Global distribution of tabanid flies, characterization and microbiome insights of *Tabanus par* (Diptera-Tabanidae)

ET Emsley

orcid.org/ 0000-0001-5804-0089

Thesis accepted in fulfilment of the requirements for the degree *Doctor of Philosophy in Science with Environmental Sciences* at the North-West University

Promoter: Dr MO Taioe
Co-promoter: Prof KE Lekota
Co-promoter: Prof MMO Thekisoe

Graduation: July 2025

DEDICATION

This thesis is dedicated to the late Paulus Isaac Emsley and Piet Koueane Mashilo. For their inspiration and lessons to overcome any obstacle that life may present.

I also dedicate this to my mother Ruth Emsley and my siblings, Cathrine Emsley and Sophy Emsley, who have been a source of endless inspiration and strength. I am deeply grateful for the love and support you have all given me—thank you from the bottom of my heart.

THANK YOU

ACKNOWLEDGEMENTS

I extend my deepest gratitude to my Lord and Savior, Jesus Christ, for granting me the opportunity and strength to embark on and complete this degree.

I am sincerely grateful to my supervisors, Dr. Moeti Taioe, Prof. Kgaugelo Lekota and Prof. Oriel Thekisoe for their invaluable guidance and mentorship throughout the course of this study. Your contributions and support have been instrumental in the completion of this project, and I am truly thankful for your patience, advice, and encouragement. Your dedication and expertise have set a remarkable example of what it means to be an academic and researcher.

A special note of thanks to Prof. Kgaugelo Lekota for your assistance with all aspects of bioinformatics and for ensuring the successful completion of the data analyses for this project. Without your support, this work would not have been possible, and words cannot fully express my gratitude.

I also extend heartfelt thanks to my mother, Maletšema Ruth Emsley, my grandmother, Emily Tsiri, and my siblings, Cathrine Emsley and Sophy Emsley, for their unwavering encouragement and support throughout my studies. Through the highs and lows of this journey, my family has been a constant source of joy and motivation, and for that, I am deeply grateful. This dissertation would not have been completed without their enduring love and backing.

Additionally, I am thankful to my colleagues at North-West University for their support, encouragement, and camaraderie throughout this project. Your kindness and encouragement have meant so much to me.

I also express my gratitude to the Agricultural Research Council (ARC) – Biotechnology Platform for providing the sequencing data used in this thesis and to the National Research Foundation (NRF) for the PhD scholarship.

May God bless you all—thank you

TABLE OF CONTENTS

DEDICATION ..... **I**

ACKNOWLEDGEMENTS ..... **II**

LIST OF ABBREVIATIONS ..... **XV**

RESEARCH OUTPUTS ..... **XVII**

ABSTRACT ..... **XVIII**

CHAPTER 1 GENERAL INTRODUCTION ***1***

1.1 Background **1**

1.2 Justification of the study **2**

1.3 Aim **3**

1.4 Objectives **3**

1.5 Thesis outline **4**

1.6 References **5**

CHAPTER 2 LITERATURE REVIEW ***7***

2.1 Taxonomy of Diptera **7**

2.2 Global distribution of Tabanidae **8**

2.3 General life cycle of tabanid flies **8**

2.3.1 Eggs 9

2.3.2 Larvae 10

2.3.3 Pupa 10

2.3.4 Adult 10

2.4 Tabanids as vectors of pathogens **11**

2.5 Microbiome profile of tabanid flies 12

2.6 Traps for the capture of tabanid flies 13

2.7 Morphological and molecular identification of tabanids 15

2.8 References 17

CHAPTER 3 THE GLOBAL PREVALENCE'S AND DISTRIBUTION OF TABANID FLIES BY SYSTEMATIC REVIEW AND META-ANALYSIS OF PUBLISHED

3.1 Introduction 23

3.2 Materials and methods 24

3.2.1 Review protocol 24

3.2.2 Search strategy for relevant studies 24

3.2.3 Inclusion and exclusion criteria 25

3.2.4 Data quality control measures 25

3.2.5 Meta-analysis 26

3.3 Results 26

3.3.1 Bibliometric data 26

3.3.2 Bibliometric data General characteristics of the included studies 28

3.3.2.1 Distribution of studies across the globe 28

3.3.3 Publication year of eligible studies 29

3.3.4 Tabanid fly identification and articles used for nomenclature 30

3.3.5 Trapping methods 31

3.3.6 Hematophagous tabanid fly data 31

3.3.6.1 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Chrysopini 33

3.3.6.2 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Diachlorini 37

3.3.6.3 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Haematopotini 39

3.3.6.4 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Philolichini 42

3.3.6.5 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Tabanini 42

3.4 Publication bias 48

3.5 Discussion 52

3.6 Limitation 56

3.7 Conclusion 56

3.8 Recommendations 57

3.9 References 58

CHAPTER 4 MORPHOLOGICAL AND GENETIC CHARACTERISTICS OF TABANUS PAR FROM SELECTED LOCALITIES IN SOUTH AFRICA 65

4.1 Introduction 65

4.2 Materials and methods 66

4.2.1 Ethical clearance 66

4.2.2 Fly sample collections 66

4.2.3 Sample collection site 67

4.2.3.1 Charters Creek 67

4.2.3.2 Hluhluwe–iMfolozi 68

4.2.3.3 Phinda Private Game Reserve 68

4.3 Preparation of samples, pinning and drying 69

4.3.1 Morphometrical analysis 70

4.3.2 DNA extraction 70

4.3.3 DNA extraction Amplification of cytochrome oxidase 1 (*CO1*) gene 71

4.3.4 Purification of PCR amplicons 71

4.3.5 Sequence alignment and phylogenetic analysis 72

4.3.6 Statistical analysis 72

4.4 Results 73

4.4.1 Morphological observation 73

4.4.2 Morphometrical observations 75

4.4.2.1 Morphometric variation across traps 76

4.4.3 Genetic observation 78

4.5 Discussion 81

4.6 Conclusion 82

4.7 Recommendations 83

4.8 References 84

CHAPTER 5 MICROBIOMES HARBOURED BY TABANUS PAR BY USING SHOTGUN METAGENOMICS 89

5.1 Introduction 89

5.2 Materials and methods 91

5.2.1 Source of *T. par* specimens 91

5.2.2 DNA extraction procedure 91

5.2.3 Metagenomic sequencing 91

5.2.4 Bioinformatics analysis 92

5.3 Results 92

5.3.1 The bacterial composition at the phylum..... 93

5.3.2 The bacterial composition at the family level 93

5.3.3 The bacterial composition at the genus level..... 93

5.3.4 The fungal composition at the phylum level..... 96

5.3.5 The fungal composition at the family level 96

5.3.6 The fungal composition at the genus level..... 96

5.3.7 The viral composition at the phylum level 99

5.3.8 The viral composition at the family level 99

5.3.9 The viral composition at the genus level..... 99

5.3.10 The bacterial plasmids composition at the phylum level..... 102

5.3.11 The bacterial plasmids composition at the family level 102

5.3.12 The bacterial plasmids composition at the genus level..... 102

5.3.13 The eukaryote composition at the phylum level..... 105

5.3.14 The eukaryote composition at the family level 105

5.3.15 The eukaryote composition at the genus/species level..... 105

5.4 Discussion 107

5.5 Conclusion..... 113

5.6 Recommendations 113

5.7 References..... 114

CHAPTER 6 GENOMIC DIVERSITY AND DISTRIBUTIONS OF BACTERIOPHAGES HARBOURED BY TABANUS PAR..... 123

6.1 Introduction 123

6.2 Materials and methods 125

6.2.1 Source of *T. par* 125

6.2.2 DNA extraction procedure 125

6.2.3 Metagenomic sequencing 125

6.2.4 Bioinformatics analysis 126

6.3 Results 127

6.3.1 Phage distribution taxa of the genomes 127

6.3.2 Taxonomic classification and characterization of complete phages 130

6.3.3 Structural and transmembrane protein annotations of ten phages in *T. par* 134

6.3.4 Placement of the phage within the phylogenetic framework 138

6.4 Discussion 143

6.5 Conclusion 146

6.5.1 Recommendations 147

6.6 References 148

CHAPTER 7 GENETIC CHARACTERISTICS OF SPIROPLASMA SPECIES HARBOURED BYTABANUS PAR USING METAGENOMICS METHOD 152

7.1 Introduction 152

7.2 Materials and methods 156

7.2.1 Source of *T.par* specimens 156

7.2.2 DNA extraction procedure 156

7.2.3 Metagenomic sequencing of *T. par* 157

7.2.4 Metagenomic, quality control, assembly and binning 157

7.3 Pangenomics and average nucleotide identity analysis 157

7.4 Results 158

7.4.1 Genome analysis of bacteria 161

7.4.2 Comparative genomics of *Spiroplasma* species 162

7.4.3 Pangenome analysis of *Spiroplasma* strains 163

7.4.4 Genome annotation of *Spiroplasma megaselia nigra* TPWB 165

7.4.5 Genome annotation of *Spiroplasma turonicum* TPWBtur 166

7.4.6 Genome annotation of *Spiroplasma* sp. TPWB 168

7.5 Discussion 169

7.6 Conclusion 172

7.7 Recommendations 172

7.8 References 173

CHAPTER 8 ANNEXURES 178

LIST OF TABLES

Table 3-1: Captures of the genus *Chrysops* (horse flies) documented between 1960-2022. 34

Table 3-2: Captures of genus Diachlorini documented between 1960 to 2022. 38

Table 3-3: Captures of the tribe *Haematopini* documented between 1960 to 2022. 40

Table 3-4: Captures of the tribe of *Philochini* documented from 1960 to 2022. 42

Table 3-5: Captures of the tribe Tabanini between 1960 to 2022. 44

Table 3-6: Captures of genus *Tabanus* between 1960 to 2022. ..**Error! Bookmark not defined.**

Table 4-1: Body part lengths and whole-body weight of female *T. par* dry specimens. 76

Table 6-1: Summary of the genomic features of phage genomes identified from the whole body of *Tabanus par.* 132

Table 6-2: Summary of the genomic features of phage genomes identified from the legs of *Tabanus par.* 133

Table 6-3: Annotation of the *Tabanus par* phage genomes identified in both whole body

Table 7-1: *Spiroplasma* species and their associated arthropod or plant hosts (Regassa and Gasparich, 2006). 153

Table 7-2: Prokaryotic metagenome assembly genomes from the whole body of *T. par* that presents good bins. 159

Table 7-3: Prokaryotic metagenome assembly genomes from the legs of *T. par* classified as good bins. 160

Table 7-4: Genomic features of *Spiroplasma* sp. strains from the whole body of *T. par.* 161

Table 7-5: Subsystem category distribution of the genome of *S. megaselia* nigra TPWB annotated using RAST. 166

Table 7-6: Subsystem category distribution of *S. turonicum* TPWB annotated using RAST. ..167

Table 7-7: : Subsystem category of *Spiroplasma*.sp TPWB annotated using RAST. 169

LIST OF FIGURES

Figure 2-1: Tabanidae hierarchy within order Diptera. 7

Figure 2-2: Life cycle of tabanid fly (adapted from Bautista-Garfias *et al*., 2021). 9

Figure 2-3: Trapping methods for tabanid flies: A)) Broce trap, (B) Malaise trap, (C) Canopy trap, (D) Blue target, (E) Vavoua trap, (F) Nzi trap, (G) Manitoba trap, (H) H-trap, and (I) Horizontally polarizing liquid trap (Baldacchino *et al.*, 2018). 14

Figure 3-1: Flow chart illustrating the process of selecting studies for inclusion in the review, following the PRISMA guidelines. 27

Figure 3-2: World map showing the number of articles from different countries which reported tabanid distribution. 28

Figure 3-3: Bar graph illustrating the distribution of selected studies over time, categorized by decade from 1950-2022, highlighting the trends in research activity. 29

Figure 3-4: Bar graph representing summary of articles utilized for the description of Tabanidae in the selected studies. 30

Figure 3-5: Illustration of pie chart showing various trapping methods employed for capturing tabanid flies. 31

Figure 3-6: Simplified systematic classification of the different tabanids (horse flies) group studies recorded. 32

Figure 3-7: The funnel plot of *Chrysops carbonarius* (Walker, 1848) studies included in the meta-analysis. 48

Figure 3-8: The funnel plot *Hybomitra bimaculate* (Macquart, 1826) studies included in meta-analysis. 49

Figure 3-9: The funnel plot of *Tabanus autumnalis* (Linnaeus, 1761) studies included in the meta-analysis. 50

Figure 3-10: The funnel plot of *Tabanus maculicornis* (Zetterstedt, 1842) studies included in the meta-analysis. 51

Figure 4-1: H-trap used to capture tabanid flies in South Africa (Taioe, 2017). 67

Figure 4-2: Map of southern Africa showing three sampled game reserves namely: Charters Creek, Hluhluwe–iMfolozi Park and Phinda Private Game Reserve. 69

Figure 4-3: Sequential images of morphological characteristics of female *Tabanus* par (Adapted from Oldroyd, 1954; Neave, 1912): A) Thoracic coloration B) Eye coloration with emerald green and golden upper facets C) proboscis, D) Pleura coloration and white hairs; E) Antennae; F) Abdomen; G) Wings H)..... 74

Figure 4-4: Morphometric measurements of female *T. par* specimens collected from Phinda Private Game Reserve. 77

Figure 4-5: Morphometric measurements of female *T.par* specimens collected from Charters Creek. 77

Figure 4-6: Morphometric measurements of female *T. par* specimens collected from Hluhluwe-Imfolozi Park. 78

Figure 4-7: Molecular phylogenetic analysis by Maximum Likelihood (ML) method. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016). 80

Figure 5-1: The Krona chart comparing the taxonomic reads distribution of bacteria at genus level in the whole body (WB) and legs (L) of *T. par.* 95

Figure 5-2: The Krona chart comparing the taxonomic reads distribution of fungi at genus level in the whole body (WB) and legs (L) of *T. par.* 98

Figure 5-3: The Krona chart comparing the taxonomic reads distribution of viruses at genus level in the whole body (WB) and legs (L) of T. par. 101

Figure 5-4: The Krona chart comparing the taxonomic reads distribution of bacterial plasmids at the genus species in the whole body (WB) and legs (L) of *T. par.* 104

Figure 5-5: The Krona chart showing the taxonomic reads distribution of eukaryotes at genus/species level in the whole body (WB) and legs (L) *of T. par.* 106

Figure 6-1: Schematic diagram illustrating life cycle of phage within a host (Chevallareau *et al.*, 2020). 124

Figure 6-2: The taxonomic assignment and distribution of phage genomes identified from the body of *T. par.* 128

Figure 6-3: The taxonomic assignment and distribution of phage genomes identified from the legs of *T. par.* 129

Figure 6-4: Structural and transmembrane protein annotations. A: *Bacillus* TPL; B: *Staphylococcus* TPL C: *Klebsiella* TPL, sequenced from legs of *Tabanus par.* The arrows represent the direction of transcription (strand + or -). 136

Figure 6-5: Structural and transmembrane protein annotations of the complete phage genomes identified in the body. A: *Acinetobacter* TPWB; B: *Arcarnobacterium* TPWB; C: *Caudovirales*; D: *Caulobacter* TPWB; E: *Staphylococcus* TPWB and F: *Stenophromonas* TPWB. The arrows represent the direction of transcription (strand + or -)..... 137

Figure 6-6: Phylogenetic placement of the ten sequenced phages from *Tabanus* par. Whole genome-based phylogeny (amino acid) was inferred using Genome-BLAST Distance Phylogeny (100 bootstraps) using formula D6, yielding an average support of 70%. 139

Figure 6-7: Phylogenetic and comparative genomic analyses of 10 *Tabanus par* phages and the 40 closest viral genomes. 140

Figure 6-8: Intergenomic similarities heat map constructed by VIRIDIC showing the average nucleotide identity between the sequenced *Tabanus par* phages. 141

Figure 6-9: The Viriclust hierarchical clustering of phages recovered from *T. par,* based on intergenomic distance calculated using protein cluster content as genomic singletons. 142

Figure 7-1: Average nucleotide identity phylogenetic tree, showing the relationship among 75 *Spiroplasma* genomes..... 162

Figure 7-2: A) Pangenomic phylogenomic tree of the 75 Spiroplasma sp., genomes, including the three MAGs determined in this study. Central dendrogram clustering of genomes is ordered by gene cluster presence/absence. Items order presence absence (D, Euclidean; L, War Ward). B) COG functional classification of core genome of the analysed 75 *Spiroplasma* genomes. 164

Figure 7-3: The pie-chart showing subsystem coverage, subsystem category distribution, and subsystem feature counts of the annotated genome of S. megaselia nigra TPWB. 165

Figure 7-4: Subsystem coverage, subsystem category distribution, and subsystem feature counts of the annotated genome of *Spiroplasma turonicum* TPWB. 167

Figure 7-5: Subsystem coverage, subsystem category distribution, and subsystem feature counts of the annotated genome of *Spiroplsama*. sp TPWB. 168

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
BLV	Bovine leukemia virus
BMR	Begg's and Mazumdar's rank
CI	Confidence interval
Cm	Centimetre
CMA	Comprehensive Meta-Analysis
CO1	Cytochrome c oxidase subunit I
COG	Orthologous groups
DNA	Deoxyribonucleic acid
EIAV	Equine infectious anaemia virus
GBDP	Genome BLAST Distance Phylogeny
KZN	KwaZulu-Natal Province
L	Legs
LSDV	Lumpy Skin Disease Virus
MAGs	Metagenome-assembled genomes
Mg	Milligram
NGS	Next-generation sequencing
NWU	North-West University
ORF	Open reading frame
PCR	Polymerase chain reaction

PPE	Pooled prevalence estimate
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analysis
SAS	Statistical Analysis System
t-LSDs	Least significant differences
T. par	*Tabanus par*
vMAGs	viral Meta-assembled genomes
WB	Whole body
WGS	Whole-genome sequencing

RESEARCH OUTPUTS

Conference papers

Emily Emsley, Moeti Taioe, Tsepo Ramatla, Kgaugelo Lekota, Oriel Thekisoe. Gut microbiome of wildtype *Tabanus par* (Walker, 1854) using metagenomics approach. Oral presentation. **21-23 August 2024, The 21st Annual Southern African Society of Veterinary Epidemiology and Medicine (SASVEPM) Congress, Lagoon Beach Hotel, Cape Town, South Africa.**

Emily Emsley, Moeti Taioe, Tsepo Ramatla, Kgaugelo Lekota, Oriel Thekisoe. Characterization of Pathogen Diversity in *Tabanus par* Using High-Throughput Sequencing. Poster presentation. **6-8 October 2024, The 52nd Annual Parasitological Society of Southern Africa (PARSA) Conference, Villa Paradiso, Hartbeespoort, South Africa.**

Emily Emsley, Moeti Taioe, Tsepo Ramatla, Kgaugelo Lekota, Oriel Thekisoe. Genomic diversity and distributions of novel bacteriophages from *Tabanus par* revealed by metagenomics sequencing. E-poster presentation. **23-25 October 2024. International Microbiological Societies (IUMS) Congress, Microorganisms for sustainable solutions: environmental and clinical implementation, Florence, Italy.**

ABSTRACT

Tabanid species, commonly known as horse flies, are blood-feeding insects belonging to the family Tabanidae. These flies are significant pests due to their painful bites and their role as biological and mechanical vectors for various medical and veterinary pathogens, including parasites, bacteria, and viruses. Despite their importance, insufficient data exists regarding their classification and identification. This study addresses this gap by conducting a systematic review and meta-analysis of the global prevalence and distribution, followed by morphological, morphometric and genetic characterization using *Tabanus par* (*T. par*) as a model species. Additionally, the study explored the microbiome of *T. par*, including bacterial, viral, fungal and eukaryotic protozoan species.

The systematic review revealed the presence of tabanid species across five continents and from 16 countries, with notable contributions from the USA, Croatia, and Bulgaria. An increasing trend in tabanid research was observed over the decades, with 209 species identified across five tribes. Morphological and genetic analyses of 30 *T. par* specimens confirmed species identification through mitochondrial *cytochrome oxidase 1* (*CO1*) gene sequencing. Furthermore, phylogenetic analysis from the *CO1* gene region showed that *T. par* collected from KwaZulu-Natal Province of South Africa are closely related to those of Mozambique as compared to those from West Africa.

Metagenomic analysis of the *T. par* revealed distinct microbial diversity in the whole body (WB) and legs (L). Bacteria from the phylum Pseudomonodota were dominant, with *Wolbachia* sp. (34%) prevalent in WB and *Burkholderia vietnamiensis* (10%) in legs. The eukaryotic microbiome was dominated by the phylum Euglenozoa, with *Trypanosoma theileri* (32%) identified as a key pathogen. Fungal diversity was marked by *Ascomycota*, with *Synchytrium* and *Fusarium proliferatum* predominating in WB and legs, respectively. Viral metagenomics identified bacteriophages from *Mycobacterium*, *Escherichia* and *Pseudomonas*, with ten complete phage genomes characterized, indicating both lytic and temperate viral lifestyles. Additionally, three draft genomes of *Spiroplasma* species were identified from *T. par*. Pangenomics analysis revealed core genes associated with translation and replication, reflecting their endosymbiotic adaptation.

In conclusion, findings of this entire study will contribute to the development of a comprehensive checklist of the family Tabanidae based on the systematic review and meta-analysis. Furthermore, data generated in this study documented a wide variety **of** microbial communities (bacterial, viral and fungal) harboured by South African *T. par* and will go a long way in playing a role in the formulation of solutions for the biological management of tabanid populations.

Key terms: *Tabanus par*, Global distribution, Morphometrics, Genetic characterization, Metagenomics, Microbiome.

CHAPTER 1 GENERAL INTRODUCTION

1.1 Background

Tabanid flies, which are also known as horseflies, clegs, marsh-flies and deerflies, are members of the Tabanidae, which are classified into four subfamilies: (i) Chrysopsinae [Bouvieromyiini, Chrysopsini and Rhinomyzini], (ii) Pangoniinae [Mycteromyiini, Pangoniini, Philolichini and Scionini], (iii) Scepsidinae (*Braunsiomyia* and *Scepsis)*. and (iv) Tabaninae [Diachlorini, Haematopini and Tabanini] (Morita *et al*., 2016). They have cosmopolitan distribution, with exceptions from the polar regions (Greenland and Iceland) and Hawaii (Han *et al*., 2019). Tabanids can be found in a wide range of habitats, from oases in deserts to water-rich regions like salt marshes and rainforests (Smit, 2019). Although most species of tabanid adults are known to feed during the day, certain species can exhibit crepuscular or nocturnal activities. Male tabanids exclusively feed on flower nectar, whereas females consume both nectar and a blood meal for oviposition from a wide variety of animals, including human beings, wild animals, amphibians, reptiles, and birds (Prisca *et al*., 2021).

Horseflies are renowned globally for their persistent and painful biting behaviour and are considered menacing pests for humans, wildlife and livestock because of their painful and irritating bite (Votýpka *et al*., 2019). They also serve as biological or mechanical vectors that transmit essential medical and veterinary pathogens including helminths, protozoa, bacteria, and viruses (Bonnet *et al*., 2020). In a recent study, Taioe (2017) detected DNA of *Besnoitia besnoiti*, *Babesia/Theileria* spp. and *Trypanosoma evansi* from tabanid flies collected in Zambia. Furthermore, several bacterial genera of veterinary, medical and environmental importance were detected from tabanid flies [*Tabanus gratus, T. par, T. taeniola* and *Atylotus agrestis* and *A. diurnus*] collected from South Africa and Zambia using 16S rRNA metabarcoding (Taioe, 2017).

Despite their importance as plant pollinators and threat to human and animal health, they have been neglected subjects of scientific research, especially in southern Africa, with most studies from the region being taxonomic (Oldroyd 1952, 1954, 1957; Usher 1965, 1972; Chainey 1983, 1987; Morita, 2008). However, due to the generalized morphology of tabanids, whereby there are few male specimens, and genital characters do not exhibit much variation at the genus as well as species level, the taxonomy of Tabanidae below

the tribal level has proven difficult in recent decades (Lessard *et al*., 2013; Morita *et al*., 2016).

Africa has over 700 tabanid species that inhabit a variety of environments. They are found in the forest, savanna, montane, coastal, and desert biomes in the Afrotropical region (Williams *et al*., 2022). The first study to describe tabanid distribution in southern Africa was by Usher in 1972. The South African hematophagous flies, particularly horse flies, have not undergone significant revision for some decades. Usher (1972), with revisions made by Chainey (1983, 1987) and Morita (2008), claimed that South Africa contains a relatively wide variety of tabanids, with about 200 species having been described from the area. There are records of over 410 species of tabanid flies in southern Africa, with KwaZulu-Natal Province (KZN) having the highest species diversity in the country with 64 species (Usher 1972; Nevill *et al*., 1979).

Tabanids are still one of the least researched Dipteran families (Mullens, 2019). This limited amount of research highlights a significant gap in understanding this diverse group of insects. Furthermore, limited studies focus on phylogenetic relationships, parasites, and symbiotic fauna harboured by tabanid flies conducted globally (Wiegman et al., 2000 and 2003). To update the modern global distribution of Tabanidae, this study sought to review data from published studies of Tabanid flies. *Tabanus par* is a hematophagous insect of the family Tabanidae, regarded as among the most significant brachycerans (Diptera: Brachycera) (Mazibuko, 2018). Recent studies on tabanid flies in Mozambique, South Africa and Zambia indicated that *T. par* appears in higher abundance than other species (Mulandane *et al*., 2020; Taioe *et al*., 2017). Hence, the current study focuses on the morphological and genetic characterization of South African *T. par* and further investigates its associated microbiota.

1.2 Justification of the study

Tabanus is one of the most prominent genera in the order Diptera and it hosts the oldest blood-feeding pestiferous flies (Mullens, 2019). However, categorisation needs to be revised due to unstable physical traits, tremendous variability, cryptic species, and a shortage of knowledge and study within the group. Correctly identifying the species of the tabanids is essential for preventing harm from the family and elucidating their vectorial involvement in spreading infectious diseases. The tabanid fauna of South Africa is

currently fragmented, and there are no definitive and reliable checklists. The absence of detailed surveys and recordings of South African tabanids limits research into their ecology, zoogeography, vector-host interactions, and disease vector importance (Mugasa et al., 2018). This limited amount of research highlights a significant gap in understanding this diverse group of insects. Despite their ecological and economic importance, such as their role as vectors of diseases in livestock, their biology, behaviour, and ecology still need to be explored, particularly in the African continent.

Molecular approaches such as metagenomics provide a distinct advantage in identifying and understanding the gut microbiota of Diptera flies, including tabanids (Bellekom, 2023). There needs to be more consolidated data for the global distribution of tabanid flies. Hence, through a systematic review and meta-analysis, the current study has unitised published articles to highlight the global distribution of tabanid flies. Furthermore, this study sought to morphologically and genetically characterise South African *T. par* fly species, the most abundant species in southern Africa, and document its microbial communities using a metagenomics approach.

1.3 Aim

This study established global distribution as well as morphological and genetic characteristics of Tabanid fly species including the microbiomes they harbour.

1.4 Objectives

1. To investigate the global prevalence's and distribution of tabanid flies by systematic review and meta-analysis of published literature.
2. To determine morphological and genetic characteristics of *Tabanus par* from selected localities in South Africa.
3. To identify microbiomes harboured by *Tabanus par* by using shotgun metagenomics.
4. Genomic diversity and distributions of bacteriophages harboured by *Tabanus par*:
5. To determine genetic characteristics of *Spiroplasma* species harboured by *Tabanus par* using metagenomics method.

1.5 Thesis outline

Chapter 1 – General introduction: This chapter contains the background of the study, followed by justification, aim, objectives and thesis outline.

Chapter 2 – Literature review: This is a detailed review of tabanid flies focusing on taxonomy, their global distribution and general life cycle, as well as morphological and genetic methods used for identification and characteriztion .

Chapter 3 – Investigate the global prevalence's and distribution of tabanid flies by systematic review and meta-analysis of published literature: This chapter is a systematic review conducted with data from published articles better to comprehend the occurrence and distribution of tabanid flies globally. Subgroup analysis determined the factors influencing distribution studies, including identification techniques, types of traps used, study period, geographical location and number of tabanid flies.

Chapter 4 – Determine morphological and genetic characteristics of *Tabanus par* from selected localities in South Africa: This chapter contains information on the morphology, morphometrics and molecular identification of *T. par.*

Chapter 5 – Identify microbiomes harboured by *Tabanus par* by using shotgun metagenomics: This chapter presents a comprehensive analysis of the microbiome associated with *T. par*, focusing on the taxonomic classification of bacteria, fungi, viruses, and plasmids identified in both the whole body and legs of the fly.

Chapter 6 – Genomic diversity and distributions of bacteriophages harboured by *Tabanus par*: This chapter reports on mining of bacteriophages within *T. par* using a metagenomics approach. It further describes the taxonomic classification of complete phage genomes and their intergenomic similarities.

Chapter 7 – Determine genetic characteristics of *Spiroplasma* species harboured by *Tabanus par* using metagenomics method: This chapter reports on bacteria within *T. par* using metagenomic approach. It describes and characterizes whole genome sequences of *Spiroplasma* bacteria.

1.6 References

1. Bellekom, B. (2023). *Community and network ecology of biting Diptera-host interactions* (Doctoral dissertation, University of Oxford).
2. Bonnet, S.I., Bouhsira, E., De Regge, N., Fite, J., Etoré, F., Garigliany, M.M., Jori, F., Lempereur, L., Le Potier, M.F., Quillery, E. & Saegerman, C. (2020). Putative role of arthropod vectors in African swine fever virus transmission in relation to their bio-ecological properties. *Viruses*, *12*(7), p.778.
3. Chainey, J.E. (1983). Afrotropical Tabanidae (Diptera): the genus *Philoliche* Wiedemann, subgenus Ommatiosteres Enderlein. *Annals of the Natal Museum*, *25*(2), pp.453-474.
4. Chainey, J.E. (1987). Afrotropical Tabanidae (Diptera): the genus *Rhigioglossa* Wiedemann, 1828 (including *Mesomyia* Macquart, 1850, as a subgenus). *Annals of the Natal Museum*, *28*(1), pp.137-159.
5. Han, Y., Ye, X., Feng, C., Zhang, K., Shih, C. & Ren, D. (2019). Diptera–true flies with two wings. *Rhythms of insect evolution: evidence from the Jurassic and Cretaceous in northern China*, pp.497-553.
6. Lessard, B.D., Cameron, S.L., Bayless, K.M., Wiegmann, B.M. & Yeates, D.K., (2013). The evolution and biogeography of the austral horse fly tribe Scionini (Diptera: Tabanidae: Pangoniinae) inferred from multiple mitochondrial and nuclear genes. *Molecular phylogenetics and Evolution*, *68*(3), pp.516-540
7. Mazibuko, X. (2018). *Species composition of and detection of haemoparasites in horse-flies (Diptera:* Tabanidae*) in south-eastern Kruger National Park*. University of Pretoria (South Africa).
8. Morita, S., 2008. A revision of the *Philoliche aethiopica* species complex (*Diptera*: Tabanidae). *African Invertebrates*, *49*(1), pp.129-158.
9. Morita, S.I., Bayless, K.M., Yeates, D.K. & Wiegmann, B.M. (2016). Molecular phylogeny of the horse flies: a framework for renewing tabanid taxonomy. *Systematic entomology*, *41*(1), pp.56-72.
10. Mugasa, C.M., Villinger, J., Gitau, J., Ndungu, N., Ciosi, M. & Masiga, D. (2018). Morphological re-description and molecular identification of Tabanidae (Diptera) in East Africa. *ZooKeys*, (769), p.117.

11. Mulandane, F.C., Snyman, L.P., Brito, D.R., Bouyer, J., Fafetine, J., Van Den Abbeele, J., Oosthuizen, M., Delespaux, V. & Neves, L. (2020). Evaluation of the relative roles of the Tabanidae and Glossinidae in the transmission of trypanosomosis in drug resistance hotspots in Mozambique. *Parasites & Vectors*, *13*, pp.1-16.
12. Mullens, B.A. (2019). Horse flies and deer flies (Tabanidae). In *Medical and veterinary entomology* (pp. 327-343). Academic Press.
13. Nevill, E.M. (1979). The experimental transmission of *Parafilaria bovicola* to cattle in South Africa using *Musca* species (subgenus *Eumusca*) as intermediate hosts.
14. Oldroyd H. (1952). The Horseflies of the Ethiopian Region. I. *Haematopota* and *Hippocentrum*. British Museum of Natural History, London.
15. Oldroyd H. (1954). The Horseflies of the Ethiopian Region. Vol. II. *Tabanus* and Related Genera. British Museum of Natural History, London.
16. Oldroyd H. (1957). The Horseflies of the Ethiopian Region. Vol. III. Subfamiles Chrysopsinae, Scepsidinae and Pangoniinae, and a Revised Classification. British Museum of Natural History, London.
17. Prisca, O.M.A., Franck, M., Roland, Z.K.C., Armel, K.A., Lendzele, S.S., Pamela, M.A., Lydie, A.Y.G., Lebel, T.J., Gustave, S., Bertrand, M.B. & François, M.J., (2021). Ecological aspects of Tabanids (Diptera: Tabanidae) in a Gabonese cattle ranch. The Wonders of Diptera—Characteristics, Diversity, and Significance for the World's *Ecosystems*.Smit, A. (2019). Abundance, composition and barcoding of Tabanidae in Kruger National Park and screening for *Besnoitia besnoiti*. University of Pretoria (South Africa).
18. Taioe, M.O., Motloang, M.Y., Namangala, B., Chota, A., Molefe, N.I., Musinguzi, S.P., Suganuma, K., Hayes, P., Tsilo, T.J., Chainey, J. & Inoue, N. (2017). Characterization of tabanid flies (Diptera: Tabanidae) in South Africa and Zambia and detection of protozoan parasites they are harbouring. *Parasitology*, 144(9), pp.1162-1178.
19. Taioe, M.O. (2017). *Molecular characterization of horse flies (Diptera: Tabanidae) and determination of their role in transmission of haemoparasites in southern Africa* (Doctoral dissertation, North-West University (South Africa), Potchefstroom Campus).

CHAPTER 2 LITERATURE REVIEW

2.1 Taxonomy of Diptera

The phylum Arthropoda is estimated to consist of between 5-10 million species, accounting for over 80% of all known extant animal species. More species have been described in the insect subgroup of arthropods than in any other taxonomic group. The insect orders with the most diversity include true flies (Diptera), over 100 000 known species. The family Tabanidae is divided into three subfamilies (Fairchild, 1969). The subfamily Pangoniinae is considered the ancestral group consisting of the genus *Stonemyia* and *Goniops*, many of which are not known to feed on blood. Most economically important tabanids are members of the subfamily Chrysopsinae and Tabaninae. Tabanid flies belong to the subfamily Chrysopsinae, also known as deerflies, consisting of the genus *Chrysops*, which includes more than 80 species. The most evolved organisms are Tabaninae members. This subfamily consists of the horse flies, represented by *Tabanus* and *Hybomitra* as well as *Haematopota* (Mullens, 2019) (Figure 2-1).

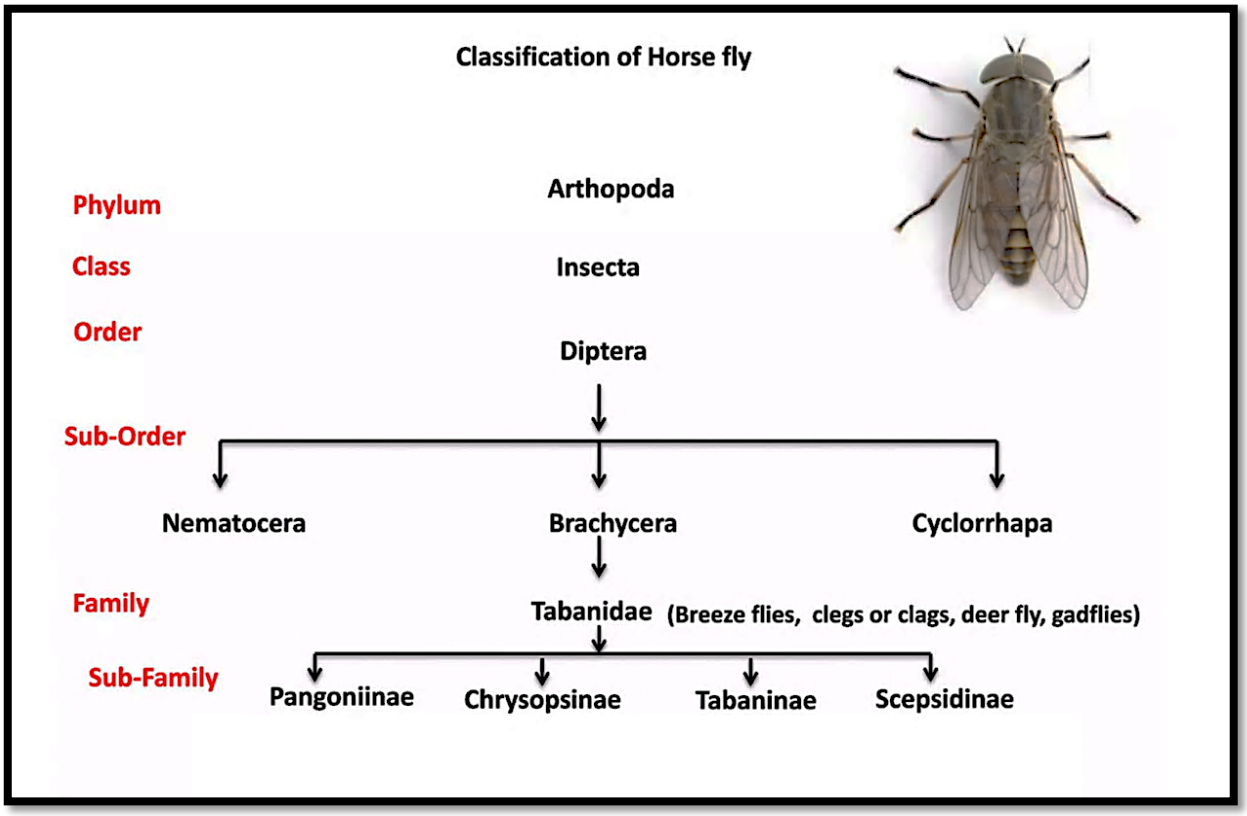


Figure 2-1: Tabanidae hierarchy within order Diptera.

2.2 Global distribution of Tabanidae

The Tabanidae family of lower brachyceran is a sizable family of flies with about 4300 species and subspecies in 133 genera worldwide, of which roughly 400 occur in southern Africa (Taioe *et al*., 2017). Globally, more than 4000 species of tabanids have been described (Bitome *et al*., 2017). In Africa, there are more than 700 species that live in a variety of settings (Kirk-Spriggs and Stuckenberg, 2009). Tabanids exhibit seasonal fluctuations and interact with their environment like any other insect family. However, the factors affecting their population may vary because they depend on hosts for their blood meal, which is necessary to produce eggs. Horseflies are seasonal insects that live in various habitats at different latitudes and altitudes (Denlinger, 2022). The diversity of tabanids is most significant in the tropics, while wet temperate regions frequently feature diverse fauna. Only about 800 species are recorded from the Afrotropical region (Chainey, 2017), of which 400 are found in southern Africa (Taioe *et a*l., 2017). They have managed to populate isolated islands like the Galápagos and the Melanesian Archipelago, which are found on every continent except Antarctica. Even at 60°N latitude, several species have large seasonal populations, but they diminish above the tree line (Mazibuko, 2018). The abundance of tabanids in an environment depends on suitable environmental conditions such as temperatures between 15 - 25°C, good luminosity, high relative humidity and availability of vertebrate hosts (Prisca *et al*., 2021).

2.3 General life cycle of tabanid flies

Crosskey *et al*., (1980) and Foil, (1989) have thoroughly studied tabanid fly reproduction in South Africa. Female tabanids are classified as autogenous, partially autogenous, or non-autogenous because they frequently require a blood meal to develop eggs. They have a complete metamorphosis cycle [egg, larvae, pupae, and adult] (Figure 2-2), and typically only have one generation annually. Breeding primarily occurs in damp environments and on semi-aquatic substrates, while some species are terrestrial, such as *Tabanus abactor*, *T. sulcifrons*, and *T. subsimilism*. Embryogenesis takes between two and 21 days, depending on the species and environmental conditions (Taioe *et al*., 2017).

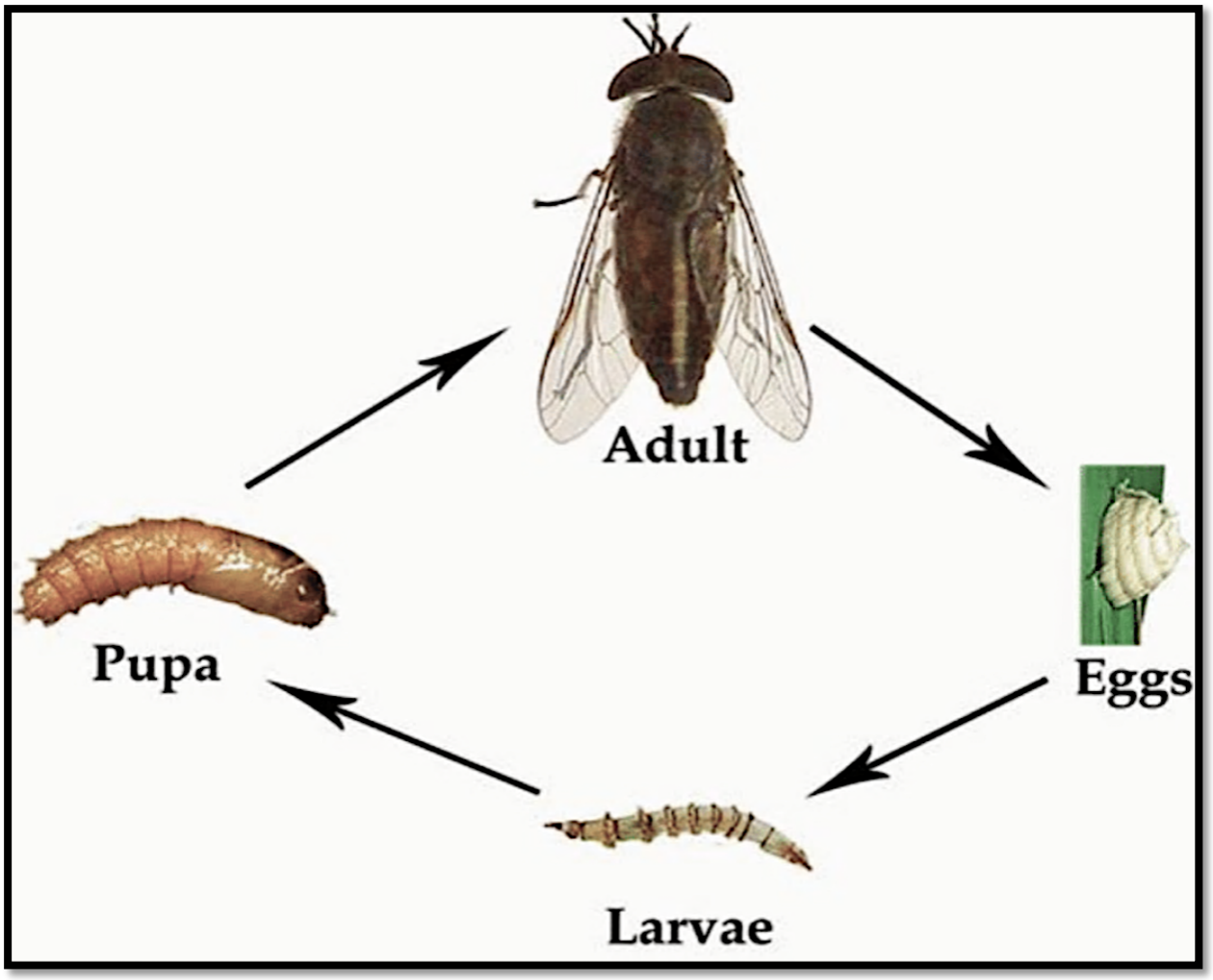


Figure 2-2: Life cycle of tabanid fly (adapted from Bautista-Garfias *et al*., 2021).

2.3.1 Eggs

The process of mating begins in the air and is completed on the ground. The female deposits egg mass produced in batches of 100 to 1000 eggs, on surfaces, such as protruding pebbles, sticks, or aquatic plants (Foil and Hogsette, 1994). Eggs are frequently covered in a glossy or chalky secretion that aids in water protection. All tabanid eggs are white to light after oviposition, then eventually darken to hues of brown or black. The length of an egg ranges from 1 to 2.5 mm, and it has a cylindrical shape. Depending on the ambient temperature, embryonic development is reported to hatch in 2-12 days. Almost all mass eggs hatch at once, and the larvae fall to the water or nearby moist soil. The majority of the females die within the first two days of oviposition, but the survivors eventually lay three small batches of eggs (King, 1910).

2.3.2 Larvae

Larvae employ a hatching spine to escape the egg casing. The emerging larvae may fall or crawl into the water. *Chrysops* spp., known as "hydrobionts," are found in locations with a lot of water. As "hemi-hydrobionts," *Tabanus* spp. favour dryer substrates. The adult larva is round when fully expanded with 11 to 12 additional segments. White in colour with a hint of grey is the colour. Dark brown to black, serrated mandibles (Husseneder *et al*., 2022). During this time, tabanid larvae feed on other insect larvae, including their species, annelids. They are both predators and cannibals. The *Chrysops*' larvae feed upon the organic matter or decaying animals in the soil. The larva penetrates the upper 2.5 to 5.0 cm of the soil, which is drier to create underground hollow mud cylinders. By doing this, the pupae are shielded from exposure when cracks develop in the mud and are also allowed to burrow deeper into the ground away from the surface when temperatures are too high (Foil and Hogsette, 1994). Within two days after moving to the surface, the pupal stage is reached. Under adverse conditions, the larva may hibernate.

2.3.3 Pupa

Depending on the climate, typically, the pupal stage lasts two to three weeks. The pupae have attached wing and leg cases, and are brown in colour, taper posteriorly, and are rounded anteriorly. Each abdominal segment is encircled by a row of spines. At the tip of the abdomen, there is a six-pointed "aster" structure that represents the pupa (Sawar, 2020).

2.3.4 Adult

The adult fly emerges from the pupal case via a slit located along the thorax of the case and can fly for several hours after drying and hardening for a few hours. In the majority of species, males emerge earlier than females, after pupation at a sex ratio of 1:1. Normally, it takes one to three weeks for a larva to complete metamorphosis. Adult horse flies have green or black eyes and are black to dark brown. In contrast to horse flies, adult deer flies have apical spurs on the hind (Sawar, 2020).

2.4 Tabanids as vectors of pathogens

Tabanids can significantly affect animals and people due to their massive size, extremely prolific reproduction, and seasonal abundance. According to Chainey, (1983), Nevill (1979), Esterhuizen, (2006), Baldacchino *et al*., (2014), and Hornok *et al*., (2014), tabanid flies are the leading carriers of most pathogenic bacteria and viruses that affect both humans and animals. High populations of these flies have a substantial negative economic impact on outdoor recreation, tourism, and livestock production. Direct negative impacts of these insects on livestock or wildlife include skin lesions caused by persistent and painful bites hundreds of bites per day per animal, discomfort, secondary infections), blood loss (up to 700 mg/insect in horseflies), visual discomfort, and cacophony (Gerry and Murillo, 2019). The indirect effect is caused by pathogens that are transmitted through the blood or lymph of their hosts. There has previously been a study of published accounts of tabanid transmission of pathogenic agents of animals, including humans (Mazibuko, 2018).

These pathogens include various viruses such as the equine infectious anemia virus (EIA) and the (BLV), Lumpy Skin Disease Virus (LSDV) and West Nile fever virus (Hennekeler *et al.,* 2008). Nevertheless, the transient and highly seasonal nature of these insects complicates the demonstration of causality. As a result, their specific contribution to disease transmission remains poorly documented. This is not true for severe and lethal diseases such as EIA. This is due to the correlation between peaks in tabanid populations and outbreaks of EIA, as well as the geographic distribution of the disease being associated with areas of high tabanid populations.

Research has shown that tabanids can harbour bacteria, resulting in diseases such as anaplasmosis and Q fever in livestock (Taioe *et al*., 2017; Keita *et al*., 2020). Additionally, other bacteria associated with tabanids, include *Anaplasma Marginale*, *Bacillus anthracis*, Brucella spp., *Clostridium chauvoei*, *Pasteurella multocida, Erysipelothryx rhusiopathiae* or *Ehrlichia risticii* through experimental transmission (Krinsky, 1976). Other pathogens transmitted by tabanids include protozoan pathogens such as apicomplexan *Besnoitia besnoiti* and various species of *Trypanosoma*, such as *T. brucei*, *T. congolense* (Savannah), with exception of T. *Theileri* (Taioe *et al.,* 2017). While Filarial nematodes reported in tabanids include *Elaeophora schneideri.* Furthermore, opportunistic

pathogens in tabanids pose a significant risk, particularly for individuals with compromised immune systems, as these pathogens can complicate diagnosis and treatment (Qiu *et al.*, 2024). Gaining insights into tabanids' ecological and behavioural traits, such as their feeding patterns and host preferences, is crucial for understanding their involvement in the spread of pathogens (Sontigun *et al.*, 2022).

2.5 Microbiome profile of tabanid flies

The microbiome characteristics of tabanids, especially horseflies, are becoming increasingly important as they have implications for understanding their ecological functions and potential effects on animal and human health. Recent research has uncovered important information regarding the composition and variety of the microbiota linked to these insects (Taioe, 2017, Litov *et al.,* 2023, Qiu *et al.,* 2024). These microorganisms exist in close symbiotic interactions with multicellular eukaryotic hosts, including nematodes, insects, plants, animals, and humans (Mafuna, 2019). The community of microbiota is primarily made up of microorganisms such as bacteria, fungi, archaea, protists, and viruses. The nature of the symbiotic relationships between microbes and their hosts can vary and include forms like commensalism, mutualism, parasitism, and potentially lethal pathogenic infections (Moya *et al.,* 2016; Sinnathamby *et al.,* 2018).

Most research on the microbiome of tabanids has focused on the midgut of tabanid species, where Tenericutes are more prevalent, while the malpighian tubules exhibit a greater presence of Bacteroides. This indicates a functional differentiation of microbiota among various anatomical locations, which could correspond to horseflies' distinct metabolic and physiological requirements in those areas (Qiu *et al*., 2024). Notably, the ovarian microbiota studies of horseflies have gained attention, which are considered to display a higher microbial diversity than the other organs, implying a specialized function of the microbiota in reproductive processes, thereby highlighting the significance of microbiota in the lifecycle of these insects (Qiu *et al*., 2024).

Due to the diverse habitats in which tabanids are frequently found, their microbiomes are likely to differ considerably based on environmental conditions and the availability of hosts. Cárdenas *et al.,* (2020) observed that tabanid flies' activity patterns and species diversity can vary with elevation and climate, which may also affect their associated

microbiome. Consequently, it is crucial to consider the role of the external features of flies; it can be inferred that the feet, wings, mouthparts, and other surface areas of a fly function as the main route for the mechanical spread of microbes (Junqueira *et al.,* 2017). The significance of investigating the microbiome of tabanid flies is rooted in its ability to improve the comprehension of these insects' ecological and epidemiological functions. By clarifying the relationships between tabanids and their microbial populations, scientists can more accurately evaluate the dangers linked to disease transmission and create more efficient control strategies (Qiu *et al*., 2024).

2.6 Traps for the capture of tabanid flies

Horseflies are notoriously difficult to control (Sarwar, 2020). Traps are considered the most efficient way for vector control and surveillance. Different trap types can capture varying numbers of individuals and species of horseflies due to differences in design and use of colour, movement, heat and light intensity or polarization (Otártics *et al.,* 2019). To increase the attractiveness of traps to horseflies, natural and artificial olfactory attractants like octanol, cow urine, and phenols are used; nevertheless, species-to-species variations in relative responses to various attractants may occur (Mihok and Mulye, 2010).

Conventional tabanid traps come in a variety of forms. These include the box trap (Wall and Doane, 1980), the canopy trap (Catts, 1970), and the Manitoba trap (Thorsteinson, 1958; Thorsteinson *et al.,* 1965), as depicted in Figure (2-3). Tabanids have also been caught using electrified grids (Phelp and Vale, 1978). Among the odour baits are ammonia (Hribar *et al.,* 1992), 1-octen-3-ol (French and Kline, 1989), and carbon dioxide (Jackson and Wilson, 1966). However, different trap types vary in their ability to capture deer flies or other tabanid subgroups but can provide diverse information about the behaviour of individual species.

A study by Skvarla *et al.* (2021) compared Malaise trap to aerial netting and reported that the latter favoured *Chrysops* spp. While Malaise traps offered a good indicator of overall tabanid diversity and species evenness. While Mihok, (2002), reported that Nzi, Ngu and H-traps are specifically designed for *Glossina* species, and Nzi-trap is designed especially for hematophagous flies, including tabanids and *Stomoxys* species. The Nzi trap has been reported to be highly effective in African regions, North America, Australia and parts

of Europe (Mazibuko, 2018). Different types of traps have been modified over the years to improve their efficiency in capturing diverse kinds of flies (Kappmeier *et al.,* 2007).



Figure 2-3: Trapping methods for tabanid flies: A)) Broce trap, (B) Malaise trap, (C) Canopy trap, (D) Blue target, (E) Vavoua trap, (F) Nzi trap, (G) Manitoba trap, (H) H-trap, and (I) Horizontally polarizing liquid trap (Baldacchino *et al.*, 2018).

2.7 Morphological and molecular identification of tabanids

Tabanids are stout-bodied and consist of several morphological traits (Nour *et al.*, 2017). Traditionally, the identification of horse fly species has primarily depended on their morphological traits to distinguish between different species. Important characteristics utilized for this identification encompass structures of the head, features of the callus, antennae, eyes, frons, and beard, as well as the coloration and patterns observed on the body, legs, and wings (Changbunjong *et al.*, 2025). Mouth parts have also been used as a classic identification method, due to diverse adaptations in their proboscis structure, crucial for their feeding habits. The morphology of the antennae also acts as an important distinguishing feature, frequently exhibiting unique Additionally, recent updates have produced detailed identification keys that categorize different species of Tabanid flies, often featuring morphological descriptions alongside habitat information, which is crucial due to the similarities in physical traits among various species (Mugasa *et al.*, 2018). While this method is straightforward and cost-effective, as it does not require advanced scientific equipment, it comes with significant limitations. Precise identification necessitates considerable expertise to recognize subtle variations, and the likelihood of mistakes is elevated, especially when specimens are compromised during field collection or due to natural factors. Incorrect identifications not only lead to misinformation but also obstruct effective disease management, as they cause inaccuracies in horse fly population control (Changbunjong *et al.*, 2018a; Rodrigues *et al.*, 2024).

Molecular genetic research has deepened our insights into Tabanid diversity and their evolutionary relationships. A thorough phylogenetic study that includes molecular data supports the existence of considerable genetic variation within this family of flies. Recent research has highlighted how genetic analyses illuminate the connections among different Tabanus species, creating several unique clades that shed light on the taxonomic and evolutionary background of the family (Taioe *et al.*, 2017). Furthermore, barcoding analyses using mitochondrial DNA have validated the efficiency of molecular techniques in differentiating Tabanid species, offering valuable information on their phylogeography and genetic distribution in various regions (Cywinska *et al.*, 2010).

Currently, morphological traits are used to identify tabanid fly species. The lack of fixed taxonomic features in horsefly species constraints on genital anatomy for

identifications below the tribal level, and usage of variable traits like body and wing colour make species identification hard in the absence of a professional taxonomist. With tabanids being sexually dimorphic, separated keys are needed to identify the male and female species. Considering horseflies have identification problems when using conventional taxonomic methods, molecular tools have now become pivotal in correctly identifying species (Banerjee *et al*. 2015). Utilizing molecular methods also reveals genetic diversity, which can be used to resolve species complexes and is considered more effective (Banerjee *et al*. 2015). Thus, DNA barcoding has gained attention as a method to address taxonomic difficulties posed by most invertebrates, particularly tabanids, at the species level (Banerjee *et al*., 2015). The mitochondrial gene encoding for cytochrome c oxidase subunit 1 (*COX1*) was amplified to a length of 658 base pairs to serve as the basis for the molecular identification of species proposed by Hebert *et al* (2018). Regardless of genital or developmental phase, “barcoding”, allows for the distinction of a wide variety of species and has the advantage of integrating taxonomic and genomic data. Nevertheless, there are certain restrictions to molecular identification, such as the insufficiency of sequences in the GenBank DNA database, notably about the Tabanids.

Molecular techniques have a distinct advantage in resolving identification challenges that arise in many other taxa, ranging from plants to mammals as well as tabanids. Similar research, nevertheless, is severely lacking not just in South Africa but also in other African nations (Taioe *et al*., 2017) Additionally, very little research has been done on the horsefly species higher-level phylogenetic relationships (Wiegman *et al*., 2000). To date, only a study of Taioe *et al*., (2017) has genetically characterized tabanid flies from South Africa and Zambia but did not conduct a detailed population genetic study or genome sequencing. A study conducted by Fu *et al*. (2021), where they compared the mitochondrial genomes of 6 tabanid flies in China, showed that the mitochondrial genome of horseflies (Diptera: Tabanidae) is highly conserved among different species and that members of this genera are paraphyletic. DNA barcoding has also been successfully used to identify different species in the Tabanidae family. For example, a research project carried out in Canada gathered tabanid specimens from several provinces, and DNA barcoding was used to differentiate 42 species from five genera, as reported by Cywinska *et al*. (2010). This approach validated the species' identities and uncovered genetic

variability within populations, emphasizing the importance of molecular tools in taxonomic investigations. Understanding the evolutionary connections between tabanid species and their ability to thrive in various ecological environments relies on crucial molecular details.

2.8 References

1. Baldacchino, F., Desquesnes, M., Duvallet, G., Lysyk, T. & Mihok, S. (2018). Veterinary importance and integrated management of Brachycera flies in dairy farms. *In Pests and vector-borne disease in the livestock industry* (pp. 1765-1771). Wageningen Academic Publishers.
2. Baldacchino, F., Desquesnes, M., Mihok, S., Foil, L.D., Duvallet, G. & Jittapalapong, S. (2014). Tabanids: neglected subjects of research, but important vectors of disease agents. Infection, *Genetics and Evolution*, 28, pp.596-615.
3. Banerjee, D., Kumar, V., Maity, A., Ghosh, B., Tyagi, K., Singha, D., Kundu, S., Laskar, B.A., Naskar, A. & Rath, S. (2015). Identification through DNA barcoding of Tabanidae (Diptera) vectors of surra disease in India. *Acta Tropica*, 150, pp.52-58.
4. Bitome-Essono, P.Y., Ollomo, B., Arnathau, C., Durand, P., Mokoudoum, N.D., Yacka-Mouele, L., Okouga, A.P., Boundenga, L., Mve-Ondo, B., Obame-Nkoghe, J. & Mbehang-Nguema, P. (2017). Tracking zoonotic pathogens using blood-sucking flies as' flying syringes. *eLife*, *6*, p.e22069.
5. Bonnet, S.I., Bouhsira, E., De Regge, N., Fite, J., Etoré, F., Garigliany, M.M., Jori, F., Lempereur, L., Le Potier, M.F., Quillery, E. & Saegerman, C. (2020). Putative role of arthropod vectors in African swine fever virus transmission in relation to their bio-ecological properties. *Viruses*, 12(7), p.778.
6. Cárdenas, P.A., Garcés, D., Prado-Vivar, B., Flores, N., Fornasini, M., Cohen, H., Salvador, I., Cargua, O. & Baldeón, M.E. (2020). Effect of *Saccharomyces boulardii* CNCM I-745 as complementary treatment of *Helicobacter pylori* infection on gut microbiome. *European Journal of Clinical Microbiology & Infectious Diseases*, *39*, pp.1365-1372.
7. Catts, E.P. (1970). A canopy trap for collecting Tabanidae.
8. Cywinska, A., Hannan, M.A., Kevan, P.G., Roughley, R.E., Iranpour, M. & Hunter, F.F. (2010). Evaluation of DNA barcoding and identification of new haplomorphs in

Canadian deerflies and horseflies. *Medical and Veterinary Entomology*, *24*(4), pp.382-410.

9. Chainey, J. (2017). Tabanidae (horse flies, deer flies and clegs). *Manual of Afrotropical Diptera*, *2*, pp.893-913.
10. Chainey, J.E. (1983). Afrotropical Tabanidae (*Diptera*): the genus *Philoliche* Wiedemann, subgenus *Ommatiosteres* Enderlein. *Annals of the Natal Museum*, 25(2), pp.453-474.
11. Changbunjong, T., Weluwanarak, T., Laojun, S. and Chaiphongpachara, T. (2025). Species classification of *Tabanus* (Diptera: Tabanidae) in Western Thailand: Integrating DNA barcoding and modern morphometrics. *Current Research in Parasitology & Vector-Borne Diseases*, *7*, p.100243.
12. Bautista-Garfias, C.R., Castañeda-Ramirez G.S., Torres-Acosta, J.F., Salinas-Estrella, E., Moshin, M. & Marcelino, L.A. (2021). Fly borne diseases In animals. *Parasitic disease*.
13. Crosskey, R.W., Cogan, B.H., Freeman, P., Pont, A.C., Smith, K.G.V. & Oldroyd, H. (1980). *Catalogue of the Diptera of the b Region*. British Museum (Natural History), Cromwell Road, London SW7 5BD.
14. Denlinger, D.L. (2022). *Insect diapause*. Cambridge University Press.
15. Esterhuizen, J. (2006). Seasonal abundance of horse flies (Diptera: Tabanidae) from two conservation areas in northeastern KwaZulu-Natal Province, South Africa. *African Entomology*, *14*(2), pp.395-397.
16. Fairchild, G.B. (1969). Climate and the phylogeny and distribution of Tabanidae. *Bulletin of the ESA*, *15*(1), pp.7-11.
17. Foil, L.D., anHogsette, J.A. (1994). Biology and control of tabanids, stable flies and horn flies*. Scientific and Technical Review of the Office International des Epizootie*s. 13(4): 1125-1158.
18. Foil, L.D., Leprince, D.J. & Church, G.E. (1989). Changes in the parity rate of *Tabanus fuscicostatus* (Diptera: Tabanidae) populations associated with controlling available hosts. *Journal of medical entomology*, *26*(4), pp.306-310.
19. French, F.E. & Kline, D.L. (1989). I-Octen-3-ol, an effective attractant for Tabanidae (*Diptera*). *Journal of Medical Entomology*, 26(5), pp.459-461.

20. Fu, Y.T., Zhang, Y., Xun, Y., Liu, G.H. & Zhao, Y. (2021). Characterization of the complete mitochondrial genomes of six horseflies (Diptera: Tabanidae). *Infection, Genetics and Evolution*, *95*, p.105054.
21. Gerry, A.C. & Murillo, A.C. (2019). Promoting biosecurity through insect management at animal facilities. In Biosecurity in animal production and veterinary medicine: from principles to practice (pp. 243-281). Wallingford UK: CABI.
22. Hebert, P.D., Ratnasingham, S., Zakharov, E.V., Telfer, A.C., Levesque-Beaudin, V., Milton, M.A., Pedersen, S., Jannetta, P. & DeWaard, J.R. (2016). Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702), p.20150333.
23. Hribar L.J., Leprince, D.J. & Foil, L.D. (1992). Ammonia as an attractant for adult *Hybomitra lasiophthalma* (*Diptera*: Tabanidae). *Journal of Medical Entomology*, 29(2), pp.346-348.
24. Van Hennekeler, K., Jones, R.E., Skerratt, L.F., Fitzpatrick, L.A., Reid, S.A. & Bellis, G.A. (2008). A comparison of trapping methods for Tabanidae (Diptera) in North Queensland, Australia. *Medical and Veterinary Entomology*, *22*(1), pp.26-31.
25. Hornok, S., Basso, W., Fedák, A., Baska, F., Dencső, L., Abonyi, T. & Dénes, B. (2014). Epidemiological investigations of bovine besnoitiosis in Hungary. *Parasites & Vectors*, *7*(Suppl 1), p.O32.
26. Husseneder, C., Bhalerao, D.R. & Foil, L.D. (2022). Was the decline of saltmarsh tabanid populations after the 2010 oil spill associated with change in the larval food web?. *Ecosphere*, *13*(7), p.e4157.
27. Jackson, J.O. & Wilson, B.H. (1966). Parasitism of tabanid (*Diptera*: Tabanidae) eggs in four areas of southern Lousiana. *Annals of the Entomological Society of America*, 59(2), pp.261-267.
28. Junqueira, A.C.M., Ratan, A., Acerbi, E., Drautz-Moses, D.I., Premkrishnan, B.N., Costea, P.I., Linz, B., Purbojati, R.W., Paulo, D.F., Gaultier, N.E. & Subramanian, P. (2017). The microbiomes of blowflies and houseflies as bacterial transmission reservoirs. *Scientific reports*, *7*(1), p.16324.
29. Kappmeier Green, K., Joint, F., Potgieter, F., Vreysen, M., Robinson, A. & Hendrichs, J. (2007). A strategy for an area-wide control campaign with an SIT (sterile insect technique) component to establish a tsetse- (*Glossina austeni* and *Glossina*

brevipalpis) free South Africa. *The Ondertepoort Journal of Veterinary Research*, 67, : pp.15-16.

30. King, H.H. (1910). Some observations on the bionomics of *Tabanus par*, Walker, and *Tabanus taeniola*, Pal. de Beauv. *Bulletin of Entomological Research*, 1(2), pp.99-104.
31. Kirk-Spriggs, A.H., & Stuckenberg, B.R. (2009). Afrotropical Diptera—rich savannas, poor rainforests. In: Bickel, D.J., Pape, T., Meier, R. (eds.). *Diptera* diversity: status, challenges and tools Brill, Leiden, *The Netherlands*. pp. 155-196.
32. Keita, M.L., Medkour, H., Sambou, M., Dahmana, H. & Mediannikov, O. (2020). Tabanids as possible pathogen vectors in Senegal (West Africa). *Parasites & Vectors*, *13*, pp.1-15.
33. Litov, A.G., Belova, O.A., Kholodilov, I.S., Kalyanova, A.S., Gadzhikurbanov, M.N., Rogova, A.A., Gmyl, L.V. & Karganova, G.G. (2023). Viromes of Tabanids from Russia. *Viruses*, 15(12), p.2368.
34. Mafuna, T. (2019). *16SrRNA-based bacterial community profiling of Haemonchus contortus infecting Dohne Merino sheep using next generation sequencing* (Doctoral dissertation, North-West University (South Africa).
35. Mazibuko, X. (2018). *Species composition of and detection of haemoparasites in horse-flies (Diptera:* Tabanidae*) in south-eastern Kruger National Park* (Doctoral dissertation, University of Pretoria).
36. Mihok, S. & Mulye, H. (2010). Response of tabanids to Nzi traps baited with octenol, cow urine and ns phenols in Canada. *Medical and Veterinary Entomology*, 24(3), pp.266-272
37. Mihok, S. (2002). The development of a multipurpose trap (the Nzi) for tsetse and other biting flies. *Bulletin of Entomological Research*, 92(5), pp.385-403.
38. Moya, A. & Ferrer, M. (2016). Functional redundancy-induced stability of gut microbiota subjected to disturbance. *Trends in microbiology*, *24*(5), pp.402-413.
39. Mullens, B.A. (2019). Horse flies and deer flies (Tabanidae). In Medical and veterinary entomology (pp. 327-343). Academic Press.
40. Nevill, E.M. (1979). The experimental transmission of *Parafilaria bovicola* to cattle in South Africa using *Musca* species (subgenus Eumusca) as intermediate hosts.

41. Nour, H.I.C.M.M. and Ali, N.O., 2017. Morphological identification of horse flies (Diptera: Tabanidae) and estimation of their seasonal abundance in Al-Showak district, Gedaref State, Eastern Sudan.
42. Otártics, M.Z., Altbäcker, V., Solymosi, K., Mátics, R., Romvári, R. & Farkas, S., 2019. Efficacy of h-traps is affected by exposure to sunshine. *Natura Croatica*, *28*(1).
43. Phelps, R.J. & Vale, G.A. (1978). Studies on populations of *Glossina morsitans moarsitans* and *G. pallidipes (Diptera:* Glossinidae) in Rhodesia. *Journal of Applied Ecology*, pp. 743-760.
44. Prisca, O.M.A., Franck, M., Roland, Z.K.C., Armel, K.A., Lendzele, S.S., Pamela, M.A., Lydie, A.Y.G., Lebel, T.J., Gustave, S. & François, M.J. (2021). Ecological Aspects of Tabanids (Diptera: Tabanidae) in a Gabonese Cattle Ranch. The Wonders of Diptera-Characteristics, Diversity, and Significance for the World's Ecosystems.
45. Otártics, M.Z., Altbäcker, V., Solymosi, K., Mátics, R., Romvári, R. & Farkas, S., (2019). EFFICACY OF H-TRAPS IS AFFECTED BY EXPOSURE TO SUNSHINE. *Natura Croatica*, *28*(1).
46. Qiu, H.Y., Lv, Q.B., Wang, C.R., Ju, H., Luo, C.F., Liu, S.S., Na, M.H., Chang, Q.C. & Jiang, J.F. (2024). Microbiota profile in organs of the horseflies (Diptera: Tabanidae) in Northeastern China. *Frontiers in Microbiology*, *15*, p.1467875.
47. Sarwar, M. (2020). Typical flies: Natural history, lifestyle and diversity of Diptera. In Life Cycle and Development of Diptera. IntechOpen.
48. Sinnathamby, G., Henderson, G., Umair, S., Janssen, P., Bland, R. & Simpson, H. (2018). The bacterial community associated with the sheep gastrointestinal nematode parasite *Haemonchus contortus*. *PLoS One*, *13*(2), p.e0192164.
49. Skvarla, M.J., Larson, J.L., Fisher, J.R. & Dowling, A.P. (2021). A review of terrestrial and canopy Malaise traps. *Annals of the Entomological Society of America*, *114*(1), pp.27-47.
50. Taioe, M.O., Motloang, M.Y., Namangala, B., Chota, A., Molefe, N.I., Musinguzi, S.P., Suganuma, K., Hayes, P., Tsilo, T.J., Chainey, J. & Inoue, N. (2017). Characterization of tabanid flies (*Diptera*: Tabanidae) in South Africa and Zambia and detection of protozoan parasites they are harbouring. *Parasitology*, 144(9), pp.1162-1178.

51. Thorsteinson, A.J.(1958). The chemotactic influence of plant constituents on feeding by phytophagous insects. *Entomologia experimentalis et applicate*, 1(1), pp.23-27.
52. Thorsteinson, A.J.,Bracken, G.K. & Hanec, W. (1965). The orientation behaviour of horse flies and deer flies. (Tabanidae, Dipteria). LII. The use of traps in in the study of orientation of tabanids in the field. *Entomologia experimentalis et applicate*, 8(3), pp.189-192.
53. Wall, W.J. & Doane Jr, O.W. (1980). Large scale use of box traps to study and control saltmarsh greenhead flies (*Diptera*: Tabanidae) on Cape Cod, Massachusetts. *Environmental Entomology,* 9(4), pp.371-375.

CHAPTER 3 THE GLOBAL PREVALENCE'S AND DISTRIBUTION OF TABANID FLIES BY SYSTEMATIC REVIEW AND META-ANALYSIS OF PUBLISHED LITERATURE

3.1 Introduction

Tabanids (horseflies) are significant ectoparasitic arthropods responsible for substantial economic losses that result from the direct or indirect effects on their hosts (Chalchisa and, Bersissa,2023). They are hematophagous flies feeding on livestock and wildlife and are responsible for spreading infectious diseases of medical and veterinary importance that can be transmitted cyclically (Lendzele *et al*., 2019), mechanically (De Liberato *et al*., 2019) or by regurgitation (Sontigun *et al*., 2022). Male and female tabanid flies are adapted to feed on nectar, however, females also feed on blood for reproductive purposes. Apart from their role as blood-sucking pests, they play a positive role as pollinators of flora in nature (Karolyi *et al*., 2014).

Tabanid flies can be found in a wide range of habitats, including grasslands, alpine meadows, tropical rainforests, deciduous forests, wet savannas, and deserts, although they are most prevalent in warm, humid areas where they can effectively build their nests (Lucas *et al*., 2020). According to Guimares *et al*. (2017), the genera *Tabanus*, *Chrysops*, and *Haematopota* are all found in temperate, subtropical, and tropical climates, but they are absent from Australia and South America.

The abundance of horse flies is associated with ecological elements that arise from environmental factors (both biotic and abiotic). Species show a variety of patterns of population fluctuation connected to the environment, flora, and host blood food supply (Schowalter, 2022). A compromise of any of the factors will have a direct or indirect effect on tabanid flies leading to changes in the distribution and prevalence of tabanid populations (Essono *et al*., 2015). Thus, a thorough understanding of the phenology and ecology of the species at the regional level is necessary for the development of management methods against various species as well as strategies that reduce the risk of fly bites (Skendzic *et al*., 2021). The role of tabanids in the transmission of pathogens to domestic and companion animals has been studied extensively worldwide (Cywinska *et al*.,2010; Baldacchino *et al*., 2014; Morita *et al*., 2016; Al-Talatha *et al*., 2018; Mullens

2019). However, little is known about the global faunal composition and species diversity of biting tabanid flies. Foil (1989) reported that there were at most 50 ongoing research programs in systematics and distribution of tabanids, 15 in economic entomology, and 5 in disease transmission. Thus, tabanid distribution surveys are necessary to estimate their worldwide distribution (Sonenshine *et al*., 2006).

Studies conducted on tabanids are based on the identification and characterization of pathogens with robust approaches like polymerase chain reaction (PCR), next-generation sequencing (NGS), DNA barcoding, and metagenomics (Barbosa *et al*. 2017). However, considering the continuous change in environmental factors new information on the seasonal population dynamics of tabanids affecting both domestic and wild animals, as well as humans are crucial. Thus, this systematic review and meta-analysis focused on determining the global prevalence and distribution of tabanid flies.

3.2 Materials and methods

3.2.1 Review protocol

This systematic review was conducted according to the updated 2020 Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA, 2020) guidelines (Page *et al*., 2021). The first step was to search for all scientific publications to find data available for the global distribution and prevalence of tabanid flies published from 1 January 1960 to 9 September 2022. Various steps included were the literature search rationale, predefined criteria for inclusion and exclusion of the relevance of the references, and the extraction of relevant data to achieve the study objective.

3.2.2 Search strategy for relevant studies

Studies were searched in the following four electronic databases: PubMed (https://pubmed.ncbi.nlm.nih.gov/, from January 1, 1960 to September 9, 2022), ScienceDirect (https://www.sciencedirect.com/, from January 1, 1960 to September 9, 2022), Google Scholar (https://scholar.google.com/, from January 1, 1960 to September 9, 2022) and Scopus (https://www.scopus.com/, January 12, 2022 to January 13, 2022). The search strategy utilized a combination of Boolean operators and keywords to ensure comprehensive coverage of the topic. The keywords used were Prevalence OR

Distribution OR Abundance OR Occurrence AND tabanid flies OR horseflies AND global OR worldwide. Using these keywords individually or in combination, aimed to capture relevant articles. Titles, abstracts, and full-text articles were then identified and downloaded for further review.

3.2.3 Inclusion and exclusion criteria

To screen for the articles that can answer our research questions, the eligibility of every article was assessed based on the pre-set criteria. Briefly, the records were screened based on questions such as: (i) Was the study written in English, peer-reviewed, and published between January 1960 to September 2022. The selection period was based on the availability and relevance of articles regarding the topic of interest (ii) Did the study report the distribution or prevalence of tabanids (iii) Did it report on field-derived samples in study areas? (iv) the sample size (>25 for enabling statistical calculations). Moreover, the review excluded (i) all experimental articles, (ii) studies not written in English, and (iii) studies with unclear sample information, i.e., collection area, size, and host not mentioned.

The studies that fulfilled the inclusion criteria were documented, and an electronic spreadsheet (Microsoft Excel) was created to chart the data extracted from these articles. The following information was recorded on the spreadsheet: author names, geographical area, sample collection year, number and sample type screened, and collection nets.

3.2.4 Data quality control measures

Two authors independently used the Joanna Briggs Institute Critical Appraisal Tools Checklist 2017 review guideline (Ramatla *et al*., 2022; Tawana *et al*., 2023) for prevalence studies to verify the methodological soundness of the research articles chosen for quantitative synthesis. The inclusion of studies was based on the assessment criteria score of four or above.

3.2.5 Meta-analysis

Meta-analysis was performed using the Comprehensive Meta-Analysis Version 4.0 (CMA) program (https://www.meta-analysis.com/). The pooled prevalence estimate (PPE) was calculated using the random-effects model with a 95% confidence interval (CI). Cochran's Q statistic and the I^2 tests were used to measure statistical heterogeneity between studies. The I^2 scores of more than 75% were regarded as a high degree of heterogeneity. The Begg's and Mazumdar rank (BMR) correlation test was used in conjunction with funnel plots to detect publication bias and check for symmetry (Begg and Mazumdar, 1994).

3.3 Results

3.3.1 Bibliometric data

A total search of 1952 articles were initially identified through Google Scholar (n = 996), Science Direct (n = 787), PubMed (n = 161), and Scopus (n = 8) as shown in Figure 1. After the removal of duplicates(n=177), 1775 articles remained. Of these, 1355 articles were excluded after assessing their titles and abstracts. The remaining 420 full articles were subsequently assessed for eligibility, and the following groups were excluded: (i) Experimental articles (n =150), (ii) Systematic reviews (n = 5); and (iii Language of publication not in English (n = 4) as well as (iv) articles with less than 4 species reported (n=122). Subsequently, a total of 139 articles that reported on the prevalence and distribution of tabanid flies were eligible for inclusion in this review study (Figure 3-1).

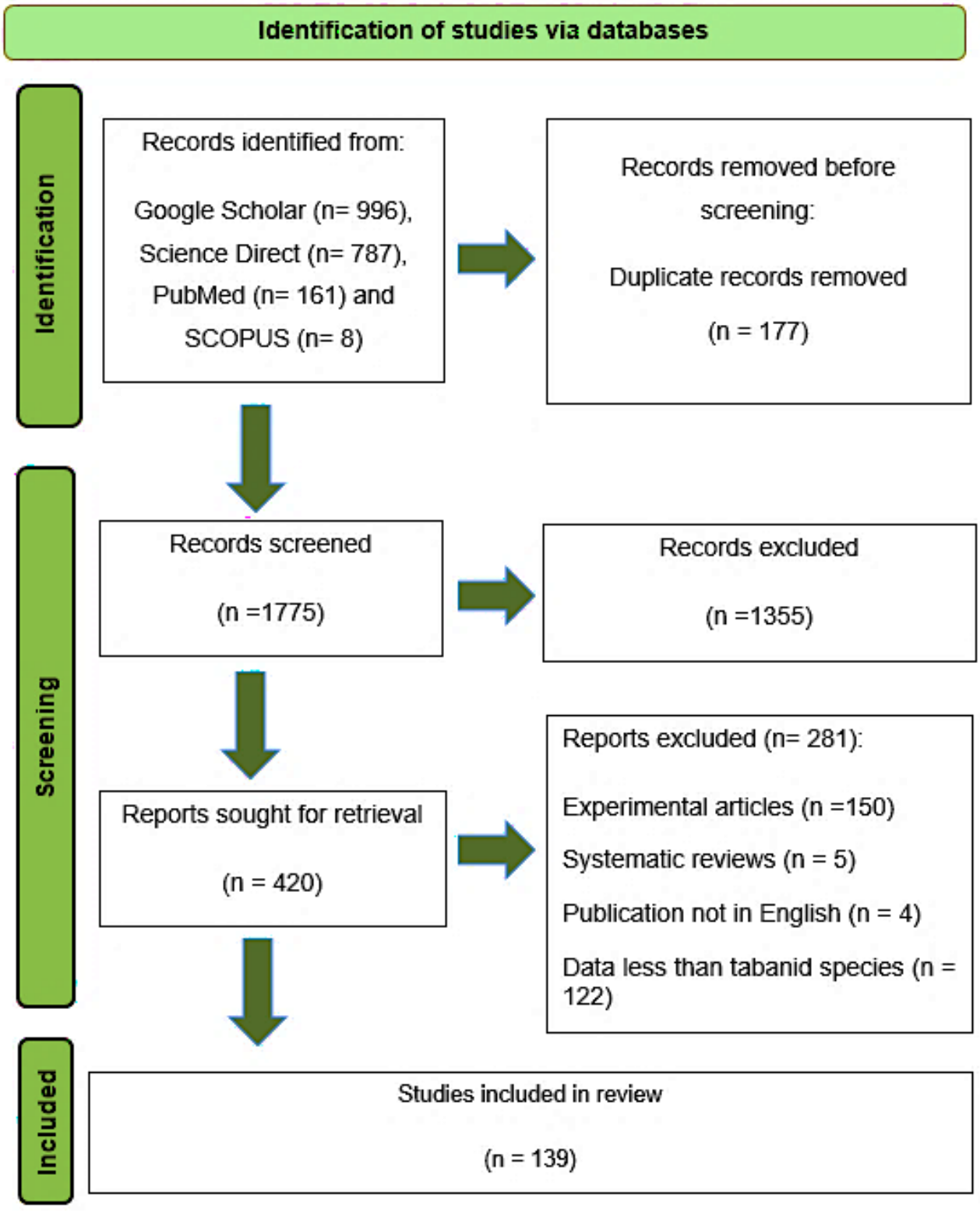


Figure 3-1: Flow chart illustrating the process of selecting studies for inclusion in the review, following the PRISMA guidelines.

3.3.2 Bibliometric data General characteristics of the included studies

3.3.2.1 Distribution of studies across the globe

Studies included in this systematic review and meta-analysis were from five continents namely, Africa Asia, Europe, North America and South America. No tabanid flies were reported from Antarctica. A total of 139 studies on Tabanidae were conducted in 45 countries, including the United States of America (n=18), Canada (n=14), Croatia (n=13), Thailand (n= 9), Brazil (n= 7), Bulgaria (n=9), Gabon (n=5), India (n= 5), Iran (n= 4), Nigeria (n=4), Tanzania (n=4), Turkey (n=4), Cameroon (n=3), France (n=3), Japan (n=3), Korea (n=3), South Africa (n=3), Poland (n=2), Russia (n=2), Jordan (n=2), Central African Republic (n=2), Australia (n=2), Algeria (n=1), Belgium (n=1), Bosnia and Herzegovina (n=1), China (n=1), Colombia (n=1), Czechoslovakia (n=1), Himalaya (n=1), Hungary (n=1), Indonesia (n=1), Iraq (n=1), Lithuania (n=1), Mozambique (n=1), Malaysia (n=1), Senegal (n=1), Serbia (n=1), Spain and France (n=1), Spain (n=1), Sweden (n=1), Switzerland (n=1), United Arab Emirates (n=1), Uruguay (n=1), Zambia (n=1) and Zimbabwe (n=1) as shown in figure (3-2).

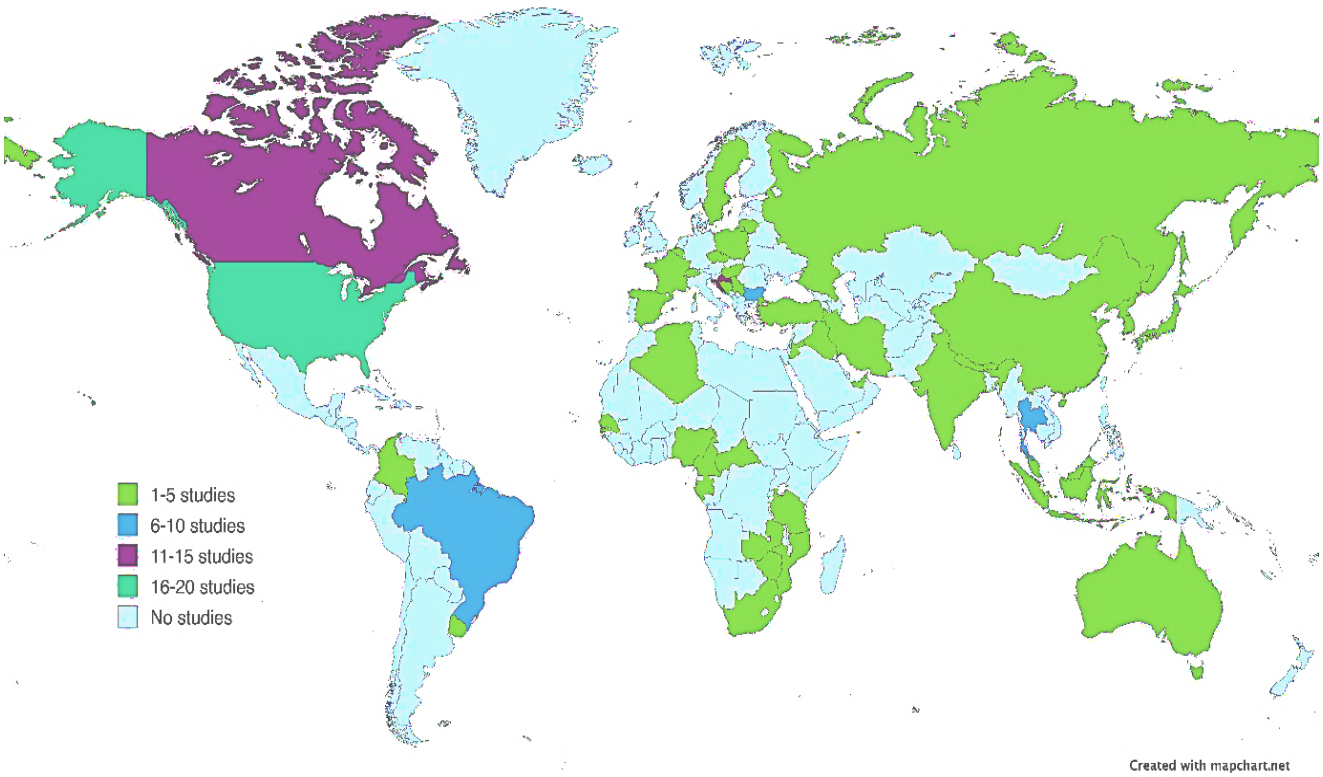


Figure 3-2: World map showing the number of articles from different countries which reported tabanid distribution.

3.3.3 Publication year of eligible studies

The characteristics of all eligible studies are shown in figure 3-3. All the studies were published from 1960 to 2022. From 1960 - 1990 period, 40 studies were reported. Most of the studies (n=50) were conducted in the last decade (2011–2022). Only n=13 studies were reported between 1991 and 2000. Since 2001, the number of published studies has increased (n=36).

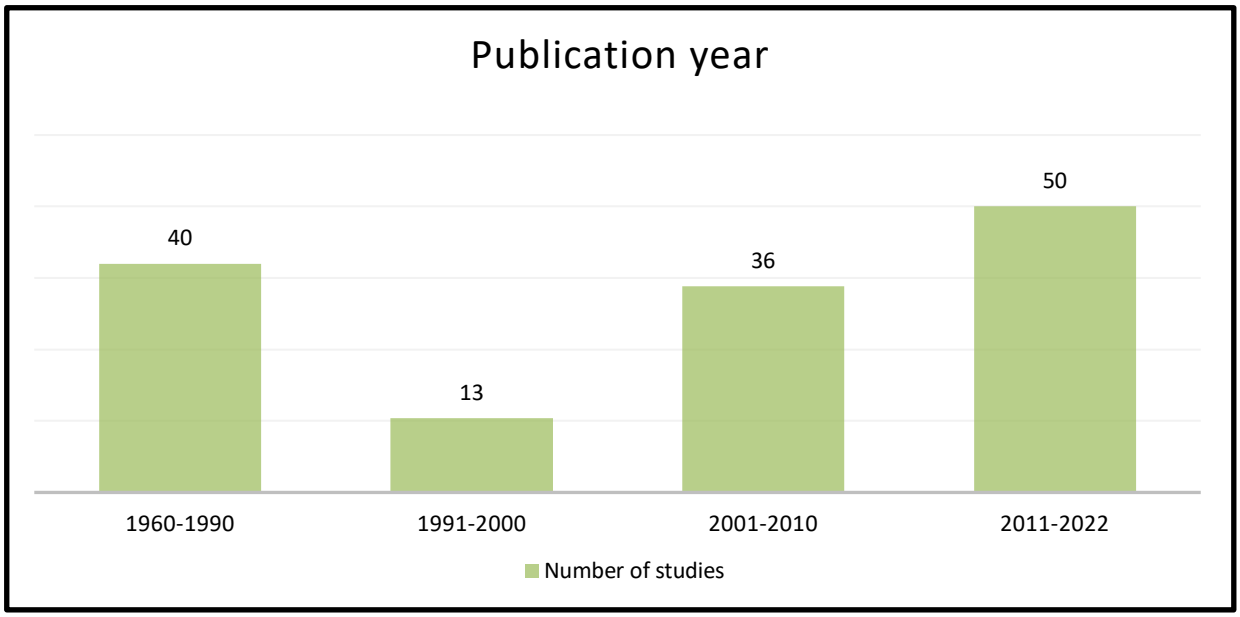


Figure 3-3: Bar graph illustrating the distribution of selected studies over time, categorized by decade from 1950-2022, highlighting the trends in research activity**.**

3.3.4 Tabanid fly identification and articles used for nomenclature

Various identification methods have been discovered since the discovery of these hematophagous flies. This review documented 124 different articles that were used as references in the taxonomical classification of tabanid flies. Of the 124 articles of reference 28 articles were analyzed (Figure 3-4). Chvala *et al*., (1972) was the most reported article of reference for taxonomic classification (n=41), followed by Chvála, (1988) and Olsufiev (1977) both with (n=12) studies. Burton, (1978) with (n=9) studies while Leclereq, (1966a) with (n=8) studies. Both studies by Oldroyd (1954 and 1977) were cited in (n=7) studies.

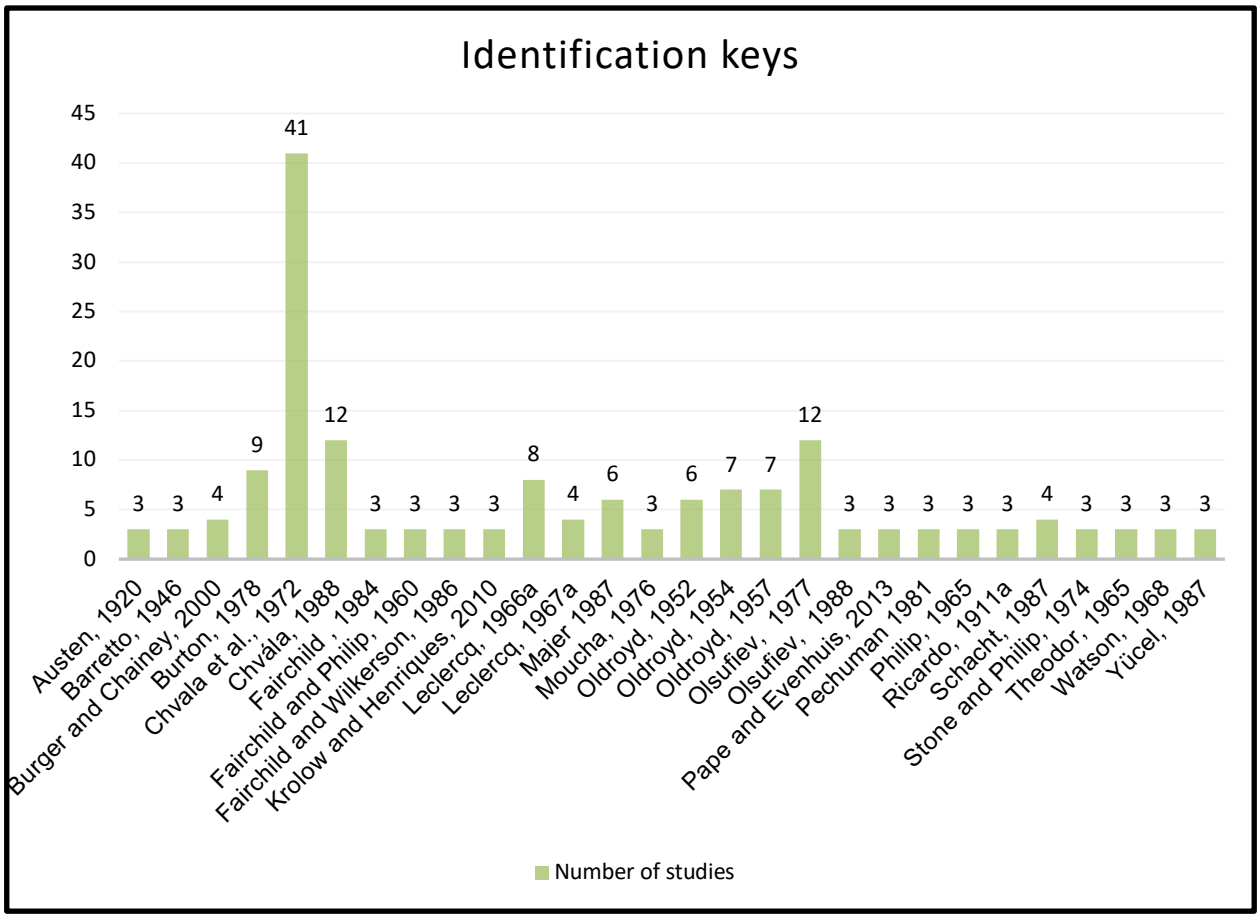


Figure 3-4: Bar graph representing summary of articles utilized for the description of Tabanidae in the selected studies.

3.3.5 Trapping methods

For the tabanid survey, 28 trapping methods were reported, of which 14 were eligible for analysis (Figure 3-5). The most reported trapping method was with sweep nets (n=40) studies, followed malaise and Nzi traps with n=27 and n= 22 studies respectively. Traps with the least studies included: mosquito magnet, epsilon and sticky trap with (n=3) studies. The remaining 14 traps documented consisted of less than 3 studies, namely: chemical trap (n=1), NGU trap (n=1), hanging trap (n=1), sevi trap (n=1), dummy trap (n=1), kerosene pool (n=1), nitse trap (n=2), New Jersey light traps (n=2), moncone trap (n=1), SLAM trap (n=1), flight intercept traps (n=1), box trap (n=1), 2-tier box trap (n=2) and biconical trap (n=2).

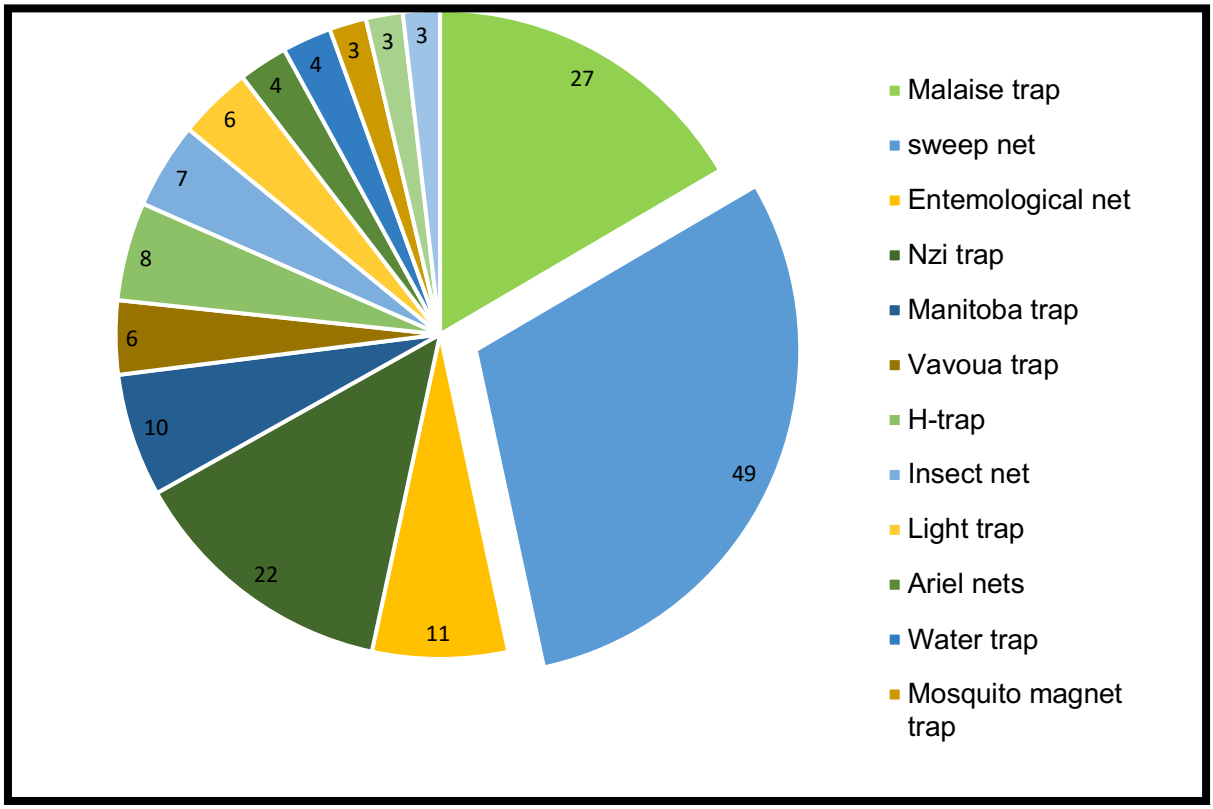


Figure 3-5: Illustration of pie chart showing various trapping methods employed for capturing tabanid flies.

3.3.6 Hematophagous tabanid fly data

A total of 209 tabanid species is reported from 139 articles included in this systemic review. The reported taxonomic subfamilies included, Pangoniinae (tribes Pangoniini,

Philolichini, Scionini), Chrysopsinae (tribes Chrysopsini), and Tabaninae (tribes Diachlorini, Haematopotini, Tabanini).

A total of 16 genera are reported the 209 species and these included, *Ancala* Enderlein, (1922); *Atylotus* Osten Sacken, (1876); *Chlorotabanus* Lutz, (1909); *Chrysops* Meigen, (1803); *Diachlorus* Osten Sacken, 1876; *Dasyrhamphis* Enderlein, 1922; *Haematopota* Meigen, (1803); *Heptatoma* Meigen, 1803; *Hybomitra* Enderlein, (1922); *Leucotabanus* Lutz, (1913); *Philipomyia* Olsufjev, (1964); *Philoliche* Wiedemann, (1820); *Poeciloderas* Lutz, (1921); *Tabanus* Linnaeus, (1758); *Therioplectes* Zeller, (1842) and *Silvius* Meigen, (1820) (Figure 3-6).

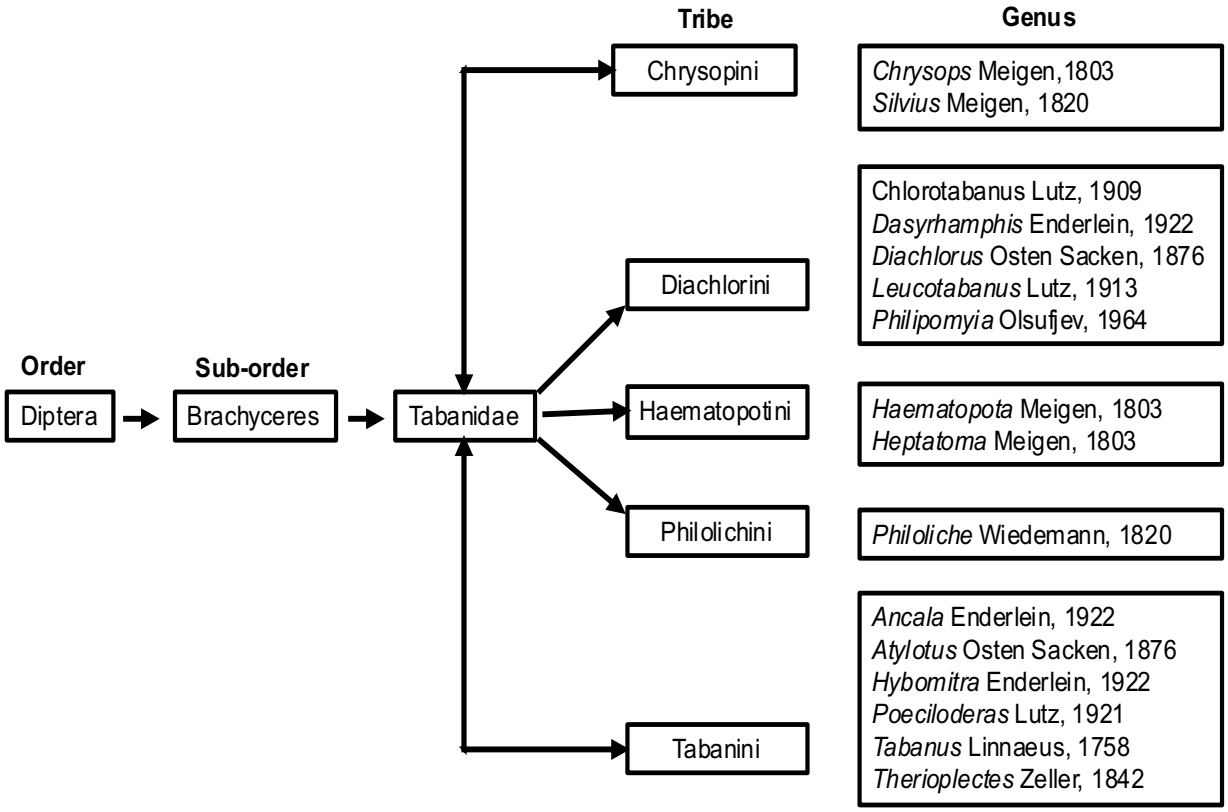


Figure 3-6: Simplified systematic classification of the different tabanids (horse flies) group studies recorded.

3.3.6.1 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Chrysopini

In the current systematic review n=42 *Chrysops* species have been reported from a total of 282 articles shown Table (3-1). Amongst the *Chrysops* species*, C. excitans* was the most prevalent species with PPE of 7.6%. While *C. Callidus,* and *C. dispar* both displayed a PPE of 5.8% and 5.8% respectively. And the least reported species with a PPE of 0.01% was *C. Sackeni* Only one species of genus *Silvius* was reported, namely, *S. alpinus* with a PPE of 0.5%.

Table 3-1: Captures of the genus *Chrysops* (horse flies) documented between 1960-2022.

	Pooled prevalence estimates			Measure of heterogeneity				Publication bias
Risk factors	No. studies	No. of flies	No. of tabanids	Prevalence 95% CI (%)	Q	I^2	Q–P	P-value
Deer flies								
Chrysops aberrans Philip, 1941	3	56305	211	0.3% (0.2–0.7)	23.132	91.354	p< 0.0001	0.296
Chrysops ater Osten Sacken, 1875	7	87272	4922	0.9% (0.1–11.3)	5811.468	99.897	p< 0.0001	0.229
Chrysops caecutiens Linnaeus 1758	26	99305	859	0.7% (0.4–1.3)	1442.749	98.267	p< 0.0001	0.073
Chrysops callidus Osten Sacken, 1875	6	103360	1681	5.8% 1.6–18.6)	2279.303	99.781	p< 0.0001	0.494
Chrysops calvus Pechuman & Teskey, 1967	4	63082	283	0.3% (0.1–1.2)	258.086	98.838	p< 0.0001	0.734
Chrysops carbonarius Walker, 1848	6	70272	305	0.3% (0.1–0.7)	151.093	96.691	p< 0.0001	0.024
Chrysops cincticornis Walker, 1848	7	73215	292	0.3% (0.1–1.0)	374.675	98.399	p< 0.0001	0.118
Chrysops cuclux Whitney, 1879	7	76743	293	0.3% (0.1–1.0)	484.394	98.761	p< 0.0001	0.548
Chrysops dimidiatus Wulp, 1885	4	2130	26	1.6% (0.3–9.5)	40.260	92.548	p< 0.0001	0.778
Chrysops dispar Fabricius, 1798	6	9573	281	5.8% (2.7–11.9)	105.934	95.280	p< 0.0001	0.452
Chrysops distinctipennis Austen, 1906	5	11139	168	1.1% (0.2–6.3)	170.121	97.649	p< 0.0001	0.462
Chrysops excitans Walker, 1850	8	40494	1595	7.6% (3.4–16.1)	1416.34	99.506	p< 0.0001	0.860

					9			
Chrysops flavipes Meigen, 1804	8	13782	192	1.9% (0.7-4.9)	258.369	97.291	p< 0.0001	0.386
Chrysops frigidus Osten and Sacken,1875	12	99969	1262	1.2% (0.05–0.27)	1821.506	99.396	p< 0.0001	0.064
Chrysops fuliginosus Wiedemann, 1821	3	84911	30586	1.7% (0.00–5.18)	446.005	99.552	p< 0.0001	0.129
Chrysops furcatus Walker, 1848	4	22197	185	6.5% (0.11–2.92)	341.328	99.121	p< 0.0001	0.734
Chrysops geminatus Wiedemann, 1828	4	53431	396	1.1% (0.3-4.1)	413.110	99.274	p< 0.0001	0.734
Chrysops indus Osten Sacken, 1875	8	77168	874	1.0% (0.4–2.7)	1213.796	99.423	p< 0.0001	0.151
Chrysops japonicus Wiedemann, 1828	3	19050	53	0.2% (0.1–0.8)	29.871	93.304	p< 0.0001	0.296
Chrysops lateralis Wiedemann, 1828	5	67366	560	0.6% (0.2–1.9)	516.184	99.225	p< 0.0001	0.462
Chrysops longicornis Macquart, 1838	7	72786	469	1.8% (0.6–5.1)	389.951	98.461	p< 0.0001	0.229
Chrysops ludens Loew, 1858	4	5438	9	0.2% (0.1–0.4)	3.275	8.407	p< 0.0001	0.734
Chrysops macquarti Philip, 1961	5	53915	260	0.8% (0.2-3.2)	436.996	99.085	p< 0.0001	0.810
Chrysops mitis Osten Sacken, 1875	8	70422	2216	2.7% (0.12-5.9)	1095.406	99.361	p< 0.0001	0.086
Chrysops montanus Osten Sacken, 1875	6	54454	696	0.9% (0.2–.3.3)	972.146	99.486	p< 0.0001	0.452
Chrysops niger Macquart, 1838	8	96169	507	0.4% (0.1–1.3)	1034.549	99.323	p< 0.0001	0.710
Chrysops parallelogrammus Zeller, 1842	5	32404	796	1.7% (0.9–3.1)	192.541	97.923	p< 0.0001	0.623
Chrysops pellucidus Fabricius, 1805	3	9166	57	0.6% (0.4–0.9)	2.356	15.103	p< 0.0001	0.962
Chrysops relictus Meigen 1820	3	52698	241	0.3% (0.2–0.7)	23.132	91.354	p< 0.0001	0.296
Chrysops rufipes Meigen 1820	3	14301	9	0.3% (0.0–9.2)	50.273	96.022	p< 0.0001	0.296
Chrysops sackeni Hine, 1903	3	46440	85	0.1% (0.0–0.5)	19.953	89.976	p< 0.0001	0.601
Chrysops shermani Hine, 1907	4	20438	1329	0.3.1% (0.14–	247.176	98.786	p< 0.0001	0.308

				0.67)				
Chrysops silaceus Austen, 1907	6	13686	978	4.1% (0.07–2.12)	614.753	99.187	p< 0.0001	0.259
Chrysops sordidus Osten Sacken, 1875	6	37450	253	0.8% (0.02–0.38)	224.774	97.776	p< 0.0001	0.706
Chrysops sp.	3	8006	447	1.2% (0.0–7.03)	128.770	98.447	p< 0.0001	0.296
Chrysops suavis Loew, 1858	4	19394	181	0.6% (0.03–0.14)	25.455	88.215	p< 0.0001	0.734
Chrysops univittatus Macquart, 1855	10	78479	2161	5.2% (0.23–11.2)	2546.880	99.647	p< 0.0001	0.810
Chrysops venus Philip, 1950	3	44253	155	0.3% (0.01–0.09)	50.402	96.032	p< 0.0001	0.289
Chrysops viduatus Fabricius, 1794	20	106423	1713	5.4% (0.7–1.5)	910.194	97.913	p< 0.0001	0.897
Chrysops viduatus Fabricius, 1794	20	30873	2563	1.0% (0.7–1.5)	97.913	97.913	p< 0.0001	0.922
Chrysops zinzala Philip, 1942	5	33166	482	0.7% (0.2–2.5)	61.241	93.468	p< 0.0001	0.462
Silvius alpinus Scopoli, 1763	10	29661	2681	0.5% (0.3-1.0)	159.896	94.371	p< 0.0001	0.474

3.3.6.2 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Diachlorini

In the current systematic review (9/209) the following Diachroini species have been reported from a total of 66 studies (Table 3-2). Of the 9 species identified *Philipomyia aprica* was the most prevalent species with a PPE of 2.8%, followed by *C. inanis* and *P. gracea* with a PPE of 2.1% and 2.0 % respectively. And the least reported species was *D. ater* with PPE of 0.1 %.

Table 3-2: Captures of genus Diachlorini documented between 1960 to 2022.

	Pooled prevalence estimates				Measure of heterogeneity			Publication bias
Risk factors	No. of studies	No. of flies	No. of tabanids	Prevalence 95% CI (%)	*Q*	I^2	*Q*–*P*	P-value
Diachlorini								
Chlorotabanus inanis Fabricius, 1787	3	73215	292	2.1% (0.7–6.2)	374.675	98.399	p< 0.0001	0.118
Dasyrhamphis anthracinus Meigen & Wiedemann, 1820	4	77168	874	1.0% (0.4–2.7)	1213.796	99.423	p< 0.0001	0.151
Dasyrhamphis ater Rossi 1970	10	45512	89	0.1% (0.1–0.3)	72.996	87.671	p< 0.0001	0.408
Dasyrhamphis carbonarius Meigen, 1820	3	8974	23	0.4 (0.0–2.6)	32.677	93.879	p< 0.0001	0.602
Dasyrhamphis umbrinus Wiedemann, 1820	6	30873	2563	1.5% (0.5–4.3)	531.446	99.059	p< 0.0001	0.573
Diachlorus curvipes Fabricius, 1805	6	29661	2681	1.6% (0.6–4.6)	312.812	98.402	p< 0.0001	0.850
Leucotabanus annulatus (Say) 1823	3	99969	1262	0.2% (0.1–0.6)	1821.506	99.396	p< 0.0001	0.064
Philipomyia aprica Meigen, 1820	14	52698	241	2.8% (1.5–5.2)	23.132	91.354	p< 0.0001	0.296
Philipomyia gracea Fabricus 1794	17	99305	859	2.0% (0.9–4.3)	1442.749	98.267	p< 0.0001	0.073

3.3.6.3 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Haematopotini

In the current systematic review (19/209) Haematopotini species were reported by a total of 135 studies of which 15 of the studies were reported for genus *Heptatoma* (Table 3-3). Of the species identified, *H. pluvialis* was the most prevalente with PPE of 5.6%, followed by *H. koryoensis*, *H. javana*, *Haematopota* sp. and *H. sinensis* with a PPE of 2.8.0%, 2.5.0%, 2.4% and 2.1% respectively., and *H. koryoensis* with a percentage of 2.1% and 2%, respectively. And the least reported species were *H. tamerlani* and *H. longeantennata* with a prevalence of 0.1%.

For genus *Heptatoma* only 1 species was reported, the *H. pellucen* with a PPE of 2.4%.

Table 3-3: Captures of the tribe *Haematopini* documented between 1960 to 2022.

Pooled prevalence estimates					Measure of heterogeneity			Publication bias
Risk factors	No. of studies	No. of flies	No of tabanids	Prevalence 95% CI (%)	Q	I^2	Q–P	p-value
Haematopota								
Haematopota albihirta Karsch, 1887	4	11975	74	0.6% (0.5–0.8)	2.272	0.000	p< 0.0001	0.343
Haematopota bigoti Gobert, 1880	9	48132	215	0.4% (0.2–1.0)	218.708	96.342	p< 0.0001	0.548
Haematopota coronata Austen, 1908	3	3293	7	0.2% (0.1–0.5)	1.095	0.000	p< 0.0001	0.601
Haematopota decora Walker, 1856	4	14164	121	0.8% (0.1–5.1)	60.589	95.049	p< 0.0001	0.308
Haematopota grandis Meigen, 1804	8	36777	123	0.2% (0.1–0.7)	164.394	95.742	p< 0.0001	0.901
Haematopota italica Meigen, 1804	22	96725	834	0.5% (0.2–1.0)	1719.555	98.779	p< 0.0001	0.910
Haematopota javana Wiedemann, 1821	7	9621	179	2.5% (1.5–4.1)	25.023	76.022	p< 0.0001	0.763
Haematopota koryoensis Shiraki, 1932	3	13328	267	2.8% (0.9–8.5)	147.368	98.643	p< 0.0001	0.983
Haematopota longeantennata Olsufiev, 1937	3	10994	6	0.1% (0.0–0.2)	3.138	36.267	p< 0.0001	0.296
Haematopota ocelligera Kröber, 1922	5	24977	47	0.3% (0.1–0.8)	33.923	88.209	p< 0.0001	0.220
Haematopota pallens Loew, 1871	3	2481	6	0.3% (0.1–0.8)	3.854	48.112	p< 0.0001	0.391
Haematopota pandazisi Kröber, 1936	7	31612	1053	0.8% (0.3–2.5)	892.300	99.328	p< 0.0001	0.179
Haematopota pluvialis Linnaeus,1758	3	93990	8268	5.6% (4.0–7.8)	5750.454	99.444	p< 0.0001	0.411
Haematopota scutellata Olsufjev & Chvála 1964	12	47891	453	0.5% (0.1–2.3)	1557.508	99.294	p< 0.0001	0.631

Haematopota sinensis Ricardo, 1911	4	9284	250	2.1% (0.3–15.3)	743.206	99.596	p< 0.0001	0.652
Haematopota sp	5	41164	8987	2.4% (0.4–12.0)	432.285	99.075	p< 0.0001	0.806
Haematopota subcylindrica Pandellé, 1883	15	70895	6042	1.6% (1.0–2.6)	3217.373	99.565	p< 0.0001	0.137
Haematopota tamerlani Szilády, 1923	3	13 328	10	0.1% (0.0–0.2)	1.840	0.000	p< 0.0001	0.296
Heptatoma pellucen Fabricius, 1776	15	72983	338	0.3% (0.2–0.6)	175.561	92.026	p< 0.0001	0.843

3.3.6.4 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Philolichini

In the current systematic review (2/209) *Philolichini* species were reported by 6 studies. The remaining species namely: *P. medialis* Oldroyd, 1957; *P. silverlocki* (Austen, 1912*); P. subfascia* Walker, 1854; *P. varipes* Ricardo, 1911; *P. zonata* Walker, 1871; *P. longirostris* Hardwicke, 1823; *P. gravoti* Surcouf, 1908; and *P. fodiens* Austen, 1908. were not included in the meta-analysis (less than three study per country). The meta-analysis resulted PPE of 7.4.% and 1.2% from genus *P. aethiopica* and *P. taprobanes*, respectively (Table 3-4).

Table 3-4: Captures of the tribe of *Philochini* documented from 1960 to 2022.

	Pooled prevalence estimates				Measure of heterogeneity			Publication bias
Risk factors	No. of studies	No. of flies	No. of tabanids	Prevalence 95% CI (%)	*Q*	I^2	*Q*–*P*	P-value
Philoliche								
Philoliche aethiopica Thunberg, 1789	3	5039	398	7.4% (1.9–24.7)	62.203	96.785	p< 0.0001	0.980
Philoliche taprobanes Walker, 1854	3	9166	94	1.2% (0.6–2.4)	10.593	81.120	p< 0.0001	0.774

3.3.6.5 Insecta, Diptera, Brachycera, Orthorrhapha, Tabanidae: Tabanini

In the current systematic review (137/209) Tabanini species have been reported which is divided into 6 genera, namely (Table3-5):

3.3.6.5.1 Genus: *Ancala*

In this review (2/09) *Ancala* species were reported by 6 studies, *Ancala* species: *A. fasciata* Fabricius, 1775 and *A. africana* Gray, 1832, with PPE of 1.0% and 0.16% respectively.

3.3.6.5.2 Genus: *Atylotus*

From a total of 77 studies, *Atylotus* species (9/209) were reported as indicated in table 5. Meta-analysis resulted in <2% prevalence for the following species: A. *miser* with PPE of 2.3%, followed by *A*. *pulchellus* with PPE of 2.2%. The least species reported was *A*. *flavoguttatus* with PPE of 0.3%.

3.3.6.5.3 Genus: *Hybomitra*

A total of 35 *Hybomitra* species were reported by 283 studies analyzed in this review (Table 5). The meta-analysis showed the most prevalent species reported in the *Hybomitra* genus was *H. lasiophthalma* with a PPE of 9.0%, Followed by *H. micans* and *H. nuda* Philip with a PPE of 4.9% and 4.1%, respectively.

3.3.6.5.4 Genus: *Therioplectes*

In total, the following 3 species were reported by 15 studies in this review: *T. tunicatus* Szilády, 1927; *T. tricolor* Zeller, 1842 and *T. gigas* Herbst, 1787 with the following PPEs 6.0%, 4.0% and 0.2% respectively.

3.3.6.5.5 Genus: *Poecilodera*

Only 1 species was reported by 3 studies included in the analysis, namely, which included *P. quadripunctatus* Fabricius, 1805 with a PPE of 43%.

Table 3-5: Captures of the tribe Tabanini between 1960 to 2022.

Pooled prevalence estimates				Measure of heterogeneity				Publication bias
Risk factors	No. of studies	No. of flies	No of tabanids	Prevalence 95% CI (%)	*Q*	I^2	*Q*–*P*	P-value
Ancala								
Ancala fasciata Fabricius, 1775	3	1259	17	1.0% (6.0–0.15.0)	1.717	0.000	p < 0.0001	0.691
Ancala africana Gray, 1832	3	64750	475	1.6% (4.0–6.8)	367.690	99.456	p < 0.0001	0.672
Atylotus								
Atylotus fulvus Meigen, 1804	16	62513	252	0.4% (0.2–0.7)	363.623	95.875	p < 0.0001	0.260
Atylotus horvathi Szilády, 1926	3	13328	32	0.4% (0.01–1.2)	16.039	87.531	p < 0.0001	0.296
Atylotus miser Szilády, 1915	3	16270	2128	2.3% (0.0–55.4)	942.416	99.788	p < 0.0001	0.418
Atylotus rusticus Linnaeus, 1767	10	47566	759	0.6% (0.2–1.8)	1198.400	99.249	p < 0.0001	0.720
Atylotus agrestis Wiedemann, 1828	8	89939	1269	1.1% (0.2–5.0)	3425.831	99.796	p < 0.0001	0.536
Atylotus loewianus Villeneuve, 1920	23	146409	6129	1.9% (0.9–3.9)	11652.454	99.811	p < 0.0001	0.634
Atylotus virgo Wiedemann, 1824	4	9249	171	1.9% (0.4–7.9)	170.596	98.241	p < 0.0001	0.734
Atylotus pulchellus Loew, 1858	4	4056	680	2.2% (0.1–41.9)	407.570	99.264	p < 0.0001	0.500
Atylotus flavoguttatus Szibidy, 1915	6	60022	1698	0.3% (0.1–1.4)	346.619	98.557	p < 0.0001	0.707
Hybomitra								
Hybomitra acuminata Loew, 1858	8	37876	123	0.3% (0.0–1.3)	396.840	98.236	p < 0.0001	0.386
Hybomitra affinis Kirby, 1837	10	177669	12681	2.3% (1.5–3.5)	1604.872	99.439	p < 0.0001	0.858

Hybomitra arpadi Szilády, 1923	7	148541	14707	3.9% (0.6–2.3)	29851.254	99.980	p < 0.0001	0.548
Hybomitra astuta Osten Sacken, 1876	4	113727	102	0.1% (0.0–0.4)	104.512	97.130	p < 0.0001	0.929
Hybomitra aterrima Meigen, 1820	3	4054	314	1.7% (0.0–49.9)	172.348	98.840	p < 0.0001	0.296
Hybomitra aurilimba Stone, 1938	4	63362	217	0.4% (0.1–1.0)	139.156	97.844	p < 0.0001	0.864
Hybomitra bimaculata Macquart, 1826	18	76730	5148	0.7% (0.3–1.5)	4596.975	99.630	p < 0.0001	0.008
Hybomitra caucasica Enderlein, 1925	6	16933	349	0.6% (0.1–2.8)	367.265	98.639	p < 0.0001	0.707
Hybomitra cincta Fabricius, 1794	4	102257	9	0.0% (0.0–0.0)	8.415	64.350	p < 0.0001	0.734
Hybomitra ciureai Séguy, 1937	25	134268	5817	2.1% (1.5–2.8)	1840.640	98.696	p < 0.0001	0.657
Hybomitra distinguenda Verrall, 1909	20	57271	928	0.9% (0.5–1.6)	1241.467	98.470	p < 0.0001	0.974
Hybomitra epistates Osten Sacken, 1878	13	190687	5928	1.6% (0.8–3.4)	6570.179	99.817	p < 0.0001	0.951
Hybomitra expollicata Pandellé, 1883	4	44281	4324	1.3% (0.1–12.2)	336.674	99.109	p < 0.0001	0.734
Hybomitra hinei Johnson, 1904	4	37263	211	1.5% (0.2–9.2)	512.964	99.415	p < 0.0001	0.357
Hybomitra illota Osten Sacken, 1876	11	188749	3897	1.2% (0.6–2.8)	4621.045	99.784	p < 0.0001	0.755
Hybomitra kaurii Chvála & Lyneborg, 1970	4	19240	438	2.4% (0.005–0.102)	555.464	99.460	p < 0.0001	0.845
Hybomitra lasiophthalma Macquart, 1838	18	318971	64081	9.0% (5.9–13.3)	24648.178	99.931	p < 0.0001	0.500
Hybomitra lundbecki Lynebrg, 195	6	34076	971	1.3% (0.7–2.4)	120.151	95.839	p < 0.0001	0.132
Hybomitra lurida Fallén, 1817	11	194692	3201	1.1 (0.8–3.0)	4831.058	99.886	p < 0.0001	0.822
Hybomitra micans Meigen, 1804	3	2774	206	4.9% (1.6-14.4)	53.103	96.234	p < 0.0001	0.296
Hybomitra microcephala Osten Sacken, 1876	4	124586	46	0.1% (0.0–0.1)	26.702	88.765	p < 0.0001	0.734
Hybomitra montana Meigen, 1820	6	31146	781	1.0% (0.3–3.7)	896.298	99.442	p < 0.0001	0.707
Hybomitra muehlfeldi Brauer, 1880	15	76293	2216	1.3% (0.7–2.5)	2231.513	99.373	p < 0.0001	0.488

Hybomitra nitidifrons Szilády, 1914	3	134548	7596	3.4% (2.3-5.1)	852.310	99.531	p < 0.0001	0.310
Hybomitra nitidifrons. confiformis Chvala et Moucha, 1971	3	29459	14	0.0% (0.0–0.2)	6.092	67.171	p < 0.0001	0.310
Hybomitra nuda Philip, 1961	3	7706	290	4.1% (1.7–9.4)	104.070	98.078	p < 0.0001	0.605
Hybomitra pechumani Teskey & Thomas, 1979	4	149213	20149	2.0% (1.2–3.3)	1736.753	99.827	p < 0.0001	0.359
Hybomitra pilosa Loew, 1858	9	51374	226	0.4% (0.2–0.9)	259.861	96.921	p < 0.0001	0.602
Hybomitra sodalis Williston, 1887	7	65043	634	1.5% (0.6–3.8)	487.163	98.768	p < 0.0001	0.955
Hybomitra solstitialis Meigen, 1820	11	50325	271	0.1% (0.0–0.6)	469.263	97.869	p < 0.0001	0.573
Hybomitra trepida McDunnough, 1921	7	150605	1394	0.5% (0.2–1.1)	556.382	98.922	p < 0.0001	0.367
Hybomitra typhus Whitney, 1904	6	37917	513	1.3% (0.4–4.4)	893.466	99.440	p < 0.0001	0.452
Hybomitra ukrainica Olsufjev, 1952	11	42881	543	0.5% (0.2–1.1)	141.059	92.911	p < 0.0001	0.436
Hybomitra zonalis Kirby, 1837	7	146682	2865	0.7/% (0.4–1.4)	1393.494	99.569	p < 0.0001	0.367
Hybomitria tripsila Wiedemann,1828	4	65964	212	0.3% (0.1–1.4)	338.880	99.115	p < 0.001	0.304
Therioplectes								
Therioplectes tunicatus Szilády, 1927	4	15927	171	6.0% (2.0–18)	76.621	96.085	p < 0.0001	0.734
Therioplectes tricolor Zeller, 1842	3	9336	31	4.0% (1.0–9.0)	8.710	77.038	p < 0.0001	0.960
Therioplectes gigas Herbst, 1787	8	36445	94	2.0% (1.0–4.0)	83.423	91.609	p < 0.0001	0.536
Poeciloderas								
Poeciloderas quadripunctatus Fabricius, 1805	3	2070	78	43.0% (23.0–79.0)	15.364	86.983	p < 0.0001	0.777

3.3.6.5.6 Genus: *Tabanus*

Genus *Tabanus* was the most reported group in the tribe of Tabanini. A total of 86 species were reported by 682 studies (Appendix I). The highest reported tabanid species was *T. occidentalis* with a PPE of 49.0%. While *T. sufis* and *T. quinquevittatus* and had PPEs of 35.1% and 32.4% respectively. Followed by *T. nigrovittatus* with PPE of 27.7%. Several species were reported with a prevalence of >20% namely: *T. pallidescens*, *T accensus*, *T. bromius* and *T. taeniola* had a prevalence of (12.4,11.2, 10.8, 10.3 and 10.3) % respectively.

Seven species showed a prevalence of >10% which included *T. par 8.6%*, *T. triangulum* 8.0%, *T. importunes* 7.7%, *T. striatus* 7.1%, T. *tergestinus* 6.5%, *T. secedens* 5.4% and *T. rubidus* 5.2%.

The least reported *Tabanus* species was *T. claripennis, T. americanus, T. briani, T. calens, T. fairchildi, T. konis, T. neovestitus T. reinwardtii* and *T. spectabilis with* a PPE of 0.1%.

3.4 Publication bias

The Begg and Mazumdar rank correlation test, Egger's regression test, and funnel plots were used in addition to one another to measure publication bias and check for symmetry. The publication bias was examined using the funnel plot. Figures 3-7 – 3-10 displays a practically symmetrical distribution of all included studies on both sides of the funnel plot, indicating a relatively low potential for publication bias of p-value 0.024 for *Chrysops carbonarius* Walker, 1848, p-value 0.008 for *Hybomitra bimaculata* Macquart, 1826, p-value 0.002 for *Tabanus autumnalis* Linnaeus, 1761, p-value 0.036 for *Tabanus maculicornis* Zetterstedt, 1842, p-value 0.014 for *Tabanus laetetinctus* Becker, 1913.

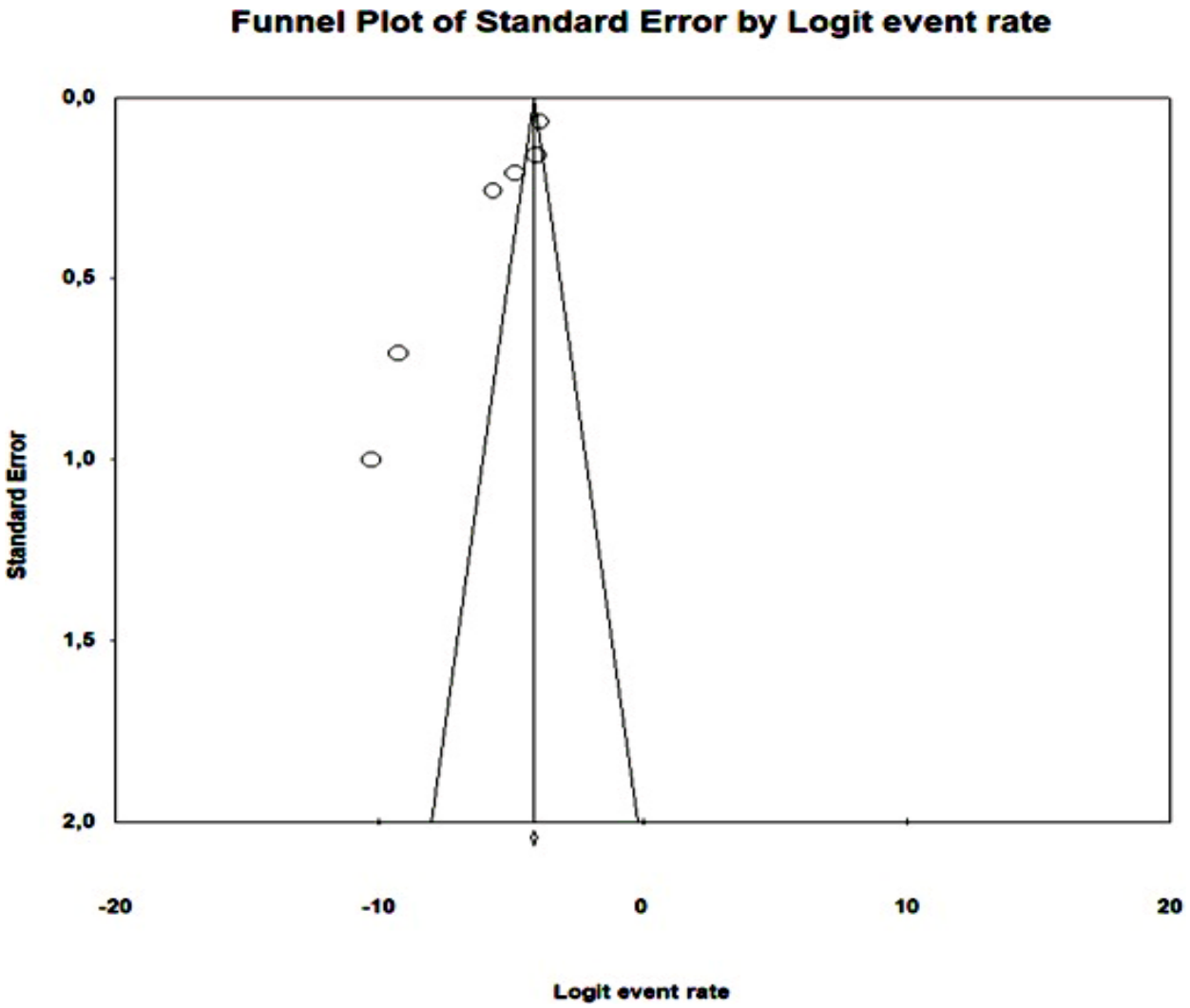


Figure 3-7: The funnel plot of *Chrysops carbonarius* (Walker, 1848) studies included in the meta-analysis

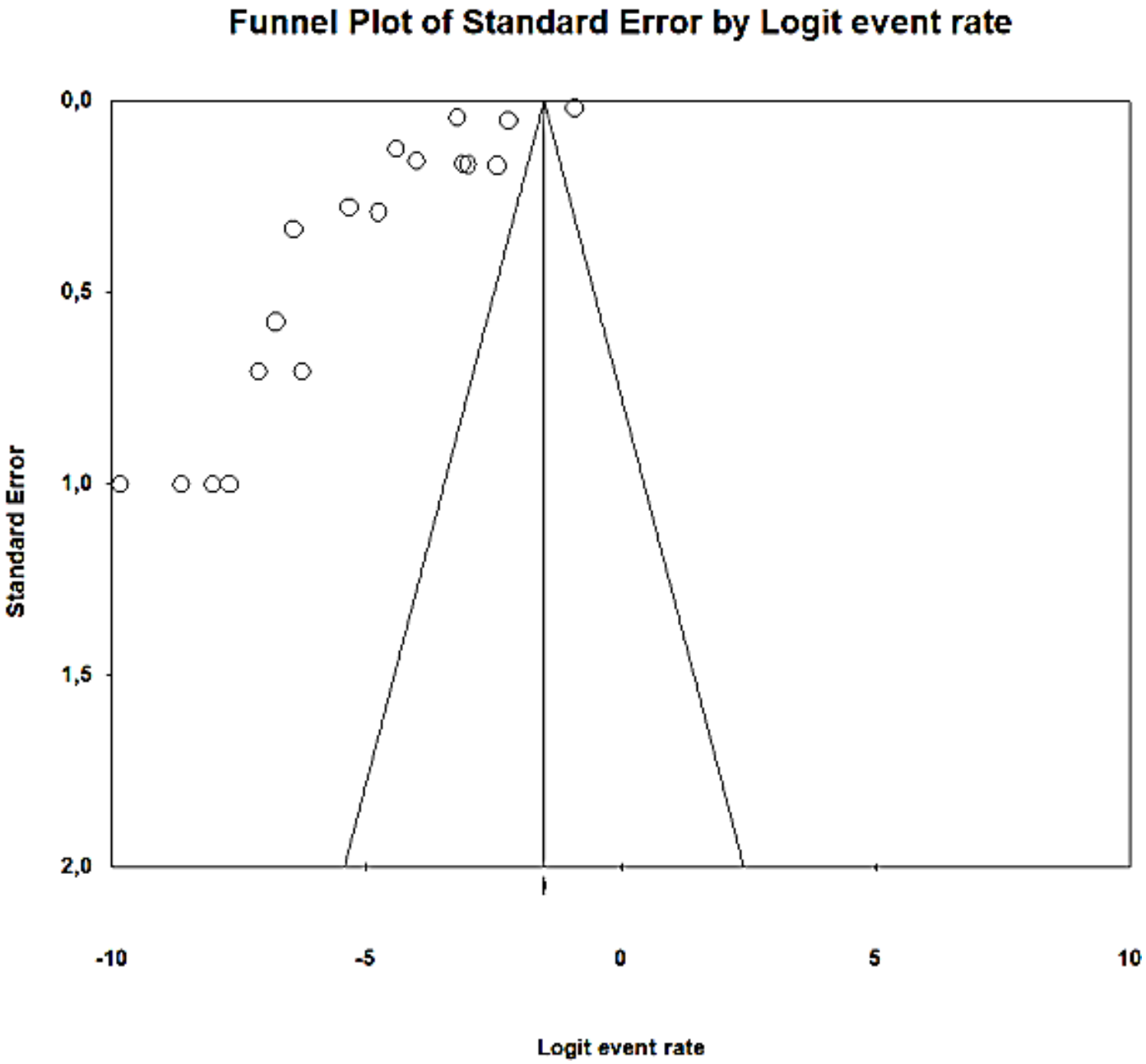


Figure 3-8: The funnel plot *Hybomitra bimaculate* (Macquart, 1826) studies included in meta-analysis.

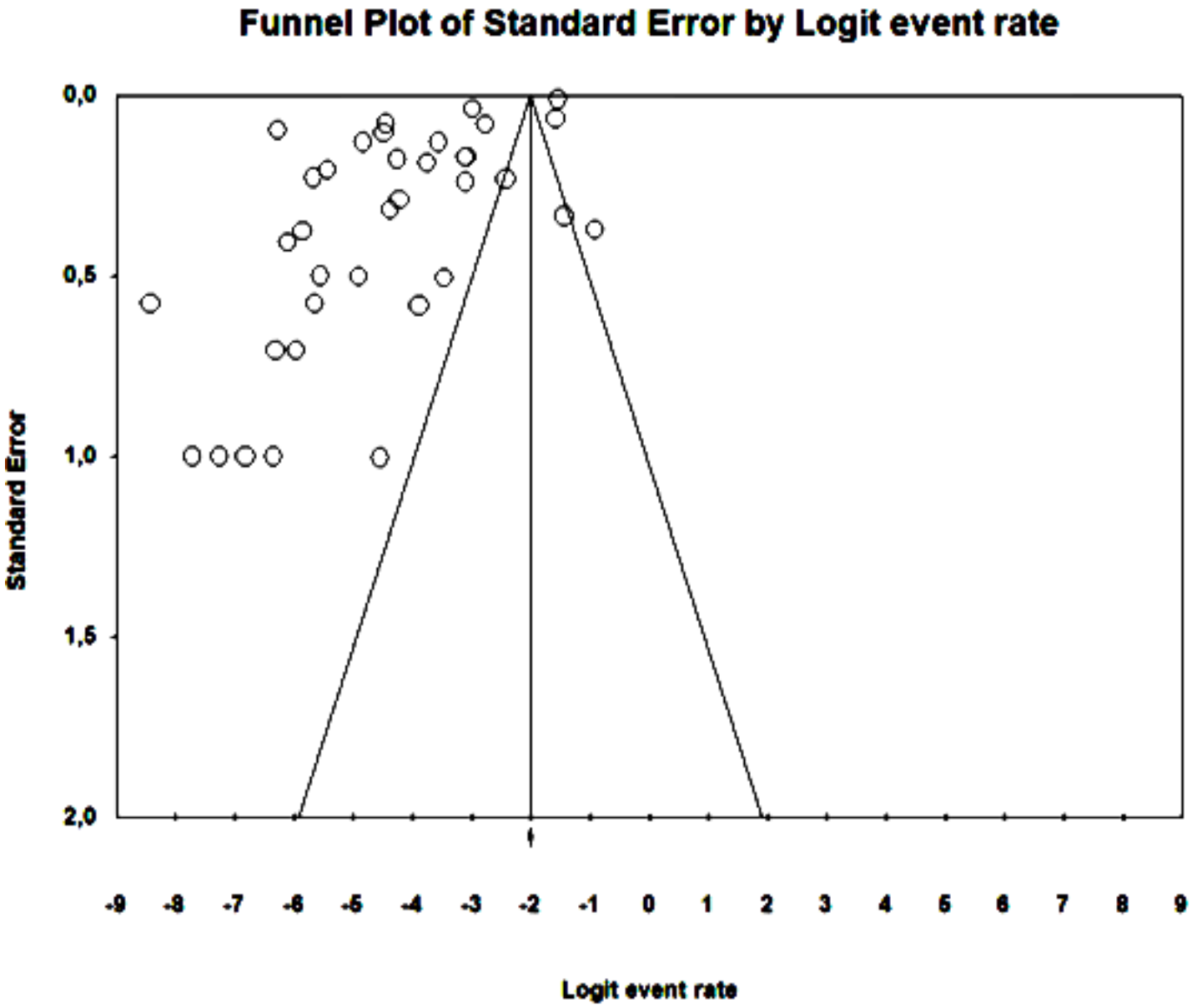


Figure 3-9: The funnel plot of *Tabanus autumnalis* (Linnaeus, 1761) studies included in the meta-analysis.

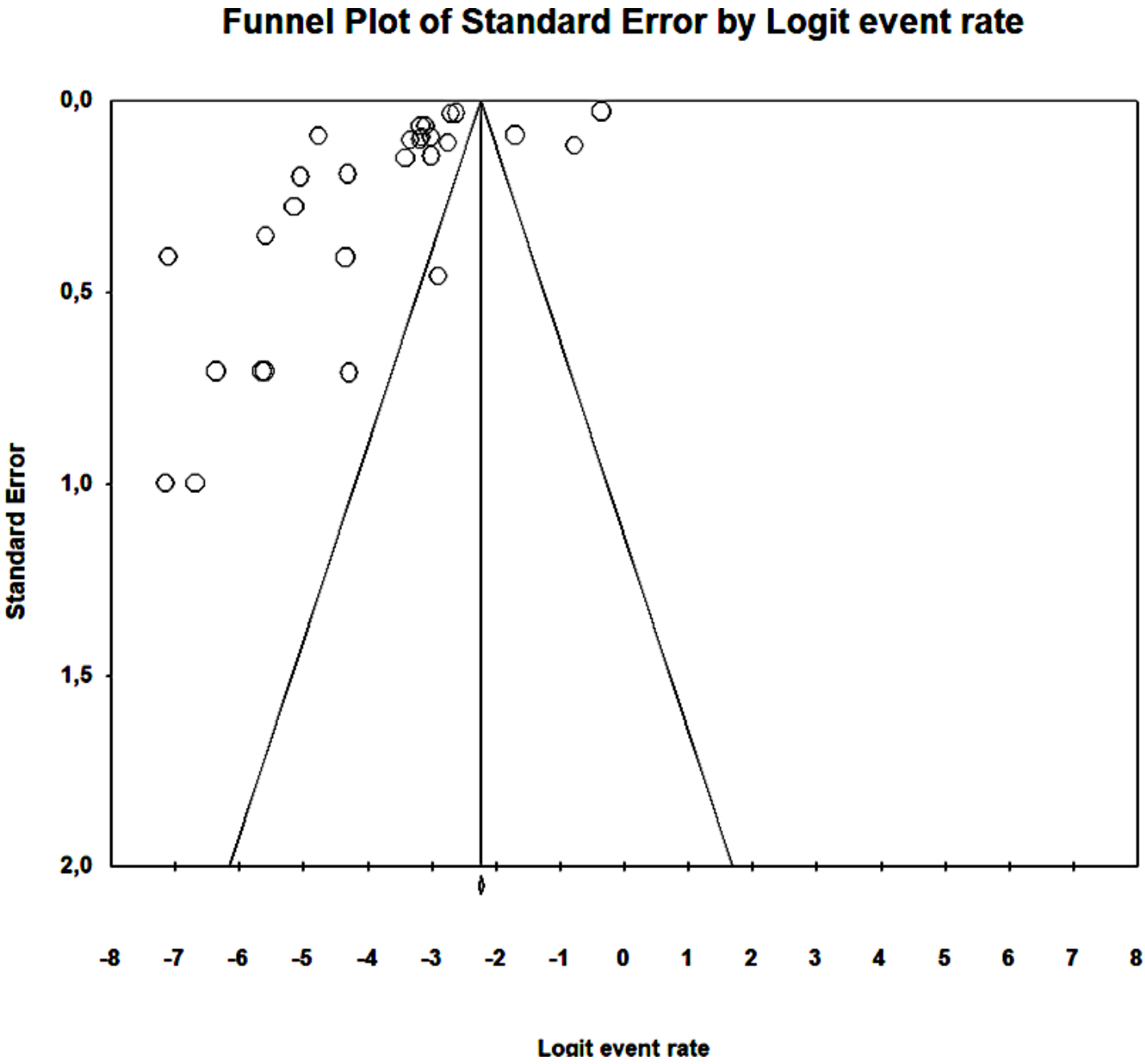


Figure 3-10: The funnel plot of *Tabanus maculicornis* (Zetterstedt, 1842) studies included in the meta-analysis.

3.5 Discussion

The aim of this systemic and meta-analysis was to collect information on the occurrence and worldwide distribution of tabanid flies reported from 1960-2022 resulting in a total of 139 studies. Most of these articles were published at a higher rate between the years 1960-1990 and 2011-2022. With the discovery of new tabanid fly species in the 1950s, the classification of subfamilies and tribes has expanded dramatically (Morita, 2008), leading to successful distinguishing between species in a variety of dipterans (Cywinska *et al*., 2010; Banerjee *et al*., 2015; Morita *et al*., 2016; Changbunjong *et al*., 2018; Mugasa *et al*., 2018). Hence, the upsurge in research on tabanid flies between 2011 and 2022 which was observed in the current review.

Although tabanid flies are the most diverse and extensively distributed pests, the range of significant tabanid species varies both geographically and temporally (Foil and Hogsette, 1994). The countries with the highest reported PPE, according to the review, were the United States, Canada, Croatia, Bulgaria, and Thailand. The least documented was reported in Cameroon, France, Japan and Korea. In contrast, various studies have been conducted reporting a higher diversity of tabanids in Cameroon (Lendzele *et al*., 2019; Sevidzem *et al*., 2021; Lendzele *et al*., 2022), Japan (Yang, 2017, Suganuma *et al*., 2024) and Korea (Choi *et al.,* 2020). Regarding studies reported in Japan and Korea, Suh *et al*., (2015) explained that low winter temperatures and precipitation have a greater impact on adult tabanid populations as compared to high temperatures do during the entire adult stage of the fly. Habitat differences account for differences in relative abundance and diversity of tabanids. There were no reports on prevalence or occurrence of tabanid flies in Antarctica and Iceland region which was expected. To date, no horseflies have been captured in most of Antarctica's Far North and South regions, though two insect species, *Belgica antarctica* Jacobs (Diptera: *Chironomidae*) and *Parochlus steinenii* Gerke (Diptera: *Chironomidae*), are native to Antarctica and are strictly limited to parts of the maritime Antarctic (Kozeretska *et al*., 2022).

Despite the widespread distribution, Tabanidae classification has historically been challenging (Morita *et al*., 2016), and their evolutionary relationships have remained poorly understood until recently. Currently, morphological keys and literature from more than 50 years ago are used to identify tabanids (Oldroyd, 1954). The extensive works of Oldroyd from the 1950s (Oldroyd, 1954, 1957, 1952) contain the most comprehensive

keys and description for African tabanid fly species. Due to their outdated status, these publication descriptions are frequently imprecise and unclear. Modifications have been made to make it simpler to identify tabanid flies morphologically and have been implemented for accurate identification which is a key component in determining the abundance of a tabanid species (Smit, 2019). In this review the most utilized article for morphological identification of tabanids was Chvala *et al*., (1972) followed by Olsufjev, (1977). Despite being renowned for describing a number of tabanid species in the early 1990s, Oldroyd, (1954 and 1957) article has been used by n=7 studies, whilst Oldroyd, (1952) was least utilized with only n = 6 studies which cited it as analyzed by the current review.

Several trapping methods are used to monitor the diversity and abundance of tabanid species. Multiple trapping techniques are employed for capturing tabanid flies since different species react differently to different types of traps. In the current study, sweep nets and Malaise traps were reported as the most utilized methods by n = 49 and n=27 studies respectively. Similarly, to studies by Stickler and walker, (1993), Krčmar *et al*., (2005), and Ringrose *et al*., (2014) where sweep nets yielded more tabanid flies (98.3%) than Malaise traps (1.7%). The third most frequently reported trap was the Nzi trap, accounting for 22% of the total. In contrast, Nzi traps were found to be more effective than Malaise traps, capturing 70.2% of the total tabanid specimens, compared to 28.9% captured by Malaise traps (Krčmar,2017). Additionally, studies conducted in Scandinavia, Sweden (Vaduva, 2015), as well as in Africa, North America, and Australia (Mihok, 2002; Mihok and Carlson, 2021; Van Hennekeler, *et al*., 2008; Baldacchino, *et al*., 2014 b) have demonstrated the efficiency of Nzi traps. According to Mihok, (2002) evaluation, Nzi traps have been found to be an effective method of capturing tabanids because they use vast amounts of blue cloth to attract flies and tiny black panels to induce fly landing. However, in this report Nzi traps were the 3rd least reported. The reason being that tabanids response to traps differs among species and locations (Chvála, *et al*., 1972).

Over 1,300 of the 4,455 Tabanidae species that have been described globally belong to the genus *Tabanus* (Fu *et al*., 2021). In this review, the overall number of members from genus *Tabanus* (86 species) was greater than all the other genera. According to Essono *et al*., (2015), genus *Tabanus* is known to be prevalent in a variety of biotopes and at various seasons. Within this genus, *T. occidentalis* was the most abundant species with

PPE of 46%. They widely distributed in the Neotropics (Baldacchino *et al*., 2014). In line with findings from a study by de Oliveira *et al*., (2023) with a prevalence of 40%. However, in contrast a study by Guimaraes *et al*., (2017) reported a prevalence of 6.62%. Various factors play a role in abundance of tabanid species globally, this can be due to the different climatic conditions in the different prospection countries.

Genus *Chrysops* (deer fly) also belongs to family Tabanidae and is one of the three genera of veterinary and medical importance. Chrysopini (338 species) is dominated by *Silvius* (40 species) and *Chrysops* (286 species). The current review reported 20.1% prevalence of genus *Chrysops* with *C. excitans* 7.6.% being the most prevalent species. This genus was the second most abundant species after *Tabanus*. According to Mazibuko, (2018) there is a scarcity of *Chrysops* with only one sepcies collected in a study in the Southern Africa. Usher's, (1972) assessment revealed that southern Africa has a comparatively low diversity of *Chrysops*, except for *C. obliqufaciatus* which was reported in Kruger National Park. However, this current review reported on the worldwide distribution, indicating that the genus *Chrysops* is more widespread in other countries including Cameroon, Nigeria, Equatorial Guinea, Democratic Republic of Congo and Rwanda.

There are at least 240 described species of genus *Hybomitra* which have been relatively well studied (Lutta, 1970; Olsufjev, 1977; Merz and Haenni, 2000). These are commonly called the hairy-eyed flies. Tribe Haematopotini (565 species.) consists primarily of *Haematopota*, the second biggest genus in family Tabanidae, sometimes known as 'clegs'. Except for the Australian and Neotropical regions, these ornately patterned flies with large faces and elongated antennae, have a distinctly mottled wing. *Haematopota* are known to be persistent biters and may specialize on infesting primates as adults (Lehane and Lehane, 1991).This tribe also consists of the genus *Heptatoma* Meigen, 1803. In this review, genera *Hybomitra* (19.7%) and *Haematopota* (9.1%) were reported as the third most prevalent. Similar to a study by Liebisch, (1987).,where *Haematopota pluvialis* 5.6% was the most prevalent in genus *Haematopota*. These findings are in congruence with a study conducted in Croatia by Krčmar *et al*., (2005), but with a higher frequency of 32.92% reported. Similar prevalence of *H. pluvialis* was reported in Europe (Dörge *et al*., 2020). It is considered as a Palearctic species, meaning that it occurs in the

largest region on earth covering areas of Europe, Central Asia (up to Pakistan and Kazakhstan) and Siberia.

Atylotus was first proposed as a subgenus of *Tabanus* by Osten (1875). There are about 70 described species in *Atylotus*. According to Wiliams *et al*., (2022), the five most frequently reported species from the Afrotropics are *A. agrestis*, *A. albipalpus* Walker, 1850*, A. diurnus*, *A. fuscipes*, and *A. nigromaculatus* Ricardo, 1900. *A. fuscipes* and *A. albipalpus* were synonymized in 1976 (Moucha, 1976) but this synonymy appears to have been largely neglected because both names are commonly used (Keita *et al*., 2020; Mugasa *et al*., 2018; Taioe *et al*., 2017). In this review*, Atylotus* species (4.8%) was represented by a smaller number of specimens as compared to (*Tabanus*, *Chyrysops*, *Hybomitra* and *Haematopota*). Similar findings were reported by Pechuman and Dearborn, (1996), who reclassified 10 species that were discovered on the savanna under the genus *Atylotus*. Results from the review identified *A. miser* (2.3%) as the most abundant species. The abundance is noticeably less than the 60% and 80% reported by Rageau *et al*. (1995) and Ovazza *et al*. (1970) respectively. Research has shown that, in contrast to *A. agretis* and *A. fuscipes,* which are typically found in Africa (Williams *et al*., 2022), the *A. miser* is more frequent in Asian countries, specifically China (Guo *et al*., 2018 and Fu *et al*., 2021) and Korea (Suh *et al*., 2015). These particular nations were found to have the fewest studies in the review, which may account for the low number of species.

The Diachlorini tribe (905 species) are horseflies with specialized adult morphology, making identification challenging (Coscarón and Papavero, 1998). The number of genera is unclear due to inconsistent use of subgenera by authors (Coscarón and Papavero, 1993; Trojan, 1994). Diachlorines are found worldwide but are most diverse in the Australasian and Neotropical regions. Numerous genera have been reported in this review from the Diachlorini tribe (4.3%, 9/209), which is regarded as the most morphologically diverse and genus-rich tribe in the Tabanidae family (Bayless *et al*., 2012). However, the varied use of subgenera by authors makes the number of genera questionable (Morita *et al*., 2008). In this review five species were identified, namely, *Philipomyia* (2 species), *Chlorotabanus* (1 species), *Leucotabanus* (1), *Dasyrhamphis* (4 species) and *Diachlorus* (1 species). *Philipomyia aprica* (2.8%) was the most prevalent in this tribe.

The horsefly genus *Therioplectes* is made up of 12 species that form a small genus group within the tribe Tabanini. This genus is limited to the Palaearctic region, with its centre in the eastern Mediterranean. The species fly in the spring and are seldom common. *Ancala* genus of horse flies (Diptera: Tabanidae) is a small group which includes only 8 known species. Other species identified from the systematic review but not in the meta- analysis included *A. nilotica* and *Euancala irrorata* (Surcouf, 1909). *Poeciloderas* is a genus of horse flies consisting of 9 species. Of all the reported genus the following had the least reported species: *Philoliche* (2 species). *Ancala* (2 species), *Heptatoma* (1), *Silvius* (1 species), *Poeciloderas* (1 species) and *Therioplectes* (3 species). The most prevalent species within these genera were *P. aethiopica* (74.0%), *A. fasciata* (10%), *S. alpinus* (0.5%), *Heptoma pellucen* (3%), *P. quadripunctatus* (43%) and *T. tunicatus* (0.6%), respectfully. Comparable outcomes were found in Al-talafha and Amr's (2024) study in the Middle East, reporting: *Silvius* (2 species), *Philoliche* (5 species), *Ancala* (2 species), and *Therioplectes* (2 species).

3.6 Limitation

Several limitations have been identified which include the following: (1) majority of the studies included samples between 1960 and 2022; (2) countries represented included only Brazil, Bulgaria, Cameroon, Canada, Croatia, Gabon, India, Iran, Japan, Korea, Nigeria, South Africa, Thailand, Turkey and the United State of America. Whilst 27 countries were underrepresented having one or two studies reported. (3) some studies reported the total number of tabanids without showing the number of individual species. (4) the studies included in the analyses showed a high level of heterogeneity, hence readers should exercise caution when interpreting the pooled analysis; (5) the bias assessment revealed the overall risk of bias from pooled studies; (6) certain article were omitted particularly studies from non-English publications or unpublished sources are excluded.(7) Limited access to raw data from primary studies restricted the ability to perform more detailed analyses or check the accuracy of reported results.

3.7 Conclusion

This systematic review and meta-analysis provided a comprehensive overview of the global distribution and taxonomic classification of tabanid flies (family Tabanidae). The findings highlight the broad geographical presence of these flies, with confirmed reports

from diverse regions including North America, Europe, Asia, and Africa. The prevalence of studies between 2011 and 2022 reflects the growing application of modern molecular techniques in entomological research. Chvala *et al*., (1972) emerged as the most frequently cited reference for taxonomic identification, underscoring its continued relevance in the field. Additionally, the sweep net was identified as the most employed method for collecting tabanid specimens. Notably, the *Tabanus* genus represented the largest number of species among the recorded tabanids. This review not only consolidates existing knowledge but also emphasizes the importance of continued surveillance and research on tabanid flies, given their ecological and potential epidemiological significance.

3.8 Recommendations

Considering the systematic review approach, additional research should be conducted in underexplored areas to gain a fuller insight into the worldwide distribution of tabanids. Additionally, upcoming research could explore Tabanids' ecological and behavioural characteristics to enhance pest management strategies and their roles within ecosystems. In the current age of technological advancements more refined molecular technologies can be utilized to enhance the precision of taxonomic classification and species identification.

3.9 References

1. Al-Talafha, H.A. & Amr, Z. (2024). Tabanidae Fauna (Order: Diptera) of the Arab Countries in the Middle East. *Ecologica Montenegrina*, *71*, pp.53-85.
2. Baldacchino, F., Desquesnes, M., Mihok, S., Foil, L.D., Duvallet, G. & Jittapalapong, S. (2014). Tabanids: neglected subjects of research, but important vectors of disease agents!. *Infection, Genetics and Evolution*, *28*, pp.596-615.
3. Banerjee, D., Kumar, V., Maity, A., Ghosh, B., Tyagi, K., Singha, D., Kundu, S., Laskar, B.A., Naskar, A. & Rath, S. (2015). Identification through DNA barcoding of Tabanidae (Diptera) vectors of surra disease in India. *Acta Tropica*, *150*, pp.52-58.
4. Barbosa, A.D., Gofton, A.W., Paparini, A., Codello, A., Greay, T., Gillett, A., Warren, K., Irwin, P. & Ryan, U. (2017). Increased genetic diversity and prevalence of co-infection with *Trypanosoma* spp. in koalas (*Phascolarctos cinereus*) and their ticks identified using next-generation sequencing (NGS). PloS one, 12(7), p.e0181279.
5. Bayless, K. M. (2012). Evolutionary Studies in Tabanomorpha (Diptera) with New Classifications of Xylophagidae and Tabanidae. http://www.lib.ncsu.edu/resolver/1840.16/7902 (Access on the 26/07/2014)Access on the 26/07/2014)
6. Begg, C.B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, pp.1088-1101.
7. Burton, J.J. (1978). *Tabanini of Thailand above the Isthmus of Kra (Diptera:* Tabanidae*)* (pp. 165-pp).
8. Chalchisa, A. & Bersissa, E. (2023). Ectoparasites and their effect on camels (Camelus dromedarius) in Ethiopia. *J Vet Res Adv*, *5*(01), pp.55-69.
9. Changbunjong, T., Sedwisi, P., Weluwanarak, T., Nitiyamatawat, E., Sariwongchan, R. & Chareonviriyaphap, T. (2018). Species diversity and abundance of Tabanus spp.(Diptera: Tabanidae) in different habitats of Thailand. *Journal of Asia-Pacific Entomology*, *21*(1), pp.134-139.
10. Choi, K.S., Kim, H.C., Chong, S.T., Kim, M.S., Klein, T.A., Kim, H.S. & Suh, S.J., (2020). Seasonal surveillance of deer and horse flies (Diptera: Tabanidae), Gyeonggi province, Republic of Korea. Journal of Asia-Pacific Entomology, 23(2), pp.315-319.
11. Chvála, M., Lyneborg, L. & Moucha, J. (1972). The Horse Flies of Europe (Diptera, Tabanidae).

12. Coscarón, S. & Papavero, N. (1993). An illustrated manual for the identification of the neotropical genera and subgenera of Tabanidae *(Diptera)* (pp. 150-pp).
13. Coscarón, S., Mancebo, O.A. and Coscarón-Arias, C.L., 1998. The preimaginal stages of Cryptotylus unicolor (Wiedemann) and Tabanus nebulosus ornativentris Kroeber (Tabanidae-Diptera-Insecta). *Memórias do Instituto Oswaldo Cruz*, *93*, pp.91-97.
14. Cywinska, A., Hannan, M.A., Kevan, P.G., Roughley, R.E., Iranpour, M. & Hunter, F.F. (2010). Evaluation of DNA barcoding and identification of new haplomorphs in Canadian deerflies and horseflies. *Medical and Veterinary Entomology*, *24*(4), pp.382-410.
15. De Liberato, C., Magliano, A., Autorino, G.L., Di Domenico, M., Sala, M. & Baldacchino, F. (2019). Seasonal succession of tabanid species in equine infectious anaemia endemic areas of Italy. *Medical and veterinary entomology*, *33*(3), pp.431-436.
16. de Oliveira Zamarchi, T.B., Henriques, A.L., Krolow, T.K., Krüger, R.F., Rodrigues, G.D., Munari, A. & Pessoa, F.A.C. (2023). Tabanidae (Diptera) captured on horses in Amazon Forest fragments of the state of Rondonia, Brazil. *Acta Tropica*, *237*, p.106734.
17. Dörge, D.D., Cunze, S. & Klimpel, S. (2020). Incompletely observed: niche estimation for six frequent European horsefly species (Diptera, Tabanidae). *Parasites & Vectors*, *13*, pp.1-10.
18. Essono, P.Y.B., Dechaume-Moncharmont, F.X., Mavoungou, J., Mba, R.O., Duvallet, G. & Bretagnolle, F. (2015). Distribution and abundance of hematophagous flies (*Glossinidae*, *Stomoxys*, and Tabanidae) in two national parks of Gabon. *Parasite*, *22*.
19. Foil, L.D. & Hogsette, J.A. (1994). Biology and control of tabanids, stable flies and horn flies. *Revue scientifique et technique-Office international des épizooties*, *13*, pp.1125-1125.
20. Foil, L.D. (1989). Tabanids as vectors of disease agents. *Parasitology today*, *5*(3), pp.88-96.
21. Fu, Y.T., Zhang, Y., Xun, Y., Liu, G.H. & Zhao, Y., (2021). Characterization of the complete mitochondrial genomes of six horseflies (Diptera: Tabanidae). *Infection, Genetics and Evolution*, *95*, p.105054.

22. Guimaraes, R.R., Júnior, R.R.G., Harlan-Rodrigues, R.S., Guimarães, R.R., Carvalho, R.W. & de Xerez, R. (2017). Seasonality and Relative Abundance of tabanids (Diptera, Tabanidae) on Marambaia Island, Rio de Janeiro, Brazil. *EntomoBrasilis*, *10*(2), pp.82-88.
23. Guo XiaoXia, G.X., Li ChunXiao, L.C., Wang Gang, W.G., Zheng Zhong, Z.Z., Dong YanDe, D.Y. & Zhao TongYan, Z.T. (2018). Species diversity and biting habit of horsefly in Heixiazi Island of Heilongjiang Province, China.
24. Karolyi, F., Colville, J.F., Handschuh, S., Metscher, B.D. & Krenn, H.W. (2014). One proboscis, two tasks: Adaptations to blood-feeding and nectar-extracting in long-proboscid horse flies (Tabanidae, *Philoliche*). *Arthropod structure & development*, *43*(5), pp.403-413.
25. Keita, M.L., Medkour, H., Sambou, M., Dahmana, H. & Mediannikov, O. (2020). Tabanids as possible pathogen vectors in Senegal (West Africa). *Parasites & Vectors*, *13*, pp.1-15.
26. Kozeretska, I., Serga, S., Kovalenko, P., Gorobchyshyn, V. & Convey, P., 2022. Belgica antarctica (Diptera: *Chironomidae*): A natural model organism for extreme environments. *Insect Science*, *29*(1), pp.2-20.
27. Krčmar, S. (2017). Comparative efficiency of three trap types for collecting host-seeking female tabanid flies (Diptera: Tabanidae). *Journal of entomological science*, *52*(1), pp.82-86.
28. Krcmar, S., Merdic, E. & Kopi, M. (2005). Diurnal periodicity in the biting activity of horsefly species in the Kopacki rit Nature Park, Croatia (Diptera: Tabanidae). *Entomologia generalis*, *28*(2), p.139.
29. Lehane, M.J. & Lehane, M.J. (1991). The blood-sucking insect groups. *Biology of Blood-Sucking Insects*, pp.193-247.
30. Lendzele, S.S., Abah, S., Nguetoum, C., Burinyuy, K.A., Koumba, A.A. & Mavoungou, J.F. (2022). Tabanid-transmitted animal trypanosomiasis in Cameroon: Evidence from a study in the tsetse free pastoral zone of Galim. *Parasite Epidemiology and Control*, *18*, p.e00253.
31. Lendzele, S.S., François, M.J., Roland, Z.K.C., Armel, K.A. & Duvallet, G. (2019). Factors influencing seasonal and daily dynamics of the genus Stomoxys Geoffroy, 1762 (Diptera: Muscidae), in the Adamawa Plateau, Cameroon. *International Journal of Zoology*, *2019*(1), p.3636943.

32. Liebisch, A. (1987). Vector biology of flies on grazing cattle in Germany. In *Summer Mastitis: A Workshop in the Community Programme for the Coordination of Agricultural Research, held at the Central Veterinary Institute, Lelystad, The Netherlands, 23–24 October 1986 Sponsored by the Commission of the European Communities, Directorate General for Agriculture, Coordination of Agricultural Research* (pp. 109-115). Dordrecht: Springer Netherlands.
33. Lucas, M., Krolow, T.K., Riet-Correa, F., Barros, A.T.M., Krüger, R.F., Saravia, A. & Miraballes, C. (2020). Diversity and seasonality of horse flies (Diptera: Tabanidae) in Uruguay. *Scientific reports*, *10*(1), p.401.
34. Lutta, A.S., 1970. Tabanids of Karelia.
35. Mazibuko, X. (2018). *Species composition of and detection of haemoparasites in horse-flies (Diptera:* Tabanidae*) in south-eastern Kruger National Park*. University of Pretoria (South Africa)
36. Merz, B. & Haenni, J.P., 2000. Morphology and terminology of adult Diptera (pp. 21-51). *Contributions to a manual of Palaearctic Diptera*, *1*.
37. Mihok, S. & Carlson, D.A. (2021). New materials for improving catches of horseflies (Diptera: Tabanidae) in Nzi traps. *Medical and Veterinary Entomology*, *35*(4), pp.580-594.
38. Mihok, S. (2002). The development of a multipurpose trap (the Nzi) for tsetse and other biting flies. *Bulletin of Entomological Research*, *92*(5), pp.385-403.
39. Morita, S.I. (2008). A phylogeny of long-tongued horse flies (Diptera: Tabanidae: Philoliche) with the first cladistic review of higher relationships within the family. *Invertebrate Systematics*, *22*(3), pp.311-327.
40. Morita, S.I., Bayless, K.M., Yeates, D.K. & Wiegmann, B.M. (2016). Molecular phylogeny of the horse flies: a framework for renewing tabanid taxonomy. *Systematic entomology*, *41*(1), pp.56-72.
41. Moucha J. (1976). Horse-flies (Diptera, Tabanidae) of the World. Národní muzeum v Praze, Přírodovědecké muzeum, Entomologické oddělení; 7..
42. Mugasa, C.M., Villinger, J., Gitau, J., Ndungu, N., Ciosi, M. & Masiga, D. (2018). Morphological re-description and molecular identification of Tabanidae (Diptera) in East Africa. *ZooKeys*, (769), p.117.
43. Mullens, B.A. (2019). Horse flies and deer flies (Tabanidae). In *Medical and veterinary entomology* (pp. 327-343). Academic Press.

44. Oldroyd H. (1954). The horse-flies (Diptera, Tabanidae) of the Ethiopian region. Vol. II: *Tabanus* and related genera. British Museum (Natural History), London; pp 341.
45. Oldroyd, H. (1952). The Horse-Flies (Diptera: Tabanidae) of the Ethiopian Region. Vol. I. *Haematopota* and *Hippocentrum*.
46. Oldroyd, H. (1957). The Horse-flies (Diptera: Tabanidae) of the Ethiopian Region. Volume III. Subfamilies *Chrysopinae*, *Scepsidinae* and *Pangoniinae* and a revised Classification.
47. Olsufjev N., (1977). Tabanidae. In: Fauna SSSR. 7, No 2, 1434, Nauka, Leningrad (RU).
48. Osten-Sacken, C.R. (1875). Prodrome of a Monograph of the Tabanidae of the United States: supplement:2, pp.365-397.
49. Ovazza, M., Mouchet, J. & Rickenbach, A. (1970). Tabanidae of Cameroon: II. Known species and their distribution. Cah. ORSTOM, ser Ent mét et Parasitol. 8(1): p.323-351.
50. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. & Chou, R. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, *372*.
51. Pechuman, L.L. & Dearborn, R. (1996). The horse flies and deer flies of Maine (Diptera, Tabanidae).
52. Rageau, J., Grenier, P. & Adam, J.P. (1955). Tabanidae du Cameroun français. *Annales de Parasitologie humaine et comparée*, *30*(3), pp.243-271.
53. Ramatla, T., Tawana, M., Mphuthi, M.B., Onyiche, T.E., Lekota, K.E., Monyama, M.C., Ndou, R., Bezuidenhout, C. & Thekisoe, O. (2022). Prevalence and antimicrobial resistance profiles of Campylobacter species in South Africa: a “One Health” approach using systematic review and meta-analysis. *International Journal of Infectious Diseases*, *125*, pp.294-304.
54. Ringrose, J.L., Abraham, K.F. & Beresford, D.V. (2014). New range records, and a comparison of sweep netting and Malaise trap catches of horse flies and deer flies (Diptera: Tabanidae) in northern Ontario. The Journal of the Entomological Society of Ontario, 145.
55. Schowalter, T.D. (2022). *Insect ecology: an ecosystem approach*. Academic press.

56. Sevidzem, S.L., Koumba, A.A., Yao-Acapovi, G.L. & Mavoungou, J.F. (2021). A nationwide survey of the tabanid fauna of Cameroon. *Parasites & Vectors*, *14*(1), p.392.
57. Skendžić, S., Zovko, M., Živković, I.P., Lešić, V. & Lemić, D. (2021). The impact of climate change on agricultural insect pests. *Insects*, *12*(5), p.440.
58. Smit, A. (2019). *Abundance, composition and barcoding of* Tabanidae *in Kruger National Park and screening for Besnoitia besnoiti*. University of Pretoria (South Africa).
59. Sonenshine, D.E. (2006). Tick pheromones and their use in tick control. *Annual Review of Entomology*, *51*(1), pp.557-580.
60. Sontigun, N., Boonhoh, W., Phetcharat, Y. & Wongtawan, T. (2022). First study on molecular detection of hemopathogens in tabanid flies (Diptera: Tabanidae) and cattle in Southern Thailand. *Veterinary World*, *15*(8), p.2089.
61. Strickler, J.D. & Walker, E.D. (1993). Seasonal abundance and species diversity of adult Tabanidae (Diptera) at Lake Lansing Park-North, Michigan. *The Great Lakes Entomologist*, *26*(2), p.3.
62. Suh, S.J., Kim, H.C., Chong, S.T., Kim, M.S. & Terry, A. (2015). Seasonal abundance of deer and horse flies (Diptera: Tabanidae) in the northern part of Gyeonggi-do, Republic of Korea. *The Korean journal of parasitology*, *53*(3), p.307.
63. Surcouf, J. M., & Ricardo, G. (1909). Étude monographique des tabanides d'Afrique:(groupe des Tabanus). Masson et cie.
64. Taioe, M.O., Motloang, M.Y., Namangala, B., Chota, A., Molefe, N.I., Musinguzi, S.P., Suganuma, K., Hayes, P., Tsilo, T.J., Chainey, J. & Inoue, N. (2017). Characterization of tabanid flies (Diptera: Tabanidae) in South Africa and Zambia and detection of protozoan parasites they are harbouring. *Parasitology*, *144*(9), pp.1162-1178.
65. Tawana, M., Onyiche, T.E., Ramatla, T. & Thekisoe, O. (2023). A 'One Health'perspective of Africa-wide distribution and prevalence of *Giardia* species in humans, animals and waterbodies: a systematic review and meta-analysis. *Parasitology*, *150*(9), pp.769-780.
66. Trojan, P. (1997). Distribution pattern of *Diachlorini* s. str.(Diptera: Tabanidae). *Fragmenta faunística*, *40*(6), pp.59-71.
67. Usher, P.J. (1972). A review of the South African horsefly fauna (Diptera: Tabanidae). *Annals of the Natal Museum*, *21*(2), pp.459-507.

68. Vaduva, G. (2020). Behavioral responses of Tabanidae and *Stomoxys calcitrans* to unbaited and baited Nzi and horse pal traps in southern Sweden. *Journal of Biology and Life Science*, *11*(2).
69. Van Hennekeler, K., Jones, R.E., Skerratt, L.F., Fitzpatrick, L.A., Reid, S.A. & Bellis, G.A. (2008). A comparison of trapping methods for Tabanidae (Diptera) in North Queensland, Australia. *Medical and Veterinary Entomology*, *22*(1), pp.26-31.
70. Williams, K.A., Smit, A., Neves, L. & Snyman, L.P. (2022). Evaluating the morphological and molecular challenges in identifying the afrotropical *Atylotus* species (Diptera: Tabanidae). *Acta Tropica*, *226*, p.106262.
71. YANG, C. (2017). Female genitalia of some tabanids inhabiting Hokkaido, Japan (Diptera: Tabanidae), Hokkaido, Jpapan.

CHAPTER 4 MORPHOLOGICAL AND GENETIC CHARACTERISTICS OF *TABANUS PAR* FROM SELECTED LOCALITIES IN SOUTH AFRICA

4.1 Introduction

There are about 4455 species of Tabanidae tabanid flies known to exist worldwide, with more than 1300 of those species occurring in the genus *Tabanus* (Changbunjong *et al., 2021*). Morphological characterization of insects primarily relies on the examination of head structures, including the callus, antennae, eyes, frons, and beard, as well as the colour and patterns of the body, legs, and wings (Changbunjong *et al*., 2021). While morphometrics include various methods such as landmark, semi-landmark, and outline-based methods, which are used for performing geometric morphometric analysis, depending on the characteristics and specifics of the specimens (Nazri *et al*., 2020).

The notoriously challenging morphological identification of Tabanidae may account for this group's marginalisation despite their vectoral status. In the 1950s, Oldroyd produced the most thorough keys and descriptions for the African Tabanidae (Oldroyd, 1952, 1954, 1957). Owing to their age, these books do not include any colour images, and the line drawings are frequently hard to understand. Additionally, the descriptions are frequently ambiguous and unclear. It is quite easy to misidentify these species if one does not have access to recognised specimens, such as those in museum collections comparison. Several evolutionary trees based on the *COI* gene have been established by recent Tabanidae studies conducted in Africa (Keita *et al*., 2020; Mugasa *et al*.,2018)

Geometric morphometrics is known to be effective for species identification in various models, including blow flies (Sontigun *et al*.,2017) mosquitoes (Martinet *et al*., 2021), atomoxetine flies (Changbunjong *et al*., 2018), and tsetse flies (Kaba *et al*., 2017). However, Changbunjong *et al*., (2021) pointed out the difficulties in distinguishing closely related *Tabanus* species based solely on morphological characteristics, underscoring the efficacy of molecular methods such as DNA barcoding for precise identification (Changbunjong *et al*., 2021). Similarly, Banerjee *et al*., (2015), demonstrated the utility of DNA barcoding in identifying Tabanidae vectors of surra disease in India, highlighting the genetic divergence within *Tabanus* species and the implications for vector control

strategies. The genetic diversity observed in *Tabanus* populations, as noted by Sanal *et al*., (2021) further underscores the importance of molecular techniques in understanding the phylogeography and population genetics of these insects.

According to Desquesnes *et al*., (2009), the importance of tabanid flies as mechanical vectors is essentially linked to their size and abundance. Therefore, effective control of diverse vectored diseases can be aided by accurate identification of tabanids, which is crucial for monitoring their potential for invasion into naïve ecosystems and vector control strategies. In livestock rearing areas, veterinarians and veterinary technicians are facing field situations where it is necessary to identify these insects, to propose adapted control measures. This chapter sought to characterize South African *T. par* using a combination of morphological features, morphometrics and genetic analyses.

4.2 Materials and methods

4.2.1 Ethical clearance

The Integrated Pest Management Scientific Committee at North-West University and the Research Ethics Committee of the Faculty of Natural and Agricultural Sciences approved the research with reference no: NWU-01300-22-A9. The research was categorized as "No Risk".

4.2.2 Fly sample collections

The fly samples were collected using H-traps baited with a 1:8 ratio mix of Octenol and Methyl phenol (*p*-Cresol), as well as with 50 ml of acetone dispensed from a 250 ml glass bottle from the selected three game reserves in northeastern KwaZulu-Natal, South Africa. The traps were placed for one week at a minimum distance of more than 100 m apart in all three game reserves and monitored daily to collect the captured flies (Figure 4-1).



Figure 4-1: H-trap used to capture tabanid flies in South Africa (Taioe, 2017).

4.2.3 Sample collection site

The sampled game reserves included Charters Creek (28°13'37"S; 032°24'1"E), Hluhluwe-Imfolozi Park (28°9'50"S; 032°12'15"E) and Phinda Private Game Reserve (27°46'39"S; 032°20'57"E) all situated in the uMkhanyakude District Municipality (Figure 4-2).

4.2.3.1 Charters Creek

Charters Creek forms part of iSimangaliso Wetland Park, situated on the western shore near St. Lucia, KwaZulu-Natal province, South Africa. The ecosystem consists of a blend of wetlands, pans and forest regions, providing diverse habitats for various wildlife species. The Savanna vegetation experiences summer and winter rainfall with a maximum average precipitation (MAP) ranging from 800 mm to 1,050 mm, and frost occurs often during the winter (Mucina and Rutherford, 2006).

4.2.3.2 Hluhluwe–iMfolozi

Hluhluwe–iMfolozi Park is one of the oldest proclaimed nature reserves in Africa located in KwaZulu-Natal province, South Africa, covering an area of 960 km² (96,000 hectares). Established in 1895, the park is renowned for its rich biodiversity and conservation efforts. Home to the Big Five and other notable species including African Wild dogs, Cheetahs, Hyenas, Giraffes and various antelope species. The areas experience summer rainfall and light winter rain. The average annual precipitation varies from 500 mm to 900 mm, and the average monthly maximum and minimum temperatures are 38.5°C and 7.8°C respectively, with no occurrence of frost (Taioe, 2017).

4.2.3.3 Phinda Private Game Reserve

Phinda Private Game Reserve, located in KwaZulu-Natal, South Africa, covers an area of approximately 298.66 km² (29,866 hectares) and is positioned between the Mkhuze Game Reserve and iSimangaliso Wetland Park. It is home to the Big Five and boasts over 436 bird species, supporting a diverse range of wildlife and plant species. The region experiences similar temperatures to Phinda Private Game Reserve, falling under the Zululand lowveld vegetation type, which extends from south of Mkhuze, Hluhluwe**,** Ulundi, and small patches in the Mpumalanga Province, as well as in Swaziland (Taioe, 2017).

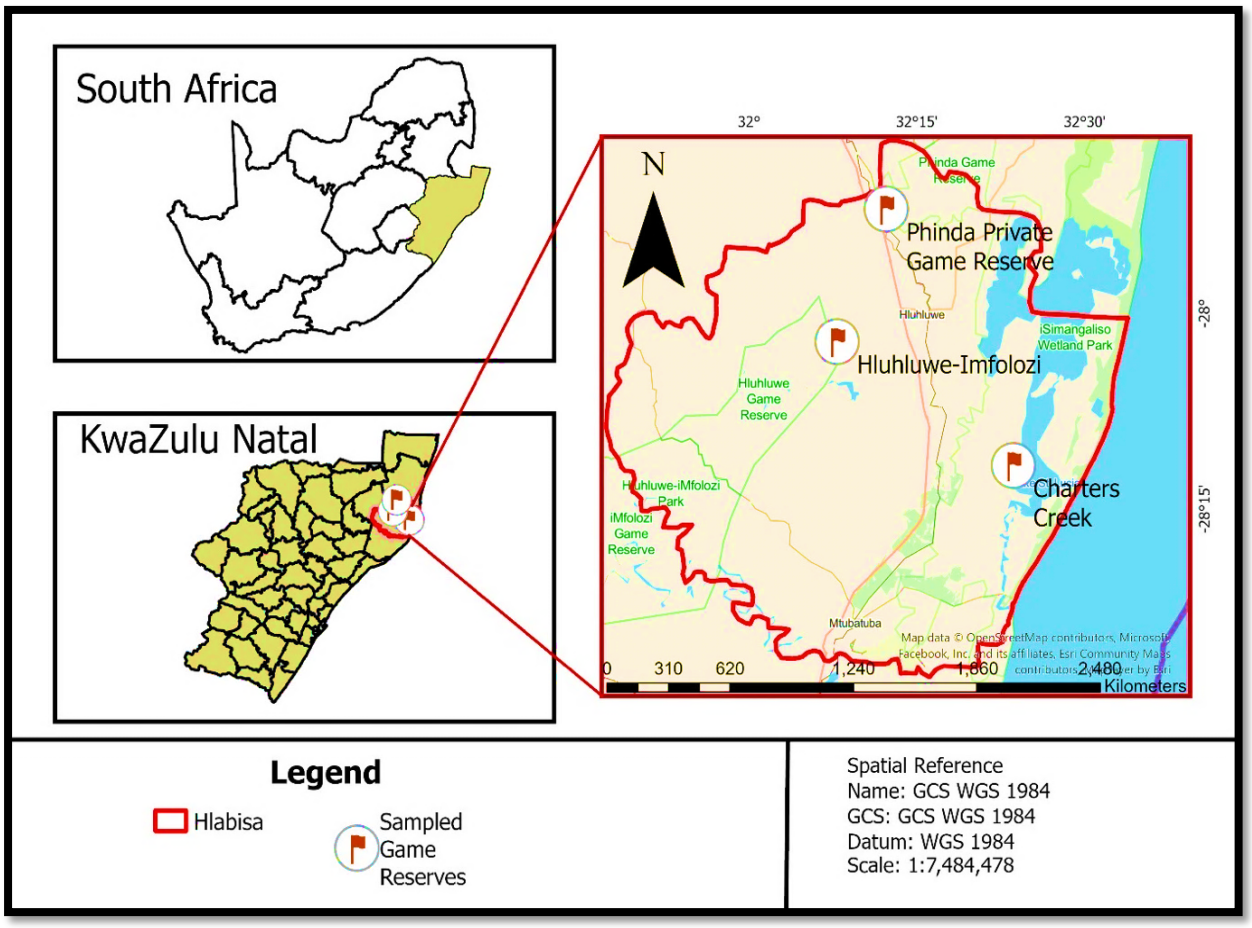


Figure 4-2: Map of southern Africa showing three sampled game reserves namely: Charters Creek, Hluhluwe–iMfolozi Park and Phinda Private Game Reserve.

4.3 Preparation of samples, pinning and drying

Collected samples were preserved in alcohol and removed as required for use. For morphological and morphometric analyses, the intact specimens with no damaged appendages were selected and pinned using 0.35 mm pins on the thorax to a Styrofoam. Each fly specimen was positioned with the wings spread and legs extended. The flies were subsequently left to dry and weighed before sequential images of each specimen were taken using the Nikon AZ 100 M multi-zoom microscope (Nikon Inc. in Tokyo, Japan). Morphological identification of *T. par.*

Species identification and selection of key morphological features to be measured were done using references from the literature of Oldroyd (1954), Yagi (1968), Veer (1999), Morita (2008), and Taioe (2017).

4.3.1 Morphometrical analysis

A total of 30 individual female *T. par* were used and each specimen was photographed from various angles using the program NIS Elements V4.0. Different body segments were captured and these included features such as complete Dorsal length, complete Ventral length, head ventral length, head Ventral width, thorax Dorsal length, thorax Dorsal breath, abdomen ventral widthE1, abdomen ventral widthG2, abdomen ventral length, wing left breath, wing right breath, wing length left, wing length right and the weight.

4.3.2 DNA extraction

Specimens were prepared according to Taioe, (2017) with minor modifications. All flies had their wings cut off to avoid having too many exoskeletons that could interfere with the enzymatic activity of the reaction steps. The specimens were cut into pieces in 1.5 ml microcentrifuge tubes using sterile scissors. Extraction was conducted using Qiagen DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The body and legs were then homogenized separately in a mixture containing 180 µl of Buffer ATL and 20 µl of proteinase K, vortexed, and incubated at 56°C until completely lysed. The incubated samples were vortexed occasionally during incubation. Thereafter, the samples were vortexed for an additional 15 seconds. Two hundred microliters of Buffer AL was added, mixed by vortexing, and incubated at 56°C for 10 minutes. Thereafter, 200 µl of absolute ethanol was added, and again, the samples were mixed by vortexing. Thereafter, the mixture was transferred into a DNeasy Mini spin column placed in a 2 ml collection tube, centrifuged at 8 000 rpm for 1 min, and the flow-through was discarded.

The 2 ml collection tube was replaced, and 500 µl of Buffer AW1 was added and centrifuged at 8 000 rpm again for a minute. The flow-through and the collection tube were discarded. The spin columns were transferred into a new 2 ml collection tube, 500 µl Buffer AW2 was added and centrifuged for 3 min at 14,000 rpm. The flow-through was discarded. DNA was eluted by transferring the spin column to a new 1.5 ml microcentrifuge tube and adding 200 µl Buffer AE to the centre of the spin column membrane. The solution was then incubated for 1 minute at room temperature (15–25°C)

and then centrifuged for 1 minute at ≥6000 x g. Thereafter extracted DNA was stored at –20°C until used.

4.3.3 DNA extraction Amplification of cytochrome oxidase 1 (*CO1*) gene

Polymerase chain reaction targeting the *CO1* gene region was conducted to genetically identify *T. par* species. This PCR was to supplement morphological analysis in characterizing the captured tabanid flies and to determine their phylogenetic position when compared to other related taxa.

The LCO1490: 5′-GGTCAACAAATCATAAAGATATTGG-3′ and HCO2189: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' primers which amplified approximately 653 base pairs (bp) of the target gene were used (Folmer *et al.*, 1994). The PCR contained a final volume of 25 µl which was composed of 12.5 µl OneTaq® Master Mix (OneTaq® 2X Master Mix with Standard Buffer [20 mM Tris/HCl pH 8.05, 22 mM KCl], 0.2 mM of each dNTP and 1.8 mM $MgCl_2$ and DNA polymerase) (New England Biolabs, Inc). Then 0.5 µM of each primer, 2 µl of template DNA and double distilled water was added to make the final volume. The PCR conditions were set as follows: initial denaturation at 94°C for 30 seconds, followed by 35 cycles of 94°C for 30 seconds, annealing at 68°C for *CO1* gene and 45 for *ITS2* region for 30 seconds, elongation at 72°C for 30 seconds and final elongation at 72°C for 1 minute. PCR products were electrophorized on a 1% agarose gel and visualised under UV light.

4.3.4 Purification of PCR amplicons

Positive PCR products were purified using the QIAquick® Gel Extraction Kit (QIAGEN, Hilden, Germany). The gels having the DNA fragments were cut with a clean sharp scalpel and the gel slices were placed into 1.5 ml Eppendorf tubes. Then 600 µl of QG buffer was added. To dissolve the sliced DNA fragments in QG buffer the Eppendorf tubes were heated at 50°C for 10 minutes on a heating block. The tubes were vortexed every 2 minutes during the incubation period. Then 200 µl of isopropanol was added to increase the DNA yield. The mixture was transferred into QIAquick spin column with a 2 ml collection tube. To bind the DNA, the spin columns were centrifuged for 1 minute at 13 000 rpm. The flow through in the collection tubes was discarded and an additional 500 µl

of QC buffer was added to the QIAquick column and centrifuged for 1 minute at 13 000 rpm. The purified DNA in the QIAquick column was then washed with 750 µl of PE buffer that contained 99% absolute ethanol and centrifuged again at 13 000 rpm for 1 minute. The flow through was discarded and the QIAquick column centrifuged to remove excess ethanol in the PE buffer. The QIAspin columns were transferred into 28 new clean 1.5 ml tubes and 50 µl of EB buffer was used to elute the DNA. The purified DNA samples were sent for Sanger sequencing at Inqaba Biotechnical Company, Pretoria, South Africa.

4.3.5 Sequence alignment and phylogenetic analysis

The obtained sequences were subjected to BLASTn to confirm species identification with deposited *T. par* sequences from the GenBank database. BLASTn results with an identity match score of 90-100% were considered significant and those matching sequences were subsequently downloaded for phylogenetic analysis. The DNA sequences from this study and those downloaded from GenBank were edited and aligned by ClustalW on MEGA 7 (Kumar *et al.*, 2016). The final alignment was trimmed to make even ends and the Best Model Fit was used to estimate the best substitution model to be used for constructing the phylogenetic tree. The General Time Reversible model was selected based on the results from the Best Model Fit and the Maximum likelihood phylogeny at 1000 bootstrap replicates was used to construct the final tree.

4.3.6 Statistical analysis

The morphometrics analysis for 3 fly traps was treated as a Randomized Block Design experiment. The data was subjected to an appropriate analysis of variance (ANOVA). The Shapiro-Wilk's test was performed on the standardized residuals to test for deviations from normality (Shapiro and Wilk, 1965). In cases where significant deviation from normality was observed due to skewness, these outliers were removed until it was normal or symmetrically distributed. (Glass *et.al*, 1972). The t-LSDs (Least significant differences) were calculated at a 5% significance level to compare means of significant source effects. The above analysis was performed using SAS version 9.4 statistical software (SAS, 1999).

4.4 Results

4.4.1 Morphological observation

Only female *T. par* flies were captured and were identified by characteristics described by Oldroyd (1954). The species were small and yellow with no specific pattern. Its mesonotum had a blackish ground colour and densely covered with yellow tomentum, resulting in a greyish-yellow thorax that contrasts with the more orange-yellowish colouration (Figure 4-3).

The eyes of *T. par* are a vivid emerald green in fresh samples and do not show any bands in both living and dry specimens. According to Neave (1912), female *T. par* has a shining golden upper large-facetted part of the eye. The proboscis had a yellow or orange. The colour of the pleura was grey or slightly yellowish, with long white hairs. The abdomen's dorsal side consisted of an orange to dark hair and does not have any markings. The first two segments of the antennae were yellow and have yellow hairs, while the remaining segments were a bright orange colour. In many specimens, the third segment of the antennae indicated to be wider. (Figure 4-3).

The lower part of the abdomen had a yellowish hue and becomes darker towards the top, lacking any distinct patterns, and was adorned with fine black and golden-yellow hairs. The wings had quite clear veins and a yellow stigma (Figure 4-3). The legs were yellow to reddish-brown colour, with tarsi being a deeper brown than the rest, but lacking any black markings anywhere. The hairs on the femora were yellowish, while those on the tibiae were a mix of black ones, with more towards the tips, creating an appearance of dark-tipped tibiae to the naked eye (Figure 4-3).



Figure 4-3: Sequential images of morphological characteristics of female Tabanus par (Adapted from Oldroyd, 1954; Neave, 1912): A) Thoracic coloration B) Eye coloration with emerald green and golden upper facets C) proboscis, D) Pleura coloration and white hairs; E) Antennae; F) Abdomen; G) Wings H)

4.4.2 Morphometrical observations

A morphometrical analysis of female *T. par* flies was carried out using samples from 30 specimens collected from three different game reserves and 10 specimens were used per locality (Table 7). The total dorsal length of *T. par* ranged from 1181,850 to 12860,420µm. The ventral length was roughly equivalent to the dorsal length, exhibiting a range 1000,450 to 12981,180µm. Nevertheless, within the samples, thorax length and breadth showed a degree of consistency, indicating structural uniformity, with a range of 3616, 48 to 4911,50 µm and a range of 3081,51 to 4772,00 µm, respectively. The abdomen length of *T. par* across the three traps exhibited moderate variability. The body weight (dried female specimen) range was from 0.008 to 0.018 mg, with a standard deviation of 0.002, suggesting a reliable measurement across individuals. Variability in wing size among female *T. par* is displayed in Table 7. The mean width of the wings was 4676,836 µm, but there was considerable variability (Std. Dev = 6692.531), indicating differences in the sizes of *T. par* wings.

Table 4-1: Body part lengths and whole-body weight of female *T. par* dry specimens.

Measurements	Total specimen	Minimum (µm)	Maximum (µm)	Mean	Standard deviation
Complete dorsal length	30	1181,850	12860,420	11016,210	2018,443
Complete ventral length	30	1000,450	12981,180	10617,625	2736,776
Head ventral length	30	1497,430	4078,850	1971,511	434,451
Head ventral width	30	344,600	4335,370	3779,876	712,276
Thorax dorsal length	30	3616,480	4911,500	4274,707	338,122
Thorax dorsal breath	30	3360,300	4651,880	3823,879	263,484
Abdomen ventral length	30	5218,750	7745,530	6479,276	525,304
Wing breath	30	2881,020	40069,090	4676,836	6692,531
Wing length	30	887,720	10654,400	9163,638	1673,945
Weight (mg)	30	0.0076	0,0183	0,013	0,0022

4.4.2.1 Morphometric variation across traps

Morphometrics comparison across the three different sampled areas are depicted in (Figure 4-4, 4-5 and 4-6). All three sampled sites Phinda Private Game Farm (PHI), Hluhluwe-Imfolozi Park (HLU) and Charters Creek (CCR) demonstrated variability in the morphometric profiles of measured *T. par*. All sampled sites exhibited similar measurements with slight variation. The total dorsal and ventral length measurements were the highest recorded characteristics across all sites with a maximum measurement of 1200,00 µm. The lengths and breadths of the left wings were typically uniform across different traps, although there were slight variations with a maximum measurement of 10000 µm. Different localities showed variation, especially in Phinda Private Game Reserve. In comparison to flies from other locations, these flies displayed smaller measures, particularly in dorsal and ventral length.

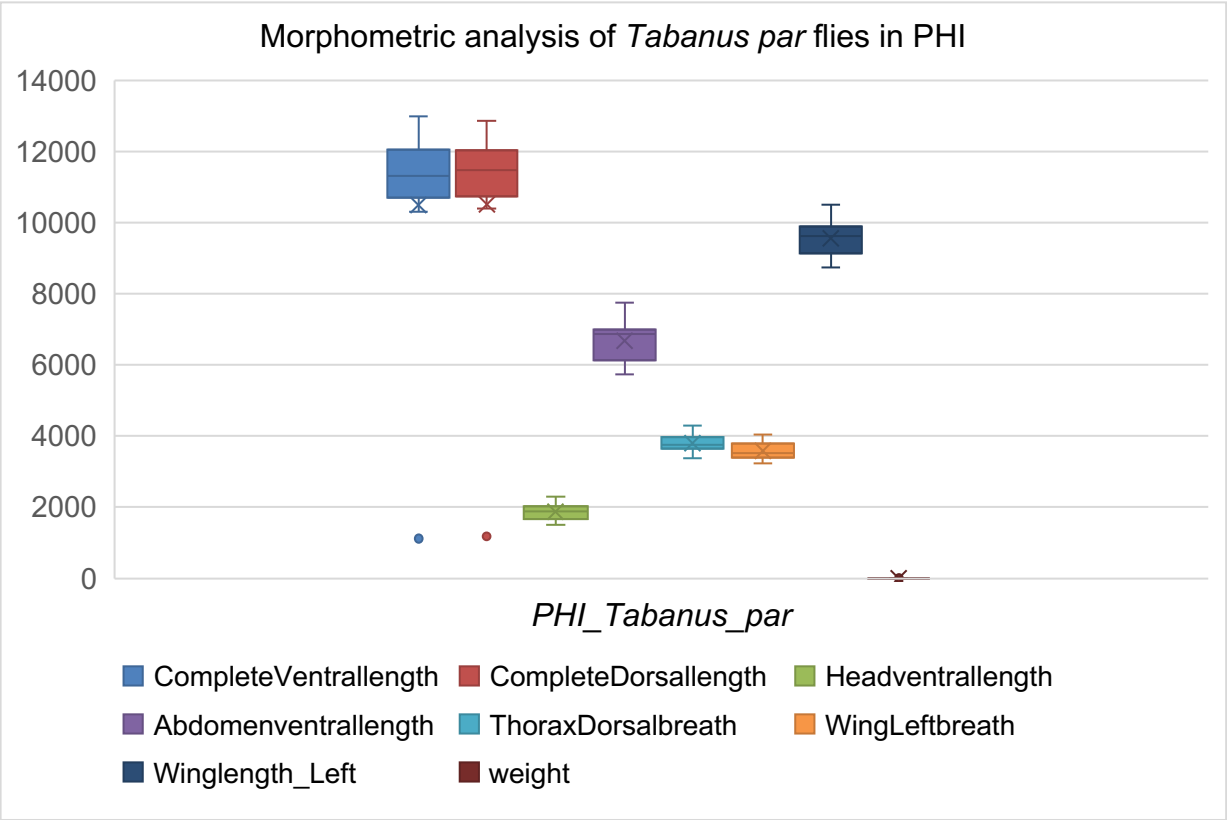


Figure 4-4: Morphometric measurements of female *T. par* specimens collected from Phinda Private Game Reserve.

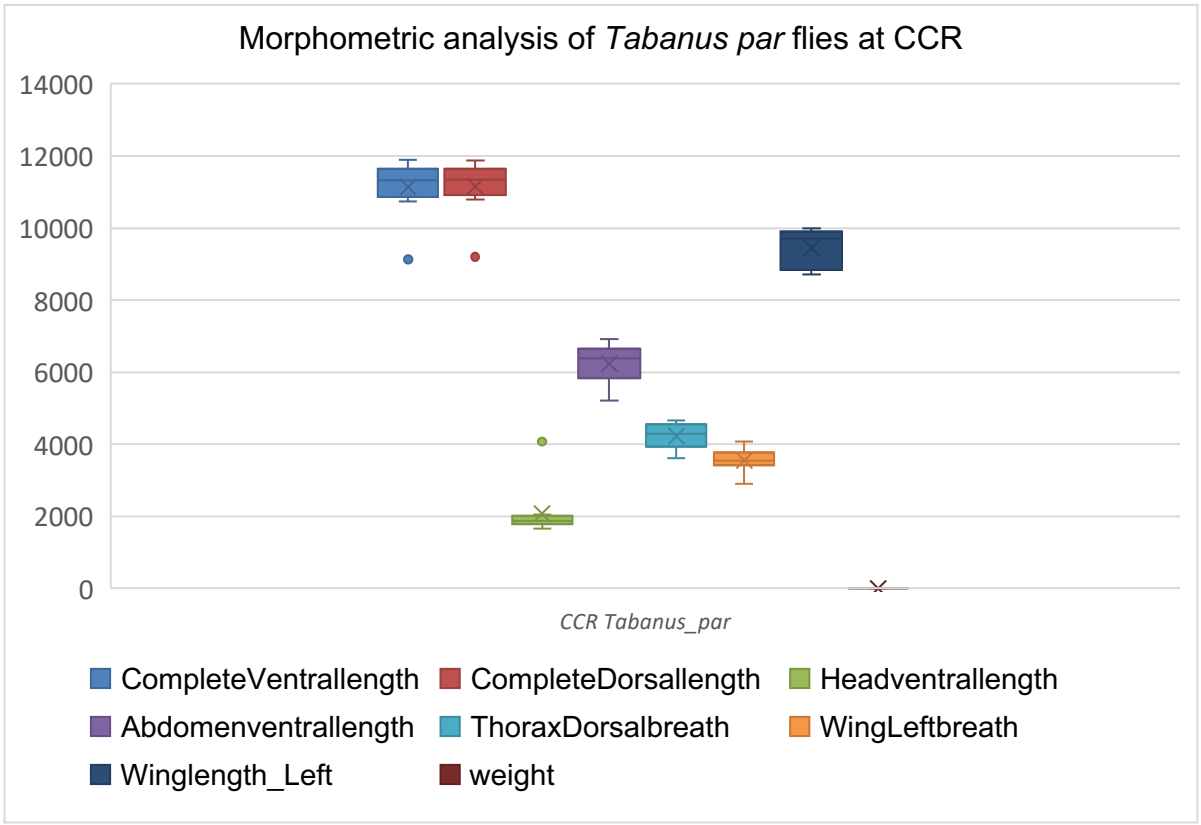


Figure 4-5: Morphometric measurements of female *T.par* specimens collected from Charters Creek.

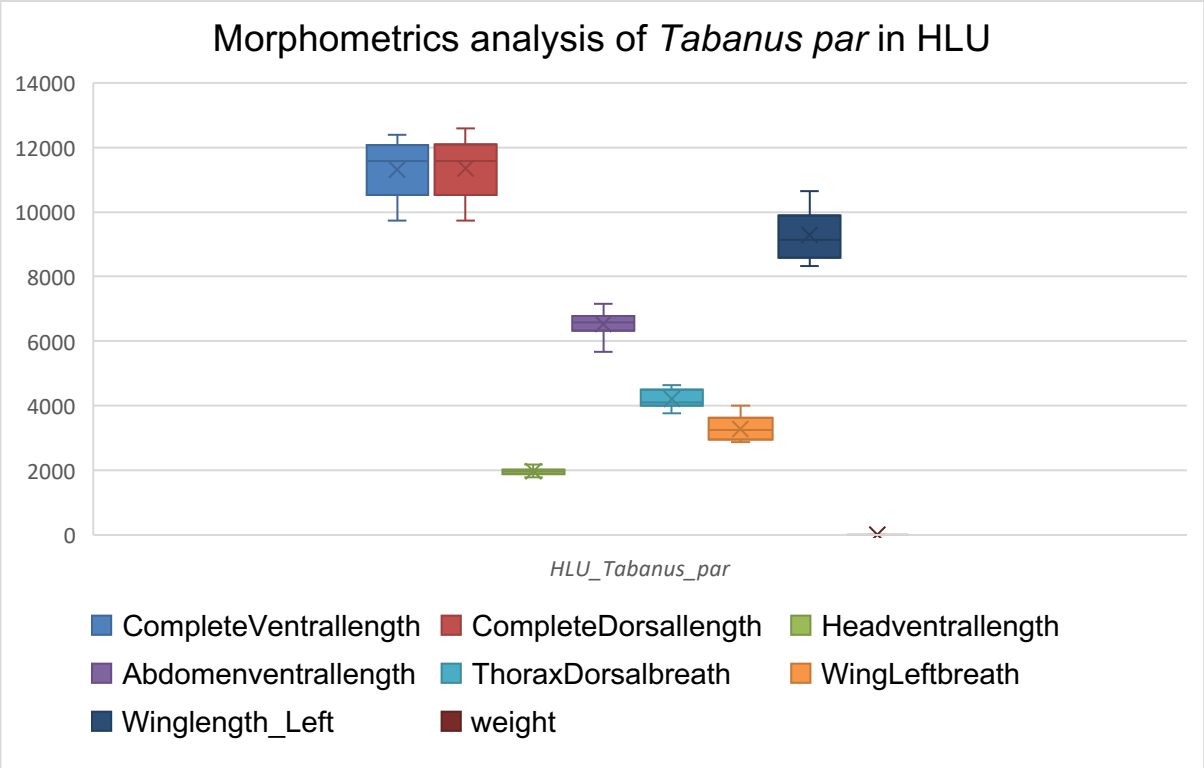


Figure 4-6: Morphometric measurements of female *T. par* specimens collected from Hluhluwe-Imfolozi Park.

4.4.3 Genetic observation

Out of the 50 samples tested by *CO1* PCR assay, 31 positive samples were sent for sequencing. However, only 13 samples were successfully sequenced. Produced accession numbers from submitted sequences include PQ636941, PQ636942, PQ636944, PQ636947, PQ636948, PQ636949, PQ636951, PQ636953, PQ636956, PQ636957, PQ636959, PQ636960, and PQ636961, which correspond to *T. par (*Appendix ll). The *COI* sequences were compared with those available in the GenBank and BOLD databases using BLAST, indicating that *T. par* sequences generated in this study exhibited over 94.87%-100% genetic similarity, aligning with their morphological species classifications.

Based on the *CO1* gene phylogenetic analysis, the *T. par* sequences from the current clustered with similar species from the Tabanini tribe, whilst members from the tribes Haematopotini formed a second clade and Philolichini was an outgroup (Figure 4-7).

Within the *T. par* group, two monophyletic clades composed of *T. par* sequences from this study and those from Mozambique (EMU21, EMU22 and EMU23), the second clade composed of sequences from Gabon HF66, HF136 and HF259) and Central African Republic (HF65, HF67, HG68) were observed at 100% bootstrap support. This suggests that *T. par* specimens from southern Africa are more genetically related as compared to those from western/central Africa. The *Atylotus* groups also clustered with conspecific genera, forming a monophyletic clade, whilst *T. boeti* formed a paraphyletic clade with other conspecific species. Members from *Ancala africana* and *T. taeniola* grouped and this might be because both flies are very large and bite more aggressively and painfully.

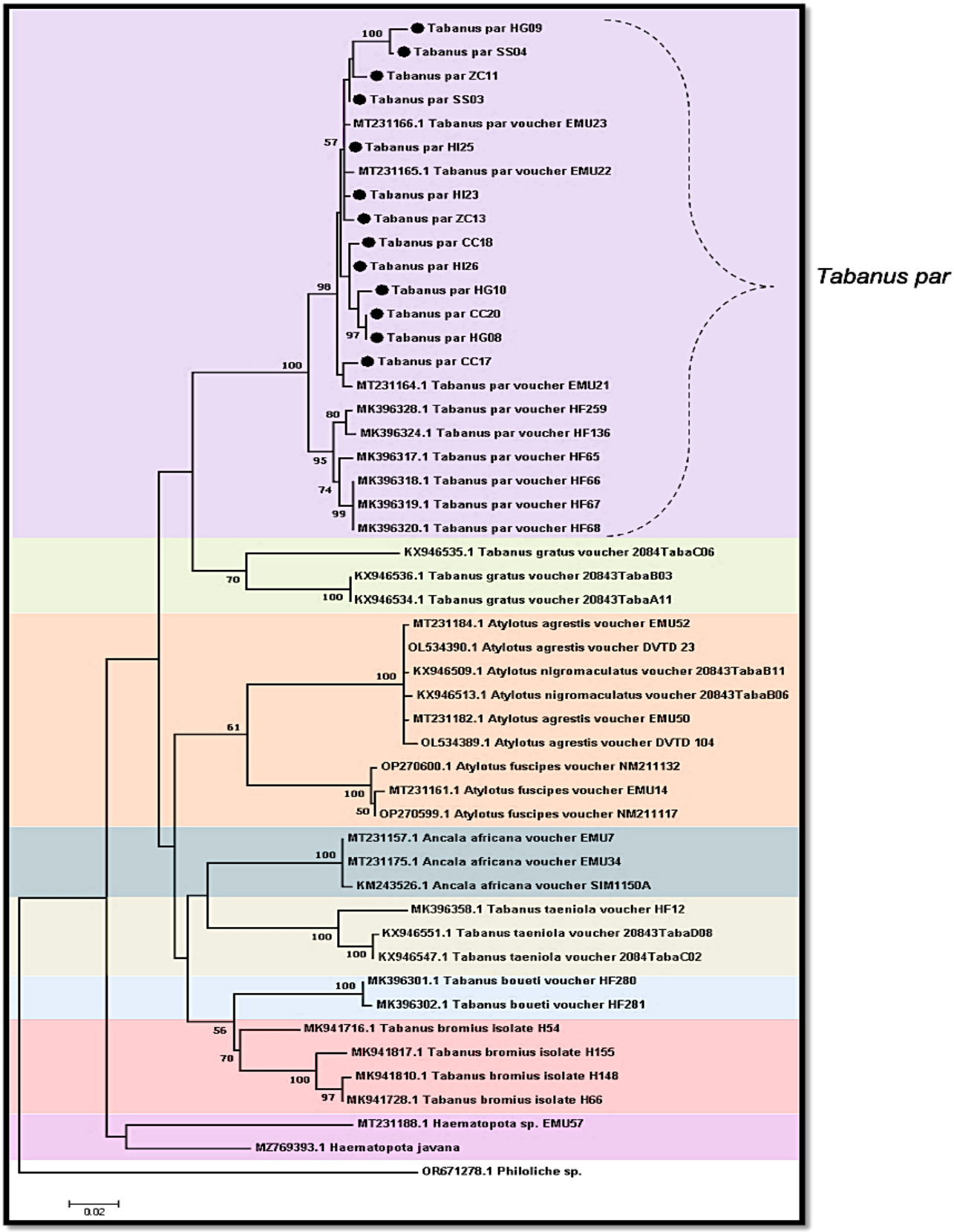


Figure 4-7: Molecular phylogenetic analysis by Maximum Likelihood (ML) method. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016).

4.5 Discussion

The study of morphology and morphometric measurements of *Tabanus* species is essential for gaining insights into their biodiversity, classification, and ecological functions. The classification of *T. par* and other species in the Tabanidae family primarily depends on both morphological and molecular traits. Morphological characteristics are especially important for differentiating closely related species, as they often show minute differences that can point to specific taxa (Changbunjong *et al*., 2021). *Tabanus* species display a broad spectrum of morphological traits that can be measured quantitatively to distinguish between species and comprehend their adjustments to different habitats (Mullens *et al*., 2022). Based on the results in this study. Morphological features *of T. par* were consistent with various studies conducted in South Africa (Taioe, 2017; Mazibuko, 2018). Taxonomic descriptions of each *T. par* fly were carried out following the classification scheme of Oldroyd (1954). Across the various collection locations (PHI, HLU, CCR), *T. par* species exhibited comparable morphological characteristics. They indicated brightly coloured with yellow-orange abdomen but appear dull in preserved specimens. *Tabanus par* is known for its emerald, green robust eyes, however, in the current study this feature is not observed due to the preservation method.

Along with morphology, other anatomical characteristics such as the structure of the mouthparts, and size are also pivotal for identification purposes. (Nour and Ali, 2017). Morphometric analyses have been used to differentiate closely related species within the *Tabanus* genus. For instance, Changbunjong *et al*., (2021) leveraged landmark data to distinguish between morphologically similar species like *T. megalops* and *T. striatus*, which frequently display overlapping morphological characteristics. However, morphometrics analysis of *T. par* has never been documented. This study documented various body parts and appendages measurements of *T. par*. Most horseflies exhibit a body length that ranges from 5 to 25 mm, with some of the largest individuals reaching up to 60 mm. In this study, the maximum dorsal and ventral body lengths were measured at 12860,420 µm and 12981,180µm respectively, which align with the results found by Taioe, (2017), Mazibuko, (2018), and Oldroyd, (1952, 1954, 1957). The longest wing length recorded was 10,654.400 µm, which matches the data presented by Oldroyd (1954). While many morphometric studies have primarily

concentrated on wing structure (Torres *et al*., 2016; Altunsoy, 2017; Sontigun *et al*., 2019; Rodrigues et al., 2024) or features of the basal callus (Changbunjong *et al*., 2021), the data presented here provide valuable scientific insights and can serve as a reference for future morphometric identification of *T. par*.

In addition to species identification based on morphology and morphometrics, *COI* gene sequences were obtained to supplement identifications of morphological groupings. A total of 13 *COI* sequences were produced and they matched with *T. par* sequences available in NCBI database using BLASTn, which supplementary conformed the accurate identification of this fly species The *T. par* clade, consisting of sequences from this study, demonstrated a close relationship with *T. par* sequences from Mozambique as compared to those Central African Republic and Gabon, however, the *T. par* was resolved with 100% support and monophylly within this group was confirmed. Similar phylogenetic findings were reported by Taioe (2017), suggesting a close relationship among tabanids within the African continent. This supports the idea that there are connections between *T. par* and related species in South Africa and other Afrotropical nations, providing important insights into their ecological distribution, genetic diversity, and potential as vectors for various pathogens.

Observations made from tree topologies of ML showed 8 major clades and a polyphyletic phylogeny was resolved between members of the tribes Tabanini, Haematopotini and Philolichini however, a monophyletic relationship was observed in each clade. A similar observation was made in previous studies, where both morphological analysis of external traits and molecular techniques targeting the CO1 gene in Tabanidae were employed (Lessard and Yeates, 2012; Lessard *et al*., 2013; Morita *et al*., 2016; Taioe *et al*., 2017; Mazibuko, 2018).

4.6 Conclusion

This study has provided a detailed investigation into the morphology, morphometrics, and molecular characteristics of *Tabanus par*, contributing valuable knowledge to the taxonomy and ecological understanding of the Tabanidae family. The morphological traits observed were consistent with previous findings in South Africa, affirming the reliability of traditional taxonomic classifications, such as those proposed by Oldroyd

(1954). Quantitative morphometric analyses of *T. par*, which had not been previously documented, offered precise measurements of body parts and appendages, enriching the morphological dataset for this species and providing a baseline for future comparative studies.

The integration of COI gene sequencing with morphological identification reinforced the taxonomic placement of *T. par*, confirming a strong genetic similarity with populations from Mozambique and highlighting the species' monophyly. Phylogenetic analyses revealed complex evolutionary relationships within the Tabanidae family, particularly among the tribes Tabanini, Haematopotini, and Philolichini. These results emphasize the importance of combining molecular tools with traditional morphological approaches for accurate species identification, especially in morphologically similar taxa.

4.7 Recommendations

Comparative research that includes morphometric and molecular information from various Tabanid species can offer an in-depth insight into the differences between species and the evolutionary patterns within the family. For phylogenetic analysis, it is essential to examine more conserved nuclear, mitochondrial, or protein-coding genes, along with larger fragments of genes. This approach can clarify important regions of the tabanid tree topology where existing morphological and molecular data are currently inadequate and require further investigation.

4.8 References

1. Acapovi, G.L., Yao, Y., N'goran, E., Dia, M.L. & Desquesnes, M. (2001). Abondance relative des tabanidés dans la région des savanes de Côte d'Ivoire. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, *54*(2), pp.109-114.
2. Ahmed, A.B., Okiwelu, S.N. & Samdi, S.M. (2005). Species diversity, abundance and seasonal occurrence of some biting flies in Southern Kaduna, Nigeria. *African Journal of Biomedical Research*, *8*(2), pp.113-118.
3. Altunsoy, F. (2017). Analysis of morphometric characteristics of different populations of *Tabanus bromius* Linne 1758 (Diptera: Tabanidae).
4. Banerjee, D., Kumar, V., Maity, A., Ghosh, B., Tyagi, K., Singha, D., Kundu, S., Laskar, B.A., Naskar, A. & Rath, S. (2015). Identification through DNA barcoding of Tabanidae (Diptera) vectors of surra disease in India. *Acta Tropica*, *150*, pp.52-58.
5. Chainey, J.E. & Oldroyd, H. (1980). Family Tabanidae. Catalogue of the Diptera of the Afrotropical Region. British Museum (Natural History), London, pp.1437.
6. Changbunjong, T., Prakaikowit, N., Maneephan, P., Kaewwiset, T., Weluwanarak, T., Chaiphongpachara, T. & Dujardin, J.P. (2021). Landmark data to distinguish and identify morphologically close *Tabanus* spp.(Diptera: Tabanidae). *Insects*, *12*(11), p.974.
7. Changbunjong, T., Sedwisi, P., Weluwanarak, T., Nitiyamatawat, E., Sariwongchan, R. & Chareonviriyaphap, T. (2018). Species diversity and abundance of *Tabanus* spp.(Diptera: Tabanidae) in different habitats of Thailand. *Journal of Asia-Pacific Entomology*, *21*(1), pp.134-139.
8. Desquesnes, M., Biteau-Coroller, F., Bouyer, J., Dia, M.L. & Foil, L. (2009). Development of a mathematical model for mechanical transmission of trypanosomes and other pathogens of cattle transmitted by tabanids. *International Journal for Parasitology*, *39*(3), pp.333-346.
9. Djiteye, A., Diarra, M., Ouattara, I. & Traore, D. (1998). Comparison of the efficacy of different traps and attractants for Tabanidae and *Stomoxys* in Mali. *The journal of protozoology research*, *8*(4), pp.263-273.

10. Dusbábek, F., Gregor, F. & Daniel, M. (1980). Note on some Diptera (Tabanidae, *Glossinidae*, *Muscidae* and related families) at the peak of dry season in Toro Game Reserve, Uganda. *Folia parasitologica*, *27*(2), pp.165-171.
11. Essono, P.Y.B., Dechaume-Moncharmont, F.X., Mavoungou, J., Mba, R.O., Duvallet, G. & Bretagnolle, F. (2015). Distribution and abundance of hematophagous flies (*Glossinidae*, *Stomoxys*, and Tabanidae) in two national parks of Gabon. *Parasite*, *22*.
12. Esterhuizen, J. (2006). Seasonal abundance of horse flies (Diptera: Tabanidae) from two conservation areas in northeastern KwaZulu-Natal Province, South Africa. *African Entomology*, *14*(2), pp.395-397.
13. Flomer, O. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotech*, *3*, pp.294-299.
14. Kaba, D., Berté, D., Ta, B.T.D., Telleria, J., Solano, P. & Dujardin, J.P. (2017). The wing venation patterns to identify single tsetse flies. *Infection, Genetics and Evolution*, *47*, pp.132-139.
15. Keita, M.L., Medkour, H., Sambou, M., Dahmana, H. & Mediannikov, O. (2020). Tabanids as possible pathogen vectors in Senegal (West Africa). *Parasites & Vectors*, *13*, pp.1-15.
16. Glass, G.V., Peckham, P.D. & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research*, *42*(3), pp.237-288.
17. Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, *33*(7), pp.1870-1874.
18. Leclercq, M. (1961). Tabanidae (Diptera, brachycera). *Exploration du Parc national de la Garamba-Mission H. de Saeger*, *5*(21).
19. Leclercq, M. (1965). Tabanidae (Diptera) du Liberia et d'autres pays Africains. *Bulletin de l'Institut Agronomique et des Stations de Recherches de Gembloux*, *33*(1).
20. Lessard, B.D., Cameron, S.L., Bayless, K.M., Wiegmann, B.M. & Yeates, D.K. (2013). The evolution and biogeography of the austral horse fly tribe Scionini

(Diptera: Tabanidae: Pangoniinae) inferred from multiple mitochondrial and nuclear genes. *Molecular phylogenetics and Evolution*, *68*(3), pp.516-540.

21. Lessard, B.D. & Yeates, D.K. (2012). Anzomyia (Diptera: Tabanidae: Pangoniinae: Scionini): a new genus of Australian and New Zealand horse fly, including the description of three new species. *Insect Systematics & Evolution*, *43*(2), pp.101-116.
22. Mazibuko, X. (2018). *Species composition of and detection of haemoparasites in horse-flies (Diptera:* Tabanidae*) in south-eastern Kruger National Park* (Doctoral dissertation, University of Pretoria).
23. Mihok, S. (2002). The development of a multipurpose trap (the Nzi) for tsetse and other biting flies. *Bulletin of Entomological Research*, *92*(5), pp.385-403.
24. Morita, S.I., Bayless, K.M., Yeates, D.K. & Wiegmann, B.M. (2016). Molecular phylogeny of the horse flies: a framework for renewing tabanid taxonomy. *Systematic entomology*, *41*(1), pp.56-72.
25. Morita, S.I. (2008). A phylogeny of long-tongued horse flies (Diptera: Tabanidae: Philoliche) with the first cladistic review of higher relationships within the family. *Invertebrate Systematics*, *22*(3), pp.311-327.
26. Mugasa, C.M., Villinger, J., Gitau, J., Ndungu, N., Ciosi, M. & Masiga, D., (2018). Morphological re-description and molecular identification of Tabanidae (Diptera) in East Africa. *ZooKeys*, (769), p.117.
27. Mullens, B.A., Trout Fryxell, R., Masonick, P.K., Yanega, D.A. & Davis, T.M., (2022). Hiding in plain sight: an abundant and widespread North American horse fly (Diptera: Tabanidae) in the *Tabanus sulcifrons* Group, *Tabanus variegatus fabricius*, redescribed. *Journal of Medical Entomology*, *59*(4), pp.1217-1235.
28. Rutherford, M.C., Mucina, L. & Powrie, L.W. (2006). Biomes and bioregions of southern Africa. *The vegetation of South Africa, Lesotho and Swaziland*, *19*, pp.30-51.
29. Nazri, A., Agbolade, O., Yaakob, R., Ghani, A.A. & Cheah, Y.K. (2020). A novel investigation of the effect of iterations in sliding semi-landmarks for 3D human facial images. *BMC bioinformatics*, *21*, pp.1-10.
30. Nei, M. & Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford university press.

31. Nour, H.I.C.M.M. & Ali, N.O. (2017). Morphological identification of horse flies (Diptera: Tabanidae) and estimation of their seasonal abundance in Al-Showak district, Gedaref State, Eastern Sudan.
32. Oldroyd H. (1952). The Horseflies of the Ethiopian Region. I. *Haematopota* and *Hippocentrum*. British Museum of Natural History, London.
33. Oldroyd H. (1954). The Horseflies of the Ethiopian Region. Vol. II. *Tabanus* and Related Genera. British Museum of Natural History, London.
34. Oldroyd H. (1957). The Horseflies of the Ethiopian Region. Vol. III. Subfamiles *Chrysopsinae, Scepsidinae* and *Pangoniinae*, and a Revised Classification. British Museum of Natural History, London.
35. Sanal Demirci, S.N., Kilic, V., Mutun, S. & Kilic, A.Y. (2021). Population genetics and phylogeography of *Tabanus bromius* (Diptera: Tabanidae). *Parasites & Vectors*, *14*, pp.1-16.
36. Shapiro, S.S. & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), pp.591-611.
37. Sontigun, N., Samerjai, C., Sukontason, K., Wannasan, A., Amendt, J., Tomberlin, J.K. & Sukontason, K.L. (2019). Wing morphometric analysis of forensically important flesh flies (Diptera: Sarcophagidae) in Thailand. *Acta Tropica*, *190*, pp.312-319.
38. Rodrigues, G.D., Centeno Filho, B.L., Morales, D.F., Dimer, R.D.F.R.M., Cavalheiro, C.D.S., Krolow, T.K., Moura, M.O. & Krüger, R.F. (2024). Discrimination of cryptic species: *Tabanus triangulum* and *Tabanus occidentalis* (Diptera: Tabanidae) differ in size and shape. *Revista Brasileira de Parasitologia Veterinária*, *33*(2), p.e020123.
39. Packer, L. (2024). A survey of keys for the identification of newly described insect genera: recommendations for authors, reviewers, editors, and publishers. *ZooKeys*, *1215*, p.65.
40. Taioe, M.O., Motloang, M.Y., Namangala, B., Chota, A., Molefe, N.I., Musinguzi, S.P., Suganuma, K., Hayes, P., Tsilo, T.J., Chainey, J. & Inoue, N. (2017). Characterization of tabanid flies (*Diptera*: Tabanidae) in South Africa and Zambia and detection of protozoan parasites they are harbouring. *Parasitology*, 144(9), pp.1162-1178.

41. Taioe, M.O. (2017). *Molecular characterization of horse flies (Diptera: Tabanidae) and determination of their role in transmission of haemoparasites in southern Africa* (Doctoral dissertation, North-West University (South Africa), Potchefstroom Campus).

42. Torres, A. & Miranda-Esquivel, D.R. (2016). Wing shape variation in the taxonomic recognition of species of *Diachlorus* Osten-Sacken (Diptera: Tabanidae) from Colombia. *Neotropical entomology*, *45*, pp.180-191.

43. Veer, V. (1999). Notes on Tabanidae (Diptera) that are Surra vectors or pestiferous with description of a new species of *Tabanus* from India. *Oriental Insects*, *33*(1), pp.247-266.

44. Yagi, A.I. (1968). *A contribution to the knowledge of the* Tabanidae *of the Sudan* (Doctoral dissertation, University of Khartoum).

CHAPTER 5 MICROBIOMES HARBOURED BY *TABANUS PAR* BY USING SHOTGUN METAGENOMICS

5.1 Introduction

Family Tabanidae is a cosmopolitan family of flies with members mostly known for their persistent hematophagous behaviour (Litov *et al*., 2013). They transmit viral, bacterial, and other pathogens (Baldacchino *et al*., 2014). When adult horseflies come into contact with microbe-rich substrates, they become contaminated. These microbes can be transferred via the fly's mouthparts, tarsi, and wings, and later spread to new locations through direct contact, blood-feeding, or grooming Baldacchino *et al*., 2014).

Tabanid flies are vectors of various pathogens, as a result, understanding their associated microbiome could offer insights into their health, behaviour and role in disease transmission (Baldacchino *et al*., 2014). These microbiomes are known as complex communities of microorganisms that vary in their occupied niches, species composition, and their influence on various environments (Baquero *et al*., 2021). The microbiome community is composed and dominated by microorganisms, including bacteria, viruses, fungi, and protozoa. In addition, they serve as mechanical vectors for a variety of pathogens that affect human and animal health, including viruses e.g., equine infectious anemia virus, lumpy skin disease, and bovine leukosis virus (Baldacchino *et al*., 2014). These viruses also include bacteriophages that specifically target the bacteria within the microbiota. These viruses include bacteriophages that specifically target bacteria within the microbiota. They can regulate bacterial populations and impact the overall microbial balance. Pathogens such as *Anaplasma marginale*, *A. phagocytophilum*, *Brucella* spp., *Listeria monocytogenes*, as well as protozoa like *Besnoitia besnoiti*, *Babesia bigemina*, *Trypanosoma* sp., *Theileria* sp., and *Leishmania donovani* are affected by these processes (Sontigun *et al*., 2022; Taioe *et al.,* 2017; Baldacchino *et al*., 2014; Werszko *et al*., 2020). *Trypanosoma* sp. is biologically transmitted by tsetse flies, but they are also mechanically transmitted by other blood feeding flies such as tabanids (Leak, 1999). In addition, certain species serve as intermediate hosts for filarial nematodes, like *Dirofilaria roemeri, Elaeophora schneideri*, and *Loa loa*, which causes loiasis (African eye worm) in humans (Baldacchino *et al*., 2014). Some of the fungal species related to tabanids include

Aspergillus, Beauveria, Metarhizium, *Cordyceps, Isaria*, and *Pandora* molds (Shang *et al*., 2015). The specific roles that fungi like Saccharomyces, Pichia, Kluyveromyces, Candida, Hanseniaspora, Debaryomyces, Metschnikowia, and Cryptococcus in relation to pest insects remain largely unknown (Nguyen *et al*., 2007; Stefanini, 2018).

These microorganisms can establish either temporary or permanent relationships with their insect hosts, resulting in both advantageous and detrimental impacts on the fitness of the insects. These associations include obligatory symbioses, which are essential for the survival and reproductive success of both the host and the microbes (Moran and Telang, 1998). For example, endosymbiotic microbes that live within host cells or tissues frequently rely on insect hosts for nutrition. In exchange, they provide fitness benefits such as food supply, increased host defense mechanism, and protection from viruses, parasites, and other environmental stressors (Mondal *et al*., 2023). However, not all microbial interactions are beneficial. some microorganisms can serve as pathogens, reducing insect viability and causing morbidity (Mondal *et al*., 2023). According to Yeoman *et al*., (2019), some *Spiroplasma* sp. have also been reported to be mutualistic, improving host defenses against fungal pathogens (Łukasik *et al*., 2013), parasitic nematodes (Jaenike *et al*., 2010) and parasitoid wasps (Xie *et al*., 2013).

Advances in sequencing technologies and bioinformatics have greatly enhanced knowledge of microbiome biodiversity. Although research has been conducted on the microbiomes of various insects, comprehensive studies on the microbiome of *T. par* remains limited. Specifically, the presence and roles of diverse microbial entities such as bacteria, fungi, viruses, and protozoa within this species have not been well-characterized. The lack of detailed information restricts comprehension of the full range of microbial interactions and their potential effects on the host. To address this research gap, this study characterized the taxonomic profiles of microbiomes harboured by *T. par* using shotgun metagenomics. This methodology provided a thorough overview of the bacterial, fungal, viral, and protozoan communities associated with *T. par*, yielding important insights into their diversity and abundance.

5.2 Materials and methods

5.2.1 Source of *T. par* specimens

Adult *T. par* specimen was collected at three nature reserves, namely Charters Creek Reserve, Hluhluwe Game Reserve, and Phinda Private Game Reserve in KwaZulu-Natal, South Africa, and identified morphologically and genetically as described in Chapter 4, section 4.2.3.

5.2.2 DNA extraction procedure

Prior to genomic DNA extractions, fly specimens were washed with 10% Tween 20 by immersion for 1 hour, then washed twice with 70% Ethanol and then rinsed with double distilled water and left to dry. The whole body (without legs) of *T. par* specimens was cut into pieces in a 1.5 ml microcentrifuge tubes using sterile scissors whereas the legs were separately crushed using a sterilized mortar and pestle and then transferred into a 1.5 ml microcentrifuge tube. Extraction was conducted using Qiagen DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to manufacturer's instructions. Detailed procedure is described in Chapter 4, section 4.2.5 with the following modifications: The body and legs were homogenised separately in a mixture containing 180 µl of Buffer ATL and 20 µl of proteinase K, vortexed and incubated at 56°C until completely lysed. Thereafter extracted DNA was stored at –20°C until used.

5.2.3 Metagenomic sequencing

Whole-genome sequencing (WGS) of the *T. par* DNA was performed at the Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa. Sequence libraries of the flies were constructed using the MGIEasy FS DNA Prep Kit (BGI, China) according to the manufacturer's protocol. The prepared libraries were sequenced using the BGI MGISEQ-2000 platform (BGI Shenzhen,China) with paired-end 150 nt strategy.

5.2.4 Bioinformatics analysis

Taxonomic classification of metagenomic reads was performed using Kaiju (v1.7.4), a protein-level alignment tool renowned for its sensitivity and efficiency in taxonomic sequence classification (Menzel *et al*., 2016). Before classification, raw sequencing reads were subjected to quality checks using FastQC (v0.11.9) (Andrews, 2010) to assess sequence quality metrics and identify potential adapter contamination. Quality filtering and adapter removal were conducted with Trimmomatic (v0.39) (Bolger *et al*., 2014) to ensure that only high-quality reads were retained for downstream analysis. Kaiju was executed in greedy mode, facilitating rapid alignment of translated query sequences against an extensive protein reference database.

The following databases were utilized for taxonomic classification: the ProGenomes Database (v1.1) for prokaryotic organisms, encompassing bacteria and archaea (Mende *et al*., 2017); the RefSeq Plasmid Database (NCBI, 2018) for bacterial plasmids; the NCBI Viral Database (v2023) for DNA viruses and bacteriophages; a curated Phage Database for specific bacteriophage sequences; and the NCBI RefSeq Eukaryotic Database (v2023) for eukaryotic sequences. The outputs from Kaiju provided taxonomic assignments at various hierarchical levels, including phylum, class, family, and genus. To visualize the results, KronaTools (v2.8) was employed to generate interactive taxonomic representations (Ondov *et al*., 2011). Further validation was conducted using BLASTn (v2.12.0) against the NCBI nucleotide database for unclassified reads to resolve any ambiguities. The relative abundances of classified taxa were calculated as percentages of the total reads, facilitating comparative analyses of DNA viruses, prokaryotes, plasmids, phages, and eukaryotes across different samples.

5.3 Results

The taxonomic classification of the sequences revealed a diverse array of microbial taxa on the whole body and legs of *T. par*. To effectively visualize and interpret these data, Krona plots were generated, depicting the relative abundances of operational taxonomic units (OTUs) at the phylum, family, and species levels. These plots, shown in figure 5-1- 5-5, provide an interactive representation of the microbial diversity and allow for a detailed examination of the composition of the microbial communities.

5.3.1 The bacterial composition at the phylum

In the analysis of the microbial communities associated with the WB-sequence reads and L-sequence reads of *T. par*, Pseudomonodota was the most abundant bacterial phylum, with a percentage of 78% in both samples. Followed by the Terrabacteria group, which was notably more abundant in the L-sequence reads, representing 17% of the bacterial community and 13% in WB-sequence reads. The Fibrobacteres, Chlorobi, and Bacteroidetes (FCB group) were found at lower abundance in the L-sequence reads samples with a percentage of 2 (Appendix III).

5.3.2 The bacterial composition at the family level

The bacterial composition at the family level was as follows: The WB-sequence reads were dominated by the bacterial family Enterobacteriaceae (49%) followed by *Pseudomonadaceae* (11%). Other notable families included Entomoplasmatales (7%) and Moraxellaceae (4%). The least assigned abundant bacterial family in the WB-sequence reads included Alphaproteobacteria (2%), Lactobacillales (2%), and Bacillales (1%). The bacterial composition associated with the L-sequence reads of *T. par* at the family level was as follows: Burkholeriaceae (36%), followed by Methylobacteriaceae and Sphingomonadaceae (7%). Enterobacteriaceae and Mycobacteriales with five percent. The remaining groups had an abundance lower than 5% (Appendix III).

5.3.3 The bacterial composition at the genus level

The bacterial communities associated with the WB and L sequence reads of *T. par* were further analysed at the genus level to determine their relative abundance. The WB-sequenced reads consisted of 6 genera, with *Escherichia* (37%) being the most abundant, followed by *Pseudomonas aeruginosa* representing 11% of the community. Other genera included *Aeromonas* (8%) and *Klebsiella pneumonia* (5%). The least abundant genera in the WB sequence reads were *Spiroplasma* and Bacillales (Figure 5-1). The bacterial composition in the L-sequenced reads were dominated by the genera *Paraburkholderia* (24%). Followed by *Methylobacterium* with an abundance of 6%. Other notable genera/species included *Escherichia coli* (3%), *Sphingomonas* sp.

AAP5, and *Moraxella osloensis* with 2%. The least abundant genus/species were *Cutibacterium acnes* and *Sphingomonas* sp. FARSP both comprising of only 1% of the bacterial community, as depicted in figure 5-1.

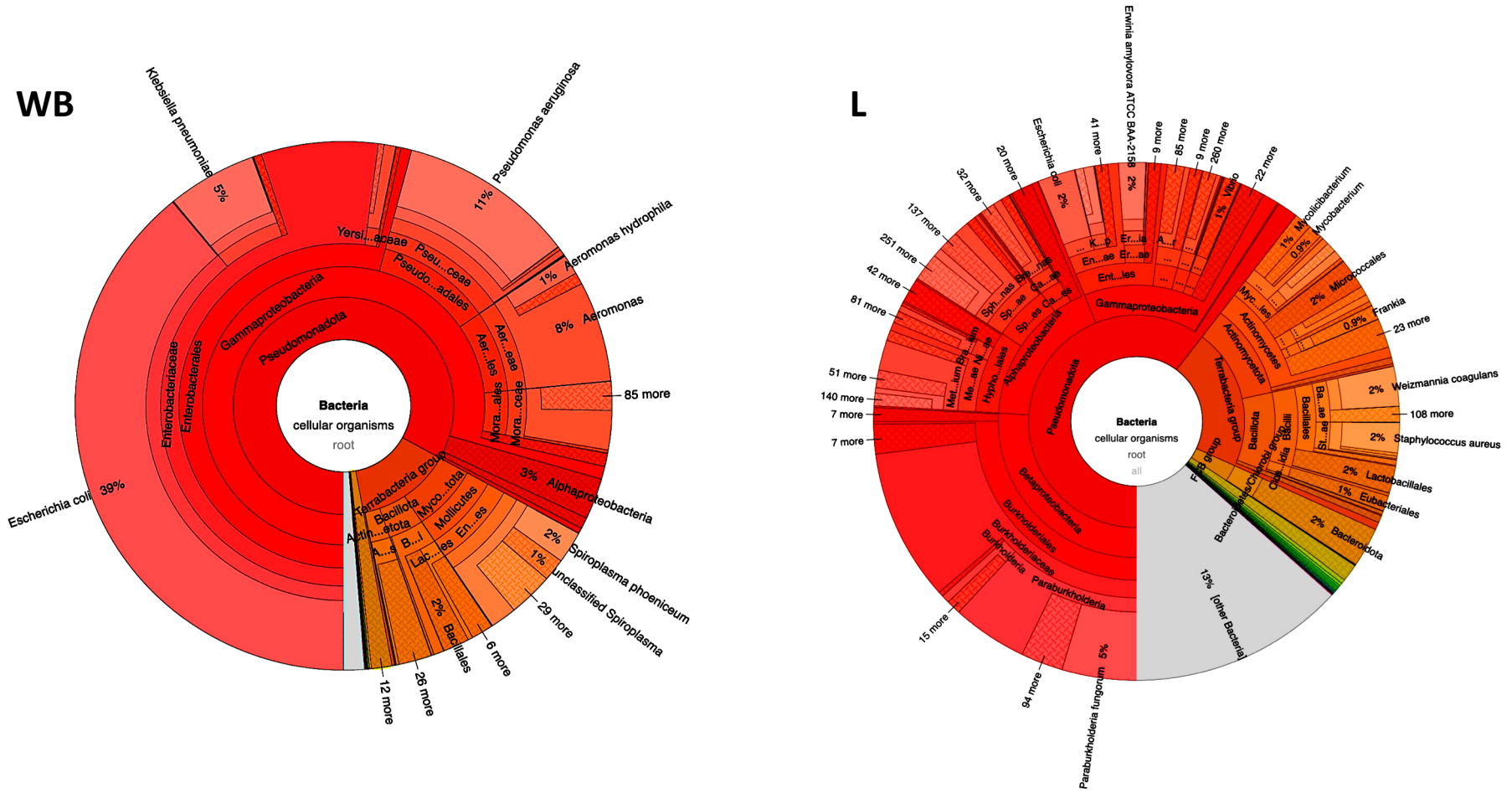


Figure 5-1: The Krona chart comparing the taxonomic reads distribution of bacteria at genus level in the whole body (WB) and legs (L) of *T. par.*

5.3.4 The fungal composition at the phylum level

The fungal composition in the WB -sequenced reads at phylum level consisted of six phyla. The most abundant phylum was Ascomycota, which comprised 44% of the fungal community. This was followed by Basidiomycota at 21%, Mucoromycota and Chytridiomycota both at 10%, and Microsporidia at 5%. The least abundant phylum was Zoopagomycota, accounting for only 1% of the community. In the L sequence reads, only four fungal phyla were identified. Ascomycota was again the dominant phylum, showing a higher abundance of 68% compared to 44% in the WB-sequenced reads. Basidiomycota remained consistent at 21%, similar to its proportion in the WB sequence reads. The Mucoromycota and Chytridiomycota were present at lower proportions, at 4% and 2%, respectively, than in the WB sequence reads (Appendix IV).

5.3.5 The fungal composition at the family level

The fungal composition at the family level of WB sequence reads indicated that Aspergillaceae was most predominant with (4%). Other families with relatively lower abundances, each making up about 1% of the total fungal composition, consisted of the following: Dipodascaceae, Herpotrichiellaceae, and Pichiaceae. The fungal composition in the L sequence reads indicated quite a diversity than in WB sequence reads. The most dominant fungi at family level were Hypocreomycetidae (33%). Plosporomycetidae, followed with (9%) and Dothideomycetidae with an abundance of (7%). The least dominant family was Agriomycetidae, accounting for 1% as depicted in Appendix IV.

5.3.6 The fungal composition at the genus level

At the genus level, the most dominant genus in WB sequence reads was *Synchytrium* (3%), followed by *Batrachochytrium*, *Nosema*, *Phycomyces*, *Pucciniomycotina*, and *Taphrinomycotina*, all with an abundance of 2%. The least abundant genera were *Candida* and *Mitosporidium*, both at 1%. In the L sequence reads, the most predominant genus was *Fusarium* (28%), followed by *Aureobasidium, Malassezia, Meira* and *Rhodotorula* with 2% (Figure 5-2). The least abundant genera at 1% were *Candida* and

Mitosporidium. In L sequence reads, the most predominant genus was *Fusarium* (28%), followed by *Aureobasidium, Malassezia, Meira* and *Rhodotorula* at 2% (Figure 5-2).

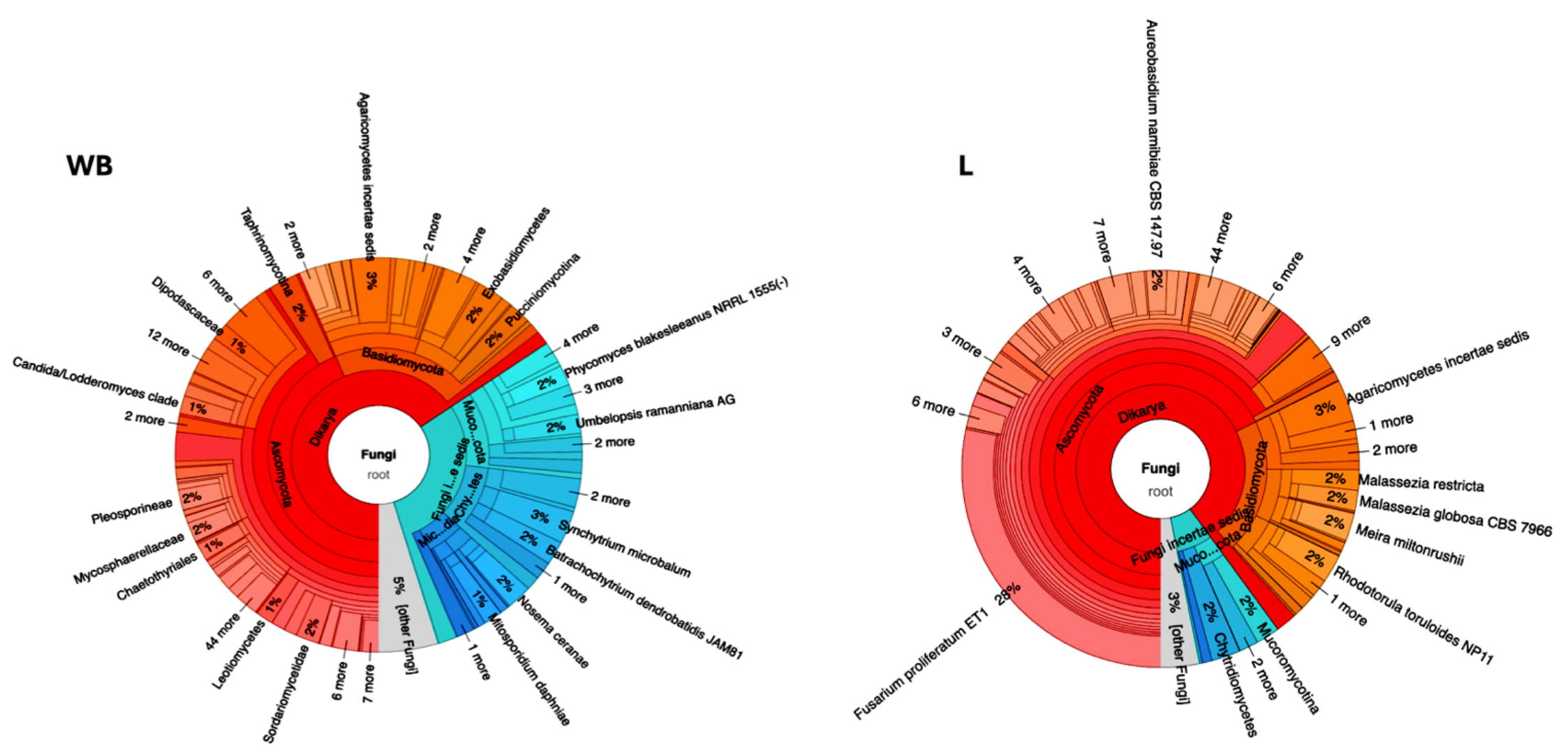


Figure 5-2: The Krona chart comparing the taxonomic reads distribution of fungi at genus level in the whole body (WB) and legs (L) of *T. par.*

5.3.7 The viral composition at the phylum level

The viral composition associated with WB and L sequenced reads at the phylum level were assessed, as depicted in Appendix V. At the phylum level in fungal composition in the WB sequence reads of *T. par* consisted of Uroviricota (40%). followed by Artverviricota, (22%), then Hofneiviricota and Nucleocytoviricota with abundance of 15% and 11%, respectively. The least abundant phylum was Preplasmiviricota, which represented only 3% of the community. In the L sequenced reads, only Ascomycota was reported as the most dominant phylum, with an abundance of 68%, compared to 44% in the WB sequence reads, followed by Basiomycota (21%), similar to that reported in WB-sequence reads. Mucoromycota and Chytridiomycota were present at 4% and 2%, respectively, lower than their proportions in the WB sequence reads.

5.3.8 The viral composition at the family level

The viral composition at the family level in the WB sequence reads consisted of over ten family groups, with the most abundant being Metaviridae, which accounted for 18% of the total. Plectoviridae followed this at 15%. Two family groups, Caulimoviridae and Phycodnaviridae, each had an abundance of 4%, while Adenoviridae comprised 3%. Additionally, four family groups had an abundance of 2%, namely, Colossusvirus, Herelleviridae, Mimiviridae, and Tevenvirinae. In contrast, the viral composition at the family level from the data sequenced from *T. par* consisted of only five family groups. Metaviridae was the most abundant, representing 37%, followed by Phycodnaviridae at 16%. Caulimoviridae was the third most abundant family at 15%, with Mimiviridae and Poxiviridae accounting for 4% and 2%, respectively (Appendix V).

5.3.9 The viral composition at the genus level

At the genus level, the most abundant genus found in the WB sequence reads was *Trichoplusia* ni TED virus, which comprised 17% of the total. This was followed by *Vespertiliovirus* with an abundance of 15%. The third most abundant genus was *Bracoviriform* with an abundance of 5%. The Ben 58058 virus accounted for 4%. Three genera, namely, *Aeromonas* phage phi18p, *Cycad* leaf necrosis virus, and Human

mastadenovirus C represented 3% of the total. *Spiroplasma* phage C74 and Vespertiliovirus SkV1CR23x each comprised 2% of the fly content. In contrast, L sequence reads showed a much higher abundance of *the Trichoplusia ni TED* virus at 36%, compared to the WB sequence data. This was followed by *Cycad* leaf necrosis virus and *Dishui lake phycodnavirus* 1, which had an abundance of 13%. Additionally, two specific species from the *Bracoviriform* genus, *B. congregatae* and *B. glomeratae* were reported with 8% and 7%, respectively (Figure 5-3).

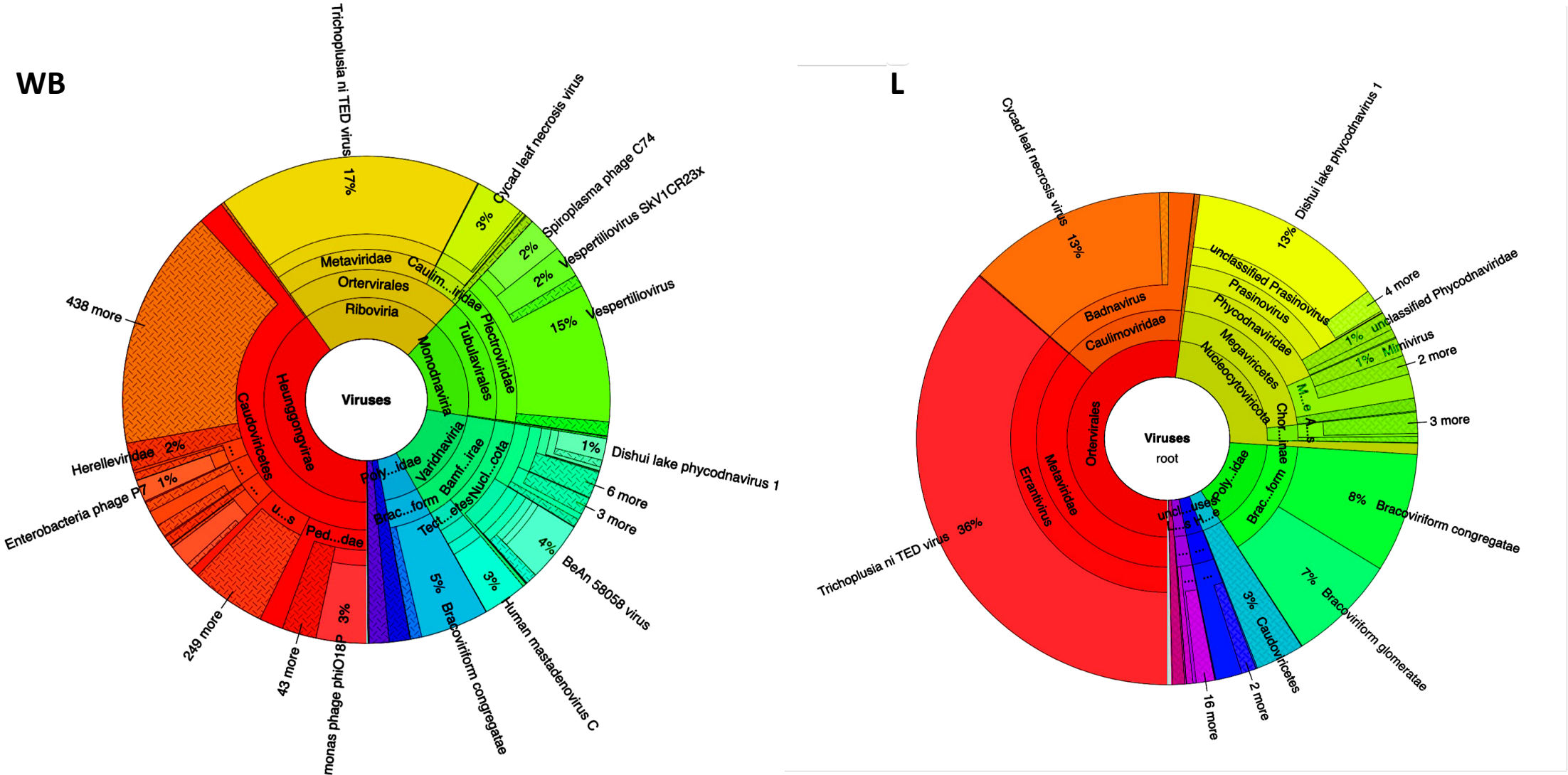


Figure 5-3: The Krona chart comparing the taxonomic reads distribution of viruses at genus level in the whole body (WB) and legs (L) of T. par.

5.3.10 The bacterial plasmids composition at the phylum level

Both WB and L sequence reads contained bacterial plasmids primarily from the phylum Pseudomonodota, the most abundant phylum in both samples (Appendix VI). Due to the dominance of Pseudomonodota, further classification was performed at the class level. In the WB sequence reads, three classes were identified within the Pseudomonodota phylum: the Alphaproteobacteria was the most abundant class, comprising 53% of the plasmids followed by Gammaproteobacteria at 21%, and Betaproteobacteria was the least abundant at 14%. In contrast, the L sequence reads showed a different distribution at the class level, reporting an abundance of Betaproteobacteria at 36%, followed by Gammaproteobacteria and Alphaproteobacteria with an abundance of 31% and 16%, respectively.

5.3.11 The bacterial plasmids composition at the family level

The bacterial plasmid composition at the family level in the WB sequence reads consisted of various family groups, with the most abundant reported in the family Anaplasmaaceae (34%) and followed by Burkholeriaceae (9%). Two family groups, namely Pseusomonadaceae and Rhizobiaceae*,* with 6%. Followed by Enterobacteriaceae with an abundance of 4%. Azospirillaceae, Comamonadaceae, and Sphingomonadaceae with an abundance of 3%. The least bacterial plasmids were Ewriniaceae and Methylobacteriaceae*,* making up 1%. In the L sequence reads, the most abundant family group was Burkholderiaceae at 35%, followed by Enterobacteriaceae (15%). The third most abundant family group was Moraxellaceae (7%). Rhizobiaceae had an abundance of 4%. Two family groups, namely Methylobacteriaceae and Sphingomonadaceae*,* had an abundance of 3% (Appendix VI).

5.3.12 The bacterial plasmids composition at the genus level

Sequence reads generated from WB sequence reads revealed various genera/species, with the most abundant being the *Wolbachia*-endosymbiont, comprising 34% of the sequences, followed by *Pseudomonas aeruginosa* at 4%. Other genera/species present included *Bacillus licheniformis, Burkholderia vietnamiensis,*

Serratia marcescens, Spirochaetia sp.*,* and an unidentified *Variovorax* sp., each accounting for 2% of the samples.

In the L sequence reads, *B. vietnamiensis* was the most abundant genus/species, representing 10% of the findings. This was followed by an unclassified *Caballeronia* at 6%. The *Escherichia coli* O85:H1 and *Acinetobacter pittii* both accounted for 5%, making them the third most abundant genera/species. Most of the remaining genera/species each had an abundance of 1% and included *Acinetobacter baumannii*, *Burkholderia arboris*, *Caballeronia* sp. M1242, *Klebsiella pneumoniae*, *Methylobacterium aquaticum*, *Paraburkholderia fungorum*, and *Sphingomonas* (Figure 5-4).

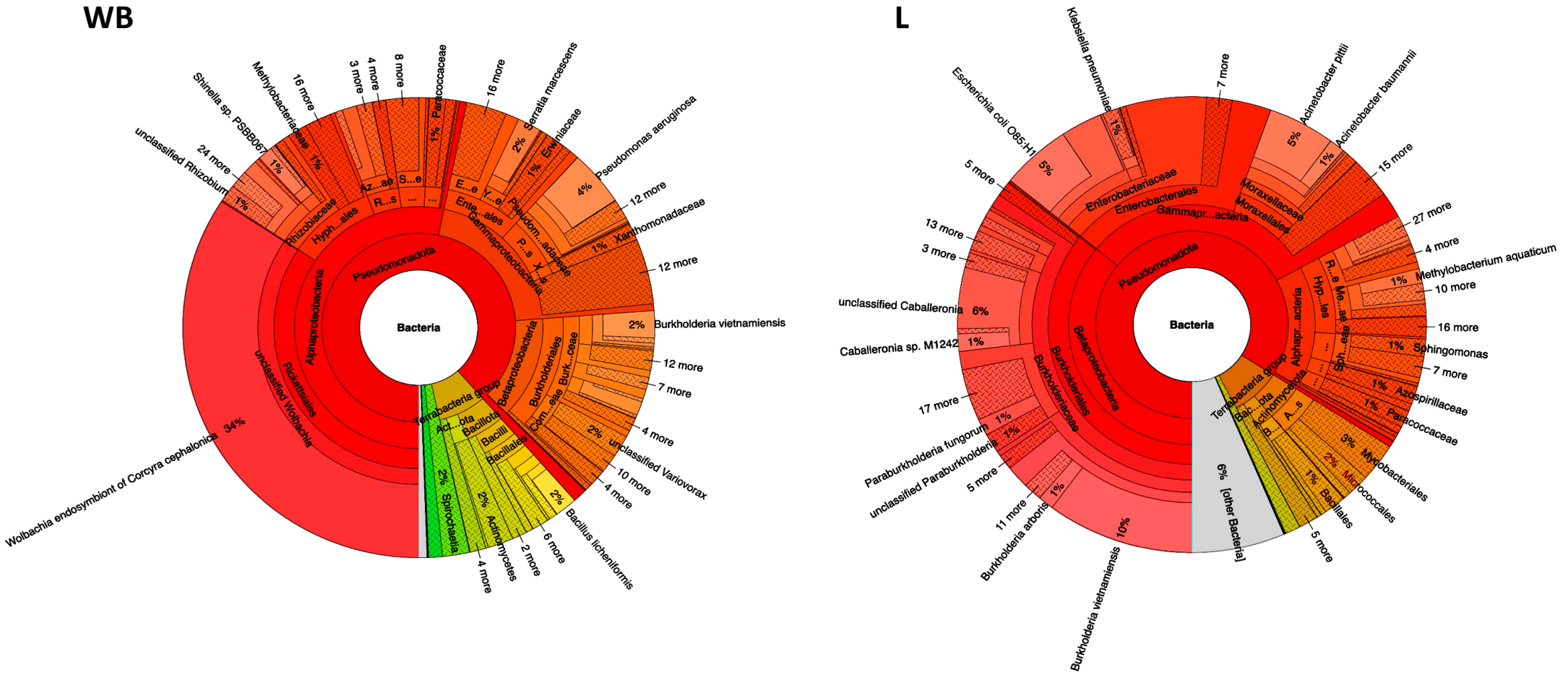


Figure 5-4: The Krona chart comparing the taxonomic reads distribution of bacterial plasmids at the genus species in the whole body (WB) and legs (L) of *T. par.*

5.3.13 The eukaryote composition at the phylum level

In the analysis of the eukaryote communities associated with the WB sequence reads of *T. par*, Euglenzoa was the most abundant eukaryote phyla, with a percentage of 79%. They were followed by the Opisthokonta, which is the fungi group that was represented by 15%. Data sampled from L sequence reads was represented by more fungi such as phyla Ascomycota (46%) and Basidiomycota (28%) followed by Mucoromycota (6%) (Appendix VII).

5.3.14 The eukaryote composition at the family level

The eukaryote composition at the family level in the WB sequence reads consisted of the family Trypanosomatidae with an abundance of 79% followed by Nosematidae at 2%. The L sequence reads consisted of various family groups, with the most abundant reported as Dothideomycetidae (17%) and Pleosporomycetidae and Hypocreomycetidae with 12% and 6%, respectively. The remaining groups had an abundance lower than 3% (Appendix VII).

5.3.15 The eukaryote composition at the genus/species level

At the genus/species level, Stercorarian trypanosomes had a high abundance, with *Trypanosoma theileri* (32%), followed by *T. melophagium* (29%). Meanwhile, *T. cruzi* and *T. grayi* had an abundance of 5% and 3%, respectively. Within the L sequence reads were more prevalent within Eukaryota, with the most abundant genus/species being *Aureobasidium melanogenum* (7%), followed by *Rhodotorula* (4%) (Figure 5-5).

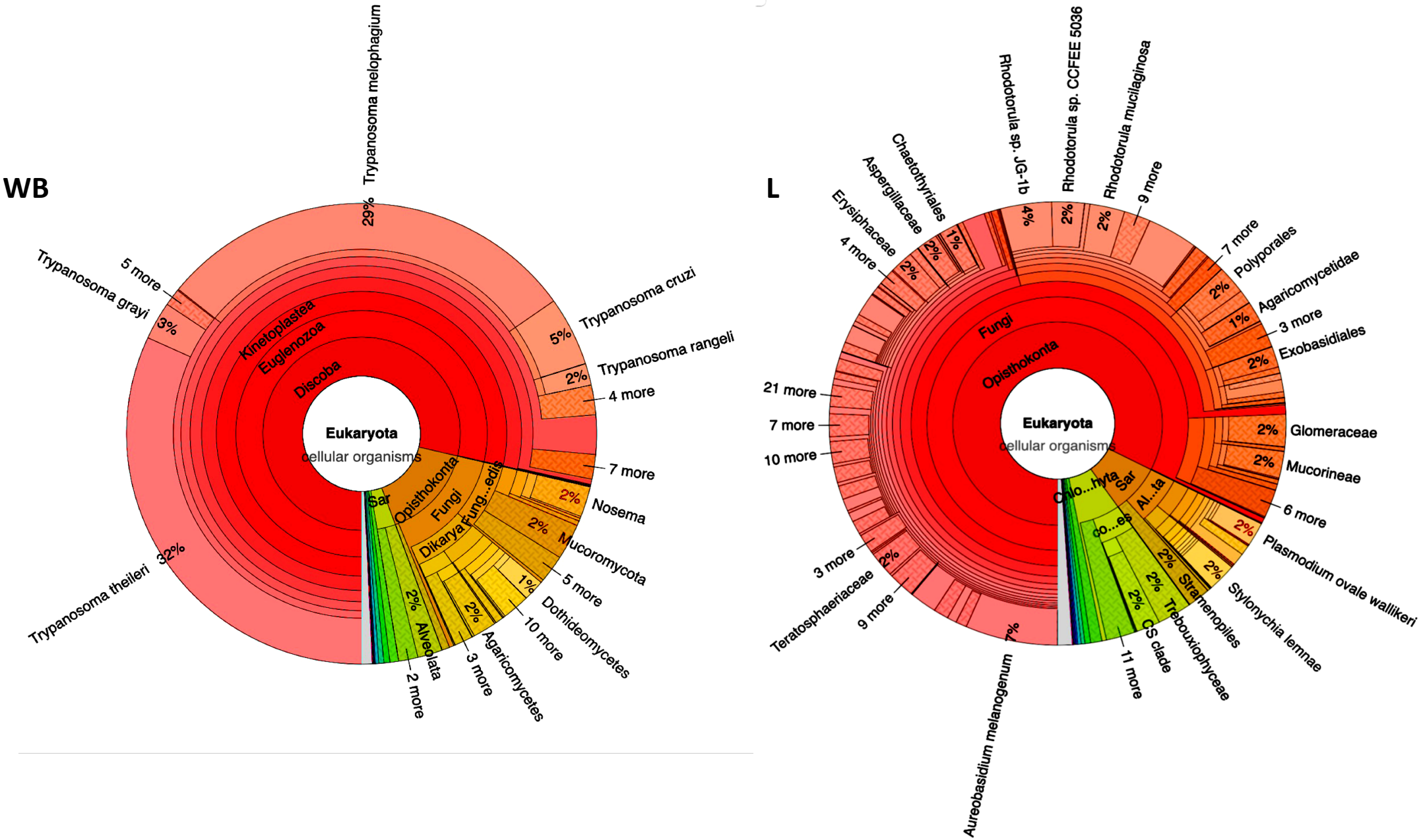


Figure 5-5: The Krona chart showing the taxonomic reads distribution of eukaryotes at genus/species level in the whole body (WB) and legs (L) *of T. par.*

5.4 Discussion

The adult dipterans possess morphological and anatomical features, such as legs, wings, mouthparts, and a variety of setae across their bodies, that create ideal surfaces for the acquisition, adhesion, or retention of microbes, facilitating the transmission of bacteria, viruses, and fungi by the vector (Stoffolano, 2019). Various structures present in adult flies have evolved to aid in the transfer of microbes from one location to another, including the proboscis and legs (Barro *et al*., 2006; Tan *et al*., 1997). A deeper understanding of the interactions between these flies and different microbes drives this rising interest. However, many microbes linked to horseflies remain undetected, and knowledge of their prevalence and diversity, whether pathogenic or non-pathogenic, remains limited.

Numerous studies have demonstrated that vector arthropods' microbiota varies according to species, sex, organs, and developmental stage. While many pathogenic viruses, bacteria, protozoa, and filarial nematodes have been isolated from tabanids, research on pathogen transmission often focuses specifically on specific pathogens rather than examining the entirety of their associated microbiomes (Baldacchino *et al*., 2013; González *et al.*, 2024). This study investigated the variations in the microbiome of *T. par* from the WB sequence reads that included the gut and head, as well as L sequence reads that included the legs.

In this study, the microbiome of *T. par* WB and L was characterized without the influence of external microbial contaminants on the surface of the specimens. The WB and L genomic DNA sequence reads were predominantly composed of the bacterial phyla Pseudomonodota/Proteobacteria, which accounted for 78% of the microbial population. Pseudomonodota is known for its diverse physiological and metabolic characteristics, allowing these bacteria to thrive in various environments, including soil, water, and the digestive systems of insects (Hernández *et al*., 2024). Regarding fungal composition, Ascomycota was the most abundant, making up 44% and 68% of the specimens, respectively. Regarding viral composition, the WB specimen was dominated by the phylum Uroviricota (40%), while the L specimen was primarily composed of Artveriricota (52%). The Uroviricota phylum is part of a group of viruses known as Riboviria, which infect various hosts, including animals, plants, fungi, and protozoa. The phylum Euglenozoa was the most prevalent in both specimens within

the domain Eukaryota, for which its relationship with tabanid flies has been extensively studied (Ganyukova *et al*., 2018; Votýpka *et al*., 2015; Taioe *et al*., 2017; Brotánková *et al*., 2022).

The microbiome found in the *T. par* exhibited a core composition at a bacterial level of *Enterobacteriaceae* (49%). Studies on Enterobacteriaceae have been reported in other tabanid flies, including *Atylotus* sp*., Haematopota turkestanica*, and *Ades aegypti* mosquitoes (Macleod *et al*., 2021; David *et al.,* 2016) and is regarded as pathogenic. Members of this family are tightly linked to viral infections and transmissions, such as conjunctivitis, urinary tract infections, respiratory and systemic infections as well as wounds and bloodstream infections (Kozlova *et al*., 2021). While Burkholeriaceae was a dominant family of bacteria in the L sequence reads Burkholeriaceae are common soil inhabitants, and this family contains pathogens of plants, humans, and animals (Mannaa *et al*., 2018). The fungal composition at family in the WB sequence reads was Aspergillaceae. Meanwhile, the L sequence reads content was dominated by Hypocreomycetidae (33%). The most abundant viruses at the family level in WB and L sequence reads was Metaviridae.

At the family level, the microbiome found in the *T. par* exhibited a core composition at a bacterial level of *Enterobacteriaceae* (49%). Studies on Enterobacteriaceae has been reported in other tabanid flies including *Atylotus* sp*., Haematopota turkestanica*, and *Ades aegypti* mosquitos (Macleod *et al*., 2021; David *et al.,* 2016) and is regarded as pathogenic. Members within the Enterobacteriaceae family are often linked to viral infections and their spread, encompassing ailments such as conjunctivitis, urinary tract infections, respiratory and systemic diseases, in addition to wound and bloodstream infections (Kozlova *et al*., 2021). While Burkholeriaceae was the dominant family of bacteria in the L-sequence reads. Burkholeriaceae are common soil inhabitants, and this family contains pathogens of plants, humans, and animals.

The fungal family composition in the WB sequence reads was primarily made up of Aspergillaceae, whereas the L sequence reads revealed that Hypocreomycetidae was more prevalent. Metaviridae emerged as the most common viral family in both WB and L sequence reads. This finding differs from Litov *et al.* (2023), which reported a variety of viral families, including Narnaviridae and Flaviviridae, across several genera. At the family level, the Eukaryota composition in WB was largely dominated by

Trypanosomatidae, akin to the observations of Odeniran *et al.* (2019), while the L reads exhibited a dominance of Dothideomycetidae

The bacterial communities associated with *T. par* WB data at the genus level revealed that *Escherichia* was the most abundant, comprising 37% of the population. In contrast, Paraburkholderia was the most abundant genus in the L sequence reads, accounting for 24%, while *Escherichia'*s prevalence was notably lower at only 3%. The high abundance of *Escherichia* has been reported in various studies (Stirrat *et al*., 1955; Perotti *et al*., 2001), particularly concerning the pathotype *E. coli* O157:H7, which is linked to severe gastrointestinal diseases in humans (Szalanski *et al*., 2004). Certain strains of *E. coli* have also been associated with various illnesses in other mammals and birds, including extraintestinal infections such as urinary tract infections, respiratory infections, meningitis, and sepsis, as well as intestinal diseases like diarrhea and dysentery (Mueller and Tainter, 2023). This mechanical transmission occurs when tabanids feed on the blood of infected hosts and subsequently transfer pathogens through their saliva or faeces during feeding.

Other bacteria of veterinary and agricultural significance included *S. phoeniceum* and several unclassified *Spiroplasma* species. Most *Spiroplasma* spp. are known to be commensals found in insects, arachnids, crustaceans, or plants, while only a limited number act as pathogens affecting these organisms. Insects are particularly abundant sources of *Spiroplasma*, which are often present in blood-feeding arthropods such as deerflies, horseflies, mosquitoes, and ticks, where they can be prevalent in the salivary glands, adipose bodies, and haemolymph (Cisak *et al*., 2015). *Spiroplasma* can also be transmitted vertically through the ovaries (Ballinger and Perlman, 2017). Previous research has indicated that *Spiroplasma* can function as an endosymbiont, providing defence against nematodes and parasitic wasps (Wang *et al*., 2024). This study suggests that *Spiroplasma* is prevalent in the gut of *T. par,* raising the possibility that this species may have a protective microbiota that is transmitted vertically.

Fungal abundance genus level in WB sequence reads was reported as *Synchytrium* (3%). This fungus thrives in environments rich in decomposing organic matter, often indicating nutrient-rich waters. They are called epi-or interbiotic, or aprobic, suggesting that they live in environments rich in organic matter but deficient in oxygen. Other fungi of veterinary importance in WB included *Batrachochytrium dendrobatidis* (Bd) and *N.*

cerenae. The pathogenic fungus *B. dendrobatidis* is responsible for causing the disease called chytridiomycosis in frogs, leading to population declines in numerous species of amphibians worldwide. Research on these fungi has never been carried out in tabanids, but a study by Beyer *et al*. (2024) showed that *Drosophila melanogaster* tested positive for Bd up to five days after exposure and could transmit Bd to conspecifics without experiencing mortality.

Meanwhile, *N. cerenae* is a parasitic intracellular fungus known to affect the midgut of honeybees. Regarding the L sequence reads, *Fusarium . proliferatum* ET1 (28%) was the most prevalent genus. This species is significant in agriculture because it produces various mycotoxins, such as fumonisin B1, moniliformin, beauvericin, and bikaverin. It can be an endophyte or pathogen on multiple plants, including economically important types like asparagus, fig, maize, onion, palm, pine, rice, and wheat. More studies need to be conducted to address the influence of this fungus on Dipteran flies (Jabal-Uriel *et al*., 2022). However, this suggests that tabanids may be drawn to environments where fungi are present, potentially influencing their feeding behaviour and habitat selection.

Viruses are the most abundant biological entity on the planet and interact with microbial communities with which they associate. Bacterial viruses, also called bacteriophages or phages, often dominate the virome of animals, and they have the potential to (re)shape bacterial communities that are crucial to the animal host (insert reference). When they bind to a susceptible host, they follow either a lytic or lysogenic replication strategy (Kasman and Porter, 2022). Research in various contexts, including their potential interactions with Arthropods, has recently been given attention, *Musa domestica* (Zang *et al.,* 2021), *Nasonia vitripennis* (Chafee *et al.,* 2011), A*. aegypti* mosquitoes (López-Cuevas *et al.,* 2022), *Acyrthosiphon pisum* (Weldon, 2015). Most of these studies were conducted on the gut of the mentioned arthropods; in the current study, phages detected in the WB-sequence reads and L-sequence reads of *T. par* were compared. At the genus level, the most abundant genus in both WB and L-sequence reads was the *Trichoplusia* ni TED virus. It is a baculovirus that infects cabbage looper moth (*Trichoplusia ni)*. The virus is part of the Baculoviridae family, which explicitly infects insects and mainly affects species of moths and butterflies in Lepidoptera.

Within the WB-sequence reads of *T. par*, viruses of veterinary importance associated with bat-borne viruses included Vespertiliovirus. While not all Vespertiliovirus are known to infect humans, bats are natural reservoirs for many viruses that can spill over to other animals, including humans, as seen with other pathogens like coronaviruses and henipaviruses. Other viruses of veterinary importance included *Spiroplasma* phage C74, although this phage occurred at a low abundance of 2%. The *T. par* fly-leg genomic DNA sequence data consisted of major pests of agricultural importance, such as, *B. congregatae* and *B. glomeratae.* These Bracoviruses are known to be associated with parasitic wasps, specifically by altering the host's immune system. Although this study offers some indications that bracoviruses could be found in tabanids (*T. par*), it remains uncertain whether this is a common occurrence or if tabanids have a notable impact on the spread of bracoviruses (Turčinavičienė *et al*., 2024).

Bacterial plasmids are small, circular, double-stranded DNA molecules that exist independently from the main chromosomal DNA within a bacterial cell. They are crucial for the host cell's metabolic processes (Al Doghaither and Gull, 2019.). In the current study, sequencing data from WB revealed that the *Wolbachia*-endosymbiont constitutes a significant part of the bacterial composition (34%). This endosymbiont has been identified in various tabanid species, as well as in other arthropods, and is known to affect host physiology, microbiota composition, and the spread of pathogens. Based on the findings of Madhave *et al*., (2020), *Wolbachia* sp. is increasingly being recognized for its promise in developing novel strategies for controlling numerous disease-causing pathogens transmitted by insects and pests that impact plants; however, its potential application in managing veterinary medicine diseases has received limited attention.

Additionally, a significant plasmid identified in the WB sequenced data is *P. aeruginosa* (4%), which is slightly lower in abundance than detected in *Musca domestica* (Housefly) 8.8% in a study by Hemmatinezhad *et al*, (2015). *Pseudomonas aeruginosa* is a complex, multi-drug-resistant pathogen that poses significant concerns for both veterinary and public health. This microorganism is associated with the development of several critical emerging diseases, such as hospital-acquired bloodstream infections and pneumonia (Qiu *et al*., 2024). The gastrointestinal tract is the main area

where *P. aeruginosa* colonizes and often serves as a common source for subsequent infections. Unfortunately, the existing control strategies for *P. aeruginosa* have largely overlooked arthropod vectors.

The sequenced data produced from the L sequence reads included a diverse range of bacterial compositions featuring distinct bacterial plasmids like *Caballeronia* sp. M124 and species of *Paraburkholderia*. In both datasets (WB and L), *Burkholderia* sp. plasmid was detected but their relative abundances differ.

Within the group Eukaryota, species of *Trypanosoma*, particularly *T. theileri*, were found to be the most abundant, accounting for a significant portion of the detected *Trypanosoma* (32%). The detection of *T. theileri* in tabanids dates back to 1887 (Rose *et al*., 1987) and has been the subject of ongoing research. Additionally, *T. theileri* DNA has been documented in *T. par* (Taioe *et al*., 2017; Votýpka *et al*., 2019; Mulandane *et al*., 2020). There have also been repeated reports of *T. theileri* in other *Tabanus* species, including *T. bromius* (Mazibuko, 2018), *T. maculicornis* (Turčinavičienė *et al*., 2024), *T. taeniola*, and *T. gratus* (Lendzele *et al*., 2022). In contrast to the pathogenic African animal trypanosomes (*T. vivax*, *T. congolense*, and *T. brucei* sp.) (Nimpaye *et al*., 2011), the Stercorarian trypanosomes species examined in this study, namely *Trypanosoma theileri*, *T. melophagium* (29%), and *T. grayi* (5%), are typically considered non-pathogenic (Kostygov *et al*., 2022). However, various reports suggest these organisms can act as opportunistic pathogens, especially when combined with viruses or piroplasmids. They may also cause mild to severe symptoms on their own, occasionally resulting in fatal loss and neonatal calf mortality (Caffarena *et al*., 2012; Martinković *et al*., 2012; Oldrieve *et al*., 2022; Desquesnes *et al*., 2022).

Unlike salivarian trypanosomes that mature in the anterior of the digestive tract and transmitted via a tsetse fly bite, the stercorarian trypanosomes develop in the posterior and are mechanically transmitted by biting insects through faeces (Radwanska *et al.*, 2018). In the fly legs, fungi were found to be more prevalent within Eukaryota, with the most abundant genera of medical significance, including *A. melanogenum* (7%) and *Rhodotorula* (4%). *Aureobasidium melanogenum*, formerly classified as *Aureobasidium pullulans*, which had not been documented in tabanids before current data was generated. This yeast-like fungus is predominantly found in freshwater

environments and is known to cause human infections. In contrast, the *Rhodotorula* fungus has been noted in various insects, particularly mosquitoes (Tobias, 2023). It has also been discovered in omnivorous crabs like *Aratus pisonii* and *Goniopsis cruentata*, as well as in clams (Wheeler *et al*., 2024). In humans, *Rhodotorula* has been observed in the microflora of patients who have oral cancer or psoriasis and is known to colonize terminally ill patients (Borkar and Shinde, 2023).

5.5 Conclusion

This study provides a comprehensive analysis of the microbiome in *T. par* and highlights the complex and diverse microbial communities associated with these dipterans. The findings reveal that adult horseflies, due to their morphological and anatomical features, serve as effective vectors for a wide range of microbes, including bacteria, viruses, fungi, and protozoa. The dominant microbial groups detected in *T. par* include the bacterial phylum Pseudomonadota, fungal family Aspergillaceae, and viral family Metaviridae, with several pathogenic and potentially pathogenic microorganisms identified, such as *E. coli* and *P. aeruginosa*. The study also uncovers the presence of *Wolbachia* and other important endosymbionts, which may influence the host's physiology and its role in pathogen transmission.

The variability in microbial composition between the head and legs of *T. par* suggests that different body parts may host distinct microbial communities, potentially influencing the fly's ecological interactions and vector capabilities. Notably, the presence of opportunistic pathogens such as *Trypanosoma theileri* emphasizes the potential for these flies to spread both zoonotic pathogens.

5.6 Recommendations

While the findings provide valuable insights into the microbiome of *T. par*, further research is needed to explore the ecological roles of these microbes, their interactions with the host, and their impact on disease transmission. Understanding the microbial dynamics of *T. par* will enhance our knowledge of how horseflies contribute to the spread of infectious diseases and inform strategies for controlling vector-borne pathogens.

5.7 References

1. Al Doghaither, H. & Gull, M. (2019). Plasmids as genetic tools and their applications in ecology and evolution. In *Plasmid*. IntechOpen.
2. Ammar, E.D., Gasparich, G.E., Hall, D.G. & Hogenhout, S.A. (2011). *Spiroplasma*-like organisms closely associated with the gut in five leafhopper species (Hemiptera: Cicadellidae). *Archives of microbiology*, *193*, pp.35-44.
3. Anbutsu, H. & Fukatsu, T. (2011). *Spiroplasma* as a model insect endosymbiont. *Environmental microbiology reports*, *3*(2), pp.144-153.
4. Simon, A. (2010). FastQC: a quality control tool for high throughput sequence data. *Version 0.10*, *1*.
5. Baldacchino, F., Desquesnes, M., Mihok, S., Foil, L.D., Duvallet, G. & Jittapalapong, S. (2014). Tabanids: neglected subjects of research, but important vectors of disease agents!. Infection, Genetics and Evolution, 28, pp.596-615.
6. Baldacchino, F., Muenworn, V., Desquesnes, M., Desoli, F., Charoenviriyaphap, T. & Duvallet, G. (2013). Transmission of pathogens by *Stomoxys* flies (Diptera, Muscidae): a review. *Parasite*, *20*.
7. Ballinger, M.J. & Perlman, S.J. (2017). Generality of toxins in defensive symbiosis: Ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*. *PLoS pathogens*, *13*(7), p.e1006431.
8. Ballinger, M.J., Gawryluk, R.M. & Perlman, S.J. (2019). Toxin and genome evolution in a *Drosophila* defensive symbiosis. *Genome Biology and Evolution*, *11*(1), pp.253-262.
9. Baquero, F., Coque, T.M., Galán, J.C. & Martinez, J.L. (2021). The origin of niches and species in the bacterial world. Frontiers in microbiology, 12, p.657986.
10. Bayer, L.V., Milano, S.N. & Bratu, D.P. (2024). The mRNA dynamics underpinning translational control mechanisms of *Drosophila melanogaster* oogenesis. Biochemical Society Transactions, p.BST20231293.
11. Bolanos, L.M., Servin-Garciduenas, L.E. & Martinez-Romero, E. (2015). Arthropod–Spiroplasma relationship in the genomic era. *FEMS Microbiology Ecology*, *91*(2), pp.1-8.

12. Bolger, A.M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), pp.2114-2120.
13. Borkar, S.G. & Shinde, K. (2023). Yeast species of diverse functionality in health sciences: A concise report. *GSC Biological and Pharmaceutical Sciences*, *25*(2), pp.149-168.
14. Brotánková, A., Fialová, M., Čepička, I., Brzoňová, J. & Svobodová, M., (2022). Trypanosomes of the *Trypanosoma theileri* group: Phylogeny and new potential vectors. *Microorganisms*, *10*(2), p.294.
15. Byer, A.M., Nguyen, K.M., Katz, T.S., Chen, R. & Briggs, C.J. (2024). Drosophila melanogaster as a model arthropod carrier for the amphibian chytrid fungus Batrachochytrium dendrobatidis. Plos one, 19(7), p.e0307833.
16. Caffarena, R.D., Casaux, M.L., Schild, C.O., Fraga, M., Castells, M., Colina, R., Maya, L., Corbellini, L.G., Riet-Correa, F. & Giannitti, F. (2021). Causes of neonatal calf diarrhea and mortality in pasture-based dairy herds in Uruguay: a farm-matched case-control study. *Brazilian Journal of Microbiology*, *52*(2), pp.977-988.
17. Chafee, M.E., Zecher, C.N., Gourley, M.L., Schmidt, V.T., Chen, J.H., Bordenstein, S.R., Clark, M.E. & Bordenstein, S.R. (2011). Decoupling of host–symbiont–phage coadaptations following transfer between insect species. *Genetics*, *187*(1), pp.203-215.
18. Cisak, E., Wójcik-Fatla, A., Zajac, V., Sawczyn, A., Sroka, J. & Dutkiewicz, J. (2015). Spiroplasma-an emerging arthropod-borne pathogen?. *Annals of Agricultural and Environmental Medicine*, *22*(4).
19. Cisak, E., Wójcik-Fatla, A., Zajac, V., Sawczyn, A., Sroka, J. & Dutkiewicz, J. (2015). *Spiroplasma* an emerging arthropod-borne pathogen?. *Annals of Agricultural and Environmental Medicine*, *22*(4).
20. David, M.R., Santos, L.M.B.D., Vicente, A.C.P. & Maciel-de-Freitas, R., (2016). Effects of environment, dietary regime and ageing on the dengue vector microbiota: evidence of a core microbiota throughout *Aedes aegypti* lifespan. *Memórias do Instituto Oswaldo Cruz*, 111(9), pp.577-587.
21. Desquesnes, M., Gonzatti, M., Sazmand, A., Thévenon, S., Bossard, G., Boulangé, A., Gimonneau, G., Truc, P., Herder, S., Ravel, S. & Sereno, D. (2022). A review on the diagnosis of animal trypanosomoses. *Parasites & Vectors*, *15*(1), p.64.

22. Desquesnes, M., Gonzatti, M., Sazmand, A., Thévenon, S., Bossard, G., Boulangé, A., Gimonneau, G., Truc, P., Herder, S., Ravel, S. & Sereno, D. (2022). A review on the diagnosis of animal trypanosomoses. *Parasites & Vectors*, *15*(1), p.64.
23. Ganyukova, A.I., Drachko, D.O., Malysheva, M.N., Frolov, A.O., Agasoi, V.V., Smirnov, P.A. & Kostygov, A.Y. (2024). High prevalence of *Trypanosoma theileri*-like trypanosomes and scarcity of monoxenous trypanosomatids in tabanids of Northwestern Russia. Protistology, 18(1), pp.4-21.
24. Ganyukova, A.I., Zolotarev, A.V., Malysheva, M.N. & Frolov, A.O. (2018). First record of Trypanosoma theileri-like flagellates in horseflies from Northwest Russia. *Protistology*, *12*(4), pp.223-230.
25. Gasparich, G.E. (2010). *Spiroplasmas* and phytoplasmas: microbes associated with plant hosts. *Biologicals*, *38*(2), pp.193-203.
26. Gasparich, G.E., Whitcomb, R.F., Dodge, D., French, F.E., Glass, J. & Williamson, D.L. (2004). The genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the Mycoplasma mycoides clade. *International Journal of Systematic and Evolutionary Microbiology*, *54*(3), pp.893-918.
27. González, M.A., Ruiz-Arrondo, I., Bravo-Barriga, D., Cervera-Acedo, C., Santibáñez, P., Oteo, J.A., Miranda, M.Á. & Barceló, C. (2024). Surveillance and screening of Stomoxyinae flies from Mallorca Island (Spain) reveal the absence of selected pathogens but confirm the presence of the endosymbiotic bacterium *Wolbachia pipientis*. *Research in Veterinary Science*, *171*, p.105206.
28. Hélias, C., Vazeille-Falcoz, M., Le Goff, F., Abalain-Colloc, M. L., Rodhain, F., Carle, P.& Chastel, C. (1998). *Spiroplasma turonicum* sp. nov. from *Haematopota* horse flies (Diptera: Tabanidae) in France. *International Journal of Systematic and Evolutionary Microbiology*, *48*(2), 457-461.
29. Hemmatinezhad, B., Ommi, D., Hafshejani, T.T. & Khamesipour, F. (2015). Molecular detection and antimicrobial resistance of *Pseudomonas aeruginosa* from houseflies (*Musca domestica*) in Iran. *Journal of Venomous Animals and Toxins including Tropical Diseases*, *21*, pp.1-5.
30. Hernández-Rodríguez, D., Hernández-Sánchez, H., Andrade-Velásquez, A., & Melgar-Lalanne, G. (2025). Characterization of lactic acid bacteria isolated

from the gastrointestinal tract of *scaptotrigona mexicana* and *scaptotrigona* pectoralis stingless bees. *Journal of Apicultural Research*, *64*(1), 99-111.

31. Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. & Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, *11*, pp.1-11.
32. Jabal-Uriel, C., Barrios, L., Bonjour-Dalmon, A., Caspi-Yona, S., Chejanovsly, N., Erez, T., Henriques, D., Higes, M., Le Conte, Y., Lopes, A.R. & Meana, A., (2022). Epidemiology of the microsporidium *Nosema ceranae* in four Mediterranean countries. *Insects*, *13*(9), p.844.
33. Jaenike, J., Unckless, R., Cockburn, S.N., Boelio, L.M. & Perlman, S.J., (2010). Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science*, 329(5988), pp.212-215.
34. Kasman, L.M. & Porter, L.D. (2022). Bacteriophages. In *StatPearls [Internet]*. StatPearls Publishing.
35. Kostygov, A.Y., Frolov, A.O., Malysheva, M.N., Ganyukova, A.I., Drachko, D., Yurchenko, V. & Agasoi, V.V. (2022). Development of two species of the *Trypanosoma theileri* complex in tabanids. *Parasites & Vectors*, *15*(1), p.95.
36. Kozlova, E.V., Hegde, S., Roundy, C.M., Golovko, G., Saldaña, M.A., Hart, C.E., Anderson, E.R., Hornett, E.A., Khanipov, K., Popov, V.L. & Pimenova, M. (2021). Microbial interactions in the mosquito gut determine Serratia colonization and blood-feeding propensity. *The ISME journal*, *15*(1), pp.93-108.
37. Ku, C., Lo, W.S., Chen, L.L. & Kuo, C.H. (2013). Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. *Genome biology and evolution*, *5*(6), pp.1151-1164.
38. Kurilkina, M.I., Zakharova, Y.R., Galachyants, Y.P., Petrova, D.P., Bukin, Y.S., Domysheva, V.M., Blinov, V.V. & Likhoshway, Y.V. (2016). Bacterial community composition in the water column of the deepest freshwater Lake Baikal as determined by next-generation sequencing. *FEMS microbiology ecology*, *92*(7), p. fiw094.
39. Lendzele, S.S., Abah, S., Nguetoum, C., Burinyuy, K.A., Koumba, A.A. & Mavoungou, J.F. (2022). Tabanid-transmitted animal trypanosomiasis in

Cameroon: Evidence from a study in the tsetse free pastoral zone of Galim. *Parasite Epidemiology and Control*, *18*, p.e00253.

40. Li, D., Liu, C.M., Luo, R., Sadakane, K. & Lam, T.W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), pp.1674-1676.
41. Litov, A.G., Belova, O.A., Kholodilov, I.S., Kalyanova, A.S., Gadzhikurbanov, M.N., Rogova, A.A., Gmyl, L.V. & Karganova, G.G. (2023). Viromes of tabanids from Russia. Viruses, 15(12), p.2368.
42. Liu, P., Li, Y., Ye, Y., Chen, J., Li, R., Zhang, Q., Li, Y., Wang, W., Meng, Q., Ou, J. & Yang, Z. (2022). The genome and antigen proteome analysis of *Spiroplasma mirum*. *Frontiers in Microbiology*, *13*, p.996938.
43. Lo, W.S., Gasparich, G.E. & Kuo, C.H. (2015). Found and lost: the fates of horizontally acquired genes in arthropod-symbiotic *Spiroplasma*. *Genome Biology and Evolution*, *7*(9), pp.2458-2472.
44. Łukasik, P., van Asch, M., Guo, H., Ferrari, J. & Charles J. Godfray, H., (2013). Unrelated facultative endosymbionts protect aphids against a fungal pathogen. *Ecology letters*, *16*(2), pp.214-218.
45. MacLeod, H.J., Dimopoulos, G. & Short, S.M. (2021). Larval diet abundance influences size and composition of the midgut microbiota of *Aedes aegypti* mosquitoes. *Frontiers in Microbiology*, *12*, p.645362.
46. Majoros, S.E. & Adamowicz, S.J. (2023). CanFlyet: Habitat Zone and Diet Trait Dataset for Diptera Species of Canada and Greenland. *bioRxiv*, pp.2023-03.
47. Majoros, S.E., Adamowicz, S.J. & Cottenie, K. (2023). Novel pipeline for large-scale comparative population genetics. *bioRxiv*, pp.2023-01.
48. Mannaa, M., Park, I., & Seo, Y. S. (2018). Genomic features and insights into the taxonomy, virulence, and benevolence of plant-associated *Burkholderia* species. *International journal of molecular sciences*, *20*(1), 121.
49. Martinković, F., Matanović, K., Rodrigues, A.C., Garcia, H.A. & Teixeira, M.M. (2012). *T rypanosoma (M egatrypanum) melophagium* in the Sheep Ked *Melophagus ovinus* from Organic Farms in Croatia: Phylogenetic Inferences Support Restriction to Sheep and Sheep Keds and Close Relationship with Trypanosomes from Other Ruminant Species. *Journal of eukaryotic microbiology*, *59*(2), pp.134-144.

50. Masson, F., Calderon Copete, S., Schüpfer, F., Garcia-Arraez, G. & Lemaitre, B. (2018). *In vitro* culture of the insect endosymbiont *Spiroplasma* poulsonii highlights bacterial genes involved in host-symbiont interaction. *MBio*, *9*(2), pp.10-1128.
51. Mazibuko, X. (2018). *Species composition of and detection of haemoparasites in horse-flies (Diptera:* Tabanidae*) in south-eastern Kruger National Park*. University of Pretoria (South Africa).
52. Mende, D. R., Letunic, I., Huerta-Cepas, J., Li, S. S., Forslund, K., Sunagawa, S., & Bork, P. (2017). proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes. Nucleic Acids Research, 45(D1), D529-D534.
53. Menzel, P., NG & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju Communicationa, 7, 1125.
54. Mondal, S., Somani, J., Roy, S., Babu, A. & Pandey, A.K. (2023). Insect microbial symbionts: Ecology, interactions, and biological significance. Microorganisms, 11(11), p.2665.
55. Moran, N.A. & Telang, A. (1998). Bacteriocyte-associated symbionts of insects. *Bioscience*, *48*(4), pp.295-304.
56. Moulder, R.W., French, F.E. & Chang, C.J. (2002). Simplified media for spiroplasmas associated with tabanid flies. *Canadian journal of microbiology*, *48*(1), pp.1-6.
57. Mueller, M. & Tainter, C.R. (2023). *Escherichia coli* infection. In *StatPearls [Internet]*. StatPearls Publishing.
58. Mulandane, F.C., Snyman, L.P., Brito, D.R., Bouyer, J., Fafetine, J., Van Den Abbeele, J., Oosthuizen, M., Delespaux, V. & Neves, L. (2020). Evaluation of the relative roles of the Tabanidae and Glossinidae in the transmission of trypanosomosis in drug resistance hotspots in Mozambique. *Parasites & Vectors*, *13*, pp.1-16.
59. NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 46(D1), D8-D13.
60. Nguyen, N.T., Borgemeister, C., Poehling, H.M. & Zimmermann, G. (2007). Laboratory investigations on the potential of entomopathogenic fungi for

biocontrol of *Helicoverpa armigera* (Lepidoptera: Noctuidae) larvae and pupae. *Biocontrol Science and Technology*, *17*(8), pp.853-864.

61. Nimpaye, H., Njiokou, F., Njine, T., Njitchouang, G.R., Cuny, G., Herder, S., Asonganyi, T. & Simo, G. (2011). *Trypanosoma vivax, T. congolense* "forest type" *and T. simiae*: prevalence in domestic animals of sleeping sickness foci of Cameroon. *Parasite: Journal de la Société Française de Parasitologie*, *18*(2), p.171.
62. Odeniran, P.O., Macleod, E.T., Ademola, I.O. & Welburn, S.C. (2019). Molecular identification of bloodmeal sources and trypanosomes in *Glossina* spp., *Tabanus* spp. and *Stomoxys* spp. trapped on cattle farm settlements in southwest Nigeria. *Medical and veterinary entomology*, *33*(2), pp.269-281.
63. Oldrieve, G.R., Venter, F., Cayla, M., Verney, M., Hebert, L., Geerts, M., Van Reet, N. & Matthews, K. (2024). Mechanisms of life cycle simplification in African trypanosomes. bioRxiv, pp.2024-07.
64. Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser with Krona. BMC Bioinformatics, 12(1), 385.
65. Ondov, B.D., Bergman, N.H. & Phillippy, A.M. (2015). Krona: interactive metagenomic visualization in a web browser. *Encyclopedia of Metagenomics*, pp.339-346.
66. Perotti, M.A., Lysyk, T.J., Kalischuk-Tymensen, L.D., Yanke, L.J. & Selinger, L.B. (2001). Growth and survival of immature *Haematobia irritans* (Diptera: Muscidae) is influenced by bacteria isolated from cattle manure and conspecific larvae. *Journal of Medical Entomology*, *38*(2), pp.180-187.
67. Qiu, H.Y., Lv, Q.B., Wang, C.R., Ju H., Luo, C.F., S.S., Na, M.H., Chang, Q.C. & Jiang, J.F. 2024. Microbiota profile in organs of horseflies (Diptera: Tabanidae) in Northeastern China. *Frontiers in Microbiology*, 15, p.1467875.
68. Radwanska, M., Vereecke, N., Deleeuw, V., Pinto, J. & Magez, S. (2018). Salivarian trypanosomosis: a review of parasites involved, their global distribution and their interaction with the innate and adaptive mammalian host immune system. *Frontiers in immunology*, *9*, p.2253.
69. Regassa, L.B. & Gasparich, G.E. (2006). *Spiroplasmas*: evolutionary relationships and biodiversity. *Front Biosci*, *11*(1), pp.2983-3002.

70. Regassa, L.B., French, F.E., Stewart, K.M., Murphy, A.C., Jandhyam, H.L. & Beati, L. (2011). A Costa Rican bacterial *Spiroplasma* biodiversity survey in tabanid flies reveals new serogroups and extends United States ranges. *International Journal of Biodiversity and Conservation*, *3*(8), pp.338-344.
71. Shang, Y., Feng, P. & Wang, C. (2015). Fungi that infect insects: altering host behavior and beyond. *PLoS pathogens*, *11*(8), p.e1005037.
72. Sontigun, N., Boonhoh, W., Fungwithaya, P. & Wongtawan, T. (2022). Multiple blood pathogen infections in apparently healthy sheltered dogs in southern Thailand. *International Journal of Veterinary Science and Medicine*, *10*(1), pp.64-71.
73. Stefanini, I. (2018). Yeast-insect associations: It takes guts. *Yeast*, *35*(4), pp.315-330.
74. Stirrat, J.H., McLintock, J., Schwindt, G.W. & Depner, K.R. (1955). Bacteria associated with wild and laboratory-reared horn flies, *Siphona irritans* (L.)(Diptera: Muscidae). *The Journal of Parasitology*, *41*(4), pp.398-406.
75. Stoffolano, J.G. (2019). Fly foregut and transmission of microbes. *Advances in insect physiology*, *57*, pp.27-95.
76. Szalanski, A.L., Owens, C.B., McKay, T. & Steelman, C.D. (2004). Detection of *Campylobacter* and *Escherichia coli* O157: H7 from filth flies by polymerase chain reaction. *Medical and Veterinary Entomology*, *18*(3), pp.241-246.
77. Taioe, M.O., Motloang, M.Y., Namangala, B., Chota, A., Molefe, N.I., Musinguzi, S.P., , K., Hayes, P., Tsilo, T.J., Chainey, J. & Inoue, N. (2017). Characterization of tabanid flies (Diptera: Tabanidae) in South Africa and Zambia and detection of protozoan parasites they are harbouring. *Parasitology*, *144*(9), pp.1162-1178.
78. Tobias, T.B., 2023. *Freshwater Fungal Diversity and Ecological Interactions With Mosquitoes and Plant Communities* (Doctoral dissertation, Western Illinois University).
79. Turčinavičienė, J., Bernotienė, R. & Petrašiūnas, A. (2024). Molecular Detection and Analysis of *Trypanosoma* (Megatrypanum) spp. Diversity in Tabanidae (Diptera) Collected in Lithuania. *Insects*, *15*(8), p.581.

80. Votýpka, J., Brzoňová, J., Ježek, J. & Modrý, D. (2019). Horse flies (Diptera: Tabanidae) of three West African countries: a faunistic update, barcoding analysis and trypanosome occurrence. *Acta tropica*, *197*, p.105069.
81. Votýpka, J., d'Avila-Levy, C.M., Grellier, P., Maslov, D.A., Lukeš, J. & Yurchenko, V. (2015). New approaches to systematics of Trypanosomatidae: criteria for taxonomic (re) description. *Trends in Parasitology*, *31*(10), pp.460-469.
82. Wang, X., Liu, Q., Wang, J., Wang, L., Tu, H. & Zhang, J. (2024). Identification and Occurrence of the Dominant Parasitic Wasp *Therophilus festivus* of Larval *Grapholita molesta. Agronomy*, *14*(10), p.2307.
83. Werszko, J., Szewczyk, T., Steiner-Bogdaszewska, Ż., Wróblewski, P., Karbowiak, G. & Laskowski, Z. (2020). Molecular detection of Megatrypanum trypanosomes in tabanid flies. *Medical and veterinary entomology*, *34*(1), pp.69-73.
84. Wheeler, C., Pearman, J.K., Howarth, J.D., Vandergoes, M.J., Holt, K., Trewick, S.A., Li, X., Thompson, L., Thomson-Laing, G., Picard, M. & Moy, C. (2024). A paleoecological investigation of recent cyanobacterial blooms and their drivers in two contrasting lakes. *Harmful Algae*, *131*, p.102563.
85. Whitcomb, R.F., Tully, J.G., Gasparich, G.E., Regassa, L.B., Williamson, D.L. & French, F.E. (2007). *Spiroplasma* species in the Costa Rican highlands: implications for biogeography and biodiversity. *Biodiversity and Conservation*, *16*, pp.3877-3894.
86. Xie, J., Butler, S., Sanchez, G. & Mateos, M. (2014). Male killing *Spiroplasma* protects *Drosophila melanogaster* against two parasitoid wasps. *Heredity*, *112*(4), pp.399-408.
87. Yao, C. (2010). Major surface protease of trypanosomatids: one size fits all?. *Infection and immunity*, *78*(1), pp.22-31.
88. Yeoman, C.J., Brutscher, L.M., Esen, Ö.C., Ibaoglu, F., Fowler, C., Eren, A.M., Wanner, K. & Weaver, D.K. (2019). Genome-resolved insights into a novel Spiroplasma symbiont of the Wheat Stem Sawfly (*Cephus cinctus*). PeerJ, 7, p.e7548.
89. Zhang, X., Wang, S., Li, T., Zhang, Q., Zhang, R. & Zhang, Z., 2021. Bacteriophage: a useful tool for studying gut bacteria function of housefly larvae, *Musca domestica*. *Microbiology Spectrum*, *9*(1), pp.10-1128.

CHAPTER 6 GENOMIC DIVERSITY AND DISTRIBUTIONS OF BACTERIOPHAGES HARBOURED BY *TABANUS PAR*

6.1 Introduction

Bacteriophages (Phages) are essential for bacterial diversity, gene change, and biogeochemical processes (Al-Shayeb *et al*., 2020). They commonly belong to the order Caudovirales with members of the Siphoviridae, Podoviridae, Myoviridae, and Tectiviridae families. They are known to form diverse interactions with eukaryotic organisms, including arthropods (Mondal *et al*., 2023), and have evolved in different natural environments (Hegarty *et al*., 2022). Their size, appearance, and genetic organization exhibit remarkable diversity (Simmonds and Aiewsakun, 2018). Phages infect bacterial communities by attacking how bacteria regulate gene expression to adapt to their surroundings (Hibstu *et al*., 2022). Research has identified distinct bacteriophage species according to their host ranges, morphology, and replication tactics (Ackermann, 2005). Like all viruses, bacteriophages often only infect a single species of bacteria or even particular strains within a species, and they are highly host-specific (Secor *et al*., 2020). Phages can also be categorized based on their life cycles, which can be chronic, lytic, or lysogenic, as shown in Figure 6-1 (Chevallereau *et al*., 2020).

Based on their life cycle, they can either be virulent or temperate (Dennehy and Abedon, 2021). Both temperate and virulent phages drive bacterial evolution, either by introducing new genetic material (temperate) or by selecting resistant strains (virulent) (Kasman and Porter, 2022). This can influence the bacterial diversity and population dynamics within the host, as phages play a crucial role in maintaining the balance of the host's microbial ecosystem (Young, 2014). Given the significance of comprehending insect-associated microbiomes and their possible uses, phages significantly impact the microbial communities that inhabit insects, affecting the microbiomes' makeup and functionality (Kolasa and Lukasik, 2023). These functions can affect the general health and behavior of the insect host, as well as facilitate horizontal gene transfer and control bacterial populations (Salazar *et al.,* 2021).

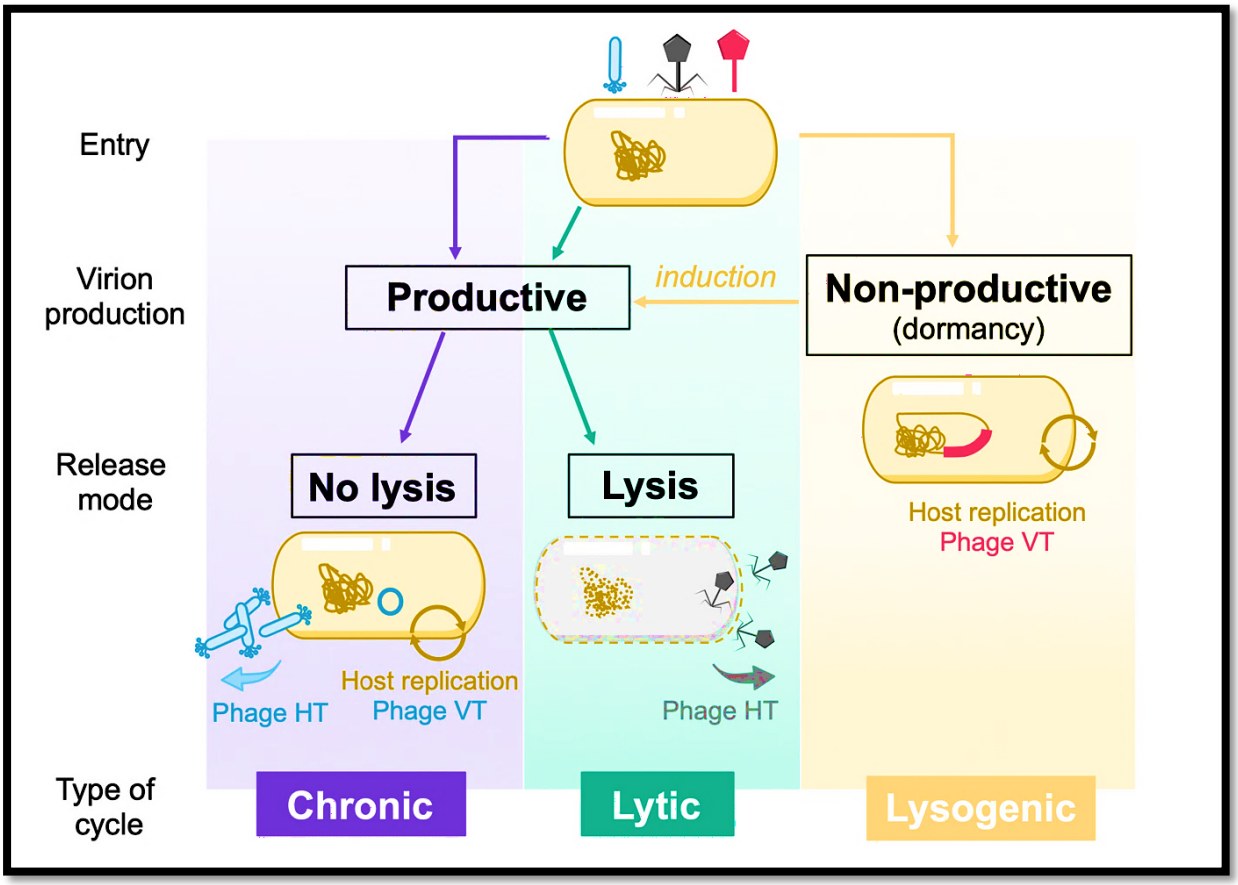


Figure 6-1: Schematic diagram illustrating life cycle of phage within a host (Chevallareau *et al.*, 2020).

To date, the discovery of phages has become of significant interest due to their impact on ecology, bacterial pathogenicity or commensalism, and their potential source of antibacterials. Advances in phage (meta)genomics have identified an unprecedented array of phage genomes, revealing a vast diversity of phages and their significant influence on microbial communities (Putzeys *et al.*, 2024). At the same time, phage genomics has revealed a significant lack of understanding regarding theoretical and functional proteins, which may be linked to novel regulatory mechanisms influencing phage-host interactions (Putzeys *et al.*, 2024).

Studies have investigated phages linked to insects such as mosquitos (Chauvatcharin *et al.*, 2006), honeybees (Merrill *et al.*, 2014), black soldier flies (Chen *et al.*, 2019), and houseflies (Zang *et al.*, 2019). In a recent study, Zang *et al.* (2022) described the bacteriophages that impact *Musca domestica*, highlighting harmful phages such as *Pseudomonas* and *Providencia*, as well as beneficial ones like *Enterobacter* and *Klebsiella*. However, research has not been conducted on phages associated with *T. par*. Understanding the interactions between phages and their bacterial hosts in hematophagous flies could provide insights into various aspects of fly biology,

including their vector competence, immune responses, and overall fitness. Additionally, exploring the diversity and function of phages in hematophagous fly microbiomes may reveal new strategies for controlling vector fly populations.

6.2 Materials and methods

6.2.1 Source of *T. par*

Adult *T. par* specimen was collected at three nature reserves, namely Charters Creek Reserve, Hluhluwe Game Reserve, and Phinda Private Game Reserve in KwaZulu-Natal, South Africa, and identified morphologically and genetically as described in Chapter 4, section 4.2.3 and section 4.4.3.

6.2.2 DNA extraction procedure

Prior to genomic DNA extractions, fly specimens were washed with 10% Tween 20 by immersion for 1 hour, then washed twice with 70% Ethanol, and rinsed with double distilled water and left to dry. The whole body (without legs) of *T. par* specimens was cut into pieces in 1.5 ml microcentrifuge tubes using sterile scissors, whereas the legs were separately crushed using a sterilized mortar and pestle and then transferred into a 1.5 ml microcentrifuge tube. Extraction was conducted using Qiagen DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Detailed procedure is described in Chapter 4, section 4.2.6 with the following modifications: The body and legs were homogenized separately in a mixture containing 180 µl of Buffer ATL and 20 µl of proteinase K, vortexed and incubated at 56°C until completely lysed. Ultimately, genomic DNA was eluted with 200 µl Buffer AE and then stored at –20°C.

6.2.3 Metagenomic sequencing

Whole-genome sequencing (WGS) of the *T. par* DNA was performed at the Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa as explained in Chapter 5, section 5.2.3.

6.2.4 Bioinformatics analysis

Using FastQC v 0.11.9 (Andrews, 2010), the quality of the raw sequencing reads was initially evaluated for overall quality scores, GC content, and possible adaptor contamination. Trimming v 0.39 (Bolger *et al*.,2014) was used to trim reads to remove adaptor sequences and low-quality bases. The following parameters were used during the trimming process: ILLUMINACLIP: TruSeq3-PE. fa:2:30:10 , LEADING:3 , TRAILING:3 , SLIDINGWINDOW:4:15 , and MINLEN:3 . MetaSPAdes v3.15.3 (Nurk *et al*., 2017) was used to assemble high-quality, trimmed reads into contigs. The assembly was carried out using the default settings, designed to handle metagenomics data's complex and varied nature. The quality and completeness of the resultant contigs were assessed further. To detect bacteriophage metagenome-assembled genomes (MAGs), VirSorter 1.0.5 (Bolduc *et al*., 2017) was utilized to analyze the filtered contigs. Using the default settings, the tool identified and categorized viral sequences from the assembled contigs. After VirSorter categorised contigs as high-confidence viral sequences.

Viral CONTigs Automatic Clustering and Taxonomy (vConTACT2) (Bolduc *et al.,* 2017) was used to classify the distribution of viral contigs generated. The assembly contigs were extracted and further analyzed using viral genomes from VIPtree and JGI phage tool (Huntemann *et al*.,2015). For each assembled phage genome, Pharokka (v1.0.0) was used for open reading frame (ORF) calling with annotation using default parameters combined with Prokka (v1.14.5) (Terzian *et al*. 2021) to ensure accurate determination of coding sequences (CDS). The presence of temperate/lytic lifestyle of genes was determined using PhageAI (Yukgehnaish *et al*. 2022).

A genome-based phylogeny was generated using VICTOR (Meier-Kolthoff and Victor, 2017). Intergenomic distances were calculated using the Genome BLAST Distance Phylogeny (GBDP) D6 formula for nucleotide sequences. To determine the relationships within each family, smaller phage genome datasets were analyzed by calculating intergenomic similarities based on nucleic acid sequences using VIRIDIC (Moraru *et al*., 2020) computed by tBLASTx scores. VICTOR and VIRIDIC can group the analyzed sequences within the genus and species based on the estimated distances. Using the viral genome, the family-level classification of phages was

performed with ViPTree (Rohwer and Edwards, 2002). Core protein analysis of core genes was calculated with VirClust (Moraru,2023).

6.3 Results

6.3.1 Phage distribution taxa of the genomes

To classify archaeal virus contigs into established or new taxa, vConTACT 2.0 was employed. This network-based tool uses whole genome gene-sharing profiles, integrating distance-based hierarchical clustering with confidence scores to facilitate virus classification, and assesses performance relative to ICTV-assigned, manually curated taxa. The distribution of viral meta-assembled genomes (vMAGs) of *T. par* in the whole body (WB) and legs (L) revealed that only 63 were associated with bacteria (bacterioviruses, formerly bacteriophages). About 95% of the dsDNA bacterioviruses belong to just one order (Caudovirales) and only three families (Myoviridae, Podoviridae, and Siphoviridae). Notably, *Mycobacterium* and *Pseudomonas* phages were reported in both meta-assembled fly parts. i.e., WB and L. Specifically, the relative abundance of *Mycobacterium* and *Pseudomonas* phages in the WB was 338% and 149%, respectively. In the fly-leg, the relative abundance was 310% and 118%, respectively. In contrast, *Bacillus* phages were more abundant (163%) in the WB of *T. par*, while *Escherichia* phages (133%) were more prevalent in the legs (Figures 6-2 and 6-3).

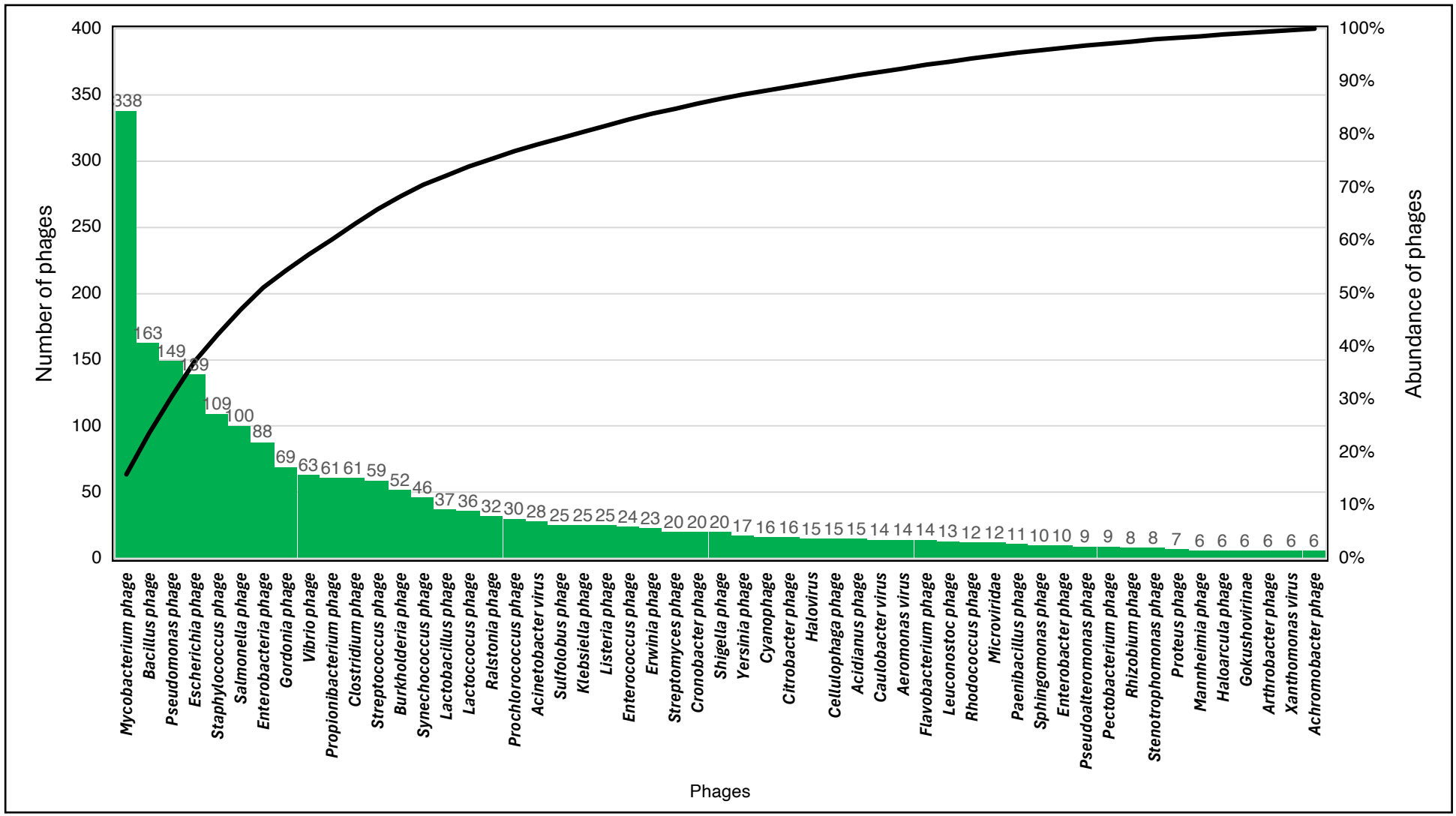


Figure 6-2: The taxonomic assignment and distribution of phage genomes identified from the body of *T. par.*

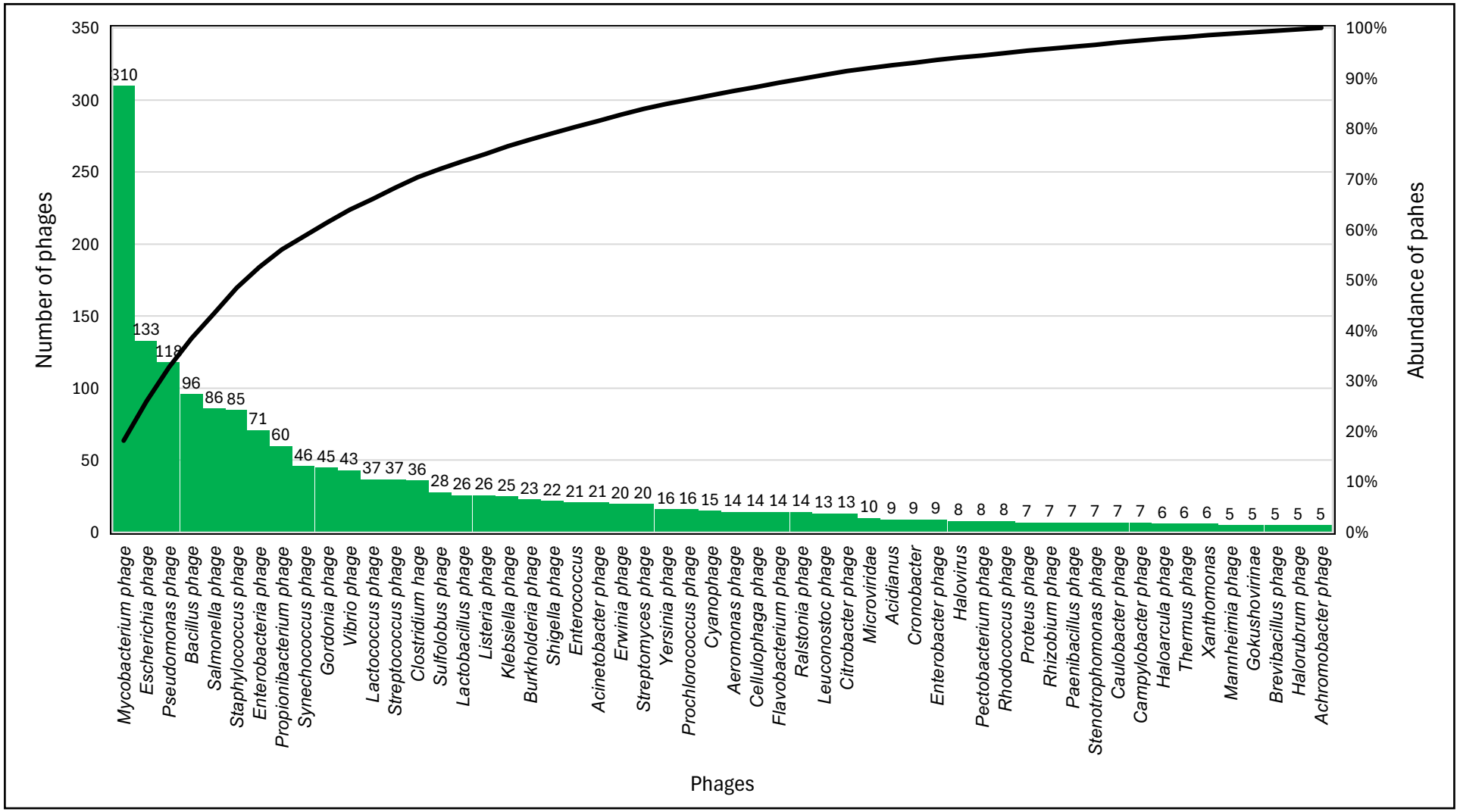


Figure 6-3: The taxonomic assignment and distribution of phage genomes identified from the legs of *T. par.*

6.3.2 Taxonomic classification and characterization of complete phages

A summary of the genomic features of the ten complete phage genomes characterized in this study from *T. par* of the WB and L is shown in Tables 9 and 10. These phages are identified into three families, with some having subfamilies. Six and four complete genomes were identified from the whole body and legs, respectively. All the identified phage genomes are linear double-stranded DNA (dsDNA). Their genome size ranges from 10.584bp to 197,841bp, with a GC content between 30.28 and 60.61% (Tables 6-1 and 6-2).

The transfer RNAs (tRNAs) were predicted in *Arcanobacterium* phage TPWB and *Caulobacter* phage TPWB, which have 5 and 6, respectively. Based on the BLASTn and PhageAI analysis, the characterized phages of the legs are classified within the class Caudoviricetes and belong to three genera: Cecivirus, Lambdavirus, and Phietavirus (as detailed in Table 6-1). BLASTn analysis revealed that the phage genomes had high sequence similarity (> 95%) to their reference phage genomes from the GenBank. Results showed that phage genomes sizes were 10,584bp; 25,801bp, 38,771bp, 44,506bp, 46,612bp, 47,313 bp, 59,598bp, 93,842bp, 170,231bp and 197,841bp of the following *Staphylococcus* phage TPL, *Stenotrophomonas* phage TPWB, *Acinetobacter* phage TPWB, *Bacillus* phage TPL, *Arcarnobacterium* phage TPWB *Staphylococcus* phage TPWB, *Caudovirales* phage TPWB, *Klebsiella* phage TPL, *Escherichia coli* TPL and *Caulobacter* phage TPWB, respectively.

The BLASTn results indicated that these phages had significant hits against the genome sequences with identifiers [DQ_908929, OM_735689, NC_029024, OM_373549, MN_101225, MF_374379], but no hits were found for *Acinetobacter* phage TPWB, *Caulobacter* phage TPWB, *Caudovirales* phage TPWB, *Staphylococcus* phage TPWB, and *Stenotrophomonas* phage TPWB, suggesting that these phages might represent novel or less characterized variants. One of the phages, *Arcarnobacterium* phage TPWB, was identified from the whole body. However, according to the Phannotate annotation implemented in the Pharroka tool, this phage was classified as a *Staphylococcus* phage, interacting with the host of *Staphylococcus pseudintermedius*. Most of the phages, including Staphylococcus phage TPL, Caulobacter phage TPWB, *Staphylococcus* phage TPWB, and *Acinetobacter* phage

TPWB, were found to be novel species (ANI < 95%) and were classified under the family Siphoviridae. These phages exhibited over 90% similarity to other members of the Siphoviridae family, indicating that while they are novel in their specific contexts, they share significant genetic similarities with existing Siphoviridae phages.

Table 6-1: Summary of the genomic features of phage genomes identified from the whole body of *Tabanus par.*

Features	***Caulobacter* phage TPWB**	***Arcarnobacterium* phage TPWB**	***Stenotrophomonas* phage TPWB**	***Caudovirales* phage TPWB**	***Staphylococcus* phage TPWB**	***Acinetobacter* phage TPWB**
Closest accession numbers*	Unspecified	OM373549	OM735689	Unspecified	Unspecified	Unspecified
Assigned accession number	PQ358087	PQ358086	PQ358089	Unspecified	PQ358088	PQ374146
Length bp	197 841	93 842	25 801	47 313	46 612	38 771
Classification	Unspecified	*Staphylococcus* phage vB_SpsS-DH5	*Stenotrophomonas*	coli c1a	Undetermined	Undetermined
Class	Undetermined	Caudoviricetes	Caudoviricetes	Undetermined	Undetermined	Undetermined
Family	Siphoviridae (98.54%)	Siphoviridae (98.89%)	Siphoviridae (98.54%)	Siphoviridae (98.6%)	Siphoviridae (98.62%)	Siphoviridae (94.05%)
Sub family	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified
Genus	Unspecified	Unspecified	Unspecified	Lambdavirus	Unspecified	Unspecified
Host interaction	Undetermined	*Staphylococcus pseudintermedius*	*Stenotrophomonas maltophilia*	Undetermined	Undetermined	Undetermined
Lifestyle	Temperate	Virulent	Temperate	Virulent	Temperate	Temperate
Nucleic acid	dsDNA	dsDNA	dsDNA	dsDNA	dsDNA	dsDNA
Genomic content (%)	60.19	30.28	37.79	60.61	37.16	36.60
CDS	364	169	47	76	77	62
tRNAs	5	6	0	3	0	0

Abbreviations: dsDNA: double-stranded deoxyribonucleic acid: tRNA: transfer RNA; GC: guanine–cytosine; TPWB: *Tabanus par* whole body. *** annotation is based on Phannotate.**

Table 6-2: Summary of the genomic features of phage genomes identified from the legs of *Tabanus par.*

Features	***Bacillus* phage TPL**	***Klebsiella* TPL**	***Escherichia* coli TPL**	***Staphylococcus* phage TPL**
Closest accession numbers	NC_029024	MN101225	MF374379	DQ908929
Assigned accession number	PQ338056	PQ338058	PQ338057	PQ338059
Length bp	44 506bp	59 598bp	170 231bp	10 584bp
Classification	*Bacillus phage 250*	*Klebsiella phage KMI11*	*Escherichia coli c1a*	Undetermined
Class	Caudoviricetes	Caudoviricetes	Caudoviricetes	Caudoviricetes
Family	Siphoviridae (98.59%)	Myoviridae (97.75%)	Myoviridae (96.05%)	Siphoviridae (96.96%)
Sub family	Undetermined	Tevenvirinae	Undetermined	Azeredovirinae
Genus	Cecivirus	N/A	Lambdavirus	Phietavirus
Host interaction	*Bacillus cereus*	*Klebsiella oxytoca; Klebsiella penumoniae*	*Escherichia coli c1a*	Undetermined
Lifestyle	Temperate	Virulent	Temperate	Virulent
Nucleic acid	dsDNA	dsDNA	dsDNA	dsDNA
Genomic content (%)	36.51	40.01	51.5	36.07
CDS	89	83	238	10
tRNAs	0	0	0	0

Abbreviations: dsDNA: double-stranded deoxyribonucleic acid: tRNA: transfer RNA; GC: guanine–cytosine; TPL: *Tabanus Par* legs. * **annotation is based on Phannotate.**

6.3.3 Structural and transmembrane protein annotations of ten phages in *T. par*

The total number of coding sequences (CDS) identified in the phage genomes varied widely, ranging from 10 to 364 determined from the *T. par* WB and L. Notably, the temperate *Bacillus* phage TPL (Table 6-3) of the fly legs has the host interaction with *B. cereus* based on *in-silico* analysis. *Bacillus* phage TPL indicated many transcriptional regulation genes, resulting in a complex set of genetic elements that control the expression of its genes. These genes are responsible for regulating various aspects of the phage life cycle, influencing its life cycle. Moreover, the moron, auxiliary metabolic gene, and host takeover are present in this phage genome.

Interestingly, a virulent *Klebsiella* phage TBL was determined from the leg and characterized as having an integrase or excision gene that interacts with *Klebsiella oxytoca* and *Klebsiella pneumoniae*. A *Staphylococcus* phage TPL is also characterized as a virulent phage, which has no interaction with available host genomes based on *in-silico* analysis. However, this phage based on Phannotate annotation is related to Staphylococcus phage 80 (DQ908929), which is known to infect *Staphylococcus aureus*. Moreover, *Staphylococcus* phage 80 is characterized as a temperate phage.

Arcanobacterium phage TPWB and *Caudovirales* phage TPWB contained tRNA genes, with 6 and 7 tRNA genes, respectively. Additionally, putative CRISPR sequences were not identified in any of the phages. Another phage of interest was the *Caulobacter* phage TPWB. It contained several transcriptional regulation genes, providing an impressive, mainly unexplored source of genetic diversity, and was the only phage to have transfer-messenger RNA (tmRNA), which is responsible for protein synthesis efficiency and cellular quality control. This circular representation delineates the organization and functional categorization of ORFs identified within phage genomes (Figure 6-4 and 6-5).

Table 6-3: Annotation of the *Tabanus par* phage genomes identified in both whole body and legs.

Description	*Bacillus* TPL	*Klebsiella* TPL	*Escherichia Coli* TPL	*Staphylococcus* TPL	*Arcarnobacterium* TPWB	*Acinetobacter* TPWB	Caudovirales TPWB	Caulobacter TPWB	*Staphylococcus* TPWB	*Stenotrophomonas* TPWB
CDS	89	83	238	10	169	62	76	364	77	47
Connector	4	0	4	0	4	1	0	0	2	2
DNA, RNA and nucleotide metabolism	8	15	7	0	17	2	5	25	11	6
Head and packaging	8	10	8	1	7	5	5	2	6	7
Integration and excision	0	1	2	0	0	1	0	1	0	0
Lysis	2	3	4	0	2	2	2	7	2	0
Moron, auxiliary metabolic gene, and host takeover	1	1	12	0	1	1	0	5	0	0
Other	4	1	14	1	5	2	2	15	4	2
Tail	2	16	17	3	7	2	6	6	4	3
Transcription regulation	6	3	3	0	1	2	0	1	3	1
Unknown function	61	34	167	5	125	44	56	302	45	26
tRNAs	0	0	0	0	6	0	3	5	0	0
CRISPRs	0	0	0	0	0	0	0	0	0	0
tmRNAs	0	0	0	0	0	0	0	1	0	0
Positive strand (%)	83.90	76.97	84.44	90.16	61.068	-	-	-	-	91.11
Negative Strand (%)	16.091	23.026	15.55	9.84	38.93	-	-	-	-	8.88

Abbreviations: ORF: open reading frame; tRNA: transfer RNA; ORFs with assigned function; tRNAs: Transfer RNA; CRISPRs: clustered regularly interspaced short palindromic repeats; tmRNA: Transfer-messenger

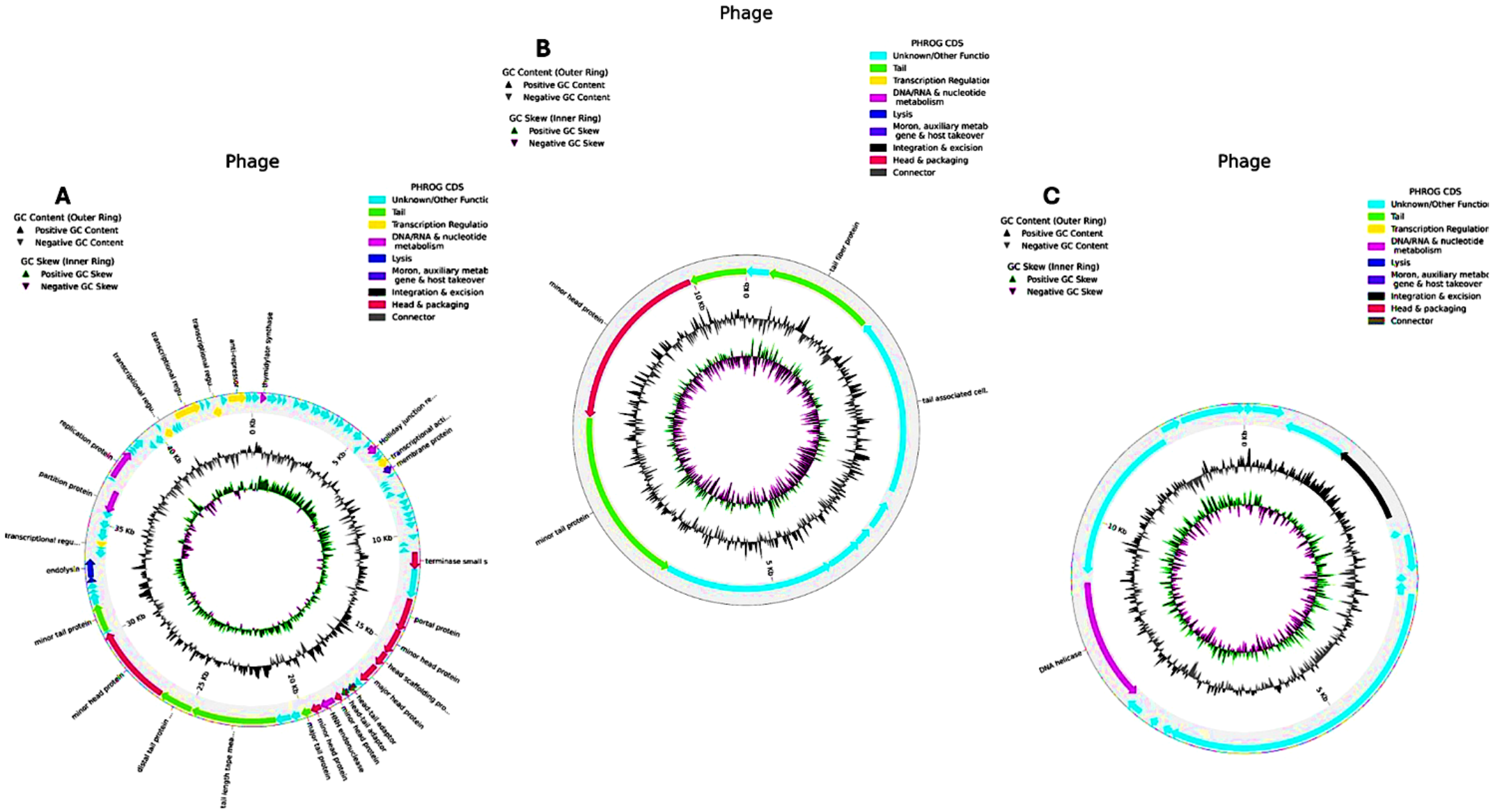


Figure 6-4: Structural and transmembrane protein annotations. A: *Bacillus* TPL; B: *Staphylococcus* TPL C: *Klebsiella* TPL, sequenced from legs of *Tabanus par.* The arrows represent the direction of transcription (strand + or -).

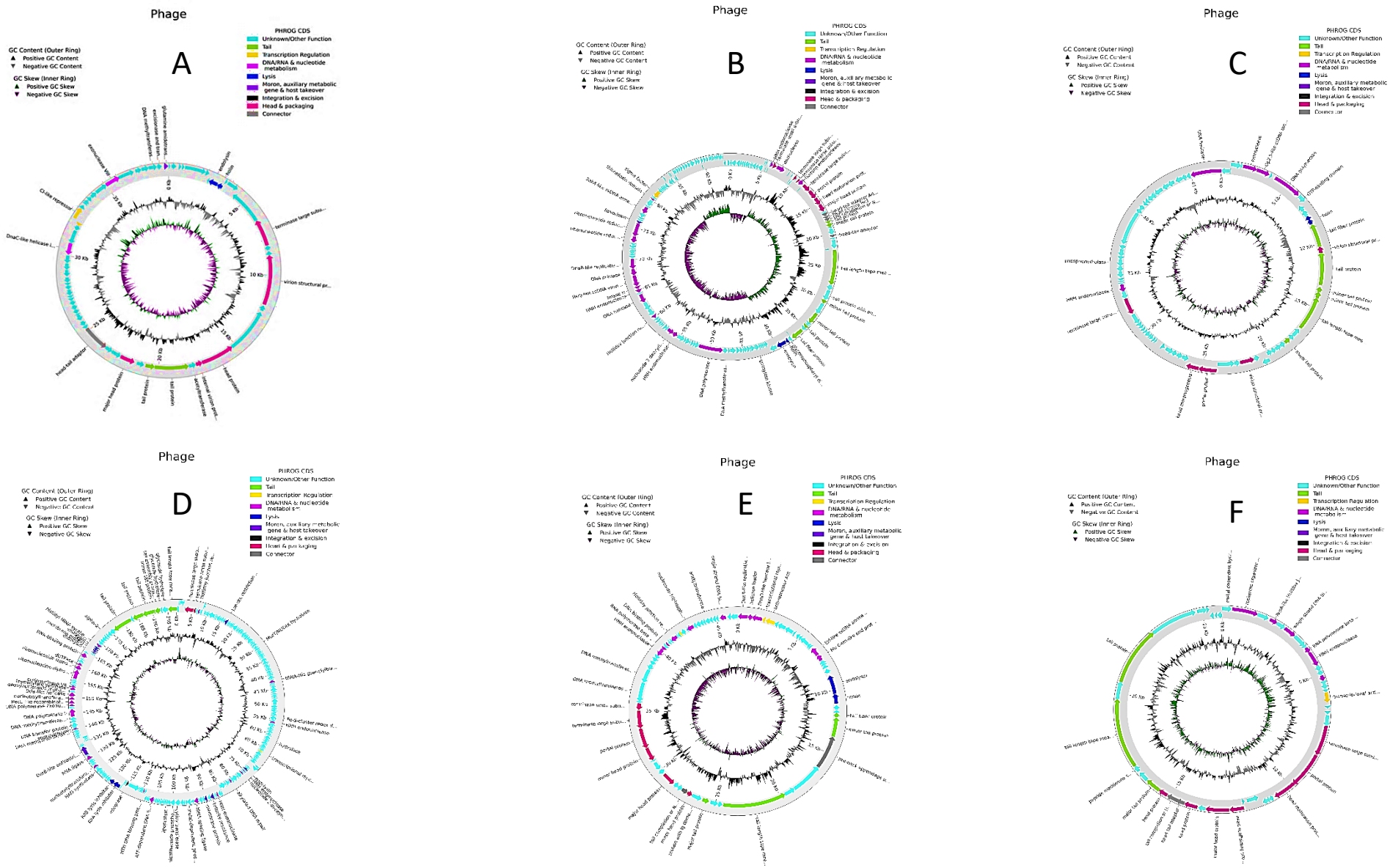


Figure 6-5: Structural and transmembrane protein annotations of the complete phage genomes identified in the body. A: *Acinetobacter* TPWB; B: *Arcarnobacterium* TPWB; C: *Caudovirales*; D: *Caulobacter* TPWB; E: *Staphylococcus* TPWB and F: *Stenophromonas* TPWB. The arrows represent the direction of transcription (strand + or -).

6.3.4 Placement of the phage within the phylogenetic framework

Proteomic (tbastx-based) and genome-wide sequence similarity-based phylogenies placed the ten phage genomes in distinct clades. This approach allowed for a comprehensive analysis of the viral proteome, aiding in identifying the specific taxonomic lineage to which the phage belongs. Forty closest reference phages were identified in the proteome tree (Figure 6-6). ViPtree classification indicated high amino acid synteny among the viral family Straboviridae (3) and Autographiviridae (1)., classified into 2 host groups, namely, Pseudomonadota (17) and Bacillota (13). Based on Phrokka analysis, some phage classifications were identical to those also determined by the Viptree database. Only *Escherichia coli* TPL, *Caulobacter* phage TPWB, *Bacillus* phage TPL, and *Staphylococcus* phages TPL and TPWB were classified into their respective genera. Further genome placement and intergenomic similarity were determined to identify the species of the phages using VIRIDIC analysis, which uses a similar algorithm to that used by the International Committee on Taxonomy of Viruses (ICTV) and Bacterial and Archaeal Viruses Subcommittee). Sequenced phages were classified using the recommended nucleotide similarity thresholds of 95% for species demarcation and 70% for genus demarcation (Figures 6-7 a and b).

Several of the individual sub-clusters in this study exhibited a nucleotide similarity of >90 using VIRIDIC (Figure 6-8), particularly the *Escherichia coli* phages. This included the *E. coli* TPL, *E. coli* strainC CP029371, *E. coli* strainB CP115178, *E. coli* strainATCC, *E. coli* strain BlamR CP115179, and *E. coli* strain_BstaR CP115180, indicating that these phages belong to the same genus. The VIRIDIC results revealed that the compared *E. coli* phages detected in this study clustered most closely with related phages. However, the *E.* coli TPL sequenced in this study has <95% to these compared *Escherichia coli* phages, which means they belong to the same genus species as also indicated by the VICTOR classification tool. However, since *E. coli* phage TPL is < 95% based on VIRIDIC, these phages belong to a new phage species. However, the similarity is considered if the threshold is 70% for the genus and 95% for species, and most of the phages indicated low sequence similarity with each other. The phages were grouped into clusters if they shared 35%-50% or more of their genes [*Arcanobacterium* phage TPWB (58%), *Staphylococcus* phage TPWB (34.5-35.5%),

Staphylococcus phage TPL (32.2-37.3), *Acinetobacter* phage TPWB (41.8%), Caudovirales phage TPWB (58.7%)], indicating that ratio of shared genes and other isolated phages was meager, proposing that these phages are novel species.

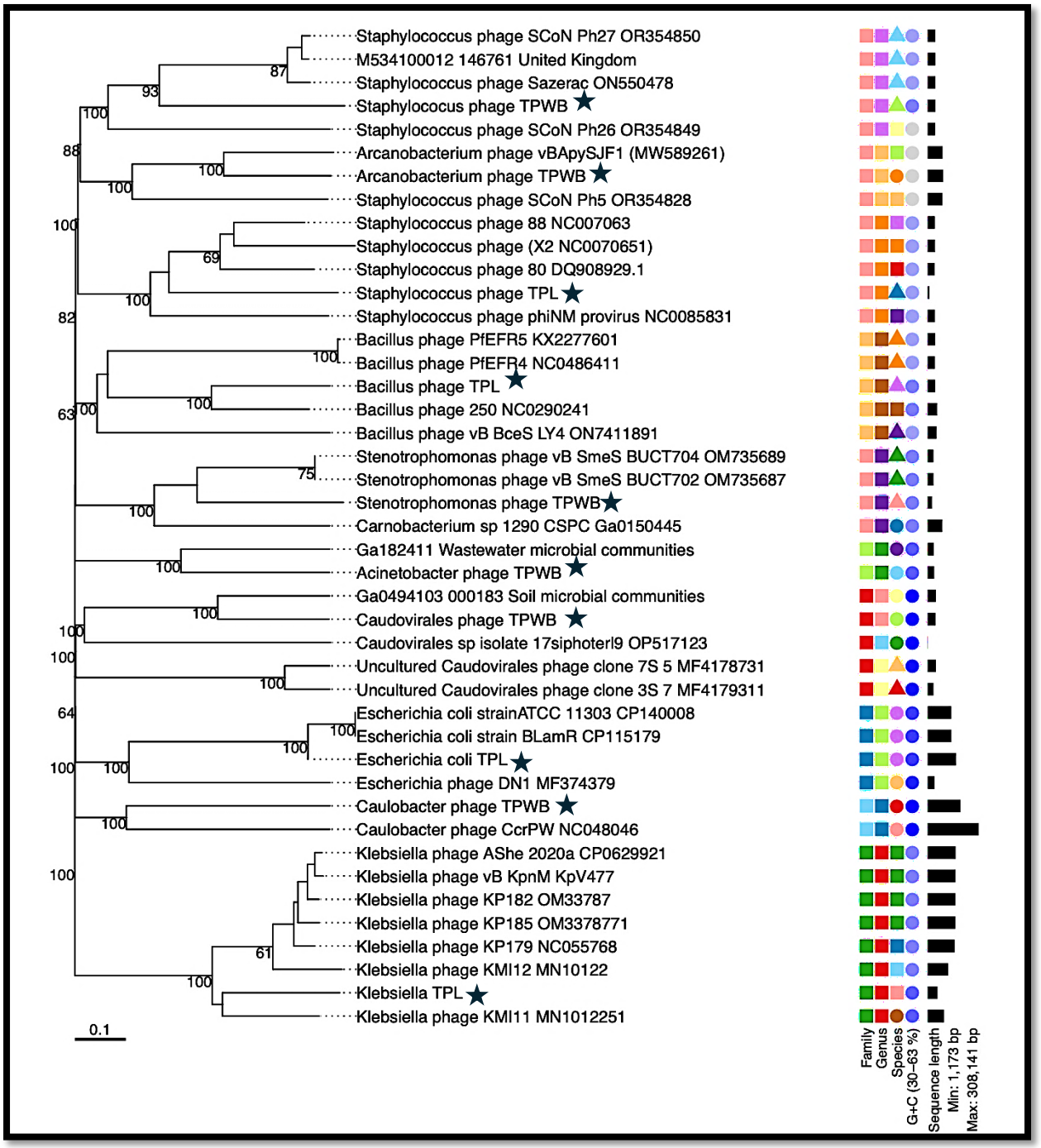


Figure 6-6: Phylogenetic placement of the ten sequenced phages from Tabanus par. Whole genome-based phylogeny (amino acid) was inferred using Genome-BLAST Distance Phylogeny (100 bootstraps) using formula D6, yielding an average support of 70%.

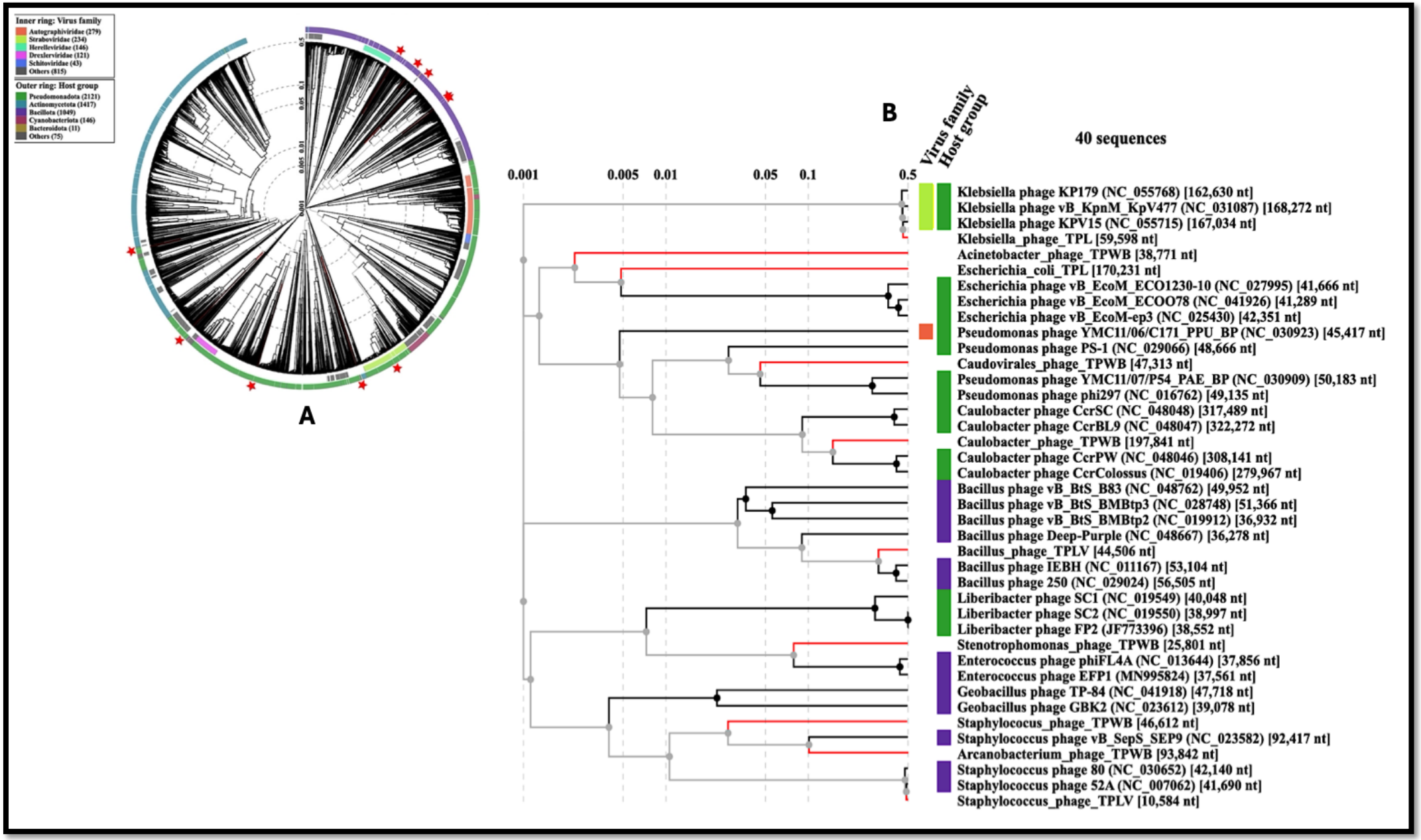


Figure 6-7: Phylogenetic and comparative genomic analyses of 10 *Tabanus par* phages and the 40 closest viral genomes.

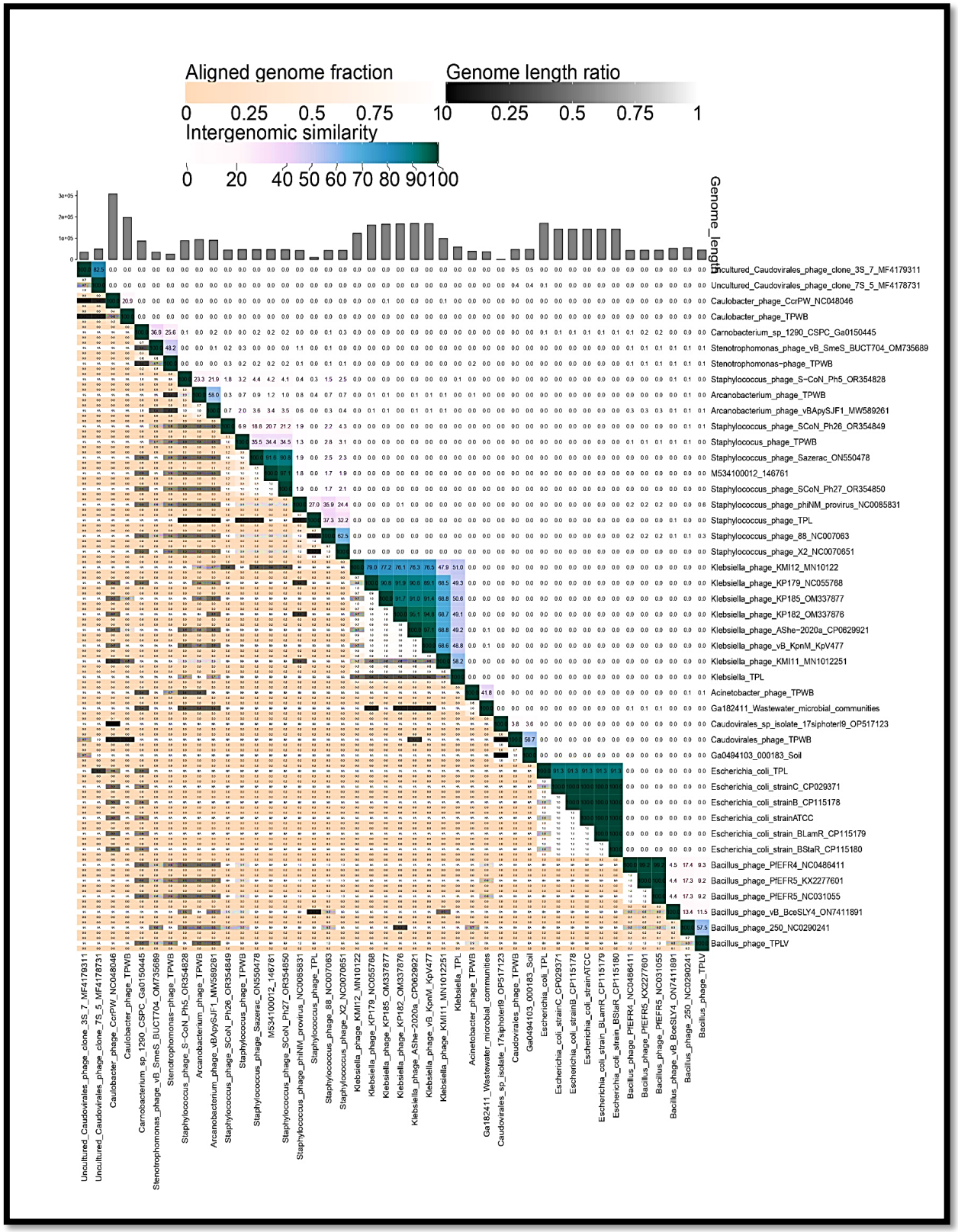


Figure 6-8: Intergenomic similarities heat map constructed by VIRIDIC showing the average nucleotide identity between the sequenced *Tabanus par* phages.

Protein-to-protein network-based Phage cloud analysis linked *T. par* phages with the phages belonging to the same genus as depicted in (Figure 9). Hierarchical clustering of the shared protein content obtained from pan-genome analyses demonstrated that 43 sub-clusters could be grouped to form 10 clusters (A-J) (Figure 6-9). Only one of these clusters is recognised as formal ICTV taxa, the phage *E. coli* TPL (Cluster C). No broader relationships were identified among the sequenced phages for the remaining clusters (A, B, D, E, F, G, H, I, and J), suggesting that these clusters represent novel or less well-characterized phage groups.

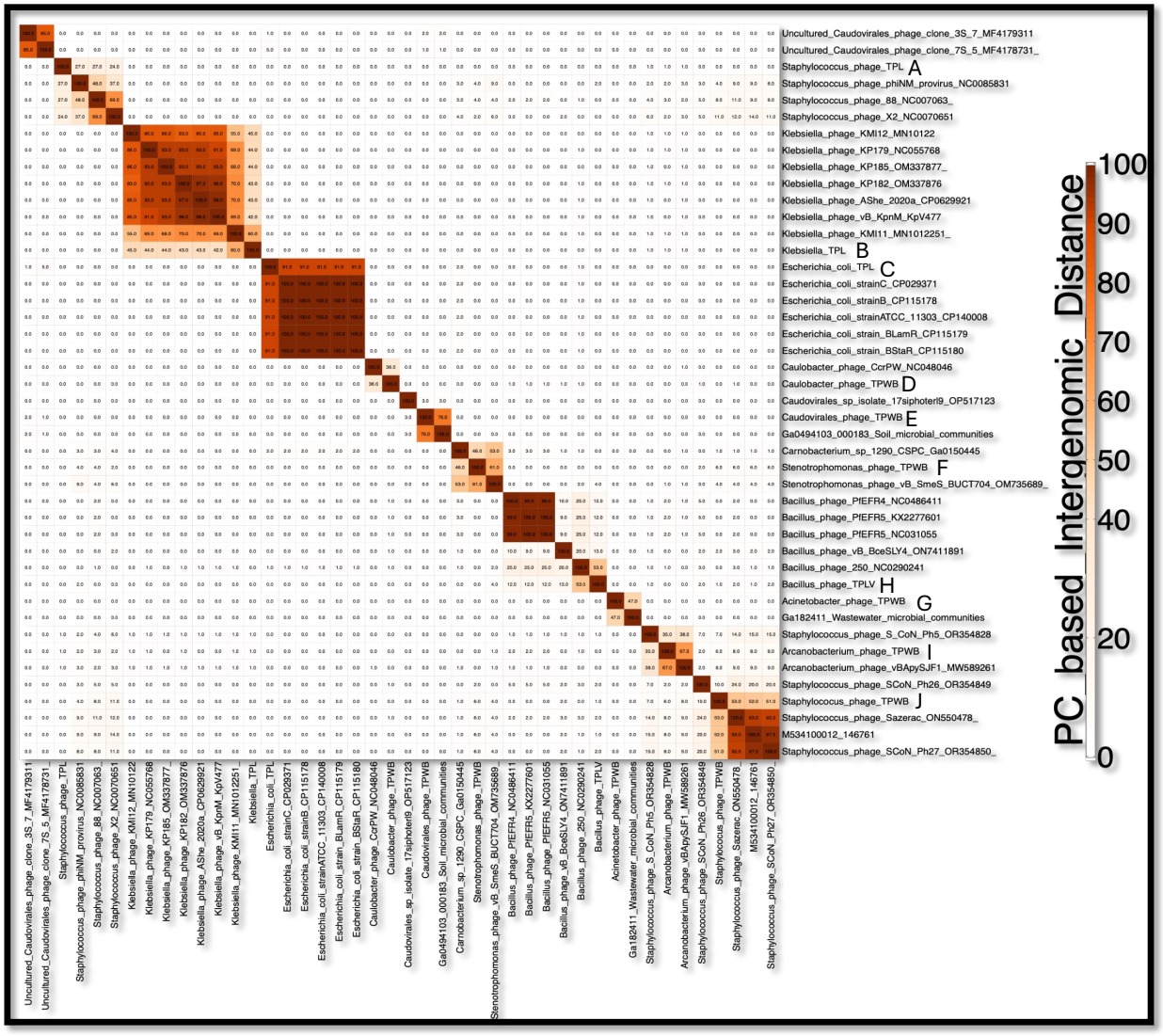


Figure 6-9: The Viriclust hierarchical clustering of phages recovered from *T. par,* based on intergenomic distance calculated using protein cluster content as genomic singletons.

6.4 Discussion

Phages are the most abundant biological entities in the virosphere itself and on earth in general (Koonin *et al*. 2015). Even though the genetic diversity of phages is unmatched (Hendrix 2003; Kristensen *et al*. 2013), probably only a minor fraction of them is known (Rohwer 2003). About insect microbiomes, phages are underexplored, and their impact on insect biology still needs to be better understood. The presence of partial phages and complete phages in the *T. par* insect was examined in this research. Using genomes sourced from various databases, it was observed that the phage genomes exhibited less than 90% intergenomic similarities, suggesting that they may be novel phages. All phages belong to the class Caudoviricetes, while phage *Klebsiella* TPL belongs to the subfamily Tevenvirinae, and *Staphylococcus* phage TPL belongs to the subfamily Azeredovirinae.

In this study, *E. coli* phage was the most identified phage. This finding aligns with its status as one of the most frequently documented and studied phages in various contexts, including human studies (Rahimzadeh *et al*., 2020; Dalmasso *et al*., 2016) and wastewater research (Balleste *et al*., 2022). Its prevalence in these studies highlights its significance in environmental and health-related research. It has also been documented in various arthropods such as common house fly (*Musca domestica*) and stable flies (*Stomoxys calcitrans*) (Cook, 2020), black soldier fly (*Hermetia illucens L)* (Cheng *et al*.,2019), and stinkbug *Plautia stali* (Koga *et al.*, 2022). In a study by Skaradzińska *et al.*, (2017), *E. coli* phage was isolated from German pig farms. Similarly, this study identified *E. coli* phage TPL from the *T. par* legs, characterized by a genome size of 170,231 bp and a GC content of 51.1%. This genomic profile is comparable to an *E. coli* phage reported in wastewater by Oslen *et al*., (2020), indicating potential similarities in phage characteristics across different environments.

This study found multiple complete phage genomes from various sections of the *T. par*, including *Klebsiella* phage TPL, *Staphylococcus* phage TPL, *Bacillus* phage TPL, and *E. coli* TPL that were sequenced from the (L). *Arcanobacterium* TPWB, *Acinetobacter* phage TPWB, *Caudovirales* phage TPWB, *Caulobacter* phage TPWB, *Stenotrophomonas* phage TPWB, and *Staphylococcus* phage TPWB were sequenced from the (WB). *In-silico* classification showed that four were classified as virulent and six as temperate phages among the identified phages.

In this study, the body and legs of *T. par* contained *Staphylococcus* phages, each exhibiting distinctive genome sizes. This indicates variability in the phage genomes across different insect parts, suggesting potential differences in phage populations or their adaptation to distinct environments within the host. The smallest phage genome identified was *Staphylococcus* phage TPL, with a genome size of 10,584 bp, found in the legs of *T. par*. Conversely, the largest genome, *Staphylococcus* phage TPWB, had a genome size of 46,612 bp and was identified from the whole-body DNA of *T. par*. Both phages belong to the Siphoviridae family, highlighting the diversity in genome sizes within this family. Phages in this family are considered to have a temperate lifestyle due to their potential for horizontal gene transfer between bacterial genomes and their incorporation as a prophage into the host genome without inducing lysis. However, in this study, the *Staphylococcus* phage TPWB showed virulent properties.

Klebsiella phages have been isolated from diverse environments globally, including wastewater, sewage, seawater, and human gastrointestinal tract samples (Herridge *et al*., 2020). In the current study, a virulent *Klebsiella* phage with a genome size of 59.598bp was detected and belonged to class Caudoviricetes and the family Myoviridae. The Myoviridae phages are associated with the lytic life cycle and are characterized by the high efficiency of bacterial lysis (Herridge *et al*., 2020), as indicated in this study. However, compared to Klebsiella phage KMI11 identified by Ku *et al*., 2019, a significantly larger genome, 98 975bp, was isolated from the same class (Caudoviricetes), isolated from *K. oxytoca*. Phages' range of activity is limited by their remarkable sensitivity in identifying and infecting host bacteria. There is some extent of diversity in host specificity; specific phages can infect more than one species of bacteria, and certain viruses can infect humans, animals, and even insects (Naureen *et al*., 2020). This sequenced *Klebsiella* phage's host was evident as it interacted with *K. pneumoniae* and *K. oxytoca* based on *in-silico* analysis. These bacteria are considered opportunistic pathogens in humans, and they are becoming increasingly associated with nosocomial infections, particularly in immunocompromised patients.

Bacillus phage TPL was identified as one of the temperate phages sequenced from *T. par* legs. The *Bacillus* phage is known to belong to the Myoviridae, Siphoviridae, Podoviridae, and Tectiviridae families (Gillis and Mahillon 2014). However, in this study, it belonged to the Siphoviridae family based on Phage AI (https://phage.ai/.) The genome

size of this phage is 59 598 bp and had a G+C% content of 35.51, similar to *Bacillus* phage, which was documented by Klumpp *et al*. (2014). *Bacillus* is a genus comprising the model bacteria *B. subtilis* and two well-known human and animal disease-causing bacterial species, *B. anthracis* and *B. cereus* (Hassim *et al*., 2024).

In this study, the phages identified from the whole body of *T. par* exhibited a degree of novelty, with only *Stenotrophomonas* phage being classified. This phage had a genome size of 25 801 bp, notably smaller than the *Stenotrophomonas* phage reported by Hans *et al,* (2022) in a mouse model from China. The host bacterium for this phage was *Stenotrophomonas maltophilia*, a species commonly found in nature and associated with human respiratory and intestinal tract infections. However, one of the phages sequenced from the whole body of *T. par* insect in this study was identified as *Arcarnobacterium* phage TPWB based on VICTOR classification, but Phannotate annotation classified it as *Staphylococcus* phage vB SpsS-DH5 with bacterial host identified as *Staphylococcus pseudintermedius*. Based on ICTV genomic data, *Arcarnobacterium* phage TPWB was closely related to *Arcarnobacterium* vBApysjf1 mw589261. According to Kasman and Porter (2022), bacteriophages, like all viruses, are highly host-specific and often only infect one or a few different strains of the same species of bacteria.

The genomes of the phages identified in this study vary significantly, ranging from 10 584 base pairs (bp) to 170 231 bp, and exhibit a G+C content ranging from 30.28% to 60.61%. Despite these variations, the phages share a high degree of genetic similarity, with 99% of their open reading frames (ORFs) being conserved across the different phages. Understanding the genetic makeup of bacteriophages requires structural annotation, which includes ORF prediction, protein classification, and transcription terminator annotation. The CDS with predicted functions among the phage genomes ranged from 10-364 annotated CDS. Lysogenic-related (integrase) genes were found in 5 genomes of both phages. Following whole phage genome sequencing and annotation, one can safely assume that these phages follow a lytic lifestyle due to the presence of endolysin. Other sequenced phages were considered temperate as they contained RIIA and RIIB lysis inhibitors. The presence of tRNA genes in phage genomes was identified in *Arcarnobacterium* phage TPWB and *Caudovirales* phage. Bailly-Bechet *et al*. (2007) state that tRNA gene abundance is linked to phage virulence. Furthermore, Kallies *et al.* (2023) stated that phages with tRNA genes recognize and require fMet for translation,

which is effective at invading bacterial cells and taking over their protein production, which plays a crucial role in bacterial protein synthesis.

Phage phylogenetic analysis and classification have always been complex because phages lack a single universally conserved gene and exhibit genomic mosaicism, an authentic attribute of phages resulting from high mutation frequency and horizontal gene transfer (Casey *et al*., 2021). Nucleotide-level analysis is the best method for categorizing closely related phages (Grose and Casjens, 2014). Viral Informatics and Phage Tree (Viptree) classification resulted in 40 closely related sequences; however, some of the phages sequenced in this study were not identical to those in the available viral database. Proteome-based phylogeny clustered the ten complete genome phages with 40 viral genomes. According to (Lammens *et a*l., 2020), VIPtree classification might be a lesser-known resource in phage research. Thus, further analyses were conducted on the nucleotide sequence, as nucleotide-level analysis is the most suitable approach for classifying closely related phages. Based on the VICTOR estimates, the current study described phages at the family level and further classified *E. coli* TPL phage as belonging to the same species as the other compared phage genomes in this study. All the sequenced phages clustered with their respective genomes, phage *E. coli* TPL indicated a monophyletic cluster with < 95 % intergenomic similarity in VIRIDIC classification, while the rest of the complete phages indicated < 50% similarity. Similar results were observed in the intergenomic heat map, suggesting the novelty of these phages.

6.5 Conclusion

This study provides the first comprehensive genomic investigation of bacteriophages associated with *T. par*, revealing a diverse and potentially novel phage community within this insect species. The identification of multiple complete and partial phage genomes from various anatomical parts of *T. par* including the L and WB sheds light on their presence, diversity, and possible functional roles.

The phages identified in this study belonged predominantly to the class *Caudoviricetes*, with representatives from families including *Myoviridae* and *Siphoviridae*, and exhibited wide genomic variability. Genome sizes ranged from 10,584 bp to 170,231 bp with GC contents between 30.28% and 60.61%, reflecting considerable genetic diversity. Many of the phages, including *E. coli* TPL, *Klebsiella*, *Staphylococcus*, and *Bacillus* phages,

exhibited intergenomic similarities below 90%, suggesting they are novel or previously uncharacterized. Phylogenetic and intergenomic analyses further supported this novelty, as most phages did not cluster closely with known reference genomes, with only *E. coli* TPL displaying species-level similarity.

Functional annotation of the genomes revealed key genetic features indicative of both lytic and temperate lifestyles, including endolysins, integrase genes, and lysis inhibitors. The presence of tRNA genes in several phages points to their potential virulence and efficiency in commandeering host translational machinery.

6.5.1 Recommendations

Further in-depth research is required to identify the precise functions and biological characteristics of these phages, concentrating on their relationships with the host microbiome and possible effects on fly physiology. Additional genomic and proteomic investigations can be conducted to reveal the molecular mechanisms that regulate phage-host interactions in *T. par* and similar vectors. In addition, it would be beneficial to conduct comparative research on various tabanid fly tribes to see if they harbour similar microbes to those identified in this study.

6.6 References

1. Ackermann, H.W. (2005). Bacteriophage classification. *Bacteriophages: Biology and applications*, pp.67-89.

2. Aiewsakun, P., Adriaenssens, E.M., Lavigne, R., Kropinski, A.M. & Simmonds, P. (2018). Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. *Journal of General Virology*, *99*(9), pp.1331-1343.

3. Al-Shayeb, B., Sachdeva, R., Chen, L.X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y. & He, C. (2020). Clades of huge phages from across Earth's ecosystems. *Nature*, *578*(7795), pp.425-431.

4. Bailly-Bechet, M., Vergassola, M. & Rocha, E. (2007). Causes for the intriguing presence of tRNAs in phages. *Genome research*, *17*(10), pp.1486-1495.

5. Bolduc, B., Jang, H.B., Doulcier, G., You, Z.Q., Roux, S. & Sullivan, M.B. (2017). vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ*, *5*, p.e3243.

6. Casey, A., Coffey, A. & McAuliffe, O. (2021). Genetics and genomics of bacteriophages: The evolution of bacteriophage genomes and genomic research. *Bacteriophages: Biology, Technology, Therapy*, pp.193-218.

7. Chen, Y., Li, X., Song, J., Yang, D., Liu, W., Chen, H., Wu, B. & Qian, P. (2019). Isolation and characterization of a novel temperate bacteriophage from gut-associated *Escherichia* within black soldier fly larvae (*Hermetia illucens*) [Diptera: Stratiomyidae]). *Archives of virology*, *164*, pp.2277-2284.

8. Chevallereau, A., Pons, B.J., van Houte, S. & Westra, E.R. (2022). Interactions between bacterial and phage communities in natural environments. *Nature Reviews Microbiology*, 20(1), pp.49-62.

9. Cook, D. (2020). A historical review of management options used against the stable fly (Diptera: Muscidae). Insects, 11(5), p.313.

10. Dennehy, J.J. & Abedon, S.T. (2021). Adsorption: phage acquisition of bacteria. Bacteriophages: biology, *technology, therapy*, pp.93-117.

11. Hassim, A. & Lekota, K.E. (2023). Isolation of Bacteriophages from Soil Samples in a Poorly Equipped Field Laboratory in Kruger National Park. In *Bacteriophages*: *Methods and Protocols*, pp. 91-103. New York, NY: Springer US.

12. Hegarty, B., Dai, Z., Raskin, L., Pinto, A., Wigginton, K. & Duhaime, M. (2022). A snapshot of the global drinking water virome: Diversity and metabolic potential vary with residual disinfectant use*. Water Research*, 218, p.118484.

13. Herridge, W.P., Shibu, P., O'Shea, J., Brook, T.C. & Hoyles, L. (2020). Bacteriophages of Klebsiella spp., their diversity and potential therapeutic uses. *Journal of medical microbiology*, 69(2), pp.176-194.

14. Hibstu, Z., Belew, H., Akelew, Y. & Mengist, H.M. (2022). Phage therapy: a different approach to fight bacterial infections. *Biologics: Targets and Therapy*, pp.173-186.

15. Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Palaniappan, K., Szeto, E., Pillay, M., Chen, I.M.A., Pati, A. & Nielsen, T. (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v. 4). *Standards in genomic sciences*, *10*, pp.1-6.

16. Kasman, L.M. & Porter, L.D. (2022). Bacteriophages. In *StatPearls [Internet]*. StatPearls Publishing.

17. Kolasa, M.R. & Łukasik, P. (2023). With a little help from my friends: the roles of microbial symbionts in insect populations and communities.

18. Lammens, E.M., Nikel, P.I. & Lavigne, R. (2020). Exploring the synthetic biology potential of bacteriophages for engineering non-model bacteria. *Nature communications*, *11*(1), p.5294.

19. Meier-Kolthoff, J.P. & Göker, M. (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics*, *33*(21), pp.3396-3404.

20. Mondal, P., Halder, P., Mallick, B., Bhaumik, S., Koley, H., Dutta, S. & Dutta, M. (2023). Controlling the bacterial load of Salmonella Typhi in an experimental mouse model by a lytic Salmonella phage STWB21: a phage therapy approach. *BMC microbiology*, *23*(1), p.324.

21. Moraru, C., Varsani, A. & Kropinski, A.M. (2020). VIRIDIC—A novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses*, *12*(11), p.1268.

22. Moraru, C. (2023). VirClust—A tool for hierarchical clustering, core protein detection and annotation of (prokaryotic) viruses. Viruses, 15(4), p.1007.

23. Mwakasungula, S., Rougeron, V., Arnathau, C., Boundenga, L., Miguel, E., Boissiere, A., Jiolle, D., Durand, P., Msigwa, A., Mswata, S. & Olotu, A. (2022). Using haematophagous fly blood meals to study the diversity of blood-borne pathogens infecting wild mammals. *Molecular Ecology Resources*, *22*(8), pp.2915-2927.

24. Nevot, A., Onju, S., Fossaert, C., Chalermwong, P., Thaisungnoen, K. & Desquesnes, M. (2021). Measures of the direct impact of hematophagous flies on feeder cattle: an unexpectedly high potential economic impact.. https://doi.org/10.21203/rs.3.rs-717630/v1

25. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome research*, *27*(5), pp.824-834.

26. Oliver, K.M., Degnan, P.H., Hunter, M.S. & Moran, N.A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*, 325(5943), pp.992-994.

27. Putzeys, L., Wicke, L., Brandão, A., Boon, M., Pires, D.P., Azeredo, J., Vogel, J., Lavigne, R. & Gerovac, M. (2024). Exploring the transcriptional landscape of phage–host interactions using novel high-throughput approaches. *Current Opinion in Microbiology*, 77, p.102419.

28. Ramashia, S. E., T. Tangulani, T., Mashau, M.E. & Nethathe, B. (2020). Microbiology quality of different dried insects sold at Thohoyandou open market, South Africa. *Food Research*, *4*(*6*), pp.2247-2255.

29. Salazar, M.M., Pupo, M.T. & Brown, A.M. (2021). Co-occurrence of viruses, plant pathogens, and symbionts in an underexplored hemipteran clade. *Frontiers in Cellular and Infection Microbiology*, *11*, p.715998.

30. Secor, P.R. & Dandekar, A.A. (2020). More than simple parasites: the sociobiology of bacteriophages and their bacterial hosts. *Mbio*, 11(2), pp.10-1128.

31. Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R.E., Mom, R., Toussaint, A., Petit, M.A. & Enault, F. (2021). PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics and Bioinformatics*, *3*(3), p.lqab067.

32. Young, R. (2014). Phage lysis: three steps, three choices, one outcome. *Journal of microbiology*, 52, pp.243-258.

33. Yukgehnaish, K., Rajandas, H., Parimannan, S., Manickam, R., Marimuthu, K., Petersen, B., Clokie, M.R., Millard, A. & Sicheritz-Pontén, T. (2022). PhageLeads: rapid assessment of phage therapeutic suitability using an ensemble machine learning approach. *Viruses*, *14*(2).

CHAPTER 7 GENETIC CHARACTERISTICS OF *SPIROPLASMA* SPECIES HARBOURED BY*TABANUS PAR* USING METAGENOMICS METHOD

7.1 Introduction

Spiroplasma species are gram-positive, helical-shaped, cell wall-less bacteria (Davis *et al*., 1972). They are considered amongst the smallest self-replicating prokaryotes identified, featuring genomes that vary in size from around 780 to 2,220 kb. Most *Spiroplasma* sp. exist with insects, arachnids, crustaceans, or plants, while only a few species act as pathogens for plants, insects, and crustaceans (Regassa and Gasparich, 2006). Certain *Spiroplasma* reside in plant phloem and are obtained and transmitted by sap-sucking insects. It was discovered that *Spiroplasma* sp. are significantly more prevalent as symbiotic organisms in insects than plants (Mushtaq *et al*., 2022). In plants, *S. citri* and *S. kunkelii* notoriously cause significant damage to citrus and corn, respectively (Whitcomb *et al*., 1986).

They have been found in a diverse array of insects (Kakizawa *et al*., 2022) (Table 7-1), which encompass orders such as Coleoptera, Diptera, Hemiptera, Homoptera, Hymenoptera, Lepidoptera and Odonata. A survey by Hackett *et al*., (1996) isolated *Spiroplasma* sp. from six orders and 14 families of insects. The six primary insect orders that serve as hosts for *Spiroplasma* sp. include: 1) *Hymenoptera* (honeybees, bees, and vespid wasps); 2) *Coleoptera* (various beetles, including the green June beetle, soldier beetle, cucumber beetle, Colorado Potato beetle, scarabaeid beetle, and fireflies); 3) *Diptera* (syrphid flies, mosquitoes, tabanids, and fruit flies); 4) *Lepidoptera* (both larval and adult stages of butterflies and moths); 5) *Homoptera* (leafhoppers); and 6) *Hemiptera* (green leaf bug). Some *Spiroplasma* sp. exhibit male-killing characteristics in fruit flies and other insect species. In *Drosophila* (fruit flies), *S. poulsonii* triggers male killing, with an effector protein known as Spaid identified as responsible for inducing male-specific apoptosis during the embryonic stage (Harumoto and Lemaitre, 2018; Harumoto, 2023). Although the majority of *Spiroplasma* sp. seem to maintain a commensal relationship with their insect hosts, certain infections can lead to deadly outcomes, such as *S. melliferum* and *S. apis,* which are pathogenic in honeybees (Fünfhaus *et al*., 2018).

Table 7-1: *Spiroplasma* species and their associated arthropod or plant hosts (Regassa and Gasparich, 2006).

Binomial name	**Host(s)**	**Disease**
Spiroplasma citri	Dicots, leafhoppers	Citrus stubborn
S. melliferum	Honey bees	Honeybee spiroplasmosis
S. kunkelii	Maize, leafhoppers	Corn stunt
Spiroplasma sp.	Rabbit ticks	Unknown
Spiroplasma sp.	Green leaf bugs	Unknown
S. insolitum	Eristalis flies, flowers	Unknown
Spiroplasma sp.	Coconut palms	Unknown
S. phoeniceum	*Catharanthus roseus*	Periwinkle disease
S. penaei	*Penaeus vannamei* shrimp	Shrimp disease
S. poulsonii	*Drosophila*	Sex ratio trait
S. floricola	Insects, flowers	Beetle lethargy
S. apis	Bees, flowers	May disease
S. mirum	Rabbit ticks	Suckling mouse cataract disease
S. ixodetis	Ixodes pacificus ticks	Unknown
S. monobiae	Monobia wasps	Unknown
S. syrphidicola	*Eristalis arbustorum* flies	Unknown

S. chrysopicola	*Crysops* sp.	Unknown
Spiroplasma sp.	Horse fly	Unknown
S. clarkii	Cotinus beetles	Unknown
S. culicicola	*Aedes* mosquitoes	Unknown
S. velocicrescens	Monobia wasps	Unknown
S. diabroticae	*Diabrotica undecimpunctata*	Unknown
S. sabaudiense	*Aedes* mosquitoes	Unknown
S. corruscae	*Ellychnia corrusca* beetles, horse	Unknown
Spiroplasma sp.	Leafhopper	Unknown
S. cantharicola	Cantharid beetle	Unknown
Spiroplasma sp.	Cantharid beetle	Unknown
Spiroplasma sp.	Mosquito	Unknown
S. turonicum	Horse fly	Unknown
S. litorale	*Tabanus nigrovittatus*	Unknown
S. lampyridicola	*Photuris pennsylvanicus* beetles	Unknown
S. leptinotarsae	*Leptinotarsa decemlineata*	Unknown
Spiroplasma sp.	*Prunus* sp. Flowers	Unknown
S. taiwanense	*Culex tritaeniorhynchus*	Unknown
S. gladiatoris	*Tabanus gladiator*	Unknown
S. chinense	*Calystegia hederaceae*	Unknown
S. diminutum	*Culex* mosquito	Unknown

S. alleghenense	Scorpion fly	Unknown
S. lineolae	Horse fly	Unknown
S. platyhelix	Dragonfly	Unknown
Spiroplasma sp.	Tiphiid wasp	Unknown
Spiroplasma sp.	Flower surface	Unknown
S. montanense	Horse fly	Unknown
S. helicoides	Horse fly	Unknown
S. tabanidicola	Horse fly	Unknown
Spiroplasma sp.	Horse fly	Unknown
S. atrichopogonis	Biting midge	Unknown

Spiroplasma sp. have been extracted from the gut of tabanids (Diptera: Tabanidae) across the globe (Hackett *et al*., 1994; Whitcomb *et al*., 1997; Whitcomb *et al*., 1990; Jandhyam *et al*., 2008). Among all insect hosts of *Spiroplasma*, tabanids (horse flies and deer flies) serve as abundant reservoirs and can harbour numerous types of mollicute infections (Reeves, 2003). The mechanism through which tabanids acquire *Spiroplasma* infections naturally remains unclear. Most *Spiroplasma* sp. are considered to have a symbiotic association (either mutual or communalistic) in tabanids as no data has been documented on pathogenetic activity in tabanids (Anbutsu and Fukatsu, 2011). The efforts in genomic sequencing not only shed light on the specific genes and metabolic pathways found in *Spiroplasma* sp. but also set the stage for future investigations into its ecological functions and relationships within the insect microbiome (Lo *et al*., 2015). Herein, this study sought to identify and characterize Spiroplasma sp. genomes associated with the African horse fly, *T. par,* using a metagenomics approach.

7.2 Materials and methods

7.2.1 Source of *T.par* specimens

Adult *T. par* specimen was collected at three nature reserves, namely Charters Creek Reserve, Hluhluwe Game Reserve, and Phinda Private Game Reserve in KwaZulu-Natal, South Africa, and identified morphologically and genetically as described in Chapter 4, section 4.4.1 and 4.4.3.

7.2.2 DNA extraction procedure

Prior to genomic DNA extractions, fly specimens were washed with 10% Tween 20 by immersion for 1 hour, then washed twice with 70% Ethanol, and then rinsed with double distilled water and left to dry. The whole body (without legs) of *T. par* specimens was cut into pieces in 1.5 ml microcentrifuge tubes using sterile scissors, whereas the legs were separately crushed using a sterilized mortar and pestle and then transferred into a 1.5 ml microcentrifuge tube. Extraction was conducted using Qiagen DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Detailed procedure is described in Chapter 4, section 4.2.6 with the

following modifications: The body and legs of the *T. par* were homogenized separately in a mixture containing 180 µl of Buffer ATL and 20 µl of proteinase K, vortexed and incubated at 56°C until completely lysed. Ultimately, genomic DNA was eluted with 200 µl Buffer AE and then was stored at –20°C until used.

7.2.3 Metagenomic sequencing of *T. par*

Sequence libraries were generated separately from the pooled DNA samples of both *T. par* body and legs. Whole-genome sequencing (WGS) of the *T. par* DNA was performed at the Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa. DNA libraries were sequenced using the MGI DNBSEQ-G400 sequencer high-throughput genome sequencing platform.

7.2.4 Metagenomic, quality control, assembly and binning

The quality of the raw sequencing data was assessed using FastQC software version 0.10.1 (Andrews, 2010). Sequencing reads were trimmed to remove low-quality bases and adapter sequences using Trimmomatic v.0.35 (Bolger *et al.*,2014), using phred +33. Meta *de novo* assembly of the trimmed sequence reads was performed using MEGAHIT v.1.2.8 (Li *et al*., 2015), and the min contig was set to 300 bp. Metagenomic binning software (MetaBAT2 version 2.12.1) (Wang *et al*., 2019) produced bin sets of MAGs, each corresponding to a putative genome. Assessment of the bins was determined using CheckM v1.1.0 (Parks *et al*.,2013). Moreover, metagenomics was determined using the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (https://www.bv-brc.org). QUAST v. v 2.3(Gurevich *et al*., 2013) was used to access the quality matrices of the MAGs. Annotation of the MAGs was conducted using Rapid Annotation using Subsystems and Technology (RAST) (Aziz *et al*., 2008) and Prodigal v2.60 (Hyatt *et al*. 2012).

7.3 Pangenomics and average nucleotide identity analysis

Pangenome analysis was conducted using 75 *Spiroplasma* strains retrieved from the publicly available genomes in GenBank that included the three sequenced strains in this study. Pangenomics analysis was conducted using Anvio-7.1 (Eren *et al*., 2021). Average nucleotide identity analysis was conducted using pyANI implemented in

Anvio-7.1 (Eren *et al*., 2021). The tRNAScan-SE (Chan and Lowe, 2019) was used to determine the number of RNAs using default parameters “-c “. Clusters were created using paralogs of the genomes and were ordered by the presence/absence of orthologs (Page *et al*., 2016). Assignment of clusters of orthologous groups (COG) was achieved using DeepNOG v. 1.2.3 (Feldbauer *et al*., 2020).

7.4 Results

Bacterial genomes detected from *T. par* whole body (WB) and leg (L) DNAMAGs identified from metagenomics were 12 bacterial genomes, of which five were determined as complete and good bins of bacteria. The remaining were classified as bad genomes based on the quality criteria of completeness <= 80%, fine consistency <= 87%, contamination >= 10%, and having a single PheS protein of reasonable size. The WGS approach led to species identification of the five complete genomes, while three belong to the *Spiroplasma megaselia nigra*, *S. turonicum* clonal population and Spiroplasma sp. WSS clonal population. Other detected genomes included the *Aeromonas* hydrophila clonal population and *Cutibacterium acnes* clonal population (Table 7-3). The remaining bacterial MAGs were considered of low quality (Appendix IX). Genomes generated from the *T. par* legs (L) MAGs, generated 45 genomes, and only five genomes (*Bacillus safensis, Escherichia coli, Lawsonella clevelandensis, Sphingomonas* sp. AAP5, and *Sphingomonas* sp. FARSPH) (Table 7-4) were classified as good bins. The remaining genomes were considered as bad bins with low quality and contaminants > 10% (Appendix X).

Table 7-2: Prokaryotic metagenome assembly genomes from the whole body of *T. par* that presents good bins.

Genome Name	Coarse consistency (%)	Completeness (%)	Contamination (%)	Hypothetical (%)	Number of contigs	DNA size (bp)	Contigs N50 (bp)	Mean Coverage
***Spiroplasma megaselia nigra* clonal population**	96.5	100	0	55.32	139	1312515	118271	28.47
***Aeromonas hydrophila* clonal population**	99.6	100	1.3	51.11	1223	6408987	98966	16.37
Cutibacterium acnes clonal population	99.2	97	7.6	41.34	436	2704685	15934	16.69
***Spiroplasma turonicum* clonal population**	95.0	100	4	80.77	937	2436589	5546	13.84
***Spiroplasma* sp. WSS clonal population**	95.1	95.8	8	87.95	1307	2335584	2097	14.17

The following bins met the quality criteria of completeness >= 80%, fine consistency >= 87%, contamination <= 10%, and have a single PheS protein of reasonable size

Table 7-3: Prokaryotic metagenome assembly genomes from the legs of *T. par* classified as good bins.

Genome Name	Coarse consistency (%)	Completeness (%)	Contamination (%)	Hypothetical (%)	Number of contigs	DNA size (bp)	Contigs N50 (bp)	Mean coverage
Bacillus safensis **clonal population**	97.8	100	3.2	40.25	61	3970712	340553	41.7
Escherichia coli **clonal population**	98.8	99.7	5.6	28.63	208	5210179	79074	203.16
Lawsonella clevelandensis **clonal population**	95.4	91.2	6.1	46.77	290	1568103	8285	168.04
Sphingomonas **sp. AAP5 clonal population**	97.3	100	9.6	46.31	115	4736919	244205	17.57
Sphingomonas **sp. FARSPH clonal population**	94.8	96.7	9.5	45.18	170	4121667	97976	59.2

The following bins met the quality criteria of completeness >= 80%, fine consistency >= 87%, contamination <= 10%, and have a single PheS protein of reasonable size

7.4.1 Genome analysis of bacteria

In this study, only *Spiroplasma* draft genomes of the *T. par* WB was utilized for comparative genomics, particularly *Spiroplasma* as it represented most of the MAGs determined. The remaining bacterial genomes were excluded from the analyses due to the potential bias arising from lower completeness and higher contamination levels. A summary of the genomic characteristics of the sequenced *Spiroplasma* sp. is presented in Table 7-4. The average genome size of *S. megaselia nigra* TPWB in this study was 1 199 640bp, while *S. turonicum* TPWB is 1 021 631bp, and *Spiroplasma* sp TPWB is 1 203 742bp. The genomes' completeness was found to be 100% for *S. megaselia nigra* TPWB and *S. turonicum* TPWB, whereas *Spiroplasma* sp. TPWB was 95.8% complete. The GC content of the *Spiroplasma* genomes ranged between 25 to 26%.

Table 7-4: Genomic features of *Spiroplasma* sp. strains from the whole body of *T. par.*

Matrices	***S. megaselia nigra* TPWB**	***Spiroplasma* sp TPWB**	***S. turonicum* TPWB**
Total Length	1 199 640	1 203 742	1 021 631
Completeness %	100	95.8	100
Num contigs	124	11	61
Number of genes*	1 528	1 069	980
N50	7 144	127 576	21 587
N75	7829	2827	16 532
N90	4133	2615	9 072
GC%	26.3	25.3	25.4
Number of RNAs	24	28	14

* Annotation is based on prodigal

7.4.2 Comparative genomics of *Spiroplasma* species

To determine the relationship of the *Spiroplasma* sp., of *T. par* to previously published *Spiroplasma* genomes. The whole genome tree topology was used to infer the evolution of the sequenced *Spiroplasma* (n = 3) with compared global genomes (n = 72) genomes available in GenBank (Appendix XI). The phylogenetic placement of the three MAGs clusters *Spiroplasma* genomes into 12 species (Figure 7-1). The sequenced strains in this study indicated that *S. turconicum* TPWBtur is closely related to *S. turconicum* Tab4c (ANI >99%). In contrast, *Spiroplasma* sp. TPWBtab has an ANI >97% related with *S. gladiatoris* TG1 and *S. tabanidicola* TAUS1. Additionally, *S. megaselia* nigra TPWB1 groups with *S. megaselia nigra*1 (ANI > 99%) (Figure 7-1).

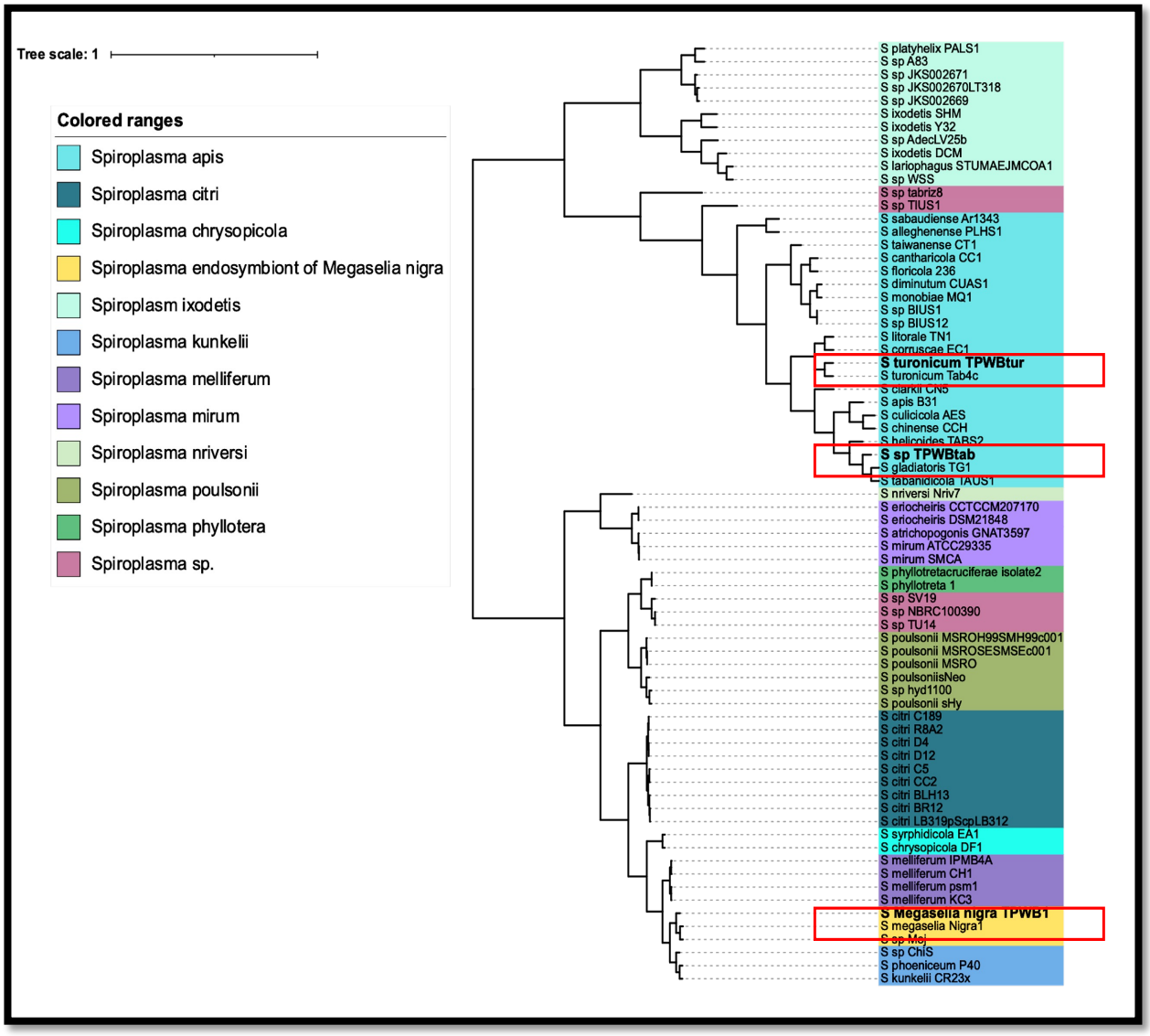


Figure 7-1: Average nucleotide identity phylogenetic tree, showing the relationship among 75 *Spiroplasma* genomes.

7.4.3 Pangenome analysis of *Spiroplasma* strains

Pangenome analysis of genetic diversity was based on 14326 gene clusters with closely related *Spiroplasma* sp. The pangenome comprising the *Spiroplasma* genomes (n = 75) consisted of 4,326 gene clusters, encompassing the genomes sequenced in this study (n = 3) (Figure 7-2 A). The clusters of orthologous groups (COGs) analysis revealed that *Spiroplasma* genomes possess a significant number of genes related to translations, ribosomal structure, and biogenesis (J), exceeding 10000 sequences, followed by genes for replication, recombination, and repair (L), which have over 4000 sequences. The presence of a significant number of these genes in *Spiroplasma* genomes indicates their importance in the biological functions and molecular mechanisms of these bacteria. The lowest COG was cell motility (N) (Figure 7-2 B).

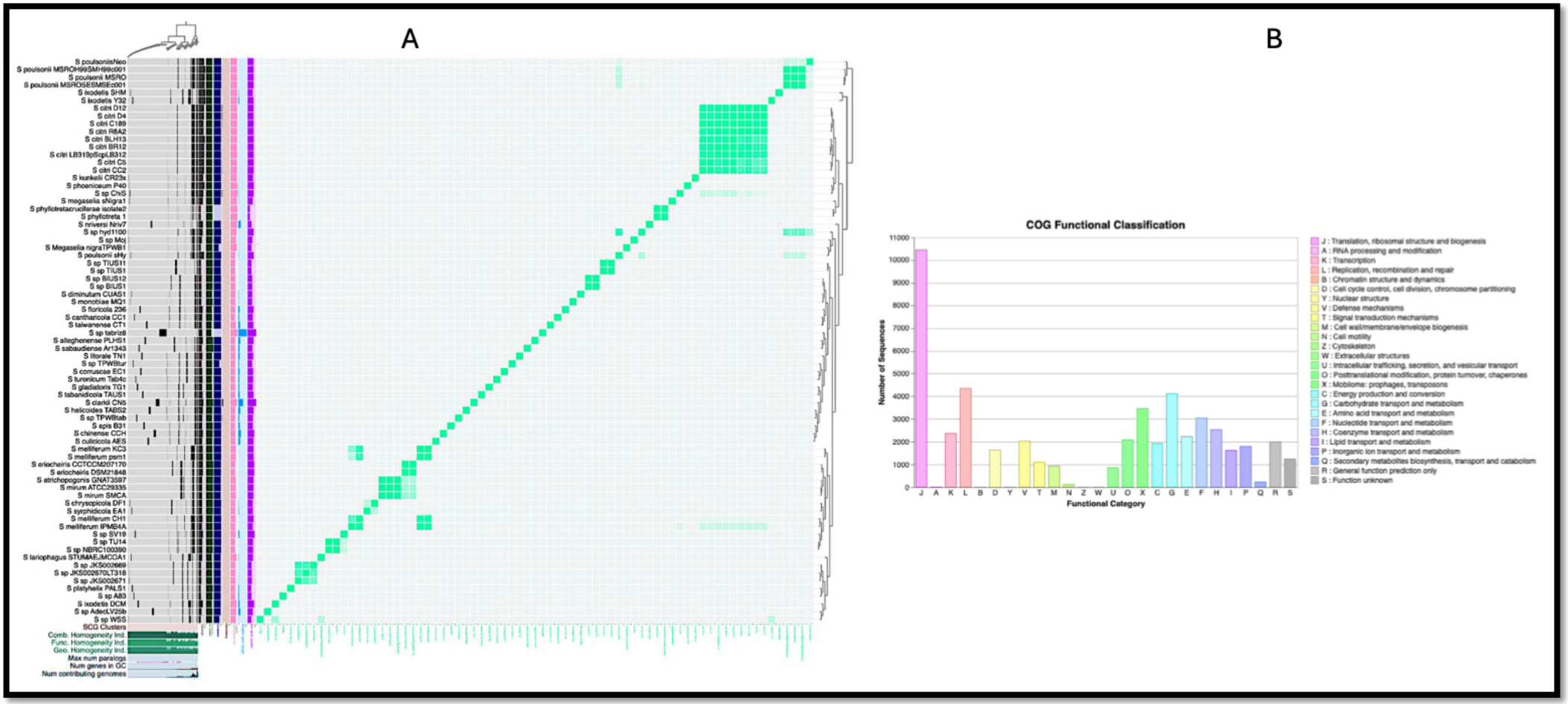


Figure 7-2: A) Pangenomic phylogenomic tree of the 75 Spiroplasma sp., genomes, including the three MAGs determined in this study. Central dendrogram clustering of genomes is ordered by gene cluster presence/absence. Items order presence absence (D, Euclidean; L, War Ward). **B)** COG functional classification of core genome of the analysed 75 *Spiroplasma* genomes.

7.4.4 Genome annotation of *Spiroplasma megaselia nigra* TPWB

The metabolic model for *S. megaselia nigra* TPWB showed the presence of core bioprocesses such as stress response, dormancy and sporulation, DNA metabolism, regulation and signalling, protein metabolism, carbohydrates metabolism, and phosphorus metabolism (Figure 7-3, Table 7-5). The subsystem category distribution of *S. megaselia nigra* TPWB genome annotation showed that 362 functioning features were present in subsystems (Table 15). Of this, 87 (24.03%) features were involved in protein metabolism, 45(12.3%) were involved in amino acid and derivative, 41(12.43%) were nucleotides and nucleotides, 39 (10.77%) were involved in cofactors, vitamins, prosthetic group and pigments and membrane features consisted of 34 (9.39%) features, and the remaining features are described below (Table 7-5).

The presence of virulence, disease, and defence with their proteins showed that the organism consists of virulence factors (their defence mechanism): Multidrug resistance efflux pump (MATE efflux family proteins), resistance to fluoroquinolones, and copper homeostasis. The bar diagram on the left indicates the percentage coverage of subsystems: green and blue bars indicate the percentage of total proteins encoded in the genome of strain *S. megaselia nigra* TPWB that could or could not be assigned to subsystem categories, respectively.

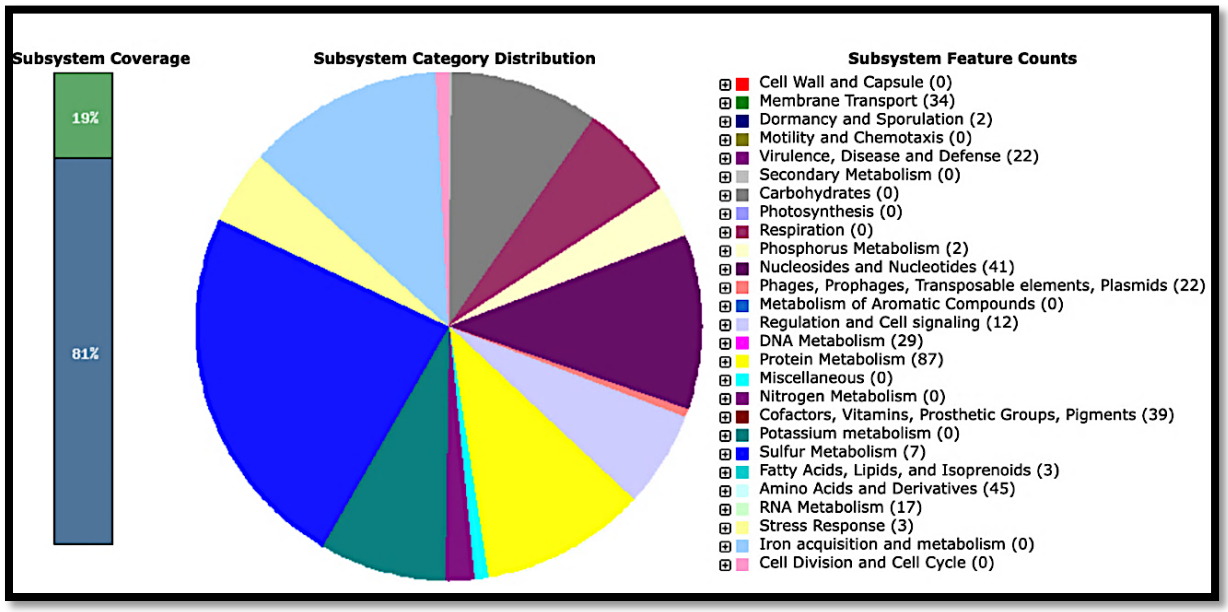


Figure 7-3: The pie-chart showing subsystem coverage, subsystem category distribution, and subsystem feature counts of the annotated genome of S. megaselia nigra TPWB.

Table 7-5: Subsystem category distribution of the genome of *S. megaselia* nigra TPWB annotated using RAST.

	Subsystem feature	Count	Percentage%
1	Membrane transport	34	9.39
2	Dormancy and sporulation	2	0.55
3	Virulence, disease, and defence	22	6.08
4	Phosphorus metabolism	2	0.55
5	Nucleosides and nucleotides	41	11.33
6	Phages, prophages, transposable elements, plasmids	22	6.08
7	Regulation and cell signalling	12	3.31
8	DNA metabolism	29	8.01
9	Protein metabolism	87	24.03
10	Cofactors, vitamins, prosthetic groups, pigments	39	10.77
11	Sulphur metabolism	7	1.93
12	Fatty acids, lipids and isoprenoids	3	0.83
13	Amino acids and derivatives	45	12.43
14	RNA metabolism	14	3.87
15	Stress response	3	0.83

7.4.5 Genome annotation of *Spiroplasma turonicum* TPWBtur

The metabolic model for *S. turonicum* showed the presence of core bioprocesses such as respiration, stress response, dormancy and sporulation, DNA metabolism, protein metabolism, carbohydrates metabolism, and phosphorus metabolism (Figure 7-4). The subsystem category distribution of *S. turonicum* TPWB genome annotation showed that 207 functioning features were present in subsystems (Table 7-6).

Of this, 52 (25.12%) features were involved in nucleosides and nucleotides; within this feature were 25 purines, 19 pyrimidines, and eight unclassified nucleosides and nucleotides. Followed by protein metabolism with 41 (19.80%), 33 (15.94%) were involved in carbohydrates, and DNA metabolism consisted of 14 (6.8%). In contrast, RNA metabolism, amino acid and derivatives, cofactors, vitamins, prosthetic group and pigments, and Virulence factors all consisted of 13 (6.3%). The virulence factors included resistance to fluoroquinolones and copper homeostasis.

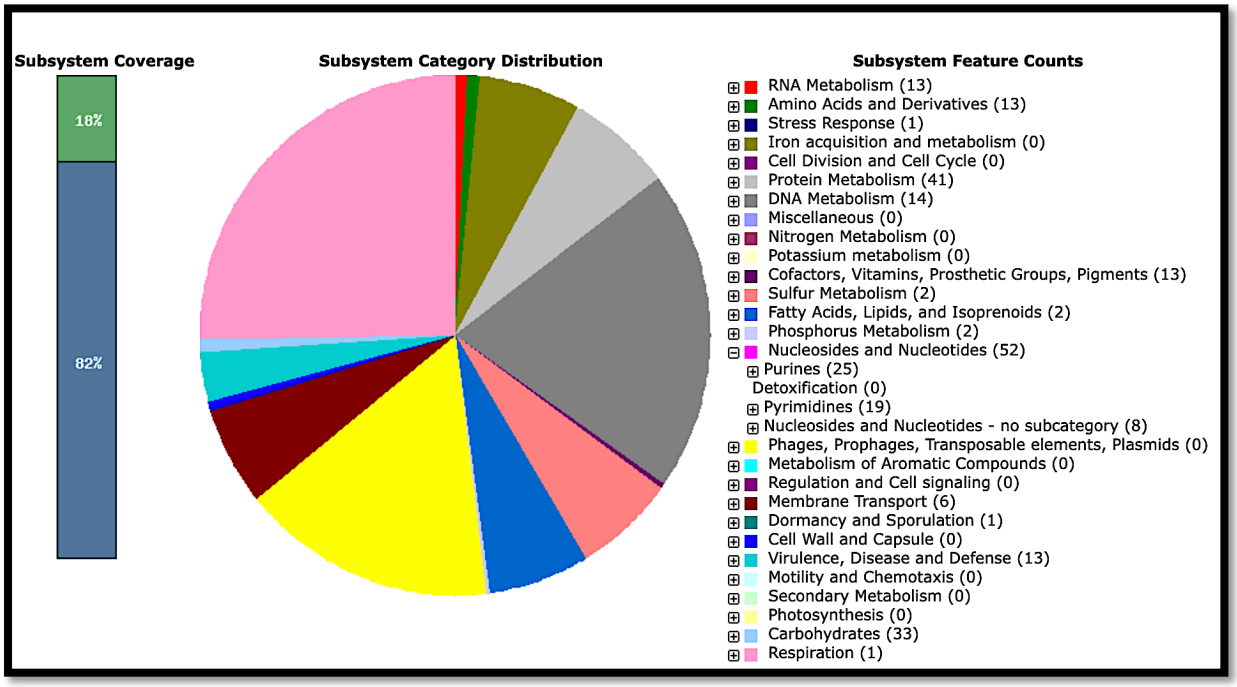


Figure 7-4: Subsystem coverage, subsystem category distribution, and subsystem feature counts of the annotated genome of *Spiroplasma turonicum* TPWB.

Table 7-6: Subsystem category distribution of *S. turonicum* TPWB annotated using RAST.

	Subsystem feature		**Count**	**Percentage%**
1	RNA metabolism		13	6.28
2	Amino acids and derivatives		13	6.28
3	Stress response		1	0.48
4	Protein metabolism		41	19.80
5	DNA metabolism		14	6.76
6	Cofactors, vitamins, prosthetic groups, pigments		13	6.28
7	Sulphur metabolism		2	0.97
8	Fatty acids, lipids and isoprenoids		2	0.97
9	Phosphorus metabolism		2	0.97
10	Nucleosides and nucleotides	Purines 25	52	25.12
		Pyrimidines 19		
		Unclassified Nucleosides 8		
11	Membrane transport		6	2.90
12	Dormancy and sporulation		1	0.48
13	Virulence, disease, and defence		13	6.28
14	Carbohydrates		33	15.94
15	Respiration		1	0.48
	Total		207	

7.4.6 Genome annotation of *Spiroplasma* sp. TPWB

The metabolic model for *Spiroplasma* sp TPWB showed the presence of core bioprocesses such as stress response, dormancy and sporulation, DNA metabolism, regulation and signalling, protein metabolism, carbohydrates metabolism, phosphorus metabolism, and others (Figure 7-5). The subsystem category distribution of S. sp TPWB genome annotation showed that 222 functioning features were present in subsystems (Table 7-7). Of this, 50 (22.5%) features were involved in protein metabolism, followed by nucleosides and nucleotides, with 47 (21.2%) and 42 (18.9%) involved in carbohydrates. RNA metabolism comprised 18 (8%), while DNA metabolism comprised 13 (7.2%). The virulence factors consisted of 13(5.9%) features, including resistance to fluoroquinolones and copper homeostasis.

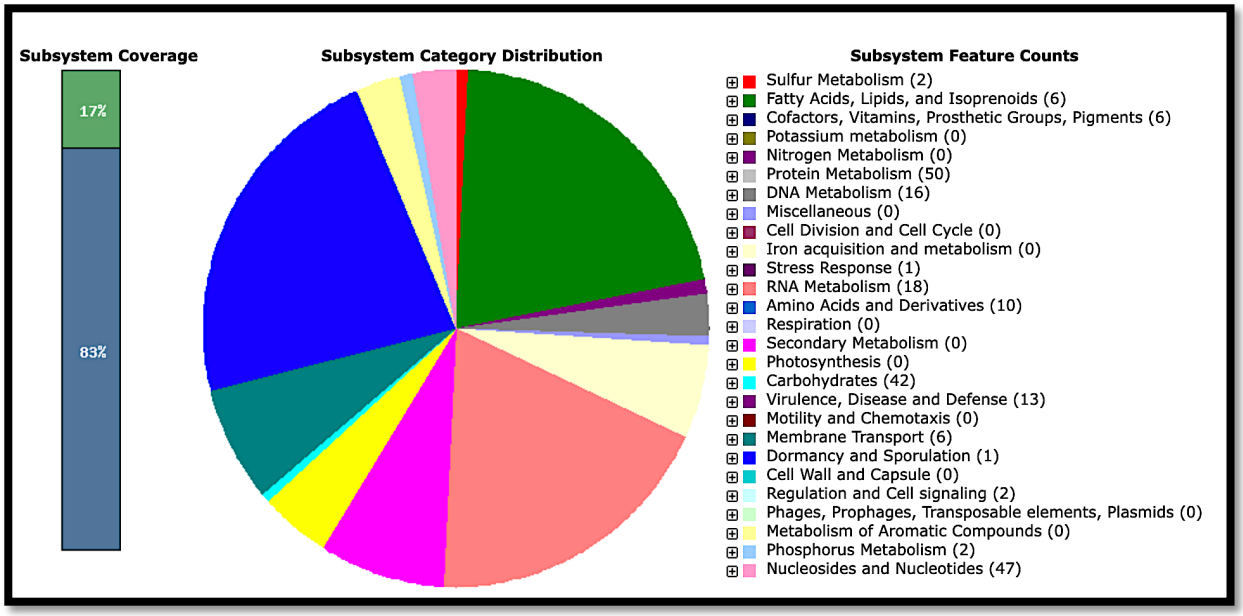


Figure 7-5: Subsystem coverage, subsystem category distribution, and subsystem feature counts of the annotated genome of *Spiroplsama*. sp TPWB.

Table 7-7: : Subsystem category of *Spiroplasma*.sp TPWB annotated using RAST.

	Subsystem feature	Count	Percentage%
1	Sulphur metabolism	2	0.90
2	Fatty acids, lipids and isoprenoids	6	2.70
3	Cofactors, vitamins, prosthetic groups, pigments	6	2.70
4	Protein metabolism	50	22.52
5	DNA metabolism	16	7.21
6	Stress response	1	0.45
7	RNA metabolism	18	8.10
8	Amino acids and derivatives	10	4.50
9	Carbohydrates	42	18.92
10	Virulence, disease, and defence	13	5.86
11	Membrane transport	6	2.70
12	Dormancy and sporulation	1	0.45
13	Regulation and cell signalling	2	0.90
14	Phosphorus metabolism	2	0.90
15	Nucleosides and nucleotides	47	21.17

7.5 Discussion

This research marks the first effort to describe and characterize the microbiome of T. par. utilizing the metagenomics approach, deriving seven genomes of bacterial strains from the whole body and 45 genomes from the legs. Draft genomes of *Spiroplasma* sp. from the whole body with > 95% completeness were further investigated in this study. Tabanids are recognized as abundant reservoirs of *Spiroplasma* (Clark *et al.*, 1985; French *et al.*, 1997). Numerous studies have explored *Spiroplasma* in horse flies (*Tabanus* sp.) and deer flies (*Chrysops* sp.) (Diptera: Tabanidae) (Hélias *et al.*, 1998; Whitcomb *et al.*, 2007; Regassa *et al.*, 2011; Qiu *et al.*, 2024). *Spiroplasma* (Spiroplasmataceae, Entomoplasmatales) is a type of low-GC, Gram-positive bacteria that lacks a cell wall and is characterized by its helical shape, setting them apart from other genera in the Mollicutes class. Most *Spiroplasma* sp. are commensals in insects or plants, while a few species act as pathogens affecting plants, insects, and crustaceans (Gasparich *et al.*, 2004).

This study identified three genomic strains, namely, *S. megaselia nigra* TPWB, *S. turonicum* TPWB, and *Spiroplasma* sp TPWB. The genome strains exhibited sizes of 1,000,000 bp, which aligns with the characteristic small genome sizes of *Spiroplasma*

documented to date, ranging from 0.78 to 2.22 Mb (Yeoman *et al*., 2019). However, the tRNA of these bacterial genomes is higher than that reported in other studies that report *Spiroplasma* genomes house 29-33 tRNA genes (Ku *et al*., 2013; Yeoman *et al*., 2019; Lo *et al*., 2015). Many *Spiroplasma* species found as symbionts in insects do not affect their hosts and are considered commensal bacteria (Gasparich, 2010). However, in some hosts, they are recognized as pathogens. Pathogenicity is associated with the bacteria's ability to breach the midgut lumen barrier, invade the hemolymph, and later colonize other host tissues, sometimes leading to the host's death (Sanat *et al*., 2017).

Spiroplasma megaselia nigra TPWB has been identified in this study. However, there is a lack of studies regarding its pathogenicity in tabanids (Moulder *et al*., 2002), despite its known pathogenic effects in scuttle flies (*Megaselia nigra*). It is recognized that *Spiroplasma* infects *M. nigra* and may provide its fly host with protection against parasitic attacks (Ballinger *et al*., 2019). This bacterium belongs to a broader category of *Spiroplasma* species recognized for their intricate relationships with insect hosts, frequently affecting host biology and ecology (Majoros and Adamowicz, 2023). The presence of *Spiroplasma* in *M. nigra* has been linked to various ecological and evolutionary dynamics, particularly about interactions between hosts and parasites, as well as symbiotic relationships. This presents an opportunity to explore particular environmental niches and behaviours of *T. par.*

Spiroplasma turonicum TPWB is not typically considered a pathogenic species in tabanids (Regassa and Gaspariach, 2006). Instead, it is often described as a commensal or mutualistic bacterium, meaning it lives in association with its host (usually an insect) without causing harm or providing benefit. *Spiroplasma turonicum* is recognized for residing in the gut microbiota of specific leafhopper species, especially those in the genus *Scaphoideus*. Although it is typically not linked to disease or harmful effects in its host, its ecological and physiological functions within the leafhopper are not entirely understood (Ammar *et al*., 2011). The pathogenic potential of *Spiroplasma* species can differ significantly based on the particular strain, the species of the host, and various environmental conditions. Nevertheless, no extensive documentation indicates that *S. turonicum* is a pathogenic species.

Spiroplasma sp. TPWB has been documented in various hosts, including rabbit ticks, coconut palms, horse flies, leafhoppers, Cantharid beetles, and Tiphiid wasps (Regassa and Gaspariach, 2006). Like *S. turonicum* TPWB and *S. melliferum* TPWB, the impact of *Spiroplasma* sp. TPWB on *T. par* or other *Tabanus* species has yet to be documented, creating an opportunity for further investigation into these bacteria. Different species of *Spiroplasma* sp. have been isolated from tabanid flies, and their phylogenetic relationships have been outlined; however, the specific function of this bacterium in tabanid flies remains unclear (Taioe, 2017). It has been observed that the levels of *Spiroplasma* spp. in the gut of arthropods are relatively low, suggesting that they do not proliferate in the gut or penetrate insect cells (Taioe, 2017). However, this was different in this study. In this study, however, this bacterium was abundant and present in *T. par* in the WB. Its precise role in these flies remains unknown and necessitates further research.

Regarding biochemical processes, *Spiroplasma* employs various metabolic strategies, such as glycolysis and the pentose phosphate pathway, but it does not possess pathways for synthesizing fatty acids and phospholipids (Liu *et al*., 2022). This metabolic adaptation results from its evolutionary background and ecological modifications, enabling it to thrive in various habitats, ranging from insect guts to plant tissues (Bolaños *et al*., 2014). Due to their unique biology and lifestyle, *Spiroplasma* genomes often contain many genes related to translation, ribosomal structure, and biogenesis. Here are a few reasons *Spiroplasma* genomes might exhibit this characteristic: (i) they are minimalist genomes, consisting of small genomes compared to other bacteria, which helps them to maintain cellular functions. This relies on highly efficient translation and ribosomal machinery (Liu *et al*.,2022). (ii) They portray a symbiotic lifestyle exhibited by various *Spiroplasma* species within insects, such as flies, bees, and leafhoppers. *Spiroplasma* faces specialized environmental challenges within these hosts, such as nutrient limitations and host immune responses. To adapt to these conditions, *Spiroplasma* sp. may require robust translational machinery to efficiently produce proteins essential for survival and interaction with the host (Anbutsu and Fukatsu, 2011). (iii) Endosymbiotic Adaptation: Some *Spiroplasma* sp. lives as endosymbionts within insect cells. In this intracellular environment, they may have limited access to external resources. They may rely on

efficient translation and ribosomal machinery to synthesize essential proteins for both them and their host (Masson *et al*.,2018).

Overall, the high number of genes related to translation, ribosomal structure, and biogenesis in *Spiroplasma* genomes likely reflects their adaptation to diverse ecological niches, including symbiotic relationships with insects and their ability to thrive under various environmental conditions. In summary, this research effectively sequenced the *Spiroplasma* genome, significantly enhancing understanding of *T. par* endosymbionts and serving as a valuable resource for investigating their relationships with insect hosts.

7.6 Conclusion

A genome survey conducted through shotgun sequencing showed that various *Spiroplasma* species are carried by *T. par*. Examining the partial genome sequences such as *S. megaselia nigra* TPWB, *S. turonicum* TPWB, and *Spiroplasma* sp. TPWB offered insights into the genetic structure and metabolic capabilities of these bacteria. Additionally, a comparative analysis with other Mollicutes indicated that having access to *Spiroplasma* genome sequences is crucial for enhancing our understanding of the evolutionary background within this class.

7.7 Recommendations

Future investigations focusing on genomic characterization and functional studies involving additional *Spiroplasma* associated with tabanids can enhance our comprehension of the diverse genetic strategies employed by these phylogenetically distinct bacteria in similar hosts. Moreover, comprehensive assessments of the symbiotic relationship (mutual, commensal, or parasitic) of each *Spiroplasma* species in tabanids are necessary to explore the interactions between the bacteria and their hosts. These findings broaden the scope of detailed functional investigations into the genetic mechanisms discovered and can clarify their functions in *Spiroplasma* biology and their interactions with the *T. par* host.

7.8 References

1. Ammar, E.D., Gasparich, G.E., Hall, D.G. & Hogenhout, S.A. (2011). *Spiroplasma*-like organisms closely associated with the gut in five leafhopper species (Hemiptera: Cicadellidae). *Archives of microbiology*, *193*, pp.35-44.
2. Anbutsu, H. & Fukatsu, T., 2011. *Spiroplasma* as a model insect endosymbiont. *Environmental microbiology reports*, *3*(2), pp.144-153.
3. Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data* [online].
4. Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M. & Meyer, F. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, *9*, pp.1-15.
5. Ballinger, M.J., Gawryluk, R.M. & Perlman, S.J. (2019). Toxin and genome evolution in a *Drosophila* defensive symbiosis. *Genome Biology and Evolution*, *11*(1), pp.253-262.
6. Bolanos, L.M., Servin-Garciduenas, L.E. & Martinez-Romero, E. (2015). Arthropod–*Spiroplasma* relationship in the genomic era. *FEMS Microbiology Ecology*, *91*(2), pp.1-8.
7. Chan, P.P. & Lowe, T.M. (2019). *tRNAscan-SE: searching for tRNA genes in genomic sequences* (pp. 1-14). Springer New York.
8. Cisak, E., Wójcik-Fatla, A., Zajac, V., Sawczyn, A., Sroka, J. & Dutkiewicz, J., (2015). *Spiroplasma*-an emerging arthropod-borne pathogen?. *Annals of Agricultural and Environmental Medicine*, *22*(4).
9. Clark, T.B., Whitcomb, R.F., Tully, J.G., Mouches, C., Saillard, C., Bové, J.M., Wroblewski, H., Carle, P., Rose, D.L., Henegar, R.B. & Williamson, D.L. (1985). *Spiroplasma melliferum*, a new species from the honeybee (Apis mellifera). *International journal of systematic and evolutionary microbiology*, *35*(3), pp.296-308.
10. Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C. & Trigodet, F. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nature microbiology*, *6*(1), pp.3-6.

11. Feldbauer, R., Gosch, L., Lüftinger, L., Hyden, P., Flexer, A. & Rattei, T.(2020). DeepNOG: fast and accurate protein orthologous group assignment. *Bioinformatics*, *36*(22-23), pp.5304-5312.
12. French, F.E., Whitcomb, R.F., Tully, J.G., Carle, P., Bové, J.M., Henegar, R.B., Adams, J.R., Gasparich, G.E. & Williamson, D.L. (1997). *Spiroplasma lineolae* sp. nov., from the horsefly *Tabanus lineola* (Diptera: Tabanidae). *International Journal of Systematic and Evolutionary Microbiology*, *47*(4), pp.1078-1081.
13. French, F.E., Whitcomb, R.F., Tully, J.G., Hackett, K.J., Clark, E.A., Henegar, R.B., Wagner, A.G. & Rose, D.L. (1990). Tabanid *spiroplasma*s of the southeast USA: new groups, and correlation with host life strategy.
14. Fünfhaus, A., Ebeling, J. & Genersch, E. (2018). Bacterial pathogens of bees. *Current opinion in insect science*, *26*, pp.89-96.
15. Gasparich, G.E., (2010). *Spiroplasmas* and phytoplasmas: microbes associated with plant hosts. *Biologicals*, *38*(2), pp.193-203.
16. Gasparich, G.E., Whitcomb, R.F., Dodge, D., French, F.E., Glass, J. & Williamson, D.L. (2004). The genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the Mycoplasma mycoides clade. *International Journal of Systematic and Evolutionary Microbiology*, *54*(3), pp.893-918.
17. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), pp.1072-1075.
18. Hackett, K.J., Whitcomb, R.F., Clark, T.B., Henegar, R.B., Lynn, D.E., Wagner, A.G., Tully, J.G., Gasparich, G.E., Rose, D.L., Carle, P. & Bove, J.M. (1996). *Spiroplasma leptinotarsae* sp. nov., a mollicute uniquely adapted to its host, the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *International Journal of Systematic and Evolutionary Microbiology*, *46*(4), pp.906-911.
19. Hackett, K.J., Whitcomb, R.F., French, F.E., Tully, J.G., Gasparich, G.E., Rose, D.L., Carle, P., Bové, J.M., Henegar, R.B., Clark, T.B. & Konai, M. (1996). *Spiroplasma corruscae* sp. nov., from a firefly beetle (Coleoptera: Lampyridae) and tabanid flies (Diptera: Tabanidae). *International Journal of Systematic and Evolutionary Microbiology*, *46*(4), pp.947-950.
20. Harumoto, T. & Lemaitre, B. (2018). Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature*, *557*(7704), pp.252-255.

21. Harumoto, T. (2023). Self-stabilization mechanism encoded by a bacterial toxin facilitates reproductive parasitism. *Current Biology*, *33*(18), pp.4021-4029.
22. Hélias, C., Vazeille-Falcoz, M., Le Goff, F., Abalain-Colloc, M.L., Rodhain, F., Carle, P., Whitcomb, R.F., Williamson, D.L., Tully, J.G., Bové, J.M. & Chastel, C. (1998). *Spiroplasma turonicum* sp. nov. from *Haematopota* horse flies (Diptera: Tabanidae) in France. *International Journal of Systematic and Evolutionary Microbiology*, *48*(2), pp.457-461.
23. Jandhyam, H., Bates, C.R., Young, T.E., Beatti, L., Gasparich, G.E., French, F.E. & Regassa, L.B. (2008). Global *Spiroplasma* biodiversity in a single host. In *17th Congress of International Organization for Mycoplasmology, Beijing, China. Abstract* (No. 206, p. 130).
24. Kakizawa, S., Hosokawa, T., Oguchi, K., Miyakoshi, K. & Fukatsu, T. (2022). *Spiroplasma* as facultative bacterial symbionts of stinkbugs. *Frontiers in Microbiology*, *13*, p.1044771.
25. Ku, C., Lo, W.S., Chen, L.L. & Kuo, C.H. (2013). Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. *Genome biology and evolution*, *5*(6), pp.1151-1164.
1. Li, D., Liu, C.M., Luo, R., Sadakane, K. & Lam, T.W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), pp.1674-1676.
26. Liu, P., Li, Y., Ye, Y., Chen, J., Li, R., Zhang, Q., Li, Y., Wang, W., Meng, Q., Ou, J. & Yang, Z. (2022). The genome and antigen proteome analysis of *Spiroplasma mirum*. *Frontiers in Microbiology*, *13*, p.996938.
27. Lo, W.S., Gasparich, G.E. & Kuo, C.H., 2015. Found and lost: the fates of horizontally acquired genes in arthropod-symbiotic *Spiroplasma*. *Genome Biology and Evolution*, *7*(9), pp.2458-2472.
28. Majoros, S.E. & Adamowicz, S.J. (2023). CanFlyet: Habitat Zone and Diet Trait Dataset for *Diptera* Species of Canada and Greenland. *bioRxiv*, pp.2023-03.
29. Masson, F., Calderon Copete, S., Schüpfer, F., Garcia-Arraez, G. & Lemaitre, B. (2018). In vitro culture of the insect endosymbiont *Spiroplasma poulsonii* highlights bacterial genes involved in host-symbiont interaction. *MBio*, *9*(2), pp.10-1128.

30. Moulder, R.W., French, F.E. & Chang, C.J. (2002). Simplified media for *spiroplasmas* associated with tabanid flies. *Canadian journal of microbiology*, *48*(1), pp.1-6.
31. Mushtaq, S., Summuna, B., Bhat, F.A., Gupta, V., Gulzar, A., Sheikh, A.H. & Muzamil, S. (2022). Vector transmitted bacterial diseases and their interaction with hemipteran insects.
32. Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A. & Harris, S.R. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial genomics*, *2*(4), p.e000056.
33. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. & Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research*, *25*(7), pp.1043-1055.
34. Qiu, H.Y., Lv, Q.B., Wang, C.R., Ju, H., Luo, C.F., Liu, S.S., Na, M.H., Chang, Q.C. & Jiang, J.F. (2022). Microbiota profile in organs of horseflies (Diptera: Tabanidae) in Nothestern China. *Front. Microbiol.* 15:1467875.
35. Reeves, W.K. (2003). *Biodiversity and behavior of selected endosymbiotes and hematophagous flies (Diptera) in the southern Appalachian Mountains, United States of America*. Clemson University.
36. Regassa, L.B. & Gasparich, G.E. (2006). *Spiroplasmas*: evolutionary relationships and biodiversity. *Front Biosci*, *11*(1), pp.2983-3002.
37. Taioe, M.O., Motloang, M.Y., Namangala, B., Chota, A., Molefe, N.I., Musinguzi, S.P., Suganuma, K., Hayes, P., Tsilo, T.J., Chainey, J. & Inoue, N. (2017). Characterization of tabanid flies (Diptera: Tabanidae) in South Africa and Zambia and detection of protozoan parasites they are harbouring. *Parasitology*, *144*(9), pp.1162-1178.
38. Wedincamp Jr, J., French, F.E., Whitcomb, R.F. & Henegar, R.B. (1997). Laboratory infection and release of *Spiroplasma* (Entomoplasmatales: Spiroplasmataceae) from horse flies (Diptera: Tabanidae). *Journal of Entomological Science*, *32*(4), pp.398-402.
39. Whitcomb, R.F., French, F.E., Tully, J.G., Carle, P., Henegar, R., Hackett, K.J., Gasparich, G.E. & Williamson, D.L. (1997). *Spiroplasma* species, groups, and subgroups from North American Tabanidae. *Current Microbiology*, *35*, pp.287-293.

40. Whitcomb, R.F., French, F.E., Tully, J.G., Gasparich, G.E., Bové, J.M., Carle, P., Clark, E.A. & Henegar, R.B. (1992). Tabanid *spiroplasma* serovars. *IOM Lett*, *2*, p.115.
41. Whitcomb, R.F., Hackett, K.J., Tully, J.G., Clark, E.A., French, F.E., Henegar, R.E., Rose, D.L. & Wagner, A.G., 1990. Tabanid *Spiroplasma* as a model for mollicute biogeography.
42. Whitcomb, R.F., Tully, J.G., Gasparich, G.E., Regassa, L.B., Williamson, D.L. & French, F.E. (2007). *Spiroplasma* species in the Costa Rican highlands: implications for biogeography and biodiversity. *Biodiversity and Conservation*, *16*, pp.3877-3894.
43. Yeoman, C.J., Brutscher, L.M., Esen, Ö.C., Ibaoglu, F., Fowler, C., Eren, A.M., Wanner, K. & Weaver, D.K. (2019). Genome-resolved insights into a novel *Spiroplasma* symbiont of the Wheat Stem Sawfly (*Cephus cinctus*). PeerJ, 7, p.e7548.

CHAPTER 8 ANNEXURES

APPENDIX I: Captures of genus *Tabanus* between 1960 to 2022.

	Pooled prevalence estimates				Measure of heterogeneity			Public ation bias
Risk factors	No. studies	Sample size	No of positives	Prevalence 95% CI (%)	*Q*	I^2	*Q*–*P*	P-value
Horse flies								
Tabanus sp.	5	4712	186	2.4% (0.09–0.64)	45.845	91.275	p< 0.0001	0.806
Tabanus abactor Philip, 1936	3	61286	25197	0.6% (0.0–79.0)	438.938	99.544	p< 0.0001	0.257
Tabanus accensus Austen, 1920	3	5418	626	11.2% (8.6–14.5)	11.853	83.126	p< 0.0001	0.500
Tabanus americanus Forster 1771	3	59711	85	0.1% (0.1–0.2)	0.070	61.319	p< 0.0001	0.601
Tabanus aratus Fabricius 1775	8	130373	289	0.7% (0.2–2.6)	562.228	98.755	p< 0.0001	0.804
Tabanus autumnalis Linnaeus, 1761	36	184583	8104	1.2% (0.7–2.0)	7419.024	99.528	p< 0.0001	0.002
Tabanus bifarius Loew, 1858	14	79776	2022	3.6% (1.8–7.1)	2421.700	99.463	p< 0.0001	0.381
Tabanus bovinus Linnaeus, 1758	18	72860	334	0.4% (0.2–0.8)	354.161	95.200	p< 0.0001	0.820
Tabanus briani Leclercq, 1962	5	25129	12	0.1% (0.0–0.3)	16.921	76.360	p< 0.0001	0.876
Tabanus bromius Linnaeus, 1758	42	195839	30774	10.3% (7.6–14.0)	23153.866	99.823	p< 0.0001	0.200
Tabanus calens Linnaeus, 1758	6	154338	473	0.1% (0.0–0.4)	530.383	99.057	p< 0.0001	0.573
Tabanus claripennis Bigot, 1892	3	1567	5	0.2% (0.1–0.5)	0.897	0.000	p< 0.0001	0.296
Tabanus cordiger Meigen, 1820	19	112 622	425	0.4% (0.2–0.9)	702.520	97.438	p< 0.0001	0.575
Tabanus darimonti Leclercq, 1964	4	5140	10	0.2% (0.1–0.7)	7.160	58.100	p< 0.0001	0.734

Tabanus diversifrons Ricardo, 1911	6	11644	256	2.4% (0.6–9.9)	423.353	98.819	p< 0.0001	0.707
Tabanus dorsilinea Wiedmann 1824	6	11361	165	1.5% (0.5–4.5)	172.498	97.101	p< 0.0001	0.798
Tabanus eggeri Szilády, 1923	7	75371	282	0.8% (0.2–3.7)	851.893	99.296	p< 0.0001	0.798
Tabanus exclusus Pandelle 1883	17	97605	4612	3.2% (1.7–5.8)	4420.366	99.638	p< 0.0001	0.266
Tabanus fairchildi Stone, 1938	4	118867	153	0.1% (0.0–0.2)	70.892	95.768	p< 0.0001	0.496
Tabanus fraternus Macquart, 1846	3	5638	70	1.5% (0.2–11.3)	79.555	97.486	p< 0.0001	0.515
Tabanus fuscofasciatus Macquart, 1838	3	2381	37	1.7% (0.2–17.0)	78.633	97.457	p< 0.0001	0.779
Tabanus gertrudae Philip, 1960	3	9166	76	1.5% (0.2–8.4)	87.745	97.721	p< 0.0001	0.394
Tabanus glaucopis Meigen 1820	21	107424	8924	4.0% (2.4–6.5)	4881.816	99.590	p< 0.0001	0.381
Tabanus gratus Loew, 1858	11	86562	5801	1.5% (0.6–3.5)	1579.474	99.367	p< 0.0001	0.755
Tabanus griseinus Philip, 1960	3	8940	2417	0.4% (0.0–57.8)	205.041	99.025	p< 0.0001	0.084
Tabanus importunus Wiedemann,1828	4	16597	2107	7.7% (0.4–62.4)	2814.003	99.893	p< 0.0001	0.279
Tabanus indianus Ricardo, 1911	3	9166	145	3.6% (0.5–22.6)	260.677	99.233	p< 0.0001	0.722
Tabanus indrae Hauser, 1939	4	9800	20	0.3% (0.1–1.4)	29.181	89.719	p< 0.0001	0.867
Tabanus jucundus Walker, 1848	3	1319	58	2.4% (0.7–8.0)	11.509	82.622	p< 0.0001	0.125
Tabanus konis Philip, 1960	3	2195	17	0.1% (0.1–1.1)	29.282	90.468	p< 0.0001	0.155
Tabanus laetetinctus Becker, 1913	5	10586	92	1.2% (0.4–3.9)	111.969	96.428	p< 0.0001	0.014
Tabanus laverani Surcouf, 1907	3	60371	1482	1.2% (0.4–3.2)	12.628	84.162	p< 0.0001	0.296
Tabanus leleani Austen, 1920	4	10169	261	2.0% (0.4–8.9)	316.475	99.052	p< 0.0001	0.308
Tabanus lineola fabricius,1974	14	216464	23203	10.8% (4.7–22.9)	18957.865	99.931	p< 0.0001	0.476
Tabanus lunatus Fabricius, 1794	9	45343	508	0.7% (0.2–2.3)	632.242	98.735	p< 0.0001	0.175
Tabanus maculicornis Zetterstedt, 1842	28	92643	5044	2.1% (1.3–3.6)	6532.407	99.587	p< 0.0001	0.036

Tabanus marginalis Fabricius, 1805	6	151891	583	0.3% (0.0–2.4)	2159.423	99.768	p< 0.0001	0.113
Tabanus megalops Linnaeus, 1758	5	2259	9	32.7% 16.9–53.7)	382.157	98.953	p< 0.0001	0.462
Tabanus miki Brauer, 1880	15	30955	179	0.7% (0.3–1.5)	291.523	95.198	p< 0.0001	0.620
Tabanus nemoralis Meigen, 1820	5	58483	4383	1.3% (0.4–4.3)	114.635	96.511	p< 0.0001	0.806
Tabanus neovestitus Kröber, 1931	4	58584	57	0.1% (0.1–0.1)	3.801	21.072	p< 0.0001	0.308
Tabanus nigripes Wiedemann, 1821	5	105223	1105	0.5% (0.2–1.0)	403.002	99.007	p< 0.0001	0.624
Tabanus nigrovittatus Macquart, 1847	3	82413	40100	27.7% (1.6–89.8)	379.577	99.473	p< 0.0001	0.088
Tabanus obscurehirtus Ricardo, 1908	4	2098	26	1.0% (0.4–3.0)	17.191	82.549	p< 0.0001	0.308
Tabanus occidentalis Linnaeus, 1758	4	5570	2016	49.6% (15.3–84.3)	368.789	99.187	p< 0.0001	0.308
Tabanus pallidescens Philip 1936	3	30165	10209	12.4% (4.0–32.5)	796.673	99.749	p< 0.0001	0.128
Tabanus par Walker, 1854	10	76968	43173	8.6% (4.5–16.1)	3029.590	99.703	p< 0.0001	0.210
Tabanus paradoxus Jaennicke, 1866	3	59444	523	0.8% (0.4–1.4)	18.118	88.961	p< 0.0001	0.886
Tabanus pertinens Austen, 1912	3	19907	81	0.4% (0.0–18.0)	277.599	99.280	p< 0.0001	0.138
Tabanus pumilus Macquart 1838	6	192287	2683	1.8% (0.9–3.9)	2808.685	99.715	p< 0.0001	0.676
Tabanus pungens Wiedemann, 1828	3	5315	12	0.3% (0.1–0.5)	2.671	25.111	p< 0.0001	0.827
Tabanus quatuornotatus Meigen, 1820	19	113591	3858	1.6% (1.0–2.6)	2184.531	99.176	p< 0.0001	0.363
Tabanus quinquevittatus Wiedemann, 1821	11	57234	27700	32.4% (24.7–41.1)	2235.014	99.553	p< 0.0001	0.436
Tabanus regularis Jaennicke, 1866	10	71871	470	0.8% (0.4–1.7)	310.232	97.099	p< 0.0001	0.591
Tabanus reinwardtii Wiedemann, 1828	6	145562	92	0.1% (0.0–0.5)	296.469	98.313	p< 0.0001	0.452
Tabanus rubidus Wiedemann, 1821	12	14933	635	5.2% (2.3–11.4)	802.952	98.630	p< 0.0001	0.303
Tabanus rupinae Austen, 1920	4	2766	110	3.9% (1.9–7.9)	30.671	90.219	p< 0.0001	0.371
Tabanus rupium Brauer, 1880	11	36214	216	0.4% (0.2–0.8)	156.405	93.606	p< 0.0001	0.436
Tabanus sackeni Fairchild, 1934	5	111723	1374	0.6% (0.4–0.1.0)	198.597	97.986	p< 0.0001	0.220

Tabanus secedens Walker, 1854	3	1890	124	5.4% (2.2–12.9)	28.774	93.049	p< 0.0001	0.296
Tabanus shannonellus Kröber, 1936	6	14060	240	1.4% (0.9–2.3)	35.103	85.756	p< 0.0001	0.183
Tabanus signatipennis Portschinsky, 1887	3	9284	19	0.6% (0.8–2.1)	19.637	69.456	p< 0.0001	0.725
Tabanus similis Macquart, 1850	8	113287	2027	2.8% (0.5–12.9)	6019.329	99.884	p< 0.0001	0.901
Tabanus sparus Whitney 1879	4	79723	2281	2.4% (1.5–3.6)	283.924	98.943	p< 0.0001	0.496
Tabanus spectabilis Loew 1858	4	9417	5	0.1% (0.00–0.4)	6.974	56.981	p< 0.0001	0.089
Tabanus spodopteroides Olsufjev, Moucha & Chvála, 1969	3	8705	26	0.5% (0.1–3.6)	35.848	94.421	p< 0.0001	0.810
Tabanus spodopterus Meigen, 1820	7	22873	423	0.8% (0.3–1.8)	229.717	97.388	p< 0.0001	0.367
Tabanus spodopterus ponticus Olsufjev, Moucha & Chvála, 1967	4	3576	88	1.8% (0.7–4.5)	37.510	92.002	p< 0.0001	0.089
Tabanus striatus Fabricius, 1787	11	18924	843	7.1% (2.6–18.1)	1424.756	99.298	p< 0.0001	0.405
Tabanus subsimilis Bellardi, 185	3	98033	4226	0.1% (0.0–20.6)	22.888	99.276	p< 0.0001	0.601
Tabanus sudeticus Zeller 1842	29	89124	3168	1.3% (0.8–2.1)	3486.979	99.197	p< 0.0001	0.970
Tabanus sufis Jaennicke, 1867	3	2709	1133	35.1% (72–78.9)	807.553	99.752	p< 0.0001	0.500
Tabanus sulcifrons Macquart, 1855	4	118867	7307	2.8% (1.1-7.0)	2851.632	99.895	p< 0.0001	0.174
Tabanus superjumentarius Whitney, 1879	4	78602	261	0.3% (1.0–1.4)	438.652	99.316	p< 0.0001	0.513
Tabanus systenus Burton, 1978	3	2195	23	1.6% (0.2–12.4)	40.702	95.086	p< 0.0001	0.504
Tabanus taeniola Palisot de Beauvois, 1807	16	79390	1460	4.5% (1.6–11.5)	3904.344	99.616	p< 0.0001	0.444
Tabanus tergestinus Egger 1859	26	126919	7479	6.5% (3.8–10.9)	10797.296	99.768	p< 0.0001	0.270
Tabanus thoracinus Palisot de Beauvois, 1807	5	10231	44	0.8% (0.2–3.3)	62.337	93.583	p< 0.0001	0.086

Tabanus tinctus Fabricius, 1794	9	47974	146	0.4% (0.1–1.1)	249.891	96.799	p< 0.0001	0.076
Tabanus triangulum Wiedemann, 1828	3	1567	51	8.0% (0.3–1.9)	39.282	90.468	p< 0.0001	0.898
Tabanus triceps Thunberg, 1827	3	9166	71	1.4% (0.2–10.9)	127.635	98.433	p< 0.0001	0.723
Tabanus trigeminus Coquillett, 1898	3	19050	44	0.2% (0.1–0.7)	20.259	90.030	p< 0.0001	0.271
Tabanus trimaculatus Palisot de Beauvois, 1806	4	104504	2606	1.0% (0.4–2.5)	1012.839	99.704	p< 0.0001	0.496
Tabanus unifasciatus Loew, 1858	18	90761	1337	0.9% (0.5–1.4)	714.946	97.622	p< 0.0001	0.150
Tabanus variabilis Loew, 1858	3	61103	10	2.2% (0.0–4.0)	56.304	96.448	p< 0.0001	0.851
Tabanus xanthocorus Burton, 1978	3	2918	10	0.5% (0.1–2.1)	9.921	79.841	p< 0.0001	0.869

APPENDIX II: Tabanid fly sequences

>PQ636941 [*Tabanus par*] [CC17] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATAATTGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGTCATCCAGGATCA
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCTTT
AATATTAGGGGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGAGCTGGA
ACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCTCATGGAGGAGGATC
AGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGCT
GTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGACCGAA
TACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACCA
GTTTTAGCTGGGGCTATTACTATACTTTTAACAGATCGAAATTTAAATACCTCATTTT
TTGACCCAGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTTG
GTCACCCTGAAGTTT

>PQ636942 [*Tabanus par*] [CC18] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATAATTGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGCCATCCAGGATCA
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCTTT
AATATTAGGAGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGGGCTGG
AACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGAT
CAGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGC
TGTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGAACGA
ATACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACC
AGTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAAAACCTCATTT
TTTGACCCGGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTT
GGTCACCCTGAAGTTTA

>PQ636944 [*Tabanus par*] [CC20]] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATAATTGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGCCATCCAGGATCA
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCTTT
AATATTAGGGGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGACTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGGGCTGG
AACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGAT
CAGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGC
TGTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGATCGA
ATACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACC
AGTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAACACCTCATTT
TTTGACCCAGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTT
GGTCACCCTGAAG

>PQ636947 [*Tabanus par*] [HG08]] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATAATTGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGCCATCCAGGATCA
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCTTT
AATATTAGGGGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGACTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGGGCTGG
AACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGAT
CAGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGC
TGTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGATCGA
ATACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACC
AGTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAACACCTCATTT
TTTGACCCAGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTT
GGTCACCCTGAAGT

> PQ636948 [*Tabanus par*] [HG09] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATGATTGGTACTTCATTGAGGATGTTAATTCGAGCTGAATTGGGCCTCCCAATCTCT
TTAATTTGTGGTGCCCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCCTT
AATATTAGGAGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGAGCTGGA
ACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGATC
AGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGCT
GTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGACCGAA
TACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACCA
GTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAATACCTCATTTT
TTGACCCGGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTT

>PQ636949 [*Tabanus par*] [HG10] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATAATTGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGCCATCCAGGATCT
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCTTT
AATATTAGGGGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGGGCTGG
AACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGAT
CAGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGC
TGTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGACCGA
ATACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACC
AGTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAATACCTCATTT
TTTGATCCAGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTT
GGTCACCCTGAAG

>PQ636951 [*Tabanus par*] [HI23] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATAATCGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGCCATCCAGGATCA
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCCTT
AATATTAGGAGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA

TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGAGCTGGA
ACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGATC
AGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGCT
GTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGACCGAA
TACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTAC

>PQ636953 [*Tabanus par*] [HI25] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATAATTGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGCCATCCAGGATCA
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCCTT
AATATTAGGAGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGAGCTGGA
ACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGATC
AGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGCT
GTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGACCGAA
TACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACCA
GTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAATACCTCATTTT
TTGACCCGGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTTG
GTCACCC

>PQ636956 [*Tabanus par*] [SS03] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATAATTGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGCCATCCAGGATCT
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCCTT
AATATTAGGAGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGAGCTGGA
ACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGATC
AGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGCT
GTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGACCGAA
TACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACCA
GTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAATACCTCATTTT

TTGACCCGGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTTG
GTCACCCTGAAGT

>PQ636957 [*Tabanus par*] [SS04] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATGATTGGTACTTCATTGAGTATGTTAATTCGAGCTGAATTGGGCCATCCTATATCT
TTAATTTGTGGTGCCCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCCTT
AATATTAGGAGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGAGCTGGA
ACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGATC
AGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGCT
GTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGACCGAA
TACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACCA
GTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAATACCTCATTTT
TTGACCCGGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTTG
GTCACCCTGAAGTT

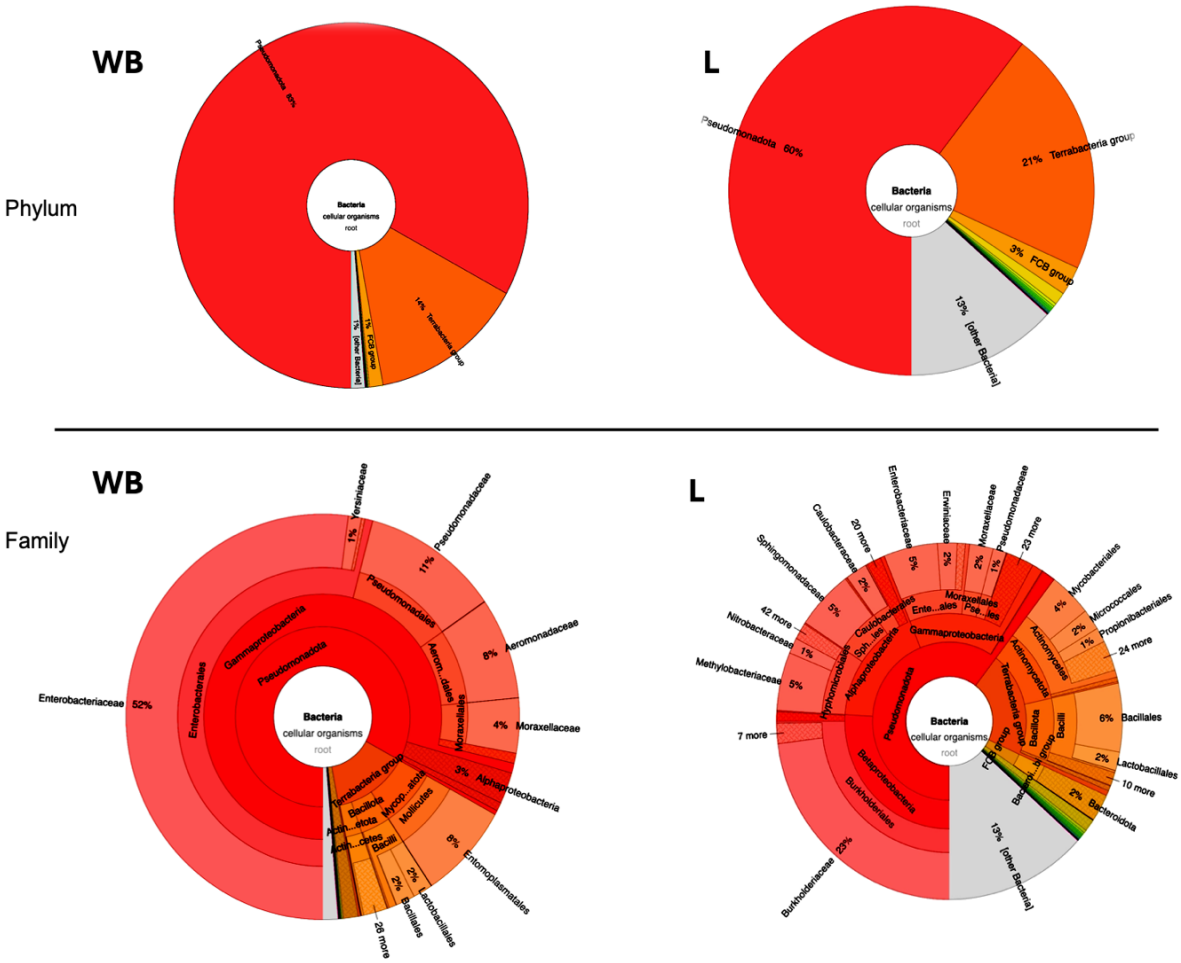
>PQ636959 [*Tabanus par]* [ZC11] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

ATATTTCTTGTTTTTAGGGGGTGAGCTGGCATAATTGGTAGTTCATTAAGTATTTTA
ATTCGAGCTGAATTGGGCCATCCAGGATCTTTAATTGGTGATGACCAAATTTATAAT
GTAATTGTAACAGCACATGCTTTTGTTATAATTTTCTTTATAGTAATACCTATTATAAT
TGGAGGATTTGGAAATTGATTAGTTCCTTTAATATTAGGGGCTCCTGATATAGCATT
TCCTCGAATAAATAATATAAGTTTTTGACTATTACCCCCCTCTCTCACTCTTTTATTA
GCCAGTAGTATAGTAGAAAATGGGGCTGGAACTGGTTGAACAGTATACCCCCCAC
TTTCTGCTGCTATTGCCCATGGAGGAGGATCAGTTGATTTAGCTATTTTTTCTCTTC
ATTTAGCAGGAATTTCTTCTATTTTAGGAGCTGTAAATTTTATTACCACTGTAATTAA
TATACGATCTACAGGAATTACATTTGATCGAATACCTTTATTTGTTTGAGCAGTAGT
AATTACTGCTATTTTATTATTATTATCATTACCAGTTTTAGCTGGAGCTATTACTATA
CTTTTAACAGATCGAAATTTAAACAC

>PQ636960 [*Tabanus par*] [ZC13] cytochrome oxidase 1(COI) gene, partial CDS; mitochondrial

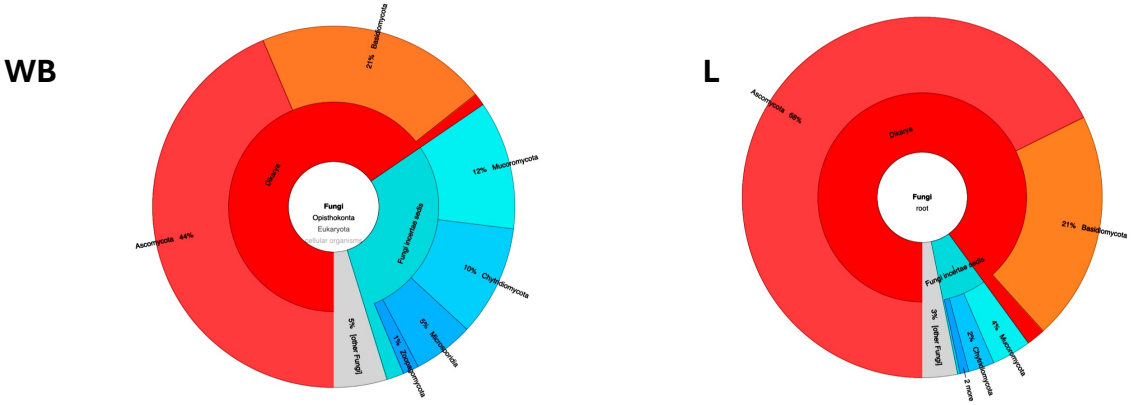
ATAGTAGGTACTTCATTAAGTATTTTAATTCGAGCTGAATTGGGCCATCCAGGATCA
TTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTTGTTATAA
TTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCCTT
AATATTAGGGGCTCCTGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGATTA
TTACCCCCCTCTCTCACTCTTTTATTAGCCAGTAGTATAGTAGAAAATGGAGCTGGA
ACTGGTTGAACAGTATACCCCCCACTTTCTGCTGCTATTGCCCATGGAGGAGGATC
AGTTGATTTAGCTATTTTTTCTCTTCATTTAGCAGGAATTTCTTCTATTTTAGGAGCT
GTAAATTTTATTACCACTGTAATTAATATACGATCTACAGGAATTACATTTGACCGAA
TACCTTTATTTGTTTGAGCAGTAGTAATTACTGCTATTTTATTATTATTATCATTACCA
GTTTTAGCTGGAGCTATTACTATACTTTTAACAGATCGAAATTTAAATACCTCATTTT
TTGACCCGGCCGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTTTTTG
GTCACCCTGAAGTT

APPENDIX III: The Krona chart comparing the taxonomic reads distribution of bacteria at phylum and family level in the whole body (WB) and legs (L) of *T. par*

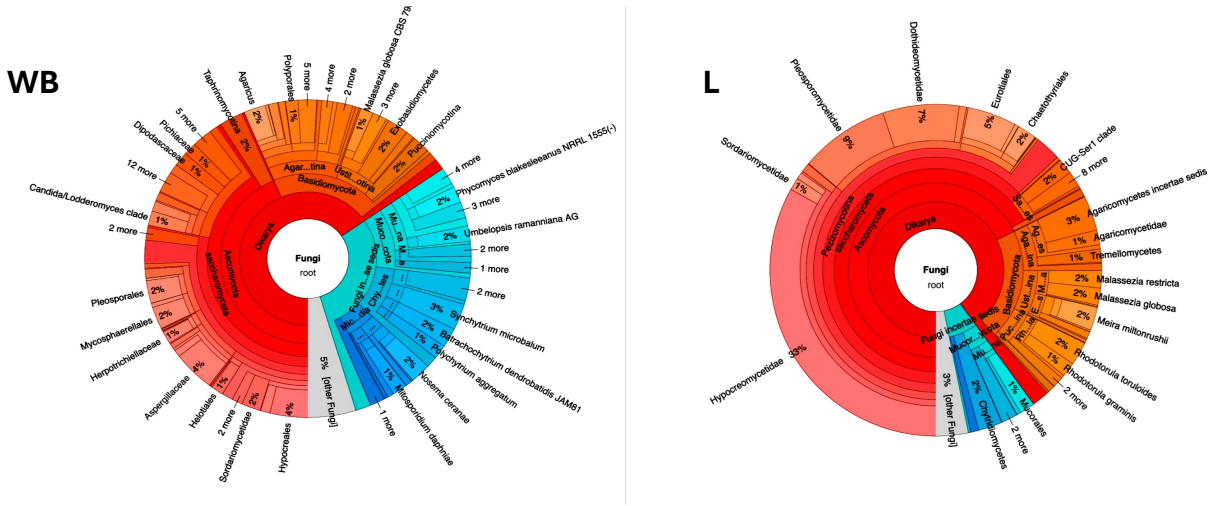


APPENDIX IV: The Krona chart comparing the taxonomic reads distribution of fungi at phylum and family level in the whole body (WB) and legs (L) of *T. par*

Phylum

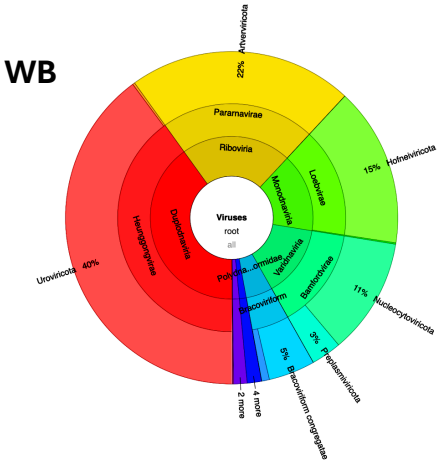


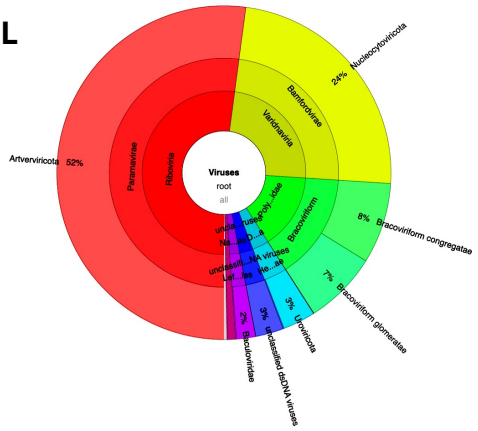
Family



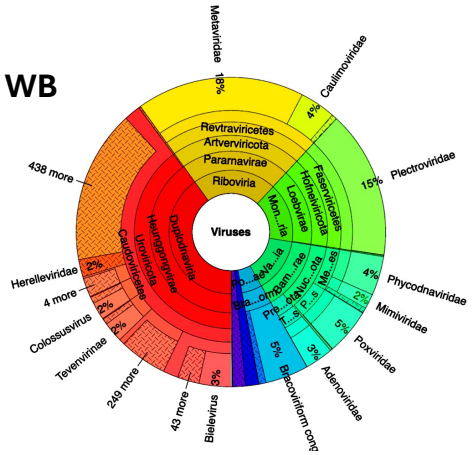
APPENDIX VI: The Krona chart comparing the taxonomic reads distribution of viruses at phylum and family level in the whole body (WB) and legs (L) of *T. par*

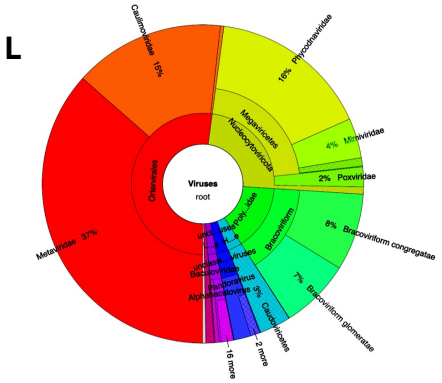
Phylum



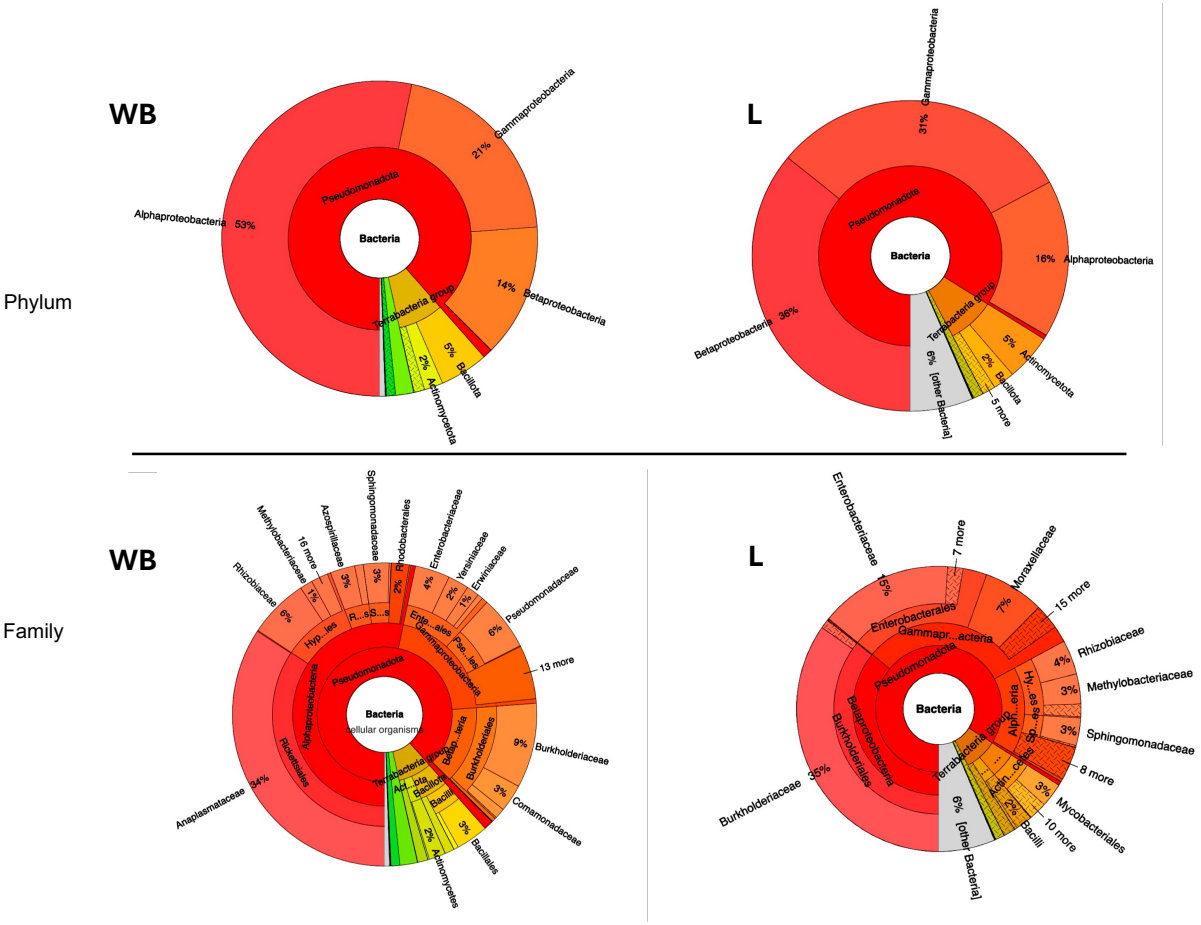


Family



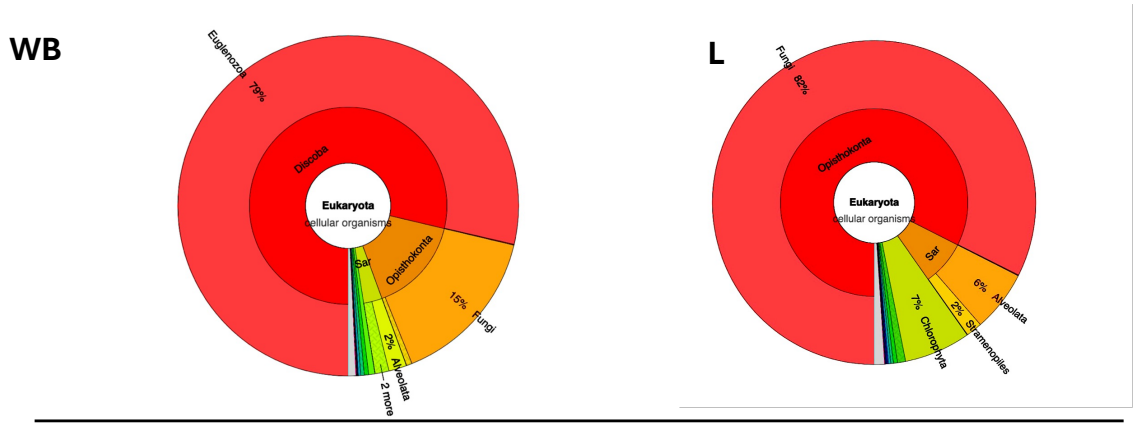


APPENDIX VII: The Krona chart comparing the taxonomic reads distribution of bacterial plasmids at phylum and family level in the whole body (WB) and legs (L) of *T. par*

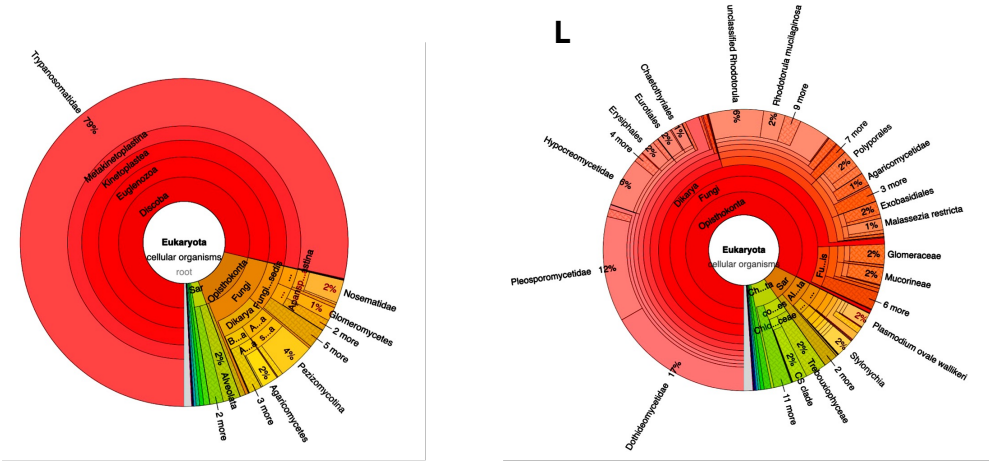


APPENDIX VIII: The Krona chart comparing the taxonomic reads distribution of eukaryotes at phylum and family level in the whole body (WB) and legs (L) of *T. par*

Phylum



Family



Appendix IX: Prokaryotic metagenome assembly genomes from the whole body of *T. par* that was classified as bad bins.

Genome ID	Coarse consistency (%)	Completeness (%)	Contamination (%)	Hypothetical (%)	Number of contigs	DNA size (bp)	Contigs N50 (bp)	Mean Coverage
Enterococcus faecalis clonal population	97.4	95.3	21.1	58.69	1950	4279164	3197	13.06
Carnobacterium sp. PL17GRE32 clonal population	94.3	95.8	18.5	79.33	2698	4888875	2191	12.12
Spiroplasma kunkelii clonal population	99.0	100	17.2	94.12	4326	6755683	1770	24.02
Bacillus cereus clonal population	92.8	100	15.7	90.38	13417	19713291	1650	13.69
Serratia marcescens clonal population	90.9	90.1	24.7	87.96	12373	17702654	1614	13.31
Acinetobacter baumannii clonal population	97.1	97.8	41.3	82.34	8347	14156682	1951	12.64
Acinetobacter bereziniae clonal population	86.4	29.2	22.2	92.62	17850	24636482	1570	12.74

The following bins met the quality criteria of completeness <= 80%, fine consistency <= 87%, contamination >= 10%, and have a single PheS protein of reasonable size.

Appendix X: Prokaryotic metagenome assembly genomes from the legs of *T. par* that were classified as bad bins.

Genome Name	Coarse consistency (%)	Fine consistency (%)	Completeness (%)	Contamination (%)	Hypothetical (%)	Contig count	DNA size (bp)	Contigs N50 (bp)	Mean Coverage
Moraxella osloensis clonal population	99.0	94.5	100	11	42.99	350	3265644	32979	38.42
Cutibacterium acnes clonal population	99.1	95.3	95.5	11.6	40.8	139	2632172	65601	18.47
Methylibium sp. T29-B clonal population	95.5	87.9	98	11.8	42.03	255	5313913	78644	48.74
Corynebacterium mycetoides clonal population	96.4	91.6	96.1	12.3	42.4	342	2526376	16142	10.05
Pedobacter sp. E01020 clonal population	91.5	91.5	58.3	0	70.9	113	677304	12692	9.68
Brevundimonas sp. GW460-12-10-14-LB2 clonal population	97.9	89.8	98	13.9	40.19	151	3644743	57657	9.22
Mycolicibacterium hodleri clonal population	97.4	89.5	99.7	12.5	51.13	176	7985037	126593	28.98
Corynebacterium accolens clonal population	95.1	90.8	89.7	12.8	40.47	337	2321768	10140	6.27

Serratia marcescens clonal population	86.5	77.8	82.6	8.6	32.74	1110	4072700	4964	7.51
Vampirovibrio chlorellavorus clonal population	88.7	83.8	100	7.7	72.72	422	3668858	13972	65.01
Micrococcus sp. 116 clonal population	99.2	92.8	98.2	16.6	38.16	453	2643031	9933	10.73
Brevundimonas sp. G8 clonal population	85.0	84.4	36.7	4	40.21	90	1556904	57302	9
Paraburkholderia insulsa clonal population	97.4	80.7	100	13.8	47.92	353	10538669	143523	112.78
Propionibacterium sp. Marseille-P4323 clonal population	97.6	90.3	92.9	18.3	34.76	225	2503922	31666	9.58
Microbacterium sp. Leaf351 clonal population	85.8	85.0	38.9	5.1	45.75	286	1173245	5628	5.82
Micrococcus luteus clonal population	92.6	92.6	4.2	0	51.38	68	153091	3119	8.04
Mesorhizobium sp. CU2 clonal population	81.1	79.7	16.7	0	48.78	351	1902152	7998	6.35
Pseudomonas aeruginosa clonal population	74.1	72.1	16.7	0	40.26	495	1993646	5767	7.4
Sphingobium sp. strain clonal population	90.3	90.3	12.5	0	64.08	121	895539	16488	11.81

Rhodothermus marinus clonal population	88.6	87.8	75	21.7	73.03	89	1417158	38553	11.51
Sphingomonas sp. 3F27F9 clonal population	85.1	79.6	25	14.3	46.3	579	3967590	12211	10.28
Ferrimonas senticii clonal population	91.7	91.6	8.3	0	99.5	2500	11872244	8459	14.9
Xanthobacter autotrophicus clonal population	79.6	73.0	56.3	25.1	47.53	894	4699056	7830	8.14
Sphingomonas adhaesiva clonal population	82.5	79.0	45.8	26.7	44.16	494	2512448	7240	8.07
Methylobacterium populi clonal population	92.3	75.6	50	25	52.68	1169	9645263	16099	8.55
Nevskia ramosa clonal population	91.1	80.5	100	38.5	52.26	168	5056092	107236	9.65
uncultured Flavihumibacter clonal population	97.1	72.4	100	49.8	61.99	577	7586068	34642	49.11
Kiritimatiella glycovorans clonal population	92.9	91.9	37.5	43.8	78.17	71	966125	26535	14.61

The following bins met the quality criteria of completeness <= 80%, fine consistency <= 87%, contamination >= 10%, and have a single PheS protein of reasonable size.

Appendix XI: *Spiroplasma* reference genomes from NCBI as well as those sequenced and obtained in this study from *T. par*.

Strains	**ID**	**Host**	**Country**	**Accession number**
S_alleghenense_PLHS1	*Spiroplasma alleghenense*	*Panorpa helena*	USA: Glady River near Bowden, West	NZ_CP031376
S_apis_B31	*Spiroplasma apis*	honeybee	Virginia	NC_022998
S_atrichopogonis_GNAT3597	*Spiroplasma mirum*	*Atrichopogon sp*	USA: Salisbury, MD	NZ_CP011855
S_cantharicola_CC1	*Spiroplasma cantharicola*	*Cantharis*	N/A	NZ_CP012622
S_chinense_CCH	*Spiroplasma chinense*	*Calystegia hederacea*	"China: Jiangsu	NZ_CP043026
S_chrysopicola_DF1	*Spiroplasma chrysopicola*	N/A	N/A	NR_025699
S_citri_BLH13	*Spiroplasma citri*	*Citrus*	USA:Ducor, CA	NZ_CP046371
S_citri_BR12	*Spiroplasma citri*	*Brittle root*	USA:Collinsville, IL	NZ_CP046368
S_citri_C189	*Spiroplasma citri*	*Citrus*	USA:Riverside, CA	NZ_CP047426
S_citri_C5	*Spiroplasma citri*	*Carrot*	USA: NW Bakersfield	NZ_CP053304
S_citri_CC2	*Spiroplasma citri*	*Chinese cabbage*	USA:Fresno	CP042472

S_citri_D12	*Spiroplasma citri*	*Sesamum indicum*	Iran	CP096808
S_citri_D4	*Spiroplasma citri*	*Neoaliturus haematoceps*	Iran	CP096246
S_citri_LB319pScpLB312	*Spiroplasma citri*	Citrus	SA:Ducor, CA	NZ_CP046373
S_citri_R8A2	*Spiroplasma citri*	*Citrus sinensis*	Morocco	NZ_CP013197
S_clarkii_CN5	*Spiroplasma clarkii*	*Cotinis nitida*	USA: Beltsville, MD	NZ_CP024870
S_corruscae_EC1	*Spiroplasma corruscae*	*Ellychnia corrusca*	USA: Beltsville, M	NZ_CP022535
S_culicicola_AES	*Spiroplasma culicicola*	N/A -	N/A	NZ_CP006681
S_diminutum_CUAS1	*Spiroplasma diminutum*	N/A	N/A	NR_025702
S eriocheiris CCTCCM207170	*Spiroplasma eriocheiris*	Eriocheir sinensis (Chinese mitten crab	N/A	GU046560
S_eriocheiris_DSM21848	*Spiroplasma eriocheiris*	*Eriocheir sinensis*	China	NZ_CP011856
S_floricola_236	*Spiroplasma floricola*	*Liriodendron tulipifera*	USA: Maryland	NZ_CP025057
S_gladiatoris_TG1	*Spiroplasma gladiatoris*	*Tabanus*	USA: Bucktown, MD	NZ_CP038013
S_helicoides_TABS2	*Spiroplasma helicoides*	*Tabanus*	USA: Bucktown, MD	NZ_CP038013
S_ixodetis_DCM	*Spiroplasma sp.*	Dactylopius coccus	Mexico	JACSER000000000
S_ixodetis_SHM	*Spiroplasma sp.*	*"Homona magnanima*	Japan:Shizuoka	NZ_AP026933
S_ixodetis_Y32	*Spiroplasma sp.*	*Ixodes pacificus*	southern Oregon"	NZ_CP127039

S_kunkelii_CR23x	*Spiroplasma kunkelii*	*Zea mays* L	Costa Rica	AY528560
S_lariophagus_STUMAEJMCOA1	*Spiroplasma sp.*	*Lariophagus distinguendus*"	Germany	NZ_JALMUW000000000
S_litorale_TN1	*Spiroplasma litorale*	*Tabanus nigrovittatus* (horsefly)	USA: Ocracoke Island, NC	NZ_CP012357
S_megaselia_sNigra1	*Spiroplasma endosymbiont of Megaselia nigra*	*Megaselia nigra*	Canada	NZ_RCOO01000121.1
S_melliferum_CH1	*Spiroplasma citri*	*Apis mellifera*	China	NZ_PHSJ00000000
S_melliferum_IPMB4A	*Spiroplasma citri*	honeybee	Chih-Horng Kuo	NZ_AMGI00000000.1
S_melliferum_KC3	*Spiroplasma citri*	N/A	France	NZ_AGBZ00000000
S_melliferum_psm1	Spiroplasma melliferum	*Apis mellifera*	USA	CP029200
S_sp_TPWBtab	*Spiroplasma sp.*	*Tabanus par* (fly)	South Africa	In this study
S_mirum_ATCC29335	*Spiroplasma mirum*	rabbit tick	N/A	NZ_CP006720
S_mirum_SMCA	*Spiroplasma mirum*	N/A	USA	DQ439670
S_monobiae_MQ1	*Spiroplasma monobiae*	wasp	N/A	NR_118705.1
S_phoeniceum_P40	*Spiroplasma sp.*	*Catharanthus roseus*	Syria	CP031088
S_platyhelix_PALS1	*Spiroplasma platyhelix*	N/A	USA	DQ307063

S_poulsonii_MSRO	*Spiroplasma poulsonii*	*Drosophila melanogaster* Oregon	Switzerland	NZ_SSBE00000000
S_poulsonii_MSROH99SMH99c001	*Spiroplasma poulsonii*	*Drosophila melanogaster* Oregon	Switzerland	SSBE01000005
S_poulsonii_sHy	*Spiroplasma poulsonii*	*Drosophila hydei*	United Kingdom	NZ_CP093047
S_poulsoniis Neo	*Spiroplasma poulsonii*	*Drosophila neotestacea*	USA: Connecticut	NZ_RAHC00000000
S_sabaudiense_Ar1343	*Spiroplasma sabaudiense*	N/A	N/A	NR_025710
S_sp_A83	*Spiroplasma sp.*	leaf litter	Brazil	JAVHXO000000000
S_sp_AdecLV25b	*Spiroplasma sp.*	*Adalia decempunctata*	United Kingdom: Liverpool	NZ_JAQYTS000000000
S_sp_ChiS	*Spiroplasma citri*	*Cichorium intybus* L	Italy: Friuli Venezia Giulia	NZ_PUUH00000000
S_sp_JKS002669	*Spiroplasma sp.*	*Trachymyrmex septentrionalis*	USA: New Jersey	NZ_JAKQXY000000000
S_sp_JKS002671	*Spiroplasma sp.*	*Trachymyrmex septentrionalis*	N/A	NZ_JAKNSO000000000
S_sp_Moj	*Spiroplasma endosymbiont of Megaselia nigra*	*Drosophila mojavensis*	USA: Catalina Island, California	MQTY00000000
S_sp_SV19	*Spiroplasma sp.*	*Stegodyphus dumicola*	Namibia	NZ_CP045455
S_sp_TIUS1	*Spiroplasma sp.*	*Tiphiidae*	SA: Beltsville, M	NZ_CP048387

S_sp_TU14	*Spiroplasma sp.*	N/A	USA: Towson, Maryland	CP017658
S_sp_WSS	*Spiroplasma sp.*	*Tabanus* par	"USA: Montana	VBWQ00000000
S_syrphidicola_EA1	*Spiroplasma syrphidicola*	N/A	N/A	CP005078
S_tabanidicola_TAUS1	*Spiroplasma tabanidicola*	"*Tabanus* sp	USA: Beltsville, MD	NZ_CP046276
S_taiwanense_CT1	*Spiroplasma taiwanense*	mosquito	N/A	CP005074
S_turonicum_Tab4c	*Spiroplasma turonicum*	*Haematopota* sp. (horse fly parasite)	France	NZ_CP012328
S_sp_TPWBtur	*Spiroplasma* sp.	*Tabanus par* (fly)	South Africa	In this study
S_Megaselia_nigraTPWB1	*Spiroplasma* endosymbiont of *Megaselia nigra*	*Tabanus par* (fly)	South Africa	In this study