CURRENT STATUS OF METATRANSCRIPTOMIC AND RELATED STUDIES IN HEMATOPHAGOUS DISEASE-TRANSMITTING VECTORS

Christina B. McCarthy¹,²

¹Centro Regional de Estudios Genómicos (CREG), Universidad Nacional de La Plata (UNLP), (1900) La Plata, Buenos Aires, Argentina.

²Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), (C1425FQB) CABA, Argentina.

Correspondence author: Dr Christina B. McCarthy, E-mail: mccarthychristina@gmail.com

Subject editor: Rui-De Xue

ABSTRACT

The incidence of numerous vector-borne diseases (VBDs) has recently increased alarmingly due to various widespread factors, including unplanned urbanization, greater human mobility, environmental changes, vector resistance to insecticides, and evolving pathogens. In this context, the World Health Organization (WHO) has repositioned effective and sustainable vector control as a key approach to prevent and eliminate VBDs. It has been shown that the microbiome influences development, nutrition, and pathogen defense in disease-transmitting vectors such as mosquitoes, sandflies, tsetse flies, triatomine bugs, and ticks. Consequently, understanding the endogenous regulation of vector biology can aid in developing effective approaches for vector control. In this respect, a metatranscriptomic approach analyzes all the expressed RNAs in an environmental sample (meta-RNAs) and can thus reveal how the metabolic activities of the microbiome influence vector biology. This review includes an extensive analysis of available literature on microbial and viral studies for some of the major hematophagous disease-transmitting arthropods, with a focus on studies that used next generation sequencing (NGS) approaches. Since a consensus terminology for these “meta-sequencing analyses” has not yet been established, a definition of these terms is presented here to provide the framework for systematically sorting the available information for each of the VBDs analyzed here to single out metatranscriptomic analyses. Finally, key gaps in knowledge were identified for some of these hematophagous disease-transmitting arthropods which will prove very useful for driving future studies.

Key words: Microbiota, Microbiome, Metabarcoding, Metagenomics, Metatranscriptomics, Vector-borne diseases, Mosquitoes, Sandflies, Tsetse flies, Triatomines, Kissing bugs, Ticks

INTRODUCTION

The importance of vector-borne diseases (VBDs) is indisputable. They represent more than 17% of all infectious diseases and cause more than 700,000 deaths each year (WHO 2017). Many of these vectors are bloodsucking arthropods that transmit disease-causing pathogens when taking a blood meal (WHO 2023a). Mosquitoes head the list with three main disease-transmitting genera: Anopheles (malaria), Aedes (mainly arboviruses, including dengue, Zika, chikungunya, yellow fever, Rift Valley and others), and Culex (West Nile virus and Japanese encephalitis virus); all three genera can vector Lymphatic filariasis (WHO 2023a). Sandflies (mainly leishmaniasis and sandfly fever), tsetse flies (African trypanosomiasis), kissing bugs (Chagas disease), and ticks (Lyme disease, relapsing fever or borreliosis, and Rickettsial diseases such as spotted fever and others), are also responsible for high levels of morbidity and mortality (WHO 2023a, 2017). Effective vaccines and medicines exist for a few of these VBDs, such as vaccines against yellow fever, Japanese encephalitis, and tick-borne encephalitis, medicines against lymphatic filariasis, human onchocerciasis and malaria, but they are currently lacking for most other VBDs. Consequently, vector control is still the most effective preventive approach against the majority of VBDs, and interventions that reduce human-vector contact and vector survival can suppress and even halt transmission (WHO 2017).

In this context, it is now accepted that all eukaryotes are meta-organisms and must be considered together with their microbiome as an inseparable functional unit (Jones 2013). The microbiome forms a dynamic and interactive micro-ecosystem that is integrated to the eukaryotic host and, as such, is crucial for its correct functioning and health (Berg et al. 2020). Consequently, an extensive understanding of the microbiome is pivotal for developing
effective vector control approaches. A significant progress in this respect has been witnessed for the past decade, particularly since the development of high throughput sequencing platforms which have transformed the field of microbial community analysis. Nevertheless, as usually happens when a certain field grows exponentially, consensus terminology is lacking. The misuse of terms such as microbiota, microbiome, metagenome, metagenomics, metabarcoding, metataxonomics and metatranscriptomics, among others, has contributed to the misinterpretation of many study results (Marchesi and Ravel 2015). In this review these terms are clearly defined (see below) based on different previous proposals, to establish a baseline and avoid confusion when interpreting the data and drawing conclusions.

The **microbiota** encompasses the living prokaryotic (Bacteria, Archaea) and eukaryotic (e.g., Protozoa, Fungi, and Algae) microorganisms present in a defined environment (Marchesi and Ravel 2015). Viruses, plasmids, prions, viroids, and free (or “relic”) DNA are not part of the microbiota because they are not living microorganisms (Dupré and O’Malley 2013).

On the other hand, the **microbiome** includes the community of microorganisms and their “theatre of activity” (Whipp et al. 1988). The latter encompasses all molecules produced by microorganisms (structural elements [nucleic acids, proteins, lipids, polysaccharides], metabolites [signaling molecules, toxins, organic, and inorganic molecules]), as well as viruses and “relic” DNA. Consequently, microorganisms, viruses, plasmids, prions, viroids, and free DNA, are all part of the microbiome (Berg et al. 2020).

**Metataxonomics** covers the large-scale analysis of sequencing data (DNA or RNA) to identify microorganisms and/or viruses from complex environmental samples. Metataxonomic studies can be undertaken using two main approaches (see below): 1) metabarcoding and 2) metagenomic/metatranscriptomic shotgun DNA/RNA sequencing as modified from (Cox et al. 2017).

**Metabarcoding** is the large-scale analysis of biodiversity (i.e., species composition within a sample) through the amplification and sequencing of homologous genes (Creer et al. 2010), such as the mitochondrial protein-coding gene cytochrome c oxidase subunit I (COI) for animal specimens (Hebert et al. 2003), the 16S ribosomal DNA gene region for bacteria and archaea (Tringe and Hugenholtz 2008), and the 18S ribosomal DNA gene region for microbial eukaryotes (Creer et al. 2010). Thus, barcoding of environmental DNA/RNA (or cDNA/eRNA) enables the simultaneous identification of many taxa within the same sample (Wikipedia 2023), and eRNA metabarcoding also provides a measure of the active or viable community (Mengoni et al. 2005).

A **metagenome** is the collection of genomes and genes of the microbiota, and is obtained through shotgun sequencing of DNA extracted from a sample (**metagenomics**). The sequencing data is first assembled and mapped, or directly mapped, to a reference database, and finally annotated. It thus provides information on the functional potential of the microbiota.

Since a metabarcoding analysis is based on the amplification and sequencing of taxonomic marker genes, it is not metagenomics (Marchesi and Ravel 2015).

While metagenomics provides information on the putative activities of a microbial community, it cannot reveal the activities that are occurring at a specific time and place, nor how those activities change in response to the environment or to biotic interactions. The challenge is to discover which of those potential functions (metagenome) are happening at a particular point in time (metatranscriptome) and, ultimately, to identify what causes the difference (Moran 2009).

Thus, **metatranscriptomics** analyses all the expressed RNAs in an environmental sample (meta-RNAs) by next generation sequencing (NGS) of the corresponding meta-cDNAs (Marchesi and Ravel 2015). Consequently, it has the potential to identify all the taxa within an environment (and not just those within a targeted lineage, as is the case with metabarcoding) and also provides a comprehensive overview of the loci that are being transcribed and of their expression levels (Galen et al. 2020). Therefore, metatranscriptomics is a more informative approach compared to metagenomics, because it not only characterizes the genetic content (as in a metagenomic analysis) but also identifies the populations that are transcriptionally active (Bashirades et al. 2016). Metatranscriptomics has been used for determining the functional profile of the microbiome, but it also has the potential to detect and classify RNA from different lineages (metataxonomics), and this latter aspect has been exploited to detect and characterize viruses (Batson et al. 2021; Marcelino et al. 2019; Ortiz-Baez et al. 2020; Westreich et al. 2019).

**METHODS**

To attain an exhaustive review of published literatures on microbial and viral studies for the hematophagous disease-transmitting arthropods included in this review (mosquitoes, sand flies, tsetse flies, triatomines, and ticks), with a focus on studies that used NGS approaches, a literature search was conducted on English databases,
mainly PubMed and Google up to 28 September 2023, using a set of terms without language or publication-type restrictions.

The first and necessary step to provide the framework for systematically sorting the available information for each of the vectors analyzed here, and then singling out the metatranscriptomic analyses, consisted in defining the terms used in this review (microbiota, microbiome, metagenome, metagenomics, metabarcoding, metataxonomics and metatranscriptomics). For this, a database search was performed using keywords that included: "Microbiota", "Microbiome", "Metabarcoding", "Metagenomic", "Metatranscriptomic", "Virome", "Metavirome", "next-generation sequencing", "high-throughput sequencing", "Vector-borne diseases". Following this database search, the title and abstract of the retrieved publications were screened to identify studies and reviews that were potentially eligible for inclusion.

Next, the full texts of likely suitable studies were retrieved and i) further assessed for eligibility, and ii) screened for other relevant studies that may not have been found in the previous step.

In this way, the full text of 26 eligible publications were thoroughly assessed to define the mentioned terms, and thus provide the necessary baseline for interpreting the data that was retrieved in the following step.

The search for each vector was performed separately. Once again, keywords were used for the database searches, that included: "Microbiota", "Microbiome", "Metabarcoding", "Metagenomic", "Metatranscriptomic", "Virome", "Metavirome", "next-generation sequencing", "high-throughput sequencing", and "Mosquitoes", "Sandflies", "Ticks", "Triatomines", "Kissing bugs", "Tssete flies", depending on the vector.

Following this, the title and abstract of the retrieved publications were screened to identify studies and reviews that were potentially eligible for inclusion.

Next, the full texts of suitable reviews were retrieved and screened to confirm: i) the eligibility of the studies that were selected and retrieved from the database search, and ii) to search for other relevant studies that may not have been found in the database searches.

Subsequently, the full texts of potentially suitable studies were retrieved and i) thoroughly assessed for eligibility, ii) screened for other relevant studies that may not have been found in the previous step, and iii) classified according to the type of analysis (e.g., metabarcoding, metavirome, metatranscriptomic) following the terminology proposed in this review. With respect to this last point, it is important to note that a thorough screening of the full texts was paramount to correctly assign the type of analysis because, due to the mentioned lack of consensus terminology, titles can be misleading. For example, titles of metavirome studies have used the terms "metatranscriptomic" (Feng et al. 2022) or "Shotgun metagenomics" (Aragão et al. 2023). Similarly, the titles of culture-dependent and -independent studies have used the terms "microbial" (Clay et al. 2008) or "microbiota" (Yadav et al. 2015), and thus had to be screened in detail to determine eligibility.

Finally, the full text of 144 suitable publications on microbial and viral studies (including studies and reviews) for mosquitoes, sandflies, tsetse flies, triatomines, and ticks, were thoroughly assessed and used for this review.

The publications that were included in this analysis are mentioned in the text and have been incorporated in the references list.

**METATRANSCRIPTOMIC STUDIES IN HEMATOPHAGOUS DISEASE-TRANSMITTING ARTHROPODS**

In accordance with the previously defined terms, in this review a metatranscriptomic study was considered as such if it used NGS to analyze all the expressed RNAs (i.e., not only taxonomic marker genes) in the microbiome (or from at least two lineages e.g., prokaryotes and viruses). Consequently, studies that used environmental RNA (eRNA) sequencing to identify only one lineage (e.g., viruses) were considered metataxonomic analyses, and are only mentioned as background information. Furthermore, due to space constraints, this review focuses on the (main) hematophagous disease-transmitting arthropods, i.e., mosquitoes, sandflies, tsetse flies, triatomines, and ticks, which include obligate and non-obligate blood feeders (Beaty and Marquardt 1996). Obligate blood feeders feed exclusively on vertebrate blood during all life stages (e.g., triatomine bugs and ticks) or only as adults (e.g., tsetse flies). Non-obligate blood feeders (e.g., mosquitoes and sandflies) consume organic materials during immature stages and, during adulthood, in addition to blood ingest sugars to obtain energy (Song et al. 2022).

In the following sections the mentioned hematophagous disease-transmitting arthropods are considered individually, and in each case, a brief introduction is included on the pathogens they transmit, vector biology, and available microbial and viral studies. Following that, a specific subsection briefly describes the metatranscriptomic studies for that vector (if there are any).
Mosquitoes (*Diptera: Culicidae*)

There are thousands of mosquito species, but the main disease-transmitting vectors with the greatest threat to public health, belong to the genera *Anopheles*, *Aedes*, and *Culex* (Beatty and Marquardt 1996; Clements 1992; WHO 2023a). Mosquitoes are responsible for transmitting some of the most dangerous pathogens, including protozoa (most importantly *Plasmodium*), filarial nematodes, and viruses (Gabrieli et al. 2021). In 2021 nearly half of the world's population was at risk of malaria, with an estimated 217 million cases and 619,000 deaths worldwide (WHO 2023b). *Culex* spp. mosquitoes transmit both arboviruses, such as West Nile virus (*Flaviviridae: Flavivirus*), and filarial parasites, and *Aedes* spp. (mainly *Aedes aegypti* and *Ae. albopictus*) transmit arboviruses of medical importance to animals and humans, including dengue (*Flaviviridae: Flavivirus*), Zika (*Flaviviridae: Flavivirus*) and chikungunya (*Togaviridae: Togavirus*) viruses (Weaver et al. 2018; WHO 2017). Some of these pathogens have been wreaking havoc for a long time, and others are emerging or resurging, and have a very real devastating potential (Weaver et al. 2018).

The mosquitoes' immature stages (larvae and pupae) are aquatic, and larvae feed on organic materials. On the other hand, adults are terrestrial and feed on plant saps and nectars, whereas females also ingest animal blood for egg development (Clements 1992). Consequently, the mosquito microbiome is (at least partly) environmentally acquired, and can be found in the midgut, salivary glands and reproductive tracts (Gao et al. 2020). The microbiome affects vector competence, host immune system signaling, and longevity, among others, and as such, is critical for mosquito development (Caragata et al. 2019; Guéган et al. 2018; Strand 2018). Due to its influence on vector-borne pathogen transmission, and potential for vector control, the mosquito microbiome has attracted increasing attention over the past decade. With this escalating interest, analyses of the mosquito microbiome using NGS approaches are generating hundreds of scientific publications every year (Dada et al. 2021b). Of these, the vast majority correspond to metataxonomic analyses that have used either DNA metabarcoding or RNA shotgun sequencing to study the different components of the microbiome separately. The metabarncoding analyses have mainly focused on bacteria (e.g., Boissière et al. 2012; Buck et al. 2016; Coon et al. 2016, 2014; Dada et al. 2021a, 2019; Díaz et al. 2021; Dickson et al. 2017; Duguma et al. 2019; Gimonneau et al. 2014; Hegde et al. 2018; Mancini et al. 2018; Muturi et al. 2016; Osei-Poku et al. 2012; Sharma et al. 2014; Trzebny et al. 2023; Villegas et al. 2018; Wang et al. 2011), but a couple have analyzed the fungal (Tawidian et al. 2021) and eukaryotic (Belda et al. 2017) components, and one metabarcoding study included both prokaryotes and eukaryotes (Thongsrippong et al. 2018).

The viral component of the microbiome has also been extensively studied by means of metataxonomic approaches that used meta-RNA shotgun sequencing (e.g., Aragão et al. 2023; Fauver et al. 2016; Feng et al. 2022; Hameed et al. 2021; Li et al. 2023; Liu et al. 2023; Ramírez et al. 2020; Sadeghi et al. 2018; Shi et al. 2015, 2017; Thongsrippong et al. 2021; Wu et al. 2023; X. Yang et al. 2023).

**Metatranscriptomic studies in mosquitoes**

Four metatranscriptomic studies in mosquitoes have been published to date (see Table 1), however, neither of these studies analyzed the metatranscriptomic data to determine the expression profile of the microbiomes. The first one was designed as a proof of concept to characterize the members of the mosquito microbiome (Chandler et al. 2015). The authors used meta-RNA shotgun sequencing on seven individual field-collected female mosquitoes from three species, *Culex pipiens* (Farajollahi et al. 2011), *Culiseta incidens* and *Ochlerotatus sierrensis* (Ledesma and Harrington 2011). Sequences from viruses, bacteria, and fungi were identified in each individual, and mosquito species identities were also verified using the sequencing data. Single stranded RNA viruses of the *Bunyaviridae* and *Rhabdoviridae* were identified, along with an unclassified genus of double-stranded RNA viruses. Further, sequences related to 8 bacterial and 13 fungal families were found across the seven samples. *Bacillus* and *Escherichia/Shigella* were identified in all samples and *Wolbachia* was identified in all *Cx. pipiens* samples, while no single fungal genus was found in more than two samples. This study underscores the advantage of using this approach to characterize the mosquito microbiome and, especially, the value of identifying all the components associated with a specific host (Chandler et al. 2015).

The next metatranscriptomic analysis was published in 2021. In this study, unbiased metatranscriptomic sequencing of 148 individual field-collected adult *Aedes*, *Culex*, and *Culiseta* mosquitoes enabled the detection of sequences from eukaryotes, prokaryotes, and 24 known and 46 novel viral species (Batson et al. 2021). The fact that individual mosquitoes were sequenced added great value to the biological information that was obtained. Among others, it was possible to compute the prevalence of each microbe and the high frequency of viral co-infections, to establish an association between animal pathogens and specific blood meals, and to speciate the host mosquito (Batson et al. 2021).
That same year another metatranscriptomic study analyzed total RNA extracted from dissected abdomens of *Ae. albopictus* females fed with sugar and human blood containing either normal or heat-inactivated serum, to evaluate the effect of heat inactivation on gene expression in the mosquitoes, and on the bacterial and viral components of their microbiome (Calle-Tobón et al. 2021). The authors found that at least 600 host genes showed a modified expression profile when mosquitoes were fed with normal vs. heat-inactivated-containing blood, and that the bacterial community changed at 6 hours post-feeding. Nevertheless, they did not observe differences in the core viral component of the mosquito microbiome. These results suggest that serum heat inactivation may have a profound effect on mosquito and microbiome metabolism. This study only described the bacterial and viral components of the microbiome.

The most recent metatranscriptomic analysis evaluated the microbiome of *Ae. albopictus* populations in Germany (Rau et al. 2022) where the mosquito specimens collected as larvae in the field from seven German locations were processed immediately after adult emergence, and adults were pooled according to sex before total RNA extraction. Sequence analysis revealed the presence of viruses, bacteria, and fungi. Some of the identified taxa had already been described in *Ae. albopictus*, such as *Wolbachia pipientis*, *Acinetobacter baumannii* or *Usinis virus*. Others had been detected previously in other mosquito species and invertebrates but not in *Ae. albopictus*, including *High Island virus*, *Guapiaçu virus* and *Elizabethkingia anophelis*. Lastly, some of the bacteria had not been identified previously in mosquitoes, including *Limnobacter humi*, *Zooglea resiniphila*, and *Chryseobacterium aureum*. The authors also found differences between males...
and females; in females more contigs were assigned to bacteria, whereas in males most contigs were assigned to viruses (Rau et al. 2022).

**Sandflies (Diptera: Phlebotominae)**

Phlebotomine sandflies can transmit various diseases. Even though the most important are the leishmaniases (Maroli et al. 2013), they also transmit viruses (Alkan et al. 2013; Depaquit et al. 2010) and bacteria (Maroli et al. 2013), although little is known about the molecular interactions of sandflies with viruses and bacteria (Telliera et al. 2018). Phleboviruses are the most significant of the sandfly-borne viruses, causing symptoms that span from short term fever to haemorrhagic fever (Alkan et al. 2013). In South America, sandflies are the most important vectors of *Bartonella bacilliformis*, the etiological agent of bartonellosis (Battisti et al. 2015; Schultz 1968).

Sandflies lay their eggs in moist environments (leaves, soil, animal burrows, and/or tree trunk niches) and immature stages feed on organic materials (Volf et al. 2002). During adulthood they feed on sugars and females also ingest blood (Beaty and Marquardt 1996). Consequently, they are exposed to a wide range of microorganisms and viruses which can become part of their microbiome (Sant’Anna et al. 2012), mainly colonizing the sandfly midgut (Telliera et al. 2018). Sandflies become infected with *Leishmania* when they engorge on host blood to develop eggs and reproduce, and as the parasite develops exclusively in the mid- and hindgut of the sandfly, it coexists and interacts with the gut microbiome (Kelly et al. 2017). Moreover, the gut microbiome has a significant impact on *Leishmania* development (Louradour et al. 2017), and on sandfly fecundity and development (Telliera et al. 2018), which is why it has gained relevance over the last decade (Tabbabi et al. 2022).

Initial approaches to study the sandfly microbiota were culture-dependent (Akhoundi et al. 2012; Dillon et al. 1996; Oliveira et al. 2000; Perira de Oliveira et al. 2001; Volf et al. 2002) but, with the advent of molecular methods, standard bacteriological methods were combined with Sanger-sequencing of clones and culture-independent methods (Campolina et al. 2020; Fraihi et al. 2017; Gouveia et al. 2008; Guernaoui et al. 2011; Gunathilaka et al. 2020; Hillesland et al. 2008; Karimian et al. 2019; Li et al. 2016; Machado et al. 2014; Maleki-Ravasan et al. 2015; Mukhopadhyay et al. 2012; Sant’Anna et al. 2012; Vivero et al. 2016) (reviewed in Tabbabi et al. 2022). Notably, since the development of high throughput platforms, fewer studies have used an NGS approach to analyze the sandfly microbiome compared to other hematophagous arthropods such as mosquitoes and ticks. A few studies have used DNA metabarcoding to describe the bacterial community (e.g., Kelly et al. 2017; Papadopoulos et al. 2020; Pires et al. 2017; Vivero et al. 2021, 2019), and the bacterial and fungal communities (Tabbabi et al. 2021), and one RNA metabarncoding study analyzed 16S rRNA transcripts (Monteiro et al. 2016).

Interestingly, even though sandfly-borne viruses have been extensively studied using traditional methods (reviewed in Ayhan and Charrel 2017; Depaquit et al. 2010; Jancarova et al. 2023), no metataxonomic approach has yet been used to analyze the viral component of the microbiome.

**Metatranscriptomic studies in sandflies**

To date, only two metatranscriptomic studies have been published for sandflies (Table 1). Neither study analyzed the metatranscriptomic data to determine the functional profile of the microbiome.

The first study analyzed the microbiome associated with field-caught adult male and female *Lutzomyia longipalpis* from an Argentine endemic (Posadas, Misiones) and a Brazilian non-endemic (Lapinha Cave, Minas Gerais) visceral leishmaniasis location (McCarthy et al. 2011). Total RNA was extracted from whole sandflies and submitted to high-throughput pyrosequencing. The diversity of bacterial, fungal, and protist transcripts that were identified mostly confirmed the sandflies’ feeding habits and behavioral patterns. Nevertheless, it also suggested that these vectors could possibly be a chance source of dispersal of various animal and plant diseases, such as coccidiosis and malaria. Gregarines (protozoan invertebrate parasites) were also identified, which suggested they could be used as an efficient control method under natural conditions (McCarthy et al. 2011).

The other study analyzed the metatranscriptomes of several adult *Phlebotomus chinensis* populations in China (Wang et al. 2022). This analysis revealed actively replicating/transcribing bacteria, RNA and DNA viruses, and eukaryotic microbes. The authors found that the microbiome represented up to 1.8% of the total non-ribosomal RNA and comprised more than 87 species, 70 of which were novel, including divergent *Flaviviridae* and Trypanosomatidae. Importantly, they identified four types of human and/or mammalian pathogens, including two phleboviruses (hedi and wuxiang viruses), one novel spotted fever group *Rickettsia*, and a member of the *Leishmania donovani* complex. This study also showed the ubiquitous presence of *Wolbachia*. 
Tsetse flies (Diptera: Glossinidae)

Tsetse flies (Glossina sp.) are the primary vector of Trypanosoma brucei, the causal agent of human and domesticated animal African trypanosomiases in sub-Saharan Africa (Wang et al. 2013).

Adult tsetse (males and females) feed exclusively on vertebrate blood and, unlike other oviparous insects, females produce only one egg per gonotrophic cycle (Tobe 1978). Offspring develop in their mother’s uterus, immediately pupate after being deposited as 3rd instar larvae (adenotrophic viviparity), and adults emerge after 30 days. A highly modified maternal accessory gland (or milk gland) provides nourishment during larvagensis (Attardo et al. 2008; Benoit et al. 2012), and maternal milk is the route used by vertically-transmitted symbiotic bacteria to colonize the developing larvae (Wang et al. 2013).

Tsetse harbor various bacterial species. The bacterial community includes 3 maternally-transmitted endosymbionts, and a taxonomically diverse but reduced assemblage acquired from the environment (Wang et al. 2013), particularly from the host skin surface during blood meals (Farikou et al. 2010; Simo et al. 2008). The simplicity of the bacterial microbiota is most probably due to the unique aspects of tsetse fly biology, which significantly limit environmental microbial exposure. Namely, the obligate vertebrate blood feeding lifestyle of adults and the live birth of progeny following intrauterine larval development (Benoit et al. 2015).

Wigglesworthia, Sodalis and Wolbachia are the three endogenous symbionts. All field-collected tsetse flies examined to date harbor the obligate Wigglesworthia, whereas infection prevalence of Sodalis in field populations varies from 0 to 85% (Farikou et al. 2011; Maudlin et al. 1990), and Wolbachia infection prevalence in field-captured tsetse differs significantly between different host species, and between populations of the same species (Alam et al. 2013; Doudoumis et al. 2012).

Tsetse’s association with Wigglesworthia is ancient (50-80 million years ago), and the significance of this mutualism has crystallized in the bacteriome structure (Aksoy et al. 1995). This specialized organ is an immunotolerant niche that only harbors Wigglesworthia within specialized epithelial cells (bacteriocytes) (Aksoy 2000, 1995). This bacterium is also found extracellularly in milk gland secretions (Attardo et al. 2008). Wigglesworthia provides its host with nutritional and immunological benefits, supplying the necessary nutrients that are lacking in the blood diet (Wang et al. 2009). Moreover, in the absence of Wigglesworthia, 1) intrauterine larval development is stunted and progeny aborted (Pais et al. 2008; Schlein 1977), and 2) larval intrauterine development produces adults with a severely compromised immune system (Weiss et al. 2012, 2011).

Sodalis is a gram-negative endosymbiont closely related to free-living Enterobacteriaceae, that is also found in other insects such as stink bugs (Kaiwa et al. 2010) and weevils (Toju et al. 2010). In contrast to Wigglesworthia, it exhibits a wide tissue tropism and can be found both intra and extracellularly in various tissues including midgut, fat body, milk gland, salivary glands and hemocoel (Balmand et al. 2013; Cheng and Aksoy 1999). Even though Sodalis lacks a clearly defined functional role within its host and is absent in several natural tsetse populations, various studies indicate that it may play a role in tsetse’s ability to vector pathogenic trypanosomes. In contrast to Wigglesworthia, which increases tsetse refractoriness to trypanosomes, Sodalis appears to favor the establishment of trypanosome infections (Wang et al. 2013; Welburn et al. 1993).

Wolbachia is a widespread alpha-proteobacteria endosymbiont that infects approximately 70% of insects, including some tsetse populations (Hilgenboecker et al. 2008). Wolbachia is only found intracellularly in tsetse germ line tissues, and can be detected in early oocyte, embryo and larvae (Balmand et al. 2013; Cheng et al. 2000). It is thus transmitted transovarially via germ line cells, in contrast to Sodalis and Wigglesworthia which are transmitted via milk gland secretions.

Other environmentally acquired bacteria are found in tsetse flies and include members of the phyla Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. They have been found consistently in different tsetse species captured in geographically distinct localities (Aksoy et al. 2014; Geiger et al. 2009; Lindh and Lehane 2011). Nevertheless, they account for less than 1% of tsetse’s bacterial gut microbiota (Aksoy et al. 2014), and their effect on the biology of tsetse flies is still unclear (Gaithuma et al. 2020).

The bacterial component of the microbiome has been studied to a certain degree using culture dependent and independent approaches (Geiger et al. 2011, 2009; Lindh and Lehane 2011), DNA metabarcoding (Aksoy et al. 2014; Doudoumis et al. 2017; Griffith et al. 2018; Tsakeng et al. 2022), and one DNA metabarcoding analysis that simultaneously gauged the bacterial component and the blood meal source (Gaithuma et al. 2020). Compared to other VBDs, the number of NGS studies is very limited.

Many tsetse flies from colonies and natural populations, also harbor a salivary gland-associated rod-shaped, enveloped DNA virus called Salivary Gland Hypertrophy Virus (SGHV) (Jaenson 1978), that can
cause hypertrophy of the salivary glands and gonadal lesions (Jaenson 1978). It is vertically transmitted via maternal milk gland secretions, or horizontally during the feeding process (Abd-Alla et al. 2011). Most SGHV-infected tsetse are asymptomatic and have no apparent loss of host fitness, but flies infected with high virus titers show reduced fecundity and lifespan, and display hypertrophied salivary glands (Abd-Alla et al. 2011; Sang et al. 1999). Infection prevalence of SGHV in field populations varies according to location and species (Malele et al. 2013). Recently, two single-stranded RNA viruses of unknown impact were isolated from a *Glossina morsitans morsitans* colony (*Glossina morsitans morsitans* iflavirus (GmmIV) and *Glossina morsitans morsitans* neveivirus (GmmNeGeV)) (Manni and Zdobnov 2021). Results revealed potential horizontal viral transmission during feeding and/or vertical viral transmission from parent to offspring (Manni and Zdobnov 2021).

Another study that analyzed public tsetse RNA-seq libraries (mainly from laboratory colonies), identified the genomes of four flaviruses (Manni and Zdobnov 2021). The flavivirus identified in *G. morsitans* (GillaVI) was found in all 136 available *G. morsitans* RNA-seq libraries, and displayed a broad tissue tropism and high abundance, reaching up to 15% of library content. Its ubiquitous distribution and presence in the reproductive tissues, intrauterine larvae, and teneral flies suggest it could be part of the initial core microbiota maternally transmitted to the progeny (Manni and Zdobnov 2021).

Nevertheless, no meta-RNA sequencing approach has been undertaken to characterize the viral component of the tsetse microbiome.

**Metatranscriptomic studies in tsetse flies**

No metatranscriptomic approach has yet been undertaken to identify the prokaryotic, eukaryotic, and viral composition of the tsetse microbiome (Table 1).

Two studies addressed tsetse-Wigglesworthia mutualism through dual RNA sequencing (Bing et al. 2017; Munoz et al. 2017). Briefly, one of these studies characterized the expression profile of the tsetse-Wigglesworthia association within the bacteriome of field captured adult tsetse (*Glossina pallidipes*) from Kenya, with the objective of understanding these interactions within the host’s natural setting (Munoz et al. 2017). The other study used colony-reared individuals to perform a dual RNA-seq analysis of the bacteriome, coupled with a metabolomic analysis of the bacteriome and haemolymph collected from normal and symbiont-cured (sterile) females (Bing et al. 2017).

Recently, a comparative transcriptomic analysis was performed between *Glossina morsitans* and *G. brevipalpis* teneral (Medina Munoz et al. 2021). Because these newly emerged adults have not yet fed, their digestive tract microbiota only consists of the core bacteria seeded through maternal milk gland secretions, namely *Wigglesworthia* and *Sodalis* (Medina Munoz et al. 2021). Although a more diverse bacterial community has been reported in the digestive tracts of adults, these environmentally acquired bacteria are lacking within teneral. Consequently, the mentioned study only compared the *Wigglesworthia*, *Sodalis* and tsetse transcriptomes (Medina Munoz et al. 2021).

As none of these studies included the eukaryotic and/or viral components of the microbiome, in this review they were not considered metatranscriptomic analyses.

**Kissing bugs (Hemiptera: Reduviidae: Triatominae)**

Triatomines, also known as kissing bugs, vector *Trypanosoma cruzi*, the etiological agent of Chagas’ disease (WHO 2023c). An estimated 6-7 million people worldwide are infected with *T. cruzi*, leading to around 12,000 deaths each year and some 75 million people at risk of infection, mainly in Latin America (WHO 2023d).

Triatomines typically live in home walls or roof cracks and peridomestic structures of rural or suburban areas. They usually feed at night, and the parasites enter the body when the person inadvertently smudges the faeces or urine into the bite, other skin breaks, the eyes or the mouth (WHO 2023c).

Triatomines feed exclusively on vertebrate blood throughout their developmental cycle and, as other hematophagous vectors, they harbor beneficial symbionts whose primary role is to supply them with nutrients that are lacking in the diet (Salcedo-Porras et al. 2020). Symbionts are extracellular, reside in the midgut and hindgut lumens (Brecher and Wigglesworth 1944; Duncan 1926; Wigglesworth 1936), and are required for the insect’s development and survival (Brecher and Wigglesworth 1944; Durvasula et al. 2008; Vallejo et al. 2009; Yassin 2005). Symbionts include *Rhodococcus rhodnii*, *Corynebacterium* sp. and *Nocardia* sp. (Salcedo-Porras et al. 2020), and are transmitted from parent to offspring by coprophagy (Salcedo-Porras et al. 2020).

The first study to use 16S rRNA gene amplification in triatomines identified only one bacterium in *Triatoma infestans* (Hypša and Dale 1997). Since then, culture-dependent (Lopez-Ordonez et al. 2018), culture-independent (da Mota et al. 2012; Gumieli et al. 2015), and high-throughput sequencing approaches have been used to gain a more comprehensive view of the bacterial component of the microbiome. The latter have mostly used bacterial metabarcoding (Brown et al. 2020; Díaz
Metatranscriptomic and disease vectors

et al. 2016; Kieran et al. 2019; Lima et al. 2018; Mann et al. 2020; McCall et al. 2018; Montoya-Porras et al. 2018; Oliveira et al. 2018; Orantes et al. 2018; Rodríguez-Ruano et al. 2018; Tarabai et al. 2023; Waltmann et al. 2019), a few have used DNA metabarcoding to identify various components simultaneously (bacteria, vertebrate hosts, parasite diversity and, in one case, triatomine bugs) (Dumonteil et al. 2020, 2018; Murillo-Solano et al. 2021), one metatranscriptomic study used shotgun pyrosequencing to describe cultivable bacteria (Carels et al. 2017), and another metatranscriptomic study used an interesting Restriction-site Associated DNA sequencing (RADSeq)-based analysis to simultaneously study the vector, the parasite, bacteria and feeding patterns (Orantes et al. 2018) (reviewed by Salcedo-Porrás et al. 2020). Various of these studies have reported low bacterial diversity in the triatomine microbiome in comparison to other insect groups (da Mota et al. 2012; Gumiel et al. 2015; Lopez-Ordóñez et al. 2018). Nevertheless, the triatomine microbiome harbors a broad spectrum of eukaryotic organisms and viruses, apart from bacteria (Song et al. 2022).

Very little is known about the viral component of the microbiome. To date, only 8 triatoma viruses have been identified and characterized: the Triatoma virus (TrV), that was discovered in a colony of field-collected Triatoma infestans (Muscio et al. 1988, 1987), and very recently seven Rhodnius prolixus viruses 1-7 (RpV1-7) (De Brito et al. 2021), which were initially discovered in transcriptome assemblies from ovarian tissues of Rhodnius prolixus (Coelho et al. 2021). Both RpVs and TrV are vertically transmitted to progeny (De Brito et al. 2021; Muscio et al. 1997). On the other hand, contigs related to viral genomes were incidentally identified in transcriptomic analyses of the salivary glands, fat bodies and testes of Rhodnius prolixus, Panstrongylus megistus and P. lignarius (Nevoa et al. 2018; Ribeiro et al. 2015; Schwarz et al. 2014), but these observations were not explored further. Finally, no metatranscriptomic analysis has yet studied the viral component of the microbiome and thus remains a pending assignment.

Metatranscriptomic studies in kissing bugs

Notably, no metatranscriptomic study has been performed in triatomines (Table 1).

Ticks (Arachnida: Ixodida)

Tick-borne pathogens cause most of the VBDs in temperate North America, Europe and Asia, and although these include viruses, bacteria, and parasites (Jongejan and Uilenberg 2004), Lyme disease is the most prevalent in the northern hemisphere (Rochlin and Toledo 2020). Some tick species may harbor numerous pathogens, whereas other species are typically associated with one major pathogen (Sanchez-Vicente et al. 2019).

There are two main tick families, Argasidae (soft ticks) and Ixodidae (hard ticks), that differ in their ecology and public health impact (Parola and Raoult 2001; Sonenshine and Roe 2014). Soft ticks have a more restricted habitat (Sonenshine 2014), feed quickly, can take several blood meals per stage (Vial 2009), and transmit fewer human pathogens than hard ticks (Parola and Raoult 2001). On the other hand, hard ticks are cosmopolitan (Sonenshine 2014), and have extended feeding periods (Sonenshine and Roe 2014) during the active feeding stages in their life cycle (larva, nymph and adult) (Parola and Raoult 2001), that facilitate the transmission of pathogens (Eisen 2018).

Ticks are obligate blood feeders and, because they are vulnerable to desiccation, they live in dark and humid conditions (e.g., in leaf litter and animal burrows) (Goddard 2005). Consequently, exposure to these habitats, combined with the process of feeding on animals that host a diverse skin microbiome, provide opportunities for ticks to obtain part of their microbiome from the environment (Burtis et al. 2019). Like all blood-sucking arthropods, ticks lack key vitamins that are necessary for their development and rely on their bacterial symbionts to overcome this dietary limitation (Bonnet and Pollet 2021). The tick microbiome thus includes vertically transmitted symbionts and the environmentally acquired commensals.

Tick microbial diversity and composition has mostly been characterized by sequencing of the 16S rRNA gene (Bonnet and Pollet 2021). Some of these studies have used culture-independent approaches and Sanger sequencing (Clay et al. 2008; Hartelt et al. 2004; Moreno et al. 2006; Schabereiter-Gurtner et al. 2003; Van Overbeek et al. 2008), but most have used bacterial DNA metabarcoding (e.g., Andreotti et al. 2011; Barraza-Guerrero et al. 2020; Beard et al. 2021; Budachetri et al. 2014; Carpi et al. 2011; Clayton et al. 2015; Gall et al. 2017; Guizzo et al. 2020; Heise et al. 2010; Lalzar et al. 2012; Narasimhan et al. 2014; Ponnusamy et al. 2014; Qiu et al. 2014; Sakamoto et al. 2020; Sperling et al. 2020; Zhang et al. 2020). One study performed DNA metabarcoding of various lineages (Bacteria, Archaea, Fungi and protists) (Landesman et al. 2019), whereas a couple of studies used shotgun metagenomics to analyze Bacteria and Archaea (Nakao et al. 2013) or only Bacteria (Díaz-Sánchez et al. 2019), and two studies analyzed bacterial transcriptomics (Hernández-Jargüín et al. 2018; Vayssier-Taussat et al.
Tick-borne viruses are a diverse group that includes members of *Flaviviridae, Bunyavirales, Orthomyxoviridae* and *Reoviridae* (Johnson et al. 2023). Probably this is why the viral component of the tick microbiome has recently been studied quite extensively through meta-RNA sequencing (e.g., Bratuleanu et al. 2023; Cai et al. 2023; Guo et al. 2022; Harvey et al. 2019; Kong et al. 2022; Liu et al. 2022; Ni et al. 2023; Pettersson et al. 2017; Tokarz et al. 2018; Xu et al. 2021; Z. Yang et al. 2023).

**Metatranscriptomic studies in ticks**

The only metatranscriptomic study in ticks to date (Table 1) did not examine the metatranscriptomic data to determine the functional profile of the microbiome. The authors used untargeted metatranscriptomics to analyze the prokaryotic, eukaryotic and viral components of the microbiome in ticks (mainly *Ixodes holocyclus* and *Haemaphysalis banchetii*) and in wildlife blood samples (from *Rattus rattus, Rattus fuscipes, Perameles lesinauta* and *Trichosurus vulpecula*) from urban and rural sites in Australia (Gofton et al. 2022). This study identified 32 unique tick-borne taxa, including 10 novel putative species. These included haemoproteozoa (*Babesia, Theileria, Hepatozoon* and *Trypanosoma* spp.), bacteria (*Borrelia, Rickettsia, Ehrlichia, Neoehrlichia* and *Anaplasma* spp.), and numerous viruses (including *Reoviridae* and a novel *Flaviviridae*-like jingmenvirus). A phylogenetic analysis of all the tick-borne microorganisms indicated that they were unique compared to their relatives from outside Australia, and no foreign tick-borne human pathogens were found (Gofton et al. 2022).

**CONCLUSION**

This review has addressed the status of metatranscriptomic and related studies in VBDs, focusing on some of the main hematophagous disease-transmitting arthropods namely mosquitoes, sandflies, tsetse flies, triatomines and ticks. The analysis was based on an extensive literature review of available microbial and viral studies for these hematophagous arthropods, and mainly focused on analyses that used high throughput sequencing approaches. Moreover, due to the lack of consensus terminology for these “meta-sequencing analyses”, as a first step, these terms were defined to establish the necessary baseline for interpreting those studies and drawing consistent conclusions, namely:

- The majority of studies that used NGS approaches to analyze the microbiome of these vectors, carried out bacterial metataxonomic analyses using DNA metabarcoding.
- Most metataxonomic studies have been carried out in mosquitoes, followed by ticks, whereas the number of analyses for triatomines, sandflies, and tsetse flies is quite limited, particularly for sandflies and tsetse.
- The number of metatranscriptomic studies is notoriously low for all these hematophagous vectors: only 4 studies in mosquitoes, 2 in sandflies, 1 in ticks, and none in triatomines and tsetse flies (Table 1). Moreover, even though metatranscriptomics has the potential to unravel the taxonomic and functional profile of a sample, these studies only focused on identifying the different components of the microbiome and did not analyze the data to determine their expression profile.

Despite the fact that it is a challenge to assign functions and correctly interpret results in metatranscriptomic studies (Moran 2009; Rozadilla et al. 2020), the benefits of identifying all the processes that are simultaneously mediated by an undisturbed microbiome are evident. Ultimately, this review has helped to single out these gaps in knowledge for the VBDs included here, and this is a major step towards addressing them in future studies.

**REFERENCES CITED**


Metatranscriptomic and disease vectors


Brecher W, Wiggleworth VB. 1941. The transmission of actinomycosis rhodni in rhodnius prolixus stål (hemiptera) and its influence on the growth of the host. *Parasitology* 35. doi:10.1017/S0031182000092168


Metatranscriptomic and disease vectors


Submitted date: November 1, 2023, Accepted Date: January 31, 2024, Published date: June 30, 2024.