
        -O chr1 all.vcf.gz
```

B.5 VCF filtering:

B.5.1 Select SNPs / remove InDels

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"

$GATK SelectVariants \
        -R hil_ref.fa \
        -V chr1_all.vcf.gz \
        --select-type-to-include SNP \
        -O chr1_snp.vcf.gz
```

B.5.2 Density plots for hard filtering parameters

Send all variants to a table for R studio to be able to create plots.

```
$GATK VariantsToTable \
        -V chr1_snp.vcf.gz \
        -F ID \
        -F QD \
        -F FS \
        -F SOR \
        -F MQ \
        -F MQRankSum \
        -F ReadPosRankSum \
        -O chr1.table
```

Run R ggplot to create density plots to know if your parameters are appropriate.

```
1 setwd("~/Desktop/")
2 library(ggplot2)
3
4 #QUALITYDEPTH
5 QD.plot <- ggplot(chr1, aes(x=QD)) +geom_density(alpha=0.2)
6
7 QD.plot
8
9 #FISHERSTRAND
0 FS.plot <- ggplot(chr1, aes(x=FS)) +geom_density(alpha=0.2)
1
2 FS.plot + scale_x_log10()
3
4 #STRANDSODDRATIO
5 SOR.plot <- ggplot(chr1, aes(x=SOR)) +geom_density(alpha=0.2)
6
7 SOR.plot
8
9 #MAPPINGQUALITY
0 MQ.plot <- ggplot(chr1, aes(x=MQ)) +geom_density(alpha=0.2)
1
2 MQ.plot
3
4 #MQRANKSUM
5 MQRankSum.plot <- ggplot(chr1, aes(x=MQRankSum)) +geom_density(alpha=0.2)
6
7 MQRankSum.plot
8
9 #READPOSRANKSUM
0 ReadPosRankSum.plot <- ggplot(chr1, aes(x=ReadPosRankSum)) +geom_density(alpha=0.2)
1
2 ReadPosRankSum.plot
```


http://etd.uwc.ac.za

B.5.3 Hard filtering

B.5.3.1 GATK hard filtering

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"

$GATK VariantFiltration \
        -V chr1_snp.vcf.gz \
        -filter "QD < 2.0" --filter-name "QD2" \
        -filter "QUAL < 30.0" --filter-name "QUAL30" \
        -filter "SOR > 3.0" --filter-name "SOR3" \
        -filter "FS > 60.0" --filter-name "FS60" \
        -filter "MQ < 40.0" --filter-name "MQ40" \
        -filter "MQRankSum < -12.5" --filter-name "MQRankSum-12.5" \
        -filter "ReadPosRankSum < -8.0" --filter-name "ReadPosRankSum-8" \
        -O chr1_snps_filtered.vcf.gz
```

B.5.3.2 Filter for missing data

```
GATK="singularity exec /tools/containers/gatk/gatk-4.2.0.0.simg gatk"

$GATK SelectVariants \
        -R hil_ref.fa \
        -V wg_Agri_snp_gatk.vcf.gz \
        --max-nocall-fraction 0.85 \
        --exclude-filtered \
        -O wg_Agri_filtered.vcf.gz

$GATK SelectVariants \
        -R hil_ref.fa \
        -V wg_Kenya_snp_gatk.vcf.gz \
        --max-nocall-fraction 0.85 \
        --exclude-filtered \
        -O wg_Kenya_filtered.vcf.gz

$GATK SelectVariants \
        -R hil_ref.fa \
        -V wg_Durban_snp_gatk.vcf.gz \
        --max-nocall-fraction 0.85 \
        --exclude-filtered \
        -O wg_Durban_filtered.vcf.gz
```

B.6 VCF Analysis:

B.6.1 PCA

```
singularity exec /tools/containers/plink/plink-1.90b4--h0a6d026_2.simg plink --bfile
plink --pca --allow-extra-chr
```

B.6.2 Nucleotide Diversity

Use VCFTools to run nucleotide diversity using sliding windows across the genome.

http://etd.uwc.ac.za

```
#chr1
singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr1_kenyan_male_fil.vcf --window-pi 20000 --out chr1_km_20kb_nd

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr1_kenyan_fil.vcf --window-pi 20000 --out chr1_k_20kb_nd

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr1_kenyan_female_fil.vcf --window-pi 20000 --out chr1_kf_20kb_nd

#chr2
singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr2_kenyan_male_fil.vcf --window-pi 20000 --out chr2_km_20kb_nd

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr2_kenyan_fil.vcf --window-pi 20000 --out chr2_k_20kb_nd

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr2_kenyan_female_fil.vcf --window-pi 20000 --out chr2_kf_20kb_nd
```

B.6.3 Tajima's D

Use VCFTools to run Tajima's D using sliding windows across the genome.

```
#chr1
singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr1_kenyan_male_fil.vcf --TajimaD 20000 --out chr1_km_20kb_td

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr1_kenyan_fil.vcf --TajimaD 20000 --out chr1_k_20kb_td

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr1_kenyan_female_fil.vcf --TajimaD 20000 --out chr1_kf_20kb_td

#chr2
singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr2_kenyan_male_fil.vcf --TajimaD 20000 --out chr2_km_20kb_td

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr2_kenyan_fil.vcf --TajimaD 20000 --out chr2_k_20kb_td

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr2_kenyan_female_fil.vcf --TajimaD 20000 --out chr2_kf_20kb_td
```

B.6.4 Heterozygosity/Inbreeding coefficient

```
singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr1_kenyan_fil.vcf --het --out chr1_kenyan_het

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr2_kenyan_fil.vcf --het --out chr2_kenyan_het

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr3_kenyan_fil.vcf --het --out chr3_kenyan_het

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr4_kenyan_fil.vcf --het --out chr4_kenyan_het

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr5_kenyan_fil.vcf --het --out chr5_kenyan_het

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr6_kenyan_fil.vcf --het --out chr6_kenyan_het

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf chr7_kenyan_fil.vcf --het --out chr7_kenyan_het

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf wg_kenyan_fil.vcf --het --out wg_kenyan_het

singularity exec /tools/containers/vcftools/vcftools-0.1.16--he941832_2.simg vcftools --vcf wg_agri_fil.vcf --het --out wg_agri_het
```

B.6.5 Unique & shared SNPs

This analysis needs to be completed on separated population GenomicsDBImport files. Otherwise, GATK forms a squared-off-matrix that establishes most likely variants across all populations. This is preferred for our initial SNP calling since BSF does not have any known SNPs to calibrate. Therefore, the matrix gives the "best" calls across all samples. However, in order to isolate unique SNPs, each population must have their own matrix.

http://etd.uwc.ac.za

```
singularity exec /tools/containers/bcftools/bcftools-1.10.2--h4f4756c_3.simg bcftools
isec -n~100 -c all wg_Agri_filtered.vcf.gz wg_Kenya_filtered.vcf.gz
wg_Durban_filtered.vcf.gz -p ISEC_AGRI_UNI

singularity exec /tools/containers/bcftools/bcftools-1.10.2--h4f4756c_3.simg bcftools
isec -n~010 -c all wg_Agri_filtered.vcf.gz wg_Kenya_filtered.vcf.gz
wg_Durban_filtered.vcf.gz -p ISEC_KENYA_UNI

singularity exec /tools/containers/bcftools/bcftools-1.10.2--h4f4756c_3.simg bcftools
isec -n~001 -c all wg_Agri_filtered.vcf.gz wg_Kenya_filtered.vcf.gz
wg_Durban_filtered.vcf.gz -p ISEC_DURBAN_UNI
```

Count the number of SNPs

```
singularity exec /tools/containers/bcftools/bcftools-1.10.2--h4f4756c_3.simg bcftools
view -v snps 0000.vcf | grep -v "^#" | wc -l

singularity exec /tools/containers/bcftools/bcftools-1.10.2--h4f4756c_3.simg bcftools
view -v snps 0001.vcf | grep -v "^#" | wc -l

singularity exec /tools/containers/bcftools/bcftools-1.10.2--h4f4756c_3.simg bcftools
view -v snps 0002.vcf | grep -v "^#" | wc -l
```

http://etd.uwc.ac.za

Appendix C: microRNA discovery pipeline

C.1 QC and trimming sample sequences:

```
#FASTQC

$/home/user/anaconda3/bin/fastqc ./.1E.fastq ./.2E.fastq ./.3E.fastq ./1P.fastq ./2P.fastq ./3P.fastq ./1L.fastq ./2L.fastq
./3L.fastq ./1M.fastq ./2M.fastq ./3M.fastq ./1F.fastq ./2F.fastq ./3F.fastq
```

C.2 Reference genome and conserved microRNA prep:

C.2.1 Concatenate reference genome files

```
#FASTA COMBINE

$~/Desktop/miRNA_study/bsf_gen_rerun/genome_combined/ cat *.fasta > hil_gencom.fa

#REMOVE WHITE SPACE

$~/Desktop/miRNA_study/bsf_gen_rerun/genome_combined/ sed 's, ,_,g' -i hil_gencom.fa
```

C.2.2 Bowtie index reference genome

```
#BOWTIE BUILD REF GENOME

$/home/user/anaconda3/bin/ bowtie-build hil_gencom.fa hil_all
```

Be sure not to specify file type at end of output ("hil_all" instead of "hil_all.fa")

C.2.3 Select appropriate conserved microRNA

```
#CONSERVED MICRORNA PULL

$/home/user/anaconda3/bin/mirbase.pl 22.1 1

#SELECT ARTHROPOD MICRORNAS

$/home/user/anaconda3/bin/extract_miRNAs.pl mature.fa aae, acy, aga, ame, bmo, cho, cma, cpi, dps, dse, dmo, dpe, dro, dwi, dya,
der, dan, dgr, hel, mse, nvi, tca > mature_arthro.fa
```


http://etd.uwc.ac.za

C.3 Map sample microRNA to reference genome:

C.3.1 Mapper.pl from MiRDeep2 software

```
$/home/user/anaconda3/bin/mapper.pl 1E.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 1E_hilreads_collapsed.fa
-t 1E_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 1L.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 1L_hilreads_collapsed.fa
-t 1L_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 1P.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 1P_hilreads_collapsed.fa
-t 1P_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 1F.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 1F_hilreads_collapsed.fa
-t 1F_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 1M.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 1M_hilreads_collapsed.fa
-t 1M_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 2E.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 2E_hilreads_collapsed.fa
-t 2E_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 2L.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 2L_hilreads_collapsed.fa
-t 2L_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 2P.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 2P_hilreads_collapsed.fa
-t 2P_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 2F.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 2F_hilreads_collapsed.fa
-t 2F_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 2M.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 2M_hilreads_collapsed.fa
-t 2M_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 3E.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 3E_hilreads_collapsed.fa
-t 3E_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 3L.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 3L_hilreads_collapsed.fa
-t 3L_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 3P.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 3P_hilreads_collapsed.fa
-t 3P_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 3F.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 3F_hilreads_collapsed.fa
-t 3F_read_vs_hil.arf

$/home/user/anaconda3/bin/mapper.pl 3M.fastq -e -h -i -j -k TGGAATTCTCGGGTGCCAAGG -l 18 -m -p hil_all -s 3M_hilreads_collapsed.fa
-t 3M_read_vs_hil.arf
```

```
#MAPPER PARAMETERS DESCRIBED

# -e      identifies input as fastq
# -h      parse to fasta format
# -i      convert rna to dan alphabet (to map against genome)
# -j      remove all entries that have letters other than "a,c,g,t,u,n,A,C,G,T,U,N"
# -k      clip 3' adapters (adapter sequence: TGGAATTCTCGGGTGCCAAGG)
# -l      discard reads shorter than 18 nucleotides long
# -m      collapse reads
# -p      map to genome using .bwt prefix files (from bowtie-build)
# -s      output processed reads to file (.fa)
# -t      output read mappings to .arf file (miRDeep2 unique file format)
# -v      ouptut progress report
```


http://etd.uwc.ac.za

C.4 Call microRNA using miRDeep2:

```
$/home/user/anaconda3/bin/miRDeep2.pl 1E_hilreads_collapsed.fa hil_gencom.fasta 1E_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 1E_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 1L_hilreads_collapsed.fa hil_gencom.fasta 1L_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 1L_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 1P_hilreads_collapsed.fa hil_gencom.fasta 1P_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 1P_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 1F_hilreads_collapsed.fa hil_gencom.fasta 1F_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 1F_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 1M_hilreads_collapsed.fa hil_gencom.fasta 1M_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 1M_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 2E_hilreads_collapsed.fa hil_gencom.fasta 2E_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 2E_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 2L_hilreads_collapsed.fa hil_gencom.fasta 2L_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 2L_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 2P_hilreads_collapsed.fa hil_gencom.fasta 2P_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 2P_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 2F_hilreads_collapsed.fa hil_gencom.fasta 2F_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 2F_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 2M_hilreads_collapsed.fa hil_gencom.fasta 2M_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 2M_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 3E_hilreads_collapsed.fa hil_gencom.fasta 3E_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 3E_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 3L_hilreads_collapsed.fa hil_gencom.fasta 3L_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 3L_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 3P_hilreads_collapsed.fa hil_gencom.fasta 3P_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 3P_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 3F_hilreads_collapsed.fa hil_gencom.fasta 3F_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 3F_nonarthro.log

$/home/user/anaconda3/bin/miRDeep2.pl 3M_hilreads_collapsed.fa hil_gencom.fasta 3M_read_vs_hil.arf none mature_arthro.fa
hairpin_arthro.fa 2 > 3M_nonarthro.log
```

C.5 Create one table from multiple life stage files:

```
#microRNA count

$~/Desktop/miRNA_study/bsf_gen_rerun/perl CountMiRNA.pl 1E.csv 1L.csv 1P.csv 1M.csv 1F.csv 2E.csv 2L.csv 2P.csv
2M.csv 2F.csv 3E.csv 3L.csv 3P.csv 3M.csv 3F.csv
```

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

```
#! /usr/bin/perl
#
# CountMiRNA.pl
# use hash tables and specific hash of hash. Python has an equivalent
#
# @ARGV is the reserved namespace in PERL for all arguments on the commandline.
#
# Usage: CountMiRNA.pl [count_file1] [count_file2] ......etc.
#
# Does not matter what order the files are specified because I hard coded the output.
# And I track the stages by the filename that was submitted.
#

my ($stage, %All);

foreach my $arg(@ARGV) {
        if ($arg =~/^(\S+)\.csv/) {
                $stage = $1;
        } else {
                die "cannot recognise $arg\n";
        }
        open(F, $arg);
        while (<F>) {
                chomp;
                my ($cnt, $id) = split(/,/);
                #print ".$cnt. .$id. .$stage.\n";
                $hash{$id}{$stage} =  $cnt;
        }
        close(F);
        if ($stage =~/^$/) {
                die "missing\n";
        }
}


my @stages = qw(1E 2E 3E 1L 2L 3L 1P 2P 3P 1F 2F 3F 1M 2M 3M);
print "1E\t2E\t3E\t1L\t2L\t3L\t1P\t2P\t3P\t1F\t2F\t3F\t1M\t2M\t3M\n";
foreach my $miR (keys %hash) {
        print "$miR\t";
        foreach my $stages(@stages) {
                if (defined $hash{$miR}{$stages}) {
                        print "\t".$hash{$miR}{$stages};
                } else {
                        print "\t0";
                }
        }
        print "\n";
}
```

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za