TRANSCRIPTOMICS PROFILE OF LONG NON-CODING RNAS AND PROTEIN-CODING GENES IN *Aedes aegypti* CELLS INFECTED WITH DENGUE VIRUS 1

SATTAM MOHAMMAD ALI OBEIDAT

UNIVERSITI SAINS MALAYSIA

2021

TRANSCRIPTOMICS PROFILE OF LONG NON-CODING RNAS AND PROTEIN-CODING GENES IN *Aedes aegypti* CELLS INFECTED WITH DENGUE VIRUS 1

by

SATTAM MOHAMMAD ALI OBEIDAT

Thesis submitted in fulfillment of the requirements
for the Degree of
Doctor of Philosophy

September 2021

ACKNOWLEDGEMENT

Praise be to Almighty Allah, the Most Gracious and Most Merciful, for His blessing and support which without this project won't be done and I would not be here today. Next, my heartiest and deep felt gratitude is for my supervisor, Associated Professor Dr. Mohd Ghows Mohd Azzam, whose patience and guidance has sincerely been the heart of the soul of this project. Also I would like to thank my co-supervisors, Dr. Muhammad Amir bin Yunus and Dr. Hasni bin Arsad for their support throughout doing the experiments in Advanced Medical & Dental Institute (AMDI).

My deeply thankful to my labmate Azali bin Azlan for his help and guidance throughout this study. Many thanks to my labmate, Ms. Nurul Islam Rosmera, Dr. Ahmad Faisal, lab technician, Mr Zulkarnain, and all other labmates and friends who helped me throughout doing this project. Finally, I would like to extend my thanks to the school of Biological Sciences and AMDI for supporting me with materials needed for the research.

DEDICATION

Dedicated to my Parents for their support, understanding and prayers throughout my entire time spent on completing this project. And to my Brothers Dr. Islam for his financial support and Eng. Jihad, and to my sisters Also to my Wife for her patience and support and to my children; Jihad, Isra'a, AbdulMalik and AbdulHameed

I won't forget to dedicate this work to my beloved late sister Wafa'a

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

%	percentage
°C	degree Celsius
µL	Microliter
µg	Microgram
Ae.	*Aedes*
ANRIL	Antisense Noncoding RNA in *INK4* Locus
BLAST	Basic Local Alignment Search Tool
bp	base pair
C	capsid
cAMP	cyclic adenosine monophosphate
cDNA	complementary DNA
CDS	protein-coding sequence
ceRNA	competitive endogenous RNA
CHIKV	Chikungunya virus
CO_2	carbon dioxide
CPE	cytopathic effect
CREB	cAMP responsive element binding protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
Cx.	*Culex*
D.	*Drosophila*
DENV	dengue virus
DENV-1	dengue virus serotype 1
DHF	dengue hemorrhagic fever
DMEM	Dulbecco's modified Eagles medium

DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphates
dpi	days post-infection
DSB	double-stranded break
dsRNA	double-stranded RNA
DSS	dengue shock syndrome
E	envelope glycoprotein
EMBOSS	The European Molecular Biology Open Software Suite
ER	endoplasmic reticulum
ESCC	esophageal squamous cell carcinoma
FBS	fetal bovine serum
FC	fold change
FDR	false discovery rate
GC	guanine-cytosine
GO	gene ontology
GTF	gene transfer format
H^+	hydrogen ion
HIV	human immunodeficiency virus
IAV	influenza A virus
kb	kilobases
lncRNA	long noncoding RNA
M.	*Mus*
miRNA	microRNA
MOI	multiplicity of infection
mL	mililiter

mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
ncRNA	noncoding RNA
ng	nanogram
NS	non-structural
nt	nucleotide
OD	optical density
ORF	open reading frame
PANDA	P21-associated Noncoding RNA-DNA Damage Activated
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
piRNA	PIWI-interacting RNA
Pol II	RNA polymerase II
poly-A	polyadenylated
pre-mRNA	precursor mRNA
prM	membrane precursor
qPCR	quantitative PCR
RIN	RNA integrity number
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RNAi	RNA interference
rpm	rotation per minute
rRNA	ribosomal RNA
SAM	Sequence Alignment/Map
SEM	standard error of the mean

SINV	Sindbis virus
siRNA	small-interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
TBE	Tris-borate EDTA
$TCID_{50}$-CPE	fifty percent tissue culture infectious dose – cytopathic effect
TLR	Toll-like receptor
TPB	tryptose phosphate broth
TPM	transcript per million
tRNA	transfer RNA
UTR	untranslated region
WHO	World Health Organization
WNV	West Nile virus
XIST	X-inactive Specific Transcript
ZIKV	Zika virus

LIST OF APPENDICES

PROFIL TRANSKRIPTOM RNA PANJANG TAK MENGEKOD DAN GEN PENGEKOD PROTEIN DALAM SEL *Aedes aegypti* YANG DIJANGKITI VIRUS DENGGI 1

ABSTRAK

Virus denggi (DENV) dari famili *Flaviviridae,* ialah virus bawaan oleh nyamuk yang menyebabkan penyakit denggi di beberapa kawasan tropika dan sub-tropika. Ia disebarkan oleh nyamuk *Aedes,* dan *Ae. aegypti* ialah vektor utama. Banyak kajian memfokuskan terhadap fungsi gen pengekod protein dalam interaksi virus-perumah. Gen tak-mengekod seperti RNA tak-mengekod panjang (lncRNA) berpotensi untuk berperanan dalam interaksi virus-perumah. Kajian sebelum ini menunjukkan bahawa lncRNAs mempunyai peranan penting dalam perkembangan, interaksi virus dan perumah, dan imuniti. Walau bagaimana pun, informasi berkenaan fungsi lncRNA dalam *Ae. aegypti* masih tidak lengkap. Oleh itu, dalam kajian ini, untuk melengkapkan set lncRNA dalam genom rujukan, pengenalpastian dan pencirian lncRNAs *Ae. aegypti* telah dijalankan menggunakan set data jujukan RNA yang mendalam dan alatan bioinformatik. Sebanyak 4689 lncRNA baharu telah dikenal pasti dan dicirikan secara sistematik, dan telah ditambah kepada genome rujukan. Daripada RBA tak-mengekod panjang yang novel ini, 2064 adalah intergenik, 2076 dan 549 masing-masing adalah intronik dan antisense. Tambahan itu, didapati bahawa ciri-ciri lncRNAs *Ae. aegypti* adalah sama dengan lncRNA yang dijumpai dalam spesies lain seperti mempunyai GC yang lebih rendah, ukuran yang lebih pendek, mempunyai ulangan jujukan yang lebih banyak dan jujukan yang terperlihara yang kurang berbanding gen pengekod protein. Selain itu, penyenaraian gen terekspresi berbeza yang tepat dan cepat adalah salah satu

target projek ini. Keputusannya, profil transkriptom sel *Ae. aegypti* yang dijangkiti virus denggi serotip 1 (DENV-1) menunjukkan bahawa kedua-dua lncRNAs dan gen pengekod protein telah berubah secara ketara (DE). Bilangan lncRNAs dan gen pengekod protein yang mengalami perubahan ekspresi yang ketara, masing-masing adalah 178 dan 310. Analisis ontologi gen (GO) menunjukkan, kebanyakan gen yang mengalami perubahan ekspresi terlibat dalam fungsi molecular, pada kedua dua gen dikawalatur naik dan turun. Didapati bahawa ianaya terlibat dalam beberapa proses dan aktiviti spesifik; ikatan, membran, dan proses biologi. Analisis terma GO menunjukkan bahawa 26 gen pengekod berkait dengan immunity, sementara 132 RNA tak pengekod panjang berkemungkinan berpotensi dalam mengawal gen pengekod protein, dan 24 darii bilangan itu berkait dengan system imun. Hasil kajian ini memberi kefahaman asas berkenaan lncRNAs *Ae. aegypti*, ciri-cirinya, dan tindak balasnya terhadap jangkitan terhadap DENV-1. Kajian ini juga memberi kefahaman dalam interaksi molekul antara virus dengan *Ae. aegypti* lncRNAs and gen pengekod protein. Selain itu, kajian ini juga boleh dijadikan asas bagi kajian fungsi lncRNA *Ae. aegypti* pada masa depan. Secara kesimpulannya, hasil daripada kajian ini, berjaya mengisi ruang dengan pengetahuan dan informasi yang baharu.

TRANSCRIPTOMICS PROFILE OF LONG NON-CODING RNAS AND PROTEIN-CODING GENES IN *AEDES AEGYPTI* CELLS INFECTED WITH DENGUE VIRUS 1

ABSTRACT

Dengue virus (DENV) of the *Flaviviridae* family is a mosquito-borne virus which cause dengue disease in several tropical and sub-tropical areas. It is transmitted by *Aedes* mosquitoes, and *Ae. aegypti* is the main vector. Many studies focused on the functions of protein-coding genes in virus-host interaction. Non-coding genes such as long non-coding RNAs (lncRNAs) have shown to have potential role in virus-host interaction. Previous studies showed that lncRNAs have important roles such as development, virus-host interaction and immunity. However, information on the functions of *Ae. aegypti* lncRNAs are still incomplete. Thus, in this research, in order to complete the set of lncRNAs in the reference genome a genome-wide identification and characterization of *Ae. aegypti* lncRNAs were performed using high depth RNA-sequencing data sets and vigorous bioinformatics tools. Total number of 4689 novel long non-coding RNAs were systematically identified and characterized, and they were added to the reference genome. Among these novel long non-coding RNAs, 2,064 were intergenic, 2076 and 549 were intronic and antisense respectively. Moreover, it was found that the characteristics of *Ae. aegypti* novel lncRNAs were similar to lncRNAs discovered in other species like lower GC content, shorter in length, more repeat sequences, and lower sequence conservation compared to protein-coding genes. Furthermore, list of accurate and vigorous differentially expressed genes was one of the targets in this project. As a result, transcriptome profile of *Ae. aegypti* cells infected

with dengue virus serotype 1 (DENV1) showed that both lncRNAs and protein-coding genes were differentially expressed (DE). The number of differentially expressed lncRNAs and protein-coding genes were 178, and 310, respectively. Gene ontology (GO) analysis revealed that most of the differentially expressed genes were involved mostly in molecular functions, in both up and down-regulated genes. It was found that they involved in some specific processes and activities, in binding, membrane, and biological processes. GO term analysis showed that 26 protein-coding genes were related to immunity, whereas 132 long non-coding RNAs were predicted to have potential in regulating protein-coding genes, and 24 of them were related to immune system. The outcomes of this research provide a basic understanding of *Ae. aegypti* lncRNAs and their characteristics and responses to DENV1. This study also provides understanding in molecular interaction between virus and *Ae. aegypti* lncRNA, and protein coding genes. Besides that, this work can be a foundation for future investigation on the function of *Ae*. *aegypti* lncRNAs. In a conclusion, outcomes of this project help to fill the gap with new knowledge and information.

CHAPTER ONE

INTRODUCTION

1.1 Background

Dengue fever (DF) is one of the most common arboviral diseases (Bäck & Lundkvist, 2013). It is threatening half of the world population (da Silva-Voorham et al., 2009). It keeps expanding its geographical areas due to the high tolerance of their mosquito vectors (Tedjou et al., 2019). DF has three stages; one mild and two severe (Kalayanarooj, 2011).

DENV is one of the viruses in *flaviviridae* family (Murugesan & Manoharan, 2020; Richards et al., 2012). It has four serotypes namely serotype 1,2,3 and 4 (DENV1-4). They are antigenically different, although they are more than 65% genetically similar (Bäck & Lundkvist, 2013). All serotypes cause full spectrum of the disease; however, clinical manifestations vary among them and all of them are circulating every few years in Malaysia (Ghazali et al., 2012; Halsey et al., 2012; Kalayanarooj, 2011; Suppiah et al., 2018; Pang & Loh, 2016).

Yellow fever mosquito, *Ae. aegypti* is the main vector of many lethal arboviruses including DENV (Paixão et al., 2018). Africa is the native habitat of this mosquito before it starts its journey to Americas, Europe, and ends in Asia in 1950 (Powell et al., 2018). *Ae. aegypti* tolerated to live in urban areas, therefore, increasing the arboviruses exposure to humans (Paixão et al., 2018).

Transcriptome is a set of all RNA molecules ranging from protein-coding genes and non-coding RNAs (ncRNAs) (Blumenberg, 2019: Domingo, 2020). Several protein-coding genes have been shown to have roles in immune system

response against bacteria, fungus, and viruses (Mukherjee et al., 2019). A previous study on transcriptome analysis of *Ae. aegypti* showed that differentially expressed genes involved in various functions in the cells (Batool et al., 2018).

Furthermore, ncRNAs also have significant roles in various biological processes, including virus-host interaction (Damas et al., 2019). There are many types of ncRNAs in eukaryotes such as, transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA) and long non-coding RNAs (lncRNAs) (Azlan et al., 2016; Blumenberg, 2019). Among these ncRNAs, this project will only focus on lncRNAs. lncRNAs are RNA molecules that are longer than 200 nucleotides (nt) (Clark & Mattick, 2011; Lekka & Hall, 2018; Rinn & Chang, 2012; Wang & Chang, 2011). Despite lacking coding potential, lncRNAs share several similar features with protein-coding transcripts, such as, capping, polyadenylation, and alternative splicing (Perry & Ulitsky, 2016; Ulitsky & Bartel, 2013).

DENV-1 was chosen for this study for several reasons. Firstly, DENV-1 and DENV-2 cause the most severe symptoms in many cases which increase the mortality in patients compared to other two serotypes. Yung et. al. 2015 reported that during DF outbreak in 2015 in Guangdong China, DENV-1 infection turned into more acute symptoms in comparison with DENV-2. Furthermore symptoms resulted from DENV1 infection found to be different from the symptoms caused by DENV-2 (Yung et al., 2015), which indicates that different serotype provokes dissimilar molecular response in infected people. Thus, is believed that the transcriptional responses in *Ae. aegypti* will vary with different DENV serotype (Bonizzoni et al., 2012; Jupatanakul et al., 2017; Miesen & Ivens, et al., 2016; Sánchez-Vargas et al.,

2009; Scott et al., 2010; Sim et al., 2012; Sim & Dimopoulos, 2010; Tsujimoto et al., 2017). Secondly, when this project started in 2017, DENV-1 was predominant and it was responsible for most cases. Additionally, a study on DF in Malaysia revealed that DENV-1 was still predominant for four years between 2015-2018 (Suppiah et al., 2018). Thirdly, DENV-1 was responsible for dengue outbreaks in different states in Malaysia. It was responsible for dengue outbreak in Negeri Sembilan in 2012, and in Sandakan, Sabah in 2016 (Ghazali et al., 2012; Gintarong et al., 2018).

Meanwhile, *Ae. aegypti* was chosen in this research because it is the main vector of DENV. Although, there are some other species in the genus *Aedes* that are vectors for DENV, *Ae. aegypti,* the primary vector is responsible for most of the outbreaks in most of countries where the disease spreads. It was postulated that *Ae. aegypti* has high tolerance and immunity against arboviruses (Alonso-Palomares et al., 2018; Goic et al., 2016).

Furthermore, protein-coding genes were studies here due to their importance in various functions in the cells. In addition to protein-coding genes, previous studies reported that, lncRNAs may play significant roles in the cells in different functions, and they may have potential roles in virus-host interactions (Du et al., 2012; Etebari et al., 2016; Gomez et al., 2013; Lazar et al., 2016; Pan et al., 2019; Winterling et al., 2014; Xu et al., 2019; Zhang et al., 2013; Zhao et al., 2016).

Indeed, there are no specific treatments for dengue, besides, the vaccine should be affective against all DENV serotypes, however, there is no sufficient vaccine yet. Meanwhile, the mosquito developed a resistance against insecticides. Thus, the mosquito increases its geographical distribution and increase the number of people who are exposure to its attack. Therefore, in the absence of efficient

controlling of both, the disease and the vector, the understanding of the mechanisms of how the mosquito respond to the infection with DENV was vital. Thus, to understand these mechanisms, it is important to study the transcriptome (including Protein-coding genes and lncRNAs) of the mosquito after infecting with the virus. Furthermore, in order to have extensive and strong list of differentially expressed genes, a complete set of lncRNAs should be available in the reference genome.

Although, there are many studies that discovered the important roles of protein-coding genes, nevertheless, the full scenes of how the virus work inside the cell are still not clear. Therefore, we performed transcriptomic analysis for *Ae. aegypti* cells infected with DENV-1. Thus, transcriptome analysis of *Ae. aegypti* cells infected with DENV1 was performed to gain an overview of transcriptional response of both protein-coding genes and lncRNAs upon virus infection. Transcriptomic aims to identify differentially expressed protein-coding transcripts as well as lncRNAs in *Ae. aegypti* cells infected with DENV-1. All in all, systematic identification and characterization of *Ae. aegypti* lncRNAs would be essential resources for future projects. Besides, the outcomes of this study will grant a better view on the functions of protein-coding genes and lncRNAs,

On the other hand, previous study has systematically identified lncRNAs using old version of genome reference (AaegL3) (Etebari et al., 2016). Therefore, the main target of this project is to perform genome-wide identification and characterization of lncRNAs in *Ae. aegypti* genome. In order to get a high quality set of novel lncRNAs, latest reference genome of *Ae. aegypti* (AaegL5, VectorBase) was used, and rigorous identification pipeline was applied. Then, the newly predicted novel lncRNAs in *Ae. aegypti* were characterized and compared to lncRNAs of other

species. To gain deeper understanding on the roles of lncRNAs in mosquito biology, developmental transcriptome analysis of *Ae. aegypti* lncRNAs was performed in this project.

Moreover, previous lncRNAs annotation in *Ae. aegypti* is limited due to two key factors: only one type was annotated and incomplete, fragmented old version of reference genome was used. Therefore, in this study, it was hypothesized that there are more novel lncRNAs in *Ae. aegypti* yet to be discovered. To ensure accuracy in lncRNA identification, therefore, the newest genome version was used and all three classes of lncRNAs – intergenic, antisense, and intronic were annotated. Meanwhile, previous studies showed alteration in genes expression post infection with DENV-2 and Zika. Thus, it was hypothesized that the infection of Aag2 cells with DENV-1 will induce the alteration in the expression of the transcripts for both protein-coding genes and lncRNAs. In addition to that, the novel lncRNAs were added to the previously annotated lncRNAs thus, in this study the differentially expressed list of lncRNAs includes novel lncRNAs.

1.2 Objectives

The main objectives of the research are as follows:

1) To perform genome-wide annotation and characterization of lncRNAs in *Ae. aegypti* genome

2) To investigate the transcriptional response of lncRNAs and protein-coding genes in *Ae. aegypti* cells upon infection with DENV-1

CHAPTER TWO

LITERATURE REVIEW

2.1 Dengue fever

DF is an arboviral disease in tropical and sub-tropical areas. Geographical distribution of DF has expanded over the years due to the invasiveness of the mosquito vectors - *Ae. aegypti* and *Ae. albopictus* (Kraemer et al., 2015; Leon Philip Lounibos & Kramer, 2016). Symptoms of DF usually take three to fourteen days after infection to be manifested in humans (Kularatne, 2015). Infected people with DF normally suffer from headache, sudden fever, muscle-joint pain, rash, nausea, and vomiting (Chen & Wilson, 2016). However, the disease can be developed into a more dangerous stage namely, dengue hemorrhagic fever (DHF), and the symptoms include damage to the lymph and blood vessels, bleeding from the nose and gums, low level of blood platelets and blood plasma leakage (Gubler, 1998). In addition to the previously mentioned symptoms, DF can be the reason of acute liver failure in endemic countries because liver is one of the common organs that is affected by the disease (Gubler, 1998).

Aedes mosquitoes get the infection after sucking sufficient blood meal from an infected person with high viral load of DENV, whether or not that person has developed DF symptoms or completely asymptomatic (Kularatne, 2015). Once the mosquito feeds on the infected blood meal, DENV starts its journey to midgut where it replicates, and then from midgut, the virus propagates to secondary tissues such as

salivary glands (Jing & Wang, 2019; Paupy et al., 2009). The mosquito will be able to transmit DENV to another person after extrinsic incubation period (EIP), which is between eight to twelve days (Tjaden et al., 2013). The EIP usually refers to an interval from the time at which a vector acquires the infectious agent until the vector is ready to pass the infectious agent to other hosts (Tjaden et al., 2013). For DENV, once the virus is taken up by a mosquito via blood-feeding, a certain period of time is necessary for the virus to divide and replicate. The virus will then travel to the mosquito midgut and spread all over the body and reaches the salivary glands. The arrival of the virus to the salivary glands enables the transmission of DENV via blood-feeding (Bäck & Lundkvist, 2013; Jing & Wang, 2019; Kularatne, 2015). It has been reported that viral load in the blood meal, temperature alteration, and virus genotype are factors that affect the EIP (Tjaden et al., 2013). As the mosquito gets the infection, it will be able to transmit the virus throughout its life span (Lambrechts et al., 2010; Wongkoon et al., 2013).

People who live in tropical and subtropical countries, and tourists who visit tropical places where DENV exists, are under the risk of the infection. Infection for the first time with one of DENV serotypes can be mild and the patient would be cured after two weeks, but second infection with another serotype, would be lethal and causes DHF and dengue shock syndrome (DSS) (Samanta & Sharma, 2015). In severe cases of DF, patients are required to be hospitalized, and they will be under a constant surveillance doctors and nurses. Maintenance of body fluids is vital during the acute stage (Lum et al., 2014). Hospitalization is very important, and it was reported that mortality rate can be sharply decreased from 20% to less than 1% after hospitalization (*Dengue and Severe Dengue*, n.d.).

2.2 Dengue virus

DENV is a positive single stranded RNA virus (+ssRNA). It is 40-60 nm in size, and it is an enveloped spherical virus (Kuhn et al., 2002; Mukhopadhyay et al., 2005). Figure 2.1 shows an illustration of the spherical shape of DENV. There are four serotypes of DENV1- 4, and all of them can cause DF (Gintarong et al., 2018; Yung et al., 2015). It has been reported that these four antigenically different serotypes share approximately 65% of their genome sequence. Furthermore, studies also showed that infection with different serotypes can result in different clinical manifestations of the disease (Halsey et al., 2012; K.-S. Lee et al., 2015). DENV is a member of *Flaviviridae* family, and its genus is flavivirus (Kuno et al., 1998). Taxonomic classification of the virus is shown in Table 2.1.

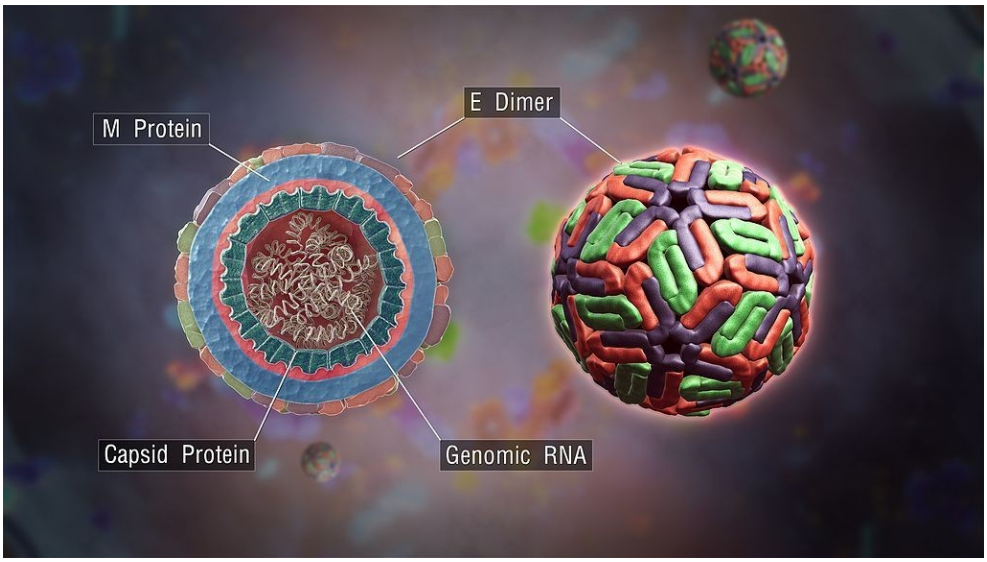

Figure 2.1 An illustration of DENV structure and its spherical shape. The image was taken and adapted from http://www.scientificanimations.com/

Table 2.1 Taxonomic rank of DENV (International Committee on Taxonomy of viruses (ICTV)**)**

Rank	Name
Unrank	Virus
Realm	Riboviria
Kingdom	Orthornavirae
Phylum	Kitrinoviricota
Class	Flasuviricetes
Order	Amarillovirales
Family	Flaviviridae
Genus	Flavivirus
Species	Dengue Virus

2.2.1 Dengue virus genome

DENV genome is around 10.7 kilobase (kb) pair (Figure 2.2). It is transcribed into ten proteins; three of them are structural and seven are non-structural (NS) proteins. DENV genome has short non-coding sequences in both 5’ and 3’ ends (Figure 2.2). Structural proteins (Membrane; M, Capsid; C, and Envelope: E) form the spherical shape of the virus. The non-structural proteins include NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5. These NS proteins have roles in virus replication in the host cells (Ashour et al., 2009; Dwivedi et al., 2017; Hannemann et al., 2013).

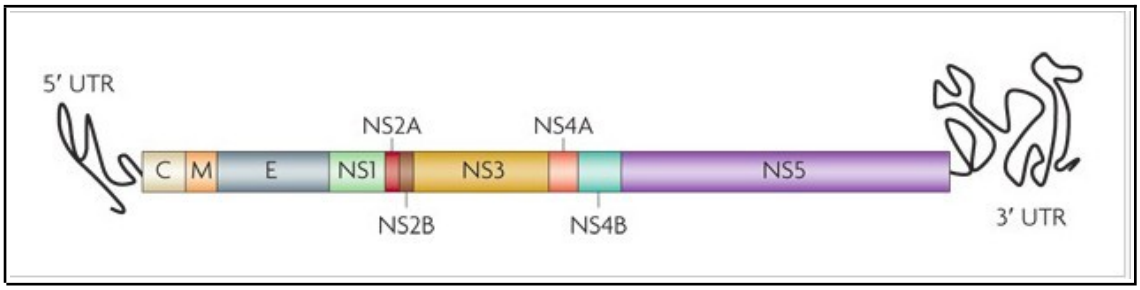

Figure 2.2 Genome of DENV which is composed of 3 structural proteins and 7 nonstructral proteins. The image was modified and adapted from Guzman et. al., 2010.

Both structural and NS proteins of DENV play important roles in virus replication, assembly and pathogenesis (Dwivedi et al., 2017; Perera & Kuhn, 2008). The E structural protein which is a dimer on the surface of the mature virus, plays a role in the attachment of the virus to the host cell (Perera & Kuhn, 2008). Meanwhile, M protein has a big role in the formation and maturation of the viral particles inside the host cells (Perera & Kuhn, 2008). On the other hand, it was found that NS proteins have important roles in immune response (Norazharuddin & Lai, 2018; Uno & Ross, 2018). One of the NS proteins, NS4a, involves in membrane attachment and enhancement of autophagy (McLean et al., 2011; Miller et al., 2007). Besides, NS4a acts as a scaffold for virus replication complex, and mutation in NS4a resulted in low level of virus replication (Lee et al., 2015; Zou et al., 2015).

Mammalian interferon (IFN) signalling pathway is critical part of the innate immune responses against viruses (Ashour et al., 2009). Type I interferon (IFN-I) is the first defense line against the infection with flaviviruses (Thurmond et al., 2018). Thus, flaviviruses have encoded a set of antagonists against this pathway in order to

succeed in infecting the host cells (Thurmond et al., 2018). NS5 is one of these antagonists, which it degrades STAT 2; an important part of type 1 interferon signalling pathway (Ashour et al., 2009). Meanwhile, NS4b is responsible for inhibiting STAT 1 pathway in innate immune system, resulting in the inhibition of interferon signalling gene (ISG) production (Muñoz-Jordan et al., 2003).

2.2.2 Dengue virus life cycle

Transmission of DENV can divided into two types: sylvatic and non-sylvatic (Holmes & Twiddy, 2003; Young et al., 2017). Sylvatic cycle involves transmission of the virus among wild animals, such as non-human primates including monkeys, and this occurs in the jungle (Guindo-Coulibaly et al., 2019). Meanwhile, non-sylvatic transmission occurs between humans and mosquito vectors in urban areas (Young et al., 2017). Previously, DENV was replicating via sylvatic cycle in Africa and Asia (Holmes & Twiddy, 2003). Meanwhile, primary life cycle occurs between human and *Aedes* mosquitoes, mainly *Ae. aegypti* (Young et al., 2017). Change from sylvatic to non-sylvatic transmission resulted in global propagation of DENV.

Viruses are obligatory intracellular parasites. They need the genome of living host cell to produce their nucleic acid and proteins (Ryu, 2017). Life cycle of almost every virus including DENV involves at least three major stages, namely entry, genome replication, and releasing from host cells (Kuhn et al., 2002; Ryu, 2017). Life cycle of DENV begins with the recognition and binding to host receptors followed penetrating into the cytoplasm using receptor-mediated endocytosis

(Agrelli et al., 2019; Schaar et al., 2008). Translocation of DENV particles inside cells is mediated by the endosomes (Mukhopadhyay et al., 2005).

Next, the acidic environment inside the host cell causes the virus envelop to fuse with the membrane of the endosomes. This leads to the release of virus RNA genome into the cytoplasm (Mukhopadhyay et al., 2005). The acidity is due to the influx of hydrogen ion (H+) inside the endosome (Rodenhuis-Zybert et al., 2010; Schaar et al., 2008; Zaitseva et al., 2010). Then, virus RNAs use the machinery of the host cell resulting in the translation of viral RNAs into polyproteins. This translation process is performed by the host ribosomes, and it takes place in rough endoplasmic reticulum (RER) (Acosta et al., 2008; Agrelli et al., 2019; Schaar et al., 2008). After that, a single polyprotein is cut into ten proteins to form both structural and NS proteins. The new RNAs are then covered by C proteins to form nucleocapsid (Acosta et al., 2008). Then, the nucleocapsids are surrounded with E and M proteins in RER forming immature virions (Acosta et al., 2008; Mukhopadhyay et al., 2005; Shrivastava et al., 2011). These immature virions then travel to Golgi apparatus where they will be matured by the cleavage of the precursor M protein (Mukhopadhyay et al., 2005). After that, the mature infectious viral particles are liberated from the host cell as shown in Figure 2.3.

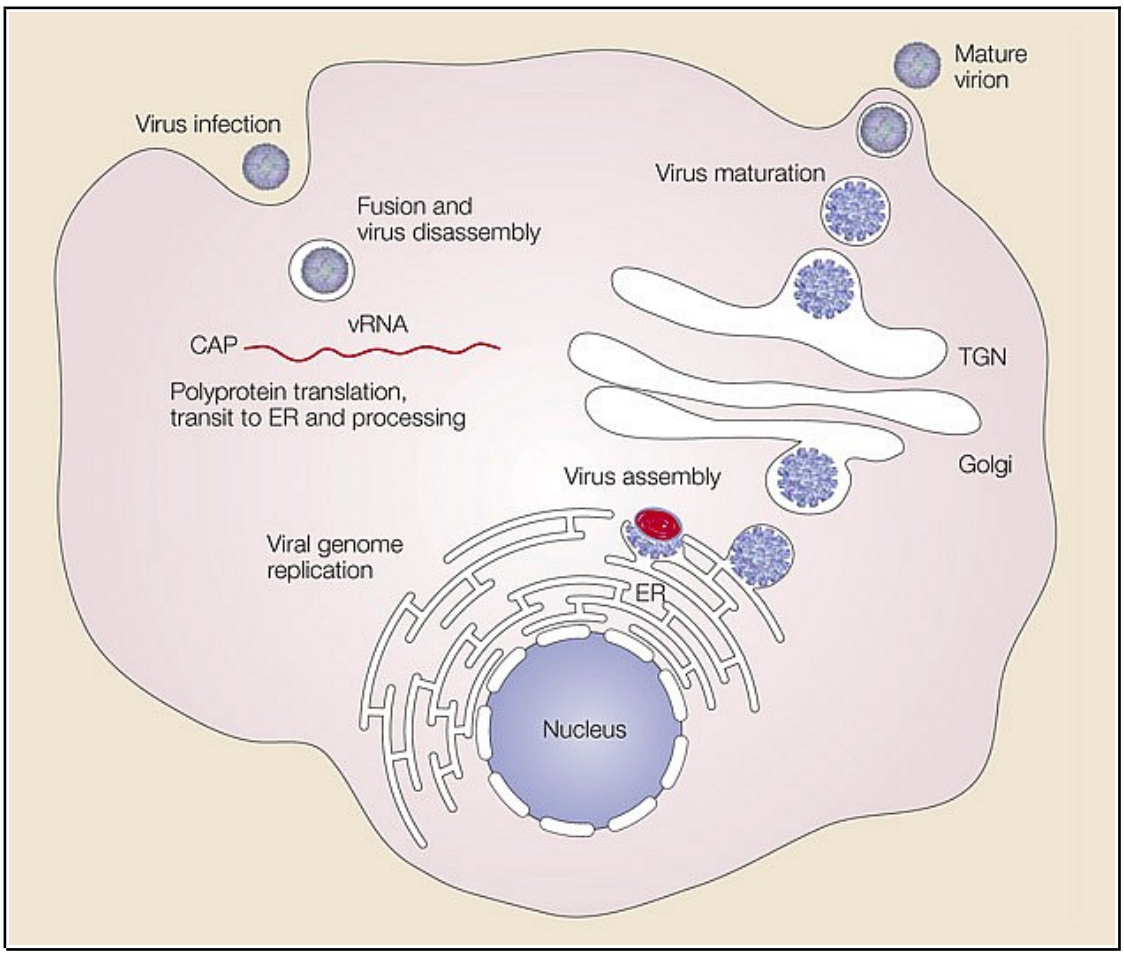

Figure 2.3 DENV life cycle that includes the three main stages namely; entry, genome replication and exit. The image was adapted from Mukhopadhyay et al., 2005.

2.3 Epidemiology and geographical distribution of dengue virus

It is a well known fact that the expansion of the mosquitoes' habitats resulted in the increment of DF cases. DENV is responsible for dengue cases in several countries in the world (Higa, 2011; Selvaretnam et al., 2016). The virus was originally found in monkeys, and it spilled over to the human 800 years ago (Pollett et al., 2018). Moreover, in the last century, DENV was restricted to a narrow zone in Africa and south-east Asia till the mid of 20th century (Shi et al., 2016). Then, it spread via nautical shipping to other regions (Semenza et al., 2014). The geographic expansion of DENV is correlated with the geographical distribution of their vector mosquitoes: *Ae. aegypti* and *Ae. albopictus*.

The expansion of *Aedes* mosquitoes can be attributed to several factors such as global warming, urbanisation, and international travel (Gubler, 1998). Figure 2.4 shows regions in the world that are susceptible for dengue epidemic

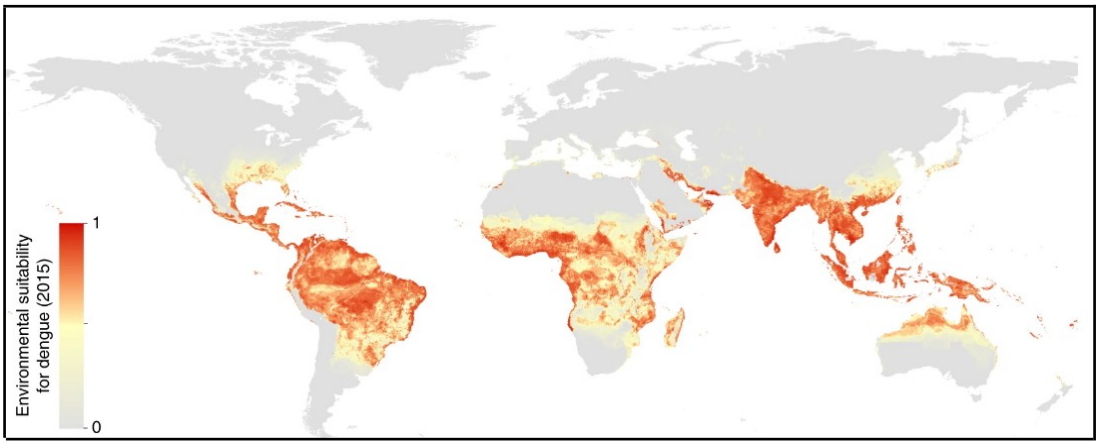

Figure 2.4 A map displaying environmental suitability for dengue in 2015. Most countries with high suitability are in the tropical areas. The distribution was predicted to expand due to global warming. The image was adapted and modified from (Messina et al., 2019)

At least 100 countries were recorded as a dengue endemic such as those in Africa, Western Pacific, the Americas, Southeast Asia, and east of the Mediterraneans (Harrington et al., 2013). Previous studies reported that there were more than 100 millions reported patients that displayed symptoms of DF every year (Bhatt et al., 2013). Moreover, within the 100 million reported cases, half million of them experience more severe DF such DHF or DSS (Bhatt et al., 2013). It has been observed that most of outbreaks of dengue were recorded in China, Singapore, and Malaysia after 2010 (Guo et al., 2017). On the other hand, the occurrence of dengue outbreaks was lower in Europe than in Asia, and among European countries, France and Portugal were the most affected countries (Guo et al., 2017).

The first dengue outbreak in Europe was recorded in 2012 in Atlantic Island of Madeira (Lourenço & Recker, 2014). However, it was claimed that those cases were not infected in the island, but they got during vacation in some of south America countries (Lourenço & Recker, 2014). *Aedes aegypti* and *Ae. albopictus* found to be responsible for transmitting the disease in these countries which had dengue outbreaks (Lwande et al., 2020). Although the outbreaks were recorded in certain European countries, the number of cases were not high (Lourenço & Recker, 2014). Thus, this might indicate that the patients didn't get the infection in European countries, and the patients were infected during their vacations in tropical or sub-tropical countries (Lourenço & Recker, 2014).

The first documented case of the severe form of dengue (DHF and DSS) were seen in Thailand and Philippines during the 1950s (Halstead & Cohen, 2015). Up to 1960s, the number of reported DF cases was lower than 1000 cases. However, the number reached to 1 million cases between 2000 to 2007 (Guo et al., 2017). The

severe form of dengue subsequently spread throughout Southeast Asia and expanded globally (Bhatia et al., 2013). Thus, it is important to control the vector mosquito in order to control DENV spreading.

2.3.1 Epidemiology of dengue virus in Malaysia

The first reported DF case in Malaysia occurred in the state of Penang in 1901, and it was believed that this case brought from other country (Skae, 1902). It was believed that *Ae. aegypti* was introduced in Asia in 1950s; thus, most probably that all cases in Asia reported before 1950 was brought by travellers (Powell et al., 2018). However, *Ae. aegypti* is responsible for most of the cases in Asia starting from 1950, and it was the main vector from that date onward (Ahmad et al., 2018). The first outbreak in Malaysia, resulted in 54 deaths out of one thousand infection cases recorded in 1973 (Wallace et al., 1980).

All DENV serotypes are detected in Malaysia, and they circulate every few years (Pang & Loh, 2016; Suppiah et al., 2018). For instance, DENV1 was found to be responsible for outbreak in Sarawak 2012, and was dominant in Malaysia in the period of 2015-2018 (Pang & Loh, 2016; Suppiah et al., 2018). Besides, it was found that DENV serotypes co-circulated in Malaysia, which means that, even though one serotype dominates, the other serotypes also can cause infections (Pang & Loh, 2016). In 2010, it was reported that DENV1 was responsible for 43.9% of cases in Negeri Sembilan compared DENV2, DENV3, and DENV4 which were responsible for 20%, 31.1% and 5% respectively (Ghazali et al., 2012). All races are equally

susceptible to DF; however, most cases are among children due to their weak immunity (Bhatia et al., 2013).

DF is a burden to people and the economy of a country including Malaysia. Many efforts were done to develop treatment or vaccines for DF in order to save the citizens' lives and economy (Gubler, D. J., 2012). One of the challenges is to develop vaccine that is efficient for all four DENV serotypes. Unfortunately, there is no proven efficient vaccine available; thus, the control of the mosquito vectors is vital to stop the disease (Izmirly, A. M., et. al., 2020).

2.4 *Aedes aegypti* the main vector of dengue virus

Aedes aegypti mosquito (Figure 2.5) is the main vector of many harmful viruses such as, ZIKV, chikungunya, and DENV, and it is considered as one of the most dangerous mosquitoes in the world (Bäck & Lundkvist, 2013; Lwande et al., 2020; Messina et al., 2019). Figure 2.5 shows an image of *Ae. aegypti* adult mosquito. *Aedes aegypti* is a member of the family *Culicidae* and the genus *Aedes* (Evenhuis, 1989). Table 2.2. shows the classification of the mosquito (Evenhuis, 1989). There are two subspecies of this mosquito; *Ae. aegypti formosus* which is believed to be the ancestor of the second subspecies which is, *Ae. aegypti aegypti* (Powell & Tabachnick, 2013). *Aedes aegypti aegypti* has been adapted to human habitat, and it can be found in urbanic and semi-urbanic areas (Ponlawat & Harrington, 2005). On the other hand, *Ae. aegypti fermosus* represents the wild type

of the mosquito, and it is found in forest and vegetated ecotones in sub-saharan in Africa (L. P. Lounibos, 1981).

Nectar is the main food for *Ae. aegypti*; however, in order to produce eggs, the female mosquitoes need sucking the human or other mammals blood (Nikbakhtzadeh et al., 2016). Thus, female *Ae. aegypti* is responsible for transmitting the viruses not the males. *Aedes aegypti* is considered as highly anthropophilic where it prefers human blood, although it sucks blood from other mammals and birds (Ponlawat & Harrington, 2005). Furthermore, Tabachnick claimed that *Ae. aegypti* is the most mosquito that prefers to suck human blood (Tabachnick, 1991).

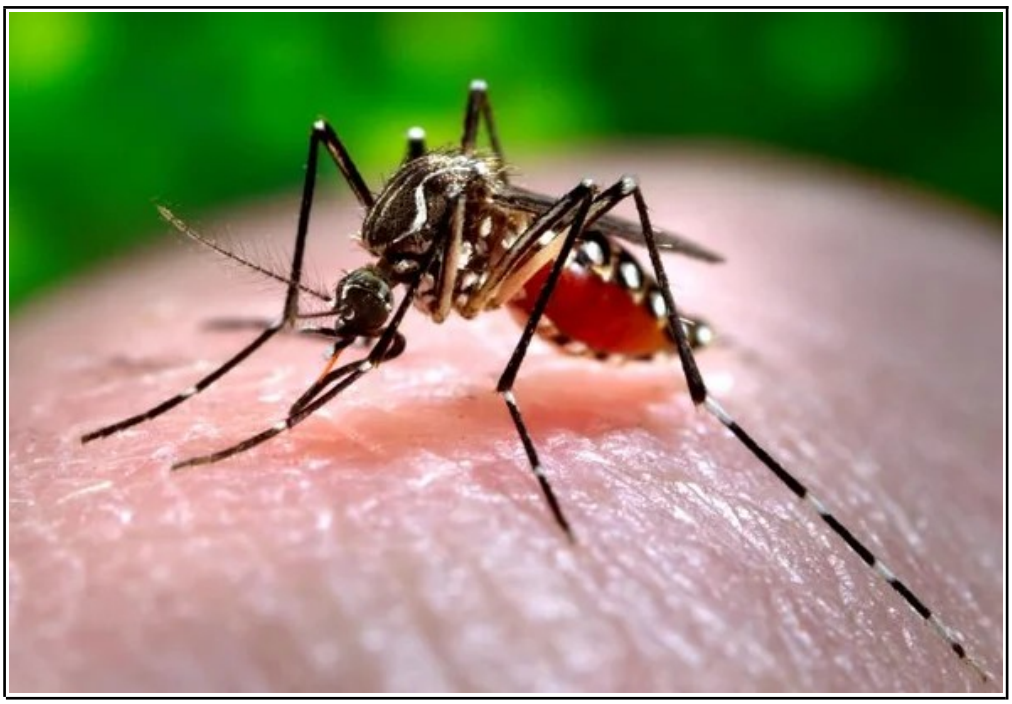

Figure 2.5 *Ae. aegypti* adult mosquito. Image was taken from Scientific American Website (https://www.scientificamerican.com/article/mosquito-borne-disease-could-threaten-half-the-globe-by-2050/)

Table 2.2 Linnaeus classification of *Ae. aegypti* (Evenhuis, N. L., & Gon III, S. M., 2007)

Taxon	Nomenclature
Kingdom	Animalia
Phylum	Arthropoda
Class	Insecta
Order	Diptera
Family	Culicidae
Genus	Aedes
Species	aegypti
Scientific name	*Aedes aegypti*
Subspecies 1	Wild species: *Ae. aegypti formosus*
Subspecies 2	Domestic species: *Ae. aegypti*

Aedes aegypti is active during daytime, early morning and before sunset (Nikbakhtzadeh et al., 2016). It is able to transmit the virus to humans after it sucks blood from infected individuals followed by one week of EIP (Bäck & Lundkvist, 2013). Once infected, *Aedes* mosquito will harbor DENV throughout its whole life (Bäck & Lundkvist, 2013). Despite being infected with DENV, *Aedes* mosquitoes do not develop fatal symptoms due to its vigorous host immune response (Miesen, Ivens, et al., 2016; Miesen, Joosten, et al., 2016; Sim & Dimopoulos, 2010). The response of the mosquito immune system toward the infection decreases the impact of the virus although it is not destroyed completely (Sánchez-Vargas et al., 2009).

2.4.1 *Aedes aegypti* life cycle

Aedes aegypti life cycle has four distinguished stages, egg, pupae, larvae and adult (Figure 2.6). In normal conditions, it takes around one week to complete one life cycle, from egg to adult (Marinho et al., 2016). However, in harsh environment such as dry seasons, it may take more than six months to complete one life cycle (Marinho et al., 2016). Female mosquito lays eggs 4-5 times, and produces up to 1000 eggs during its life span, an average of 200 eggs each time (Marinho et al., 2016). The eggs are black, and elongated oval in shape. The adult females can live up to 10–35 days in average, while males can live for three to five days, based on environmental conditions such as temperature (Getachew et al., 2015; Marinho et al., 2016).

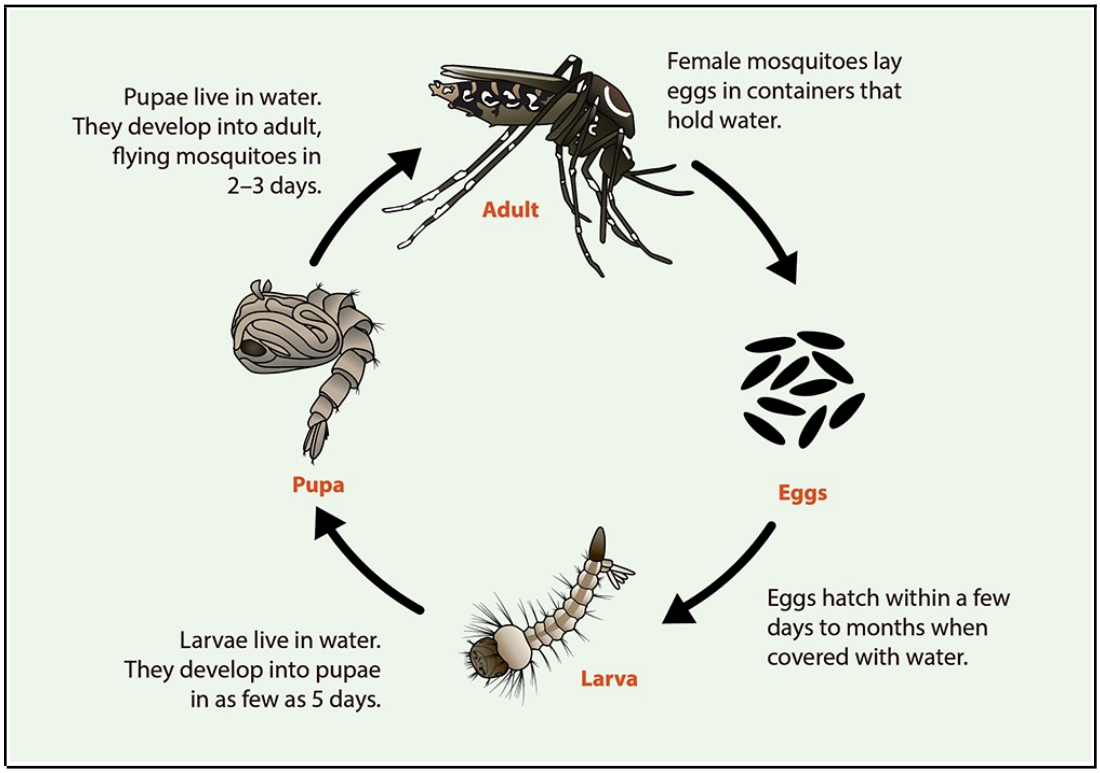

Figure 2.6 Life cycle of *Ae. aegypti.* Image was taken from Central for Diseases Control (CDC) website (https://www.cdc.gov/mosquitoes/about/life-cycles/aedes.html).

2.4.2 *Aedes aegypti* cell line

Cell lines are cultures of animal cells that can be continuously propagated in the lab (Bols et al., 1994). They are initially originated from primary cell cultures that are generated directly from the cells, tissues, or organs of animals (Bols et al., 1994; Kaur & Dufour, 2012). Unlike primary cell cultures, cell lines survive after multiple rounds of passaging, and this is the reason why they are called immortal (Kaur & Dufour, 2012). Cell lines have made a revolution in molecular biology due to many reasons such as cost-effective, continuous propagation, easily culturing and homogeneity (Kaur & Dufour, 2012). Cell lines are the most cost-effective and ethically viable resources for performing scientific research. They were idealized for their phenotypic and genotypic homogeneity after culturing them several time (Kaur & Dufour, 2012).

In the last decades, the viability of cell lines explain why nearly all scientific developments have been based on cell lines (Mirabelli et al., 2019). They are being used for antibody production, vaccine development, drug testing, genetic function studies, cancer cell growth studies and artificial skin development (Kaur & Dufour, 2012; Mirabelli et al., 2019). New cell lines are being continuously developed to discover biomarkers for cancer, skin diseases, lung dysfunctions and better visualization of tumor cells (Lizbeth et al., 2017).

Understanding mosquito immune responses can provide valuable tools for the development of novel mosquito control strategies. Continuous insect cell lines have been established to study insect innate immunity, and they have been used as research tools because they constitute more homogeneous, sensitive, and

reproducible systems (Walker et al., 2014). More recently, Aag2, an embryonic-derived *Aedes aegypti* cell line, began to be commonly used as a model for studies of mosquito immunity (Dubey et al., 2017; Göertz et al., 2019; Miesen, & Ivens, et al., 2016; Scott et al., 2010; Sim & Dimopoulos, 2010; Varjak et al., 2017). It has been found that Aag2 cells are immune-competent, which can mimic the immune response of the whole adult Ae. aegypti mosquitoes (Scott et al., 2010; Sim & Dimopoulos, 2010).

Several studies reported that Aag2 cells are immune-competent, and they are good model to study virus-host interaction and insect immunity (Miesen & Ivens, et al., 2016; Sim & Dimopoulos, 2010). Moreover, previous researches reported that the cells showed immune response against Gram-positive and negative bacteria, fungi, DENV, and Sindbis virus, in a very similar way to that described for whole mosquitoes; making them a good model for insect immune studies (Dubey et al., 2017; Göertz et al., 2019; Miesen et al., 2015; Miesen, & Ivens, et al., 2016; Scott et al., 2010; Sim & Dimopoulos, 2010; Varjak et al., 2017).

For example, it was reported that infection of Aag2 cells with DENV2 resulted in the activation of Toll pathway and the downregulation of numerous genes involved in immune signalling and anti-microbial peptides (AMPs) (Sim & Dimopoulos, 2010). Gram-negative bacteria such as *Escherichia coli* grew better when co-cultured with DENV serotypes in Aag2 cells (Sim & Dimopoulos, 2010). This finding suggests that DENV2 infection somehow inhibits the immunity of Aag2 cells, resulting in the decreased of AMPs' production inside cells to combat bacterial infection (Sim & Dimopoulos, 2010).

Furthermore. Aag2 cells were also used to study small RNA mediated immunity against viruses (Miesen et al., 2015; Miesen, Ivens, et al., 2016; Miesen, Joosten, et al., 2016). Viral infection in *Aedes* mosquitoes triggers the expression of many classes of small RNAs (both viral or host-derived) such as small interfering RNAs (siRNAs), PIWI-interacting RNAs (piRNAs), and microRNAs (miRNAs) (Miesen et al., 2015). Viral-derived siRNAs are among the key components of antiviral RNA defence to limit virus replication in insects and invertebrates (Miesen et al., 2015; Miesen, Ivens, et al., 2016; Miesen, Joosten, et al., 2016). Besides siRNAs, piRNAs had been shown to be involved in immunity against exogenous viruses in *Aedes* mosquitoes. In *Ae. aegypti,* the number of genes in PIWI family has expanded compared to other organisms (Miesen et al., 2015). It was shown that Piwi5 and Ago3 are essential in the production of virus-derived piRNAs from Sindbis virus (Miesen et al., 2015). The biogenesis of piRNAs from Sindbis virus ultimately reduces the replication of this virus in *Ae. aegypti* cells (Miesen et al., 2015). Taken together, previous studies suggest that *Ae. aegypti* cell line (Aag2) is a good model system to study insect immunity, small RNA-mediated immunity, and virus-host interaction.

2.5 RNA-sequencing

The core molecular biology dogma explains the transfer of genetic knowledge from the genes to functions of cells and organisms. This involves a two-step process: first, transcription of DNA into RNA, by RNA polymerase enzymes; second, translation of messenger RNA (mRNA) into protein (Blumenberg, 2019). Transcriptome is the result of the transcription. It is the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell. Transcriptome analysis is known as transcriptomics (Blumenberg, 2019). RNA-sequencing (RNA-seq) is a technique that is capable of examining the quantity and sequences of RNAs up to a single nucleotide resolution in a sample (Blumenberg, 2019). Through RNA-seq, one is able to analyze genome-wide transcriptome of gene expression pattern encoded within a particular genome.

Transcriptome analysis allows scientists to connect the information encoded in the genome with its expression level. RNA-seq provides information on which genes are expressed or inactivated in a given condition. This allows scientists to deeply understand the biology of a cell in response to certain changes or diseases (Blumenberg, 2019). For example, transcriptome analysis can highlight cells, tissues or organs in which a gene is expressed, which provides insights on its role in those tissues. Besides expression level, RNA-seq also captures information on alternative splicing – a process in which a single gene produces more than one mRNA transcripts (Z. Wang et al., 2009). Analysis of alternative splicing is important because certain diseases or cancer trigger different splicing events of certain genes (Z. Wang et al., 2009).