Deanship of Graduate Studies

Al- Quds University

Amplicon-based Next generation sequencing for Identification of Sand fly and leishmania parasites

Mohammad Hashem Hosen Altarade

M.Sc. Thesis

Jerusalem- Palestine

1440 \ 2019

Amplicon-based Next generation sequencing for Identification of Sand fly and leishmania parasites

Prepared By:

Mohammad Hashem Hosen Altarade

B.SC in Medical Laboratory science

Al-Quds University

Palestine

Supervisor:

Prof. Ziad Abdeen

Dr. Abed Nasereddin

A thesis Submitted in Partial fulfillment of requirements for the degree of Master of Biochemistry and Molecular Biology from Biochemistry and Molecular Biology Department Deanship of Graduate Studies/ Al- Quds University

1440 \ 2019

Al Quds University

Deanship of Graduate Studies

Biochemistry and Molecular Biology Department

Thesis Approval

Amplicon-based Next generation sequencing for Identification of Sand fly and leishmania parasites

Prepared By: Mohammad Hashem Hosen Altarade

Registration No: 21512256

Supervisor: Prof. Ziad Abdeen

Co-supervisor: Dr. Abed Nasereddin

Master thesis submitted and accepted, Date: 18/5/2019

The names and signatures of the examining committee members are as follows.

1-Head of Committee: Prof. Ziad Abdeen signature .....

2- Co-Supervisor: Dr. Abed Nasereddin signature .....

3-Internal Examiner: Dr. Ibrahim Abbasi signature .....

4-External Examiner: Dr. Robin Abu Ghazaleh signature .....

Jerusalem- Palestine

1440 \ 2019

Dedication

To my mother and father...

To my brothers and sisters...

To my friends...

To all beloved people who supported, assisted and encouraged me.

Mohammad Hashem Hosen Altarade

Declaration:

I certify that this thesis submitted for the degree of Master, is the results of my own research, except where otherwise acknowledged, and that this study (or any part of the same) has not been submitted for a higher degree to any other university of institution.

Signed: 

(Mohammad Hashem Hosen Altarade)

Date: 18/5/2019

Acknowledgements

All thanks be to Allah

I would like to thank my thesis adviser Dr. Abed Nasereddin for his supervision directions and assistance discussion and kindness throughout this study.

My thanks also go to all those who help me at Alabraj Research lab especially for Dr. Suheir Ereqat and Ahmad Abdelkader. Especially thanks to Mr. Samer Sawalhah, Palestinian entomology department Ministry of health. I would like also to acknowledge Al-Quds University for allowing me to get a master degree in Biochemistry and Molecular biology.

Also, I would like to thank my family and my friends for their support.

Mohammad Hashem Hosen Altarade

Table of Contents

Dedication ..... **I**

Declaration: ..... **i**

Acknowledgements ..... **ii**

List of abbreviations ..... **vii**

Abstract ..... **viii**

Chapter One: Introduction **1**

1.1. Leishmaniasis 1

1.2. Leishmaniasis worldwide 2

1.3. Leishmaniasis in Palestine 3

1.4. Leishmaniasis treatment. 4

1.5. Leishmania Parasite 5

1.6. Parasite Life cycle 6

1.7. Sand Fly 7

1.8. Molecular technology 9

1. 9. Study Objectives 9

Chapter Two: Materials and Methods **10**

2.1. Sand fly sample collection 10

2.2 DNA Extraction 10

2.3 Design of NGS specific primers 11

2.4 Amplicon PCR Amplification 13

2.5.1 Leishmania primers optimization 13

2.5.2 Sand fly primer Optimization 15

2.5.3 PCR products purification and NGS 15

2.6 Multiplex PCR and Sand fly primers as limit factor 16

2.7 Selection of Leishmania species specific probes 18

2.8 Optimizing the quality copy number for Leishmania 19

2.9 Analysis of field Sand flies: 20

2.10 Bioinformatics analysis 22

Workflow design: 22

Sand fly virtual Probes 23

2.11. Statistical analysis and multiple sequence alignment 24

Chapter Three: Results **25**
3.1 PCRs optimization 25
3.2 Optimization with different global worldwide *Leishmania* species 30
3.3 Species specific probes design 31
3.4 Detection of minimal Leishmania ITS-PCR copy number 33
3.5 Field Samples 34
3.6 NGS data and statistical analysis 36
Chapter Four: Discussion **40**
Study limitation 44
Recommendations 44
REFRENCES **45**
61الملخص .5

List of Figures

Figure 1.1: The epidemiology of leishmaniasis diseases worldwide. 3
Figure 1.2: Geographical distribution of leishmaniasis disease in the Palestinian West Bank region between 1994 and 2015. 4
Figure 1.4: Sand fly vector of *Leishmania* spp.....8
Figure 2.1:18S, ITS1 primer composition and adapter sequence strategies work. 12
Figure 2.2: Serial dilution for Leishmaia DNA. 14
Figure 3.2: PCR analyses of 18S rRNA gene of different sandflies used as in appendix 1. 27
Figure 3.3: ITS1 with 403 bp band size, and 279 bp for 18S that measured by TapeStation Analysis Software. 27
Figure 3.4: DNA sequencing for some Sand fly species as representative identification probes. Red dot: identical nucleotide, blue and black is the differences, dash is indels 28
Figure 3.6: ITS1 result of different Leishmania samples used as in appendix 3.....31

List of tables

Table 1.1: Leishmania species that cause Old world leishmaniasis. 5
Table 2.1: ITS1, 18S Amplicon PCR forward and reverse primers sequence.....13
Table 2.3: PCR reaction mix for 18S rRNA. 15
Table 2.4: PCR indices mixture for barcoding each amplicon. 16
Table 2.5: Multiplex PCR reaction mix for ITS1 and 18S rRNA. 17
Table 2.6: ITS1 PCR mixture. 18
Table 2.7: PCR reaction mixture. 19
Table 2.8: 18S rRNA PCR reaction mixture. 20
Table 2.9 : multiplex PCR mixture. 21
Table 2.10: specific sequence virtual probes for sand flies. 23
Table 2.11: Species specific sequence virtual probes for Leishmania. 24
Table 3.1: Next generation sequencing analysis result. 29
Table 3.2: NGS analysis result for eight Leishmania samples. 32
Table 3.5: Cross tabulation between NGS method for Phlebotomus genera and traditional method result. 37
Table 3.6: Cross tabulation between NGS method for Sergentomyia genera and traditional method result. 37
Table 3.7: Frequency of sand flies that have leishmania parasite.....38
Table 3.8: Cross tabulation between NGS and traditional method result for Leishmania 39

List of appendices

Appendix 1: Different reference sand fly species used in the study. 50
Appendix 2: Different reference sand fly species with their gene bank accession numbers used in the study. 51
Appendix 3 : Different Reference leishmania species with WHO codes used in the study. 52
Appendix 4: NGS analysis result for 173 sand flies samples. 53
Appendix 5: work flow for NGS data analysis. 60

List of abbreviations

WHO	World Health Organization
CL	Cutaneous leishmaniasis
MCL	Mucocutaneous leishmaniasis
VL	Visceral leishmaniasis
L	Leishmania
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
dmPCR	Direct Multiplex PCR
rRNA	Ribosomal ribonucleic acid
NTD	Neglected tropical diseases
DCL	Diffuse cutaneous leishmaniasis
PKDL	Post-kala-azar dermal leishmaniasis
LCL	Localized cutaneous leishmaniasis
Spp	Species
°C	Degree Celsius
ITS1	Internal transcribed spacer 1
SFNGSF	Sand fly next generation sequencing Forward primer
SFNGSR	Sand fly next generation sequencing Reverse primer
H	Hour
mA	Milliamps
NGS	Next generation sequencing
UV	Ultraviolet
LRU	Leishmaniasis Research Unit
TE	Tris Ethylenediaminetetraacetic acid
EB	elution buffer

Abstract

Leishmaniasis is a disease caused by *Leishmania* protozoa. There are different disease forms of leishmaniasis affecting people. The disease is transferred by the bite of infected sandflies. Identification of the *Leishmania* and sand fly species, especially in endemic areas, is important for Leishmaniasis disease control. Sand flies consist of more than 500 species, but only a few are medically important. Up until now many studies used the traditional method for sand fly species identification which relies upon morphological taxonomy of the phlebotomine Sandflies. This method has many disadvantages, for example, it requires expert entomologists to differentiate between the morphological features of sand fly species to avoid erroneous classification. Furthermore, special storage conditions for samples are needed, and the process is time consuming when dealing with large sample sizes, while delicate handlings during dissection is necessary. Until know no vaccines or safe drugs to prevent leishmaniasis infection are available, but by improving the system of sand fly and *Leishmania* parasite identification, progress in leishmaniasis disease control can be achieved.

The aim of the current study was to differentiate between the most common species of sand fly and to detect *Leishmania* DNA within the sand fly for species identification with high specificity and sensitivity using next generation sequencing. Sand flies (171) were collected from Tubas district, northern Palestine, using CDC light traps. Genomic DNA was extracted from all of them, and universal a multiplex PCR assay was setup up for sand fly 18S and Leishmanial ITS1 gene amplification. The PCR was designed to allow subsequent use of NGS Illumina platform adaptors for DNA sequencing. International reference strains of *Leishmania* and Sandflies were used in NGS system optimization for the NGS date were analyzed using galaxy online bioinformatics program, which showed the system ability to identify all reference (9) and collected sandflies (171) including the two genera: *Phlebotomus* (94.1%) and *Sergentomyia* genera (5.9%). *Phlebotomus sergenti* sand fly was the dominant species in our collection (86%).

The results were in accordance to the classical microscopic method with p value <0.001. *L. tropica* was identified in (8/171) 4.7% in the collected sandflies. Previously it was reported that Tubas is endemic region for *L. tropica* parasites.

In conclusion, the method is able to perform satisfactory high-throughput screening in ecological samples. These results will help in detecting the transmission of several

potential vectors that vary in their spatial and geographical distribution, which could explain the high prevalence of Leishmaniasis cases in specific endemic regions.

Chapter One:

1. Introduction

1.1. Leishmaniasis

Leishmaniasis is a Neglected Tropical Diseases (NTD) caused by a parasite and spread by the bite of infected sand flies. There are three major clinical forms for Leishmaniasis disease, cutaneous leishmaniasis (CL), visceral leishmaniasis (VL) and mucocutaneous leishmaniasis (MCL) (Steverding, 2017; Sundar & Rai, 2002). The most common type and least fatal form is CL, it is characterized by ulcerative skin lesions on site of insect bite, and it is caused by *Leishmania major, L. tropica, L. aethiopica, L. peruviana, L. guyanensis, L. panamensis, L. mexicana, L. braziliensis, and L. amazonensis* (Salam, Al-Shaqha, & Azzi, 2014).

The second form is visceral leishmaniasis or kala-azar in Asia, the most severe form and mostly fatal in developing countries if untreated (Desjeux, 2001). Africa, mainly Ethiopia and East Africa like Kenya, Sudan, Uganda have the highest number of VL caused by *L. infantum.* (Berman, 2006). This type caused by *L. donovani* and *L. infantum* parasites. The parasite spread to internal organs in infected patient like the liver, spleen, and bone marrow. Clinical symptoms of VL include fever, splenomegaly, hepatomegaly, progressive anemia, substantial weight loss, pancytopenia, and increase levels of a certain

immunoglobulin in the blood (Hypergammaglobulinemia) which complicated by serious infections (Sundar & Rai, 2002). The disease is fatal if not early diagnose and treated well.

Mucocutaneous Leishmaniasis (MCL) is an uncommon form of the disease and it's similar to the cutaneous form but MCL is characterized by destruction of mucous membranes of the nose and mouth, and the symptoms usually need time to appear on patient between one and five years after the skin lesions. It is caused by *L. panamensis*, and *L. braziliensis* (Salam et al., 2014).

1.2. Leishmaniasis worldwide

There are around 12 million patients suffer from leishmaniasis in the world, according to the World Health Organization – WHO, and more than 300 million people are at risk of infection (Fokialakis et al., 2007). In each year 200,000-400,000 VL cases worldwide reported and > 90% of these cases found in six countries: India, Bangladesh, Brazil, Ethiopia, Sudan, and South Sudan. The number of deaths reach to 10% after diagnosed with VL (Alvar et al., 2012). Half of these cases found in India alone (Sundar et al., 2001). The estimated annual incidence of CL is around 0.7-1.3 million new cases in worldwide and more than 95% of cases are found in the Mediterranean basin, the Middle East, Americas and Central Asia Figure (1.1). In 2015 over two-thirds of new cases found Especially in Afghanistan, Brazil, Algeria, Colombia, the Islamic Republic of Iran and Syria (Alvar et al., 2012; Organization, 2018,March 14). Multiple factors make the disease out of control including lack of effective vaccines, increasing travels between countries, difficulties in vectors control, lack of awareness, and drug increasing resistance (Khraiwesh et al., 2016).