GENOME PLASTICITY AND GENETIC EXCHANGE

IN *LEISHMANIA TROPICA*

Stefano Iantorno

Gonville and Caius College

This dissertation is submitted for the title of

Doctor of Philosophy

at the University of Cambridge

June 30th, 2015

Declaration

I hereby declare that this dissertation is entirely the product of my own work and contains nothing that is the product of work done in collaboration with others except when explicitly stated here and in the main text.

The sequence data that was used in this thesis was produced by the core Sequencing production teams at the Wellcome Trust Sanger Institute.

None of the work presented has been submitted for the purpose of obtaining another degree. This dissertation does not exceed 60,000 words in length, as required by the School of Biological Sciences.

Stefano Iantorno, June 2015

To my parents

ACKNOWLEDGEMENTS

Although the work presented in this dissertation is my own, this thesis is the product of concerted efforts by many different people. I thank my supervisors, David Sacks, Michael Grigg, Matt Berriman, and James Cotton for their unfailing support and critical insight in both the experimental and analytic portions of my thesis, and for always being able to provide a high level perspective to guide the direction of the research. Among those who helped carry this research project forward at NIH, I am greatly indebted to Audrey Romano and Ehud Inbar for their humour and for their assistance with a range of experimental procedures, and to Kim Beacht for assistance with the mouse work. Phillip Lawyer's wisdom provided countless opportunities to deepen my knowledge of medical entomology and his technical skills proved essential in maintaining the laboratory colonies for the sand fly feeding assays. Kashinath Ghosh also provided assistance with the sand fly colonies and sand fly midgut dissections. Alain Debrabant offered crucial assistance with transfection procedures and with genetic modification of parasite strains. At the Wellcome Trust Sanger Institute, Caroline Durrant's firm grasp on statistics was essential to some of the more complicated analyses. Adam Reid provided guidance with the RNA-seq analyses. Mandy Sanders oversaw all sequencing procedures and facilitated communications with the core sequencing teams. All of the analyses performed in this thesis would not have been possible if not for the essential work done by Alan Tracey and Karen Brooks in the pathogen genome finishing team

improving the *L. tropica* reference assembly. Lastly, I thank the NIH-OXCAM program for generous funding and for providing excellent career development opportunities throughout my graduate studies.

SUMMARY

Leishmania is a genus of unicellular eukaryotic parasites responsible for a wide range of human diseases, from cutaneous (CL) and mucocutaneous leishmaniasis (MCL) to life-threatening visceral leishmaniasis (VL). *Leishmania tropica* is responsible for significant CL in endemic areas in North and East Africa, the Middle East, and the Indian subcontinent, and has also been associated with a variant form of VL called viscerotropic leishmaniasis. Significant heterogeneity has been observed in *L. tropica* in both clinical course of disease and in response to treatment, and published data suggests there is great genetic diversity within this species. RNA-seq analysis of 12 clinical isolates of *Leishmania tropica* revealed considerable intraspecific differences in gene expression. Comparison with whole-genome sequence data generated from the same 12 isolates using a new reference genome assembly suggests that most variation in gene expression is explainable by variation in copy number at the level of individual genes, or at the level of whole chromosomes. Most field isolates appear to be near diploid, but some degree of aneuploidy is seen in all isolates. Cloning of single cells from 4 of these isolates showed variable ploidy within the same clinical isolate, a condition that in *Leishmania* has been called mosaic aneuploidy. The most significant differentially expressed genes in this set of isolates code for membrane-bound transporter proteins, which are known to be involved in uptake of nutrients and drug compounds from the extracellular environment. We identify copy number variation in these genes suggesting that a certain degree of plasticity is observed in natural

populations of *Leishmania*, creating the conditions necessary for rapid downregulation or upregulation of different transporter proteins over a limited number of mitotic generations in the presence of environmental stressors. Such an evolutionary phenomenon could be important in mediating decreased susceptibility to drug treatment in endemic areas. To further understand how such large genetic variation can be generated and the role of genetic exchange in shaping the genomic landscape in this important pathogen, we have carried out a controlled laboratory cross between one isolate collected in Israel and one collected in Lebanon. Ten hybrid lines were recovered from crosses we performed in sand flies. The present study provides the first in-depth, complete description of structural genome changes and recombination occurring during hybridization in an artificial cross of *Leishmania tropica*. The implications of this structural variation for parasite evolution in natural populations in response to drug pressure due to increased elimination efforts will be discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Amplified fragment length polymorphism	AFLP
Bayesian information criterion	BIC
Copy Number Variant	CNV
Cutaneous leishmaniasis	CL
Differentially expressed	DE
Discriminant analysis of principal component	DAPC
False discovery rate	FDR
Fluorescent *in situ* hybridization	FISH
Generalized linear model	GLM
Hardy Weinberg	HW
Human African trypanosomiasis	HAT
Human leukocyte antigen	HLA
Identical by descent	IBD
Leishmaniasis recidivans	LR
Linkage disequilibrium	LD
Lipophosphoglycan	LPG
Long runs of homozygosity	LROH
Mucocutaneous leishmanias	MCL
Multi-dimensional scaling	MDS
Multi-locus enzyme electrophoresis	MLEE
Multi-locus sequence analysis	MLSA
Multi-locus sequence typing	MLST
Neglected tropical diseases	NTD
Neighbour-Joining	NJ
Polymerase chain reaction	PCR
Post-kala-azar dermal leishmaniasis	PKDL
Principal components analysis	PCA
Promastigote secretory gel	PSG
Pulsed-field gel electrophoresis	PFGE
Quantitative trait locus	QTL
Random amplified polymorphic DNA	RAPD
Read depth	RD
Reactive oxygen species	ROS
Relative log expression	RLE
Single nucleotide polymorphism	SNP
Spliced leader	SL
T-cell receptor	TCR
Transcription start site	TSS
Trimmed mean of M-values	TMM
Untranslated region	UTR
Visceral leishmaniasis	VL
Viscerotropic leishmaniasis	VTL
Whole-genome sequencing	WGS
Identical by descent	IBD