Genetic and metabolic mechanisms of amphotericin B resistance in *Leishmania* parasites

Andrew William Pountain

MBiochem, MRes

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Institute of Infection, Immunity and Inflammation
School of Life Sciences
College of Medical, Veterinary and Life Sciences

University of Glasgow

December 2017

Abstract

Pathogenic protozoa of the genus *Leishmania* pose a significant burden on global health. Control relies on a limited range of chemotherapeutic options, with amphotericin B of increasing importance. Understanding how resistance may emerge to this drug is therefore of some concern. As amphotericin B acts through its affinity for leishmanial sterols, altered sterol composition has been described. However, little is known about the genetic basis of these changes.

In this thesis, selection and characterisation of four novel amphotericin B-resistant *L. mexicana* lines are described. Changes to parasite drug sensitivity and fitness were profiled, as well as alterations in metabolism. This revealed heterogeneity between lines, suggesting that many changes arise stochastically during selection of resistance. In one line, no fitness costs to infectivity and replication in primary macrophages were found. Hypersensitivity to the drug pentamidine, however, was a consistent phenotype.

All parasites demonstrated altered sterol composition. Genetic and transcriptomic profiling revealed associated changes in sterol biosynthesis genes, with mutation in sterol C5-desaturase observed in one line and loss of sterol C24-methyltransferase expression apparent in the other three. Broader analysis revealed extensive and apparently stochastic transcriptome remodelling, associated with chromosomal copy number changes. Deletion of the miltefosine transporter was found in one line, associated with miltefosine cross-resistance.

Through reintroduction of candidate genes into resistant lines, a role for these changes was demonstrated in both amphotericin B resistance and pentamidine hypersensitivity. Reintroduction of sterol biosynthesis genes, but not the miltefosine transporter, was associated with at least partial restoration of wild-type sterol composition. Finally, changes to the sterol C24-methyltransferase locus were investigated, revealing that structural variations are the basis of expression changes, possibly mediated by repetitive sequences at this locus.

These data demonstrate multiple routes to drug resistance, and suggest that at least some of these routes allow retention of parasite infectivity. Evidence of structural instability at the locus of sterol C24-methyltransferase, associated with altered sterol composition and drug resistance, is of particular concern.

Table of Contents

List of Tables

List of Figures

List of Accompanying Material

Supplementary file 2-1: Recipes of media used. Word document.

Supplementary file 2-2: Determination of qPCR primer pair efficiency and specificity. Word document.

Supplementary file 2-3: Custom Python script to calculate chromosome ploidy statistics. Text file, in Python 3 syntax. Open using any standard text editor, such as Microsoft WordPad.

Supplementary file 2-4: Custom Python script to calculate per-gene haploid ratios. Text file, in Python 3 syntax. Open using any standard text editor, such as Microsoft WordPad.

Supplementary file 2-5: Custom Python script for hard filtering for variants with genotypes differing by allele types present (explained in annotation). Text file, in Python 3 syntax. Open using any standard text editor, such as Microsoft WordPad.

Supplementary file 3-1: P-values for gas chromatography-mass spectrometry data. Word document.

Supplementary file 3-2: IDEOM file containing all data from the liquid chromatography-mass spectrometry experiment. Excel document. For full details of IDEOM, see Creek *et al*. 2012b. Note that data are unfiltered, and P-values given in this file are uncorrected.

Supplementary file 3-3: Filtered metabolite list, with results of statistical analysis. Excel document. The filtering strategy is described in Chapter 3. RT refers to peak retention time, PeakIDs are assigned to each unique peak. Mean absolute values are given for all resistant lines, wild-type and the pooled sample, as well as fold-changes in comparison to wild-type (FC). P-values (Pval) were calculated by ANOVA and then adjusted using the Benjamini-Hochberg multiple comparisons correction (Padj). Tukey post-hoc tests were performed for

each resistant line relative to wild type. The final column, “Tukeys” represents the number of Tukey tests for a metabolite with $P < 0.05$.

Supplementary file 4-1: NGS alignment statistics. Word document. Parameters were calculated using Samtools (Li *et al*. 2009).

Supplementary file 4-2: List of all variants, unfiltered, as identified by Freebayes (Garrison & Marth 2012) and annotated using Snpeff (Cingolani *et al*. 2012). Plain text document, in the Variant Call Format (vcf) version 4.2. See https://samtools.github.io/hts-specs/VCFv4.2.pdf for details (accessed 02/03/2018). Open using any standard text editor, such as Microsoft WordPad.

Supplementary file 4-3: List of variants differing in genotype between lines. Excel document. Genotypic changes between heterozygous genotypes (such as reference/alternative to reference/reference/alternative) were not included. For each line, the genotype is given (GT), along with the total depth at that locus (DP), and gene annotation including description and gene ontology term (GO), where appropriate. For full details on each variant, see Supplementary file 4-2.

Supplementary file 4-4: Haploid ratios derived from whole genome sequencing data. Excel document. For each gene, the haploid ratio (the length-normalised coverage of a gene divided by the median length-normalised per-gene coverage across the chromosome as a whole) is given for each line. For the resistant lines, fold-change (FC) compared to wild-type is also shown. Gene descriptions and gene ontology (GO) terms are given.

Supplementary file 4-5: Normalised per gene counts in the RNA-seq data. Plain text format, tab separated. Counts of fragments mapping to each gene were normalised using the size factor method in DESeq2 (Love *et al*. 2014).

Supplementary file 4-6: Differential expression analysis of RNA-seq data. Excel document. Each tab represents the differential expression analysis of an individual resistant line in comparison to wild-type parasites, as determined by DESeq2 (Love *et al*. 2014). For each gene, the base mean normalised count is given, along with the $\log_2$-transformed fold-change, the standard error on this,

and statistical information including adjusted P-values. Gene descriptions and gene ontology (GO) terms are given.

Supplementary file 4-7: Measurement of sterol C24-methyltransferase protein expression by western blot. Word document. Full methods and results are included.

Supplementary file 5-1: Full transcript sequences of BLAST hits for sterol C24-methyltransferase, derived from *de novo* assembly of RNA-seq data. Plain text document in fasta format. Open using any standard text editor, such as Microsoft WordPad.

Supplementary file 5-2: 5'-UTR sequences of BLAST hits for sterol C24-methyltransferase, derived from *de novo* assembly of RNA-seq data. Plain text document in fasta format. Open using any standard text editor, such as Microsoft WordPad.

Supplementary file 5-3: 3'-UTR sequences of BLAST hits for sterol C24-methyltransferase, derived from *de novo* assembly of RNA-seq data. Plain text document in fasta format. Open using any standard text editor, such as Microsoft WordPad.

Supplementary file 5-4: Sequence of intergenic region between sterol C24-methyltransferase gene copies, as determined by Sanger sequencing. Plain text document in fasta format. Note that positions 1-130 and 3019-3079 are derived from the 3'-end and 5'-end of flanking coding sequences. Open using any standard text editor, such as Microsoft WordPad.

Additional legends for supplementary information: Word document. For those files that do not contain legends (Supplementary files 2-3, 2-4, 2-5, 3-2, 4-2, 4-5, 5-1, 5-2, 5-3, 5-4), additional legends have been supplied here.

Acknowledgement

Let me begin by thanking Mike and Nick, my primary and secondary supervisors, for the guidance over the last few years. Mike has always been ready with input and suggestions but has allowed me the freedom to explore my own ideas and independent development. Nick has been a great help in getting me started with next generation sequencing data, and in answering my questions along the way.

Secondly, thanks to all those in the Barrett lab and beyond who have introduced me to new techniques, shared their thoughts, allowed me to bounce ideas off them and provided a good laugh when needed: Kevin, Clément, Federica, Fiona, Dharmendra, Dan, Fernando, Al, Kathryn, Anne, Edu, Rai, Isabel, Emily and I'm sure many others I'm forgetting. Also the guys at Polyomics who've played such a big role in generating the data in this thesis, particularly Stefan for all his helpfulness and diligence over the sterol analysis, Gavin for answering my questions on LC-MS, Jing, Julie, David and Pawel for discussions on NGS and of course generating the data in the first place.

I came into this PhD project as part of the Wellcome Trust four-year program – so thanking the Wellcome Trust itself is probably in order, but also Darren and Olwyn, who guided me so much in first year and have still given me help where sought. Add to that the other Wellcome students in my cohort – Joanne, Andrei, Síle, Hannah and Alejandro (also my flatmate of three years) – and those great ancestors of the programme who have shown the way – Pete, Kevin again, Nina, and Hussein, to name a few. Also to those who provided me the opportunity to try out placements with them in first year – Pablo and Annette, and Willie, who got me into working with genomic data in the first place, patiently starting me on some Perl programming. And of course Julie, who taught me the ropes of *Leishmania* culture in my initial placement with Mike, as well as Roy, who helped me out then and provided so much of the backdrop to this project. As an aside, I would like to acknowledge Jane Blythe, my school teacher and malaria PhD graduate, whose talk on apicomplexan biology and brain-changing toxoplasmosis I credit as an early pointer towards this field.

Then of course, there's family. Thanks to my parents, Chris and Joyce, and brother David for the support and for giving me a base to escape to from time to

time. Virgin Trains have profited handsomely from my trips back to see you. My Mum, the Strathclyde biochemistry graduate in whose copy of the *New Scientist* I first read about antigenic switching in trypanosomes, has had more than a little to do with me ending up studying parasitology in Glasgow.

Last, but probably most necessarily of all, my wife Carly. Thank you for the support, for listening, for making time for Skype conversations despite the awkward time difference, and most of all for showing me that as stressful and tiring a PhD can be, it's nothing compared to medical school.

Author's Declaration

I declare that, unless otherwise stated, all results presented in this thesis are my own work.

Andrew William Pountain

Definitions/Abbreviations

3’	3 prime end of DNA
5’	5 prime end of DNA
Å	Ångström (10^{-10} m)
ABC	Adenosine triphosphate-binding cassette
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
AmB	Amphotericin B
BLAST	Basic local alignment search tool
BMDM	Bone marrow-derived macrophages
c	Centi (10^{-2})
cDNA	Complementary DNA
cHOMEM	Complete HOMEM
CHX	Cycloheximide
Ci	Curie
CL	Cutaneous leishmaniasis
CNV	Copy number variation
CoA	Coenzyme A
cRPMI	Complete RPMI
Ct	Cycle threshold
ddNTP	Dideoxynucleotide triphosphate
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dTTP	Deoxythymine triphosphate
dUTP	Deoxyuracil triphosphate
FBS	Fetal bovine serum
FC	Fold change
g	Gravitational force
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	Gas chromatography
GFP	Green fluorescent protein
GO	Gene ontology
GOX	Glucose oxidase
HILIC	Hydrophilic interaction liquid chromatography
HMG-CoA	3-hydroxy-3-methylglutaryl coenzyme A
Hsp	Heat shock protein
IC_{50}	Concentration at which a 50% inhibition (e.g. of growth) is observed
InDel	Mutation involving a small insertion or deletion
k	Kilo (10^{3})
KEGG	Kyoto encyclopaedia of genes and genomes
l	Litres
LC	Liquid chromatography
Luria-Bertani	LB

m	Metres
M	Molar (moles/litre)
Mb	Megabase pairs
M-CSF	Macrophage colony-stimulating factor
MDR	Multidrug resistance
MILT	Miltefosine
min	Minutes
MMEJ	Microhomology-mediated end joining
mRNA	Messenger RNA
MRP	Multidrug-resistant protein
MS	Mass spectrometry
MSI	Metabolomics Standards Initiative
MT	Miltefosine transporter
m/z	Mass-to-charge ratio
n	Nano (10^{-9})
NAD^+	Nicotinamide adenine dinucleotide
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NGS	Next generation sequencing
NIST	National institute of standards and technology
Pa	Pascal (unit of pressure)
PAT	Potassium antimonyl tartrate
PBS	Phosphate-buffered saline
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PENT	Pentamidine
PMM	Paromomycin
PPP	Pentose phosphate pathway
QC	Quality control
qPCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative reverse transcription-polymerase chain reaction
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing-based transcriptomics
ROS	Reactive oxygen species
rpm	Rotations per minute
RPMI	Roswell Park Memorial Institute (medium)
rRNA	Ribosomal ribonucleic acid
RT	Reverse transcriptase
s	Seconds
Sb^{III}	Trivalent antimony
Sb^{V}	Pentavalent antimony
SC5D	Sterol C5-desaturase
SIDER	Small interspersed degenerate retrotransposon

SNP	Single nucleotide polymorphism
SL	Spliced leader
SMT	Sterol C24-methyltransferase
SREBP	Sterol regulatory element binding protein
TCA	Tricarboxylic acid
tRNA	Transfer ribonucleic acid
TLR	Toll-like receptor
TORC1	Target of rapamycin complex 1
UTR	Untranslated region
UV	Ultraviolet
V	Volts
VL	Visceral leishmaniasis
WGS	Whole genome sequencing
WHO	World Health Organisation
^{o}C	Degrees centigrade
δ	Difference between two values
μ	Micro (10^{-6})

1 Introduction

1.1 Leishmaniasis and *Leishmania* parasites

Leishmania is a genus of kinetoplastid protozoa of the order Trypanosomatida. Kinetoplastids include a number of insect-transmitted pathogenic organisms, particularly within the genera *Leishmania* and *Trypanosoma*, and are characterised by the presence of a flagellum and a single mitochondrion, containing a network of circular DNA known as the kinetoplast. Situated within the phylum Euglenozoa, these organisms display highly divergent eukaryotic biology but pose a significant threat to human and veterinary health, being responsible for the leishmaniases (the collective term for clinical manifestations of *Leishmania* infection), human African trypanosomiasis and Chagas disease in humans, as well as veterinary leishmaniasis (particularly in canids) and animal African trypanosomiasis, known as nagana. Understanding the biology of these parasites and their response to drugs is therefore of considerable importance in control of the diseases they cause.

1.1.1 The global burden of the leishmaniases

Leishmania parasites are responsible for a diversity of pathologies, the leishmaniases, distributed across tropical and subtropical regions worldwide. At least 20 species are known to be infective to humans, and new human-infectious species are still being discovered, as shown by the recent addition of members of the divergent *L. enriettii* complex (Figure 1-1) (Cotton 2017). Transmission is through the bite of infected female phlebotomine sandflies, and globally at least 98 countries are known to be endemic, with 350 million people thought to be at risk and an estimated prevalence of 12 million cases (Akhoundi *et al*. 2016). Diverse clinical manifestations are observed, but are usually grouped into cutaneous and visceral forms of the disease.

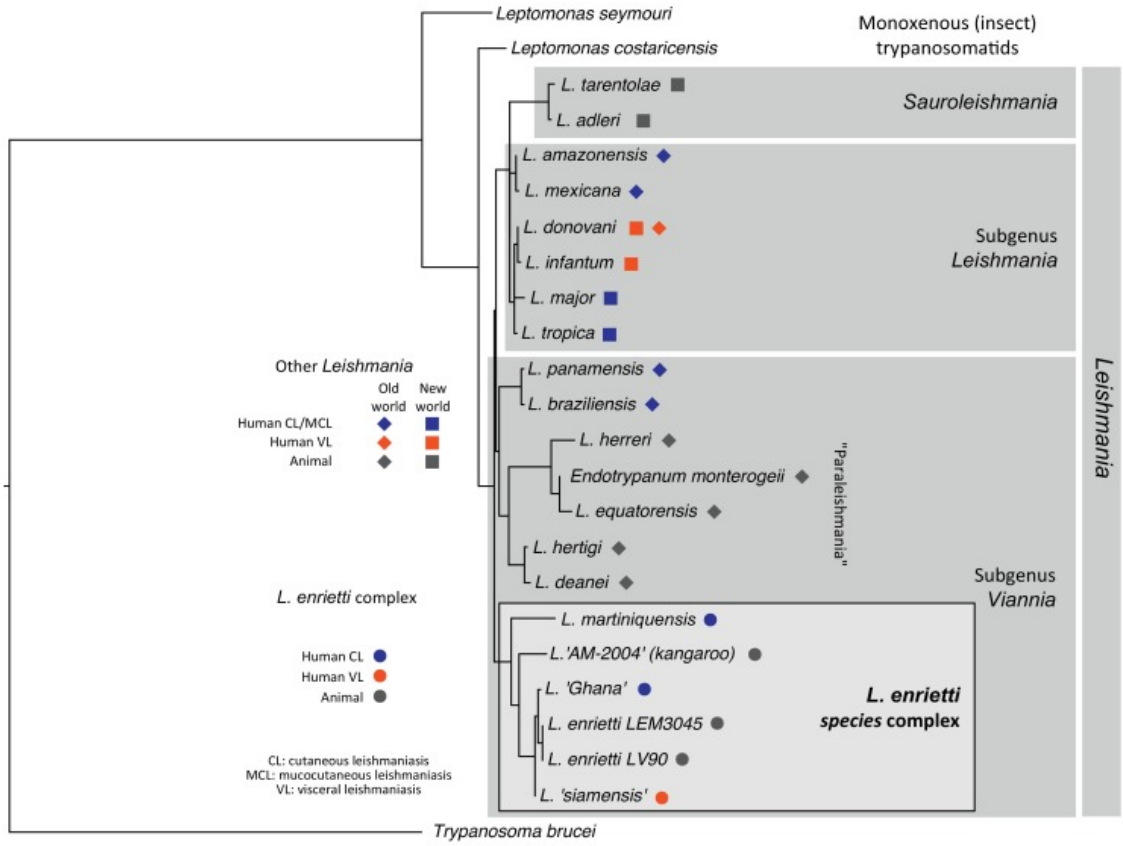


Figure 1-1: The phylogenetic tree of *Leishmania* species. The evolutionary relationship between several human and animal-infectious species of *Leishmania* parasites are shown, including the newly added *L. enriettii* species complex. Representative members of *Leptomonas* and *Trypanosoma* genera are shown for comparison (source: Cotton 2017).

Cutaneous leishmaniasis (CL) has an annual incidence of an estimated 0.7-1.2 million cases, with 70-75% of cases observed in 10 countries (Afghanistan, Algeria, Colombia, Brazil, Iran, Syria, Ethiopia, North Sudan, Costa Rica and Peru) (Alvar *et al*. 2012). Case numbers have increased notably in Syria and neighbouring countries experiencing conflict and displacement in the Middle East region (Al-Salem *et al*. 2016). CL arises through infection with a number of species, including *L. aethiopica*, *L. major* and *L. tropica* in the Old World and *L. amazonensis*, *L. colombiensis*, *L. mexicana*, *L. venezuelensis* and various members of the *Viannia* subgenus including *L. braziliensis* and *L. panamensis* in the New World (Akhoundi *et al*. 2016). Members of the *L. enriettii* complex and *L. infantum* have been found to cause infection in both the Old and New Worlds (Akhoundi *et al*. 2016), with *L. infantum* likely having been introduced to the New World relatively recently as a result of human activity (Kuhls *et al*. 2011). Cutaneous infection may result in self-healing lesions, but complications can arise and, particularly in infection by members of the *Viannia* subgenus,

destruction of mucous membranes of the nose, mouth and throat cavities may be observed, a highly disfiguring and disabling condition known as mucocutaneous leishmaniasis.

Visceral leishmaniasis (VL), also known as kala azar, has an annual estimated incidence of 0.2-0.4 million cases, with greater than 90% occurring in India, Bangladesh, Sudan, South Sudan, Ethiopia and Brazil (Alvar *et al*. 2012). VL arises normally due to infection by closely related species *L. donovani* and *L. infantum*, although cases resulting from *L. colombiensis* and *L. enriettii* complex infection have been observed (Akhoundi *et al*. 2016). Pathology typically includes irregular fever, weight loss, spleno- and hepatomegaly, and anaemia, and the disease is typically fatal if left untreated. While *L. infantum* infection usually arises through zoonosis (with dogs being a major reservoir) (Dantas-Torres 2007) in both the Old World and Latin America, *L. donovani* is associated with anthroponotic transmission in the Old World, and is further linked with a disfiguring post-treatment complication known as post-kala azar dermal leishmaniasis (PKDL). Recent efforts, as part of the WHO kala azar elimination programme, have served to greatly reduce the burden of VL in South Asia, with Bangladesh, India and Nepal reporting an overall decrease in mortality of 85% over the duration of the programme and Nepal achieving elimination of the disease (World Health Organisation 2015). However, not all infections result in VL, with factors such as nutritional status and coinfection with HIV and intestinal parasites reported to influence disease progression, and the role of asymptomatic infection in transmission is of some concern for the sustainability of reductions in disease burden (Das *et al*. 2014).

1.1.2 The life cycle of *Leishmania*

Leishmania parasites display a dixenous life cycle between phlebotomine sandflies and vertebrate hosts (Figure 1-2A). Sandflies of the genus *Phlebotomus* are largely responsible for human infection in the Old World, *Lutzomyia* in the New World (Akhoundi *et al*. 2016). Infection results from the bite of a female sandfly, depositing metacyclic promastigote-stage parasites at the site of the bite during feeding. These parasites, which are observed in the anterior midgut of sandflies, are characterised by a long flagellum and short, slender body. On infection, metacyclic parasites are taken up by phagocytic cells through

endocytosis and rapidly undergo transformation to rounded amastigote forms lacking an external flagellum. Macrophages and dendritic cells have both been established as host cells during infection, but neutrophils have also been implicated early in disease progression. In fact, early macrophage infection may result from uptake of infected neutrophils, referred to as the 'Trojan Horse' model of infection (Laskay *et al*. 2003). Within macrophages, parasites reside in acidic parasitophorous vacuoles deriving from endosome and lysosome fusion events (McConville & Naderer 2011), and are able to hijack host cell function through a variety of mechanisms. These include modulation of host translation and immune activation through cleavage of host mammalian target of rapamycin 1 by *Leishmania* protease GP63 (Jaramillo *et al*. 2011). While amastigote infection and replication eventually results in cell death and release of parasites, host cell survival is initially promoted through modulation of pro-apoptotic pathways (Akarid *et al*. 2004). After amastigote release, cycles of reinfection, proliferation and release in other host cells will occur.

During feeding on an infected host, sandflies take up infected macrophages and free amastigotes from the skin as part of the bloodmeal. The relatively high pH and low temperature trigger differentiation into procyclic promastigotes, a motile, proliferative stage that possesses an anterior flagellum (Bates 2007) (Figure 1-2B). After several cycles of replication within the bloodmeal, confined within the peritrophic matrix of the midgut, differentiation into highly motile, elongated forms called nectomonads allows escape from the bloodmeal and migration to the anterior end of the midgut (Bates 2007). Once at the stomodeal valve between the midgut and the foregut, nectomonads differentiate into shorter, proliferative leptomonads. Leptomonad production of a promastigote secretory gel at this valve plays an important role in transmission by forcing egress of this gel prior to feeding, hence enabling release of infective parasites (Bates 2007). Leptomonads also differentiate into non-proliferative infective metacyclic forms, as well as secondary adherent forms known as haptomonads. While many of these details are conserved, there is some variation among species, with *Leishmania* parasites of the subgenus *Viannia* showing departure from this cycle in particular (Bates 2007).

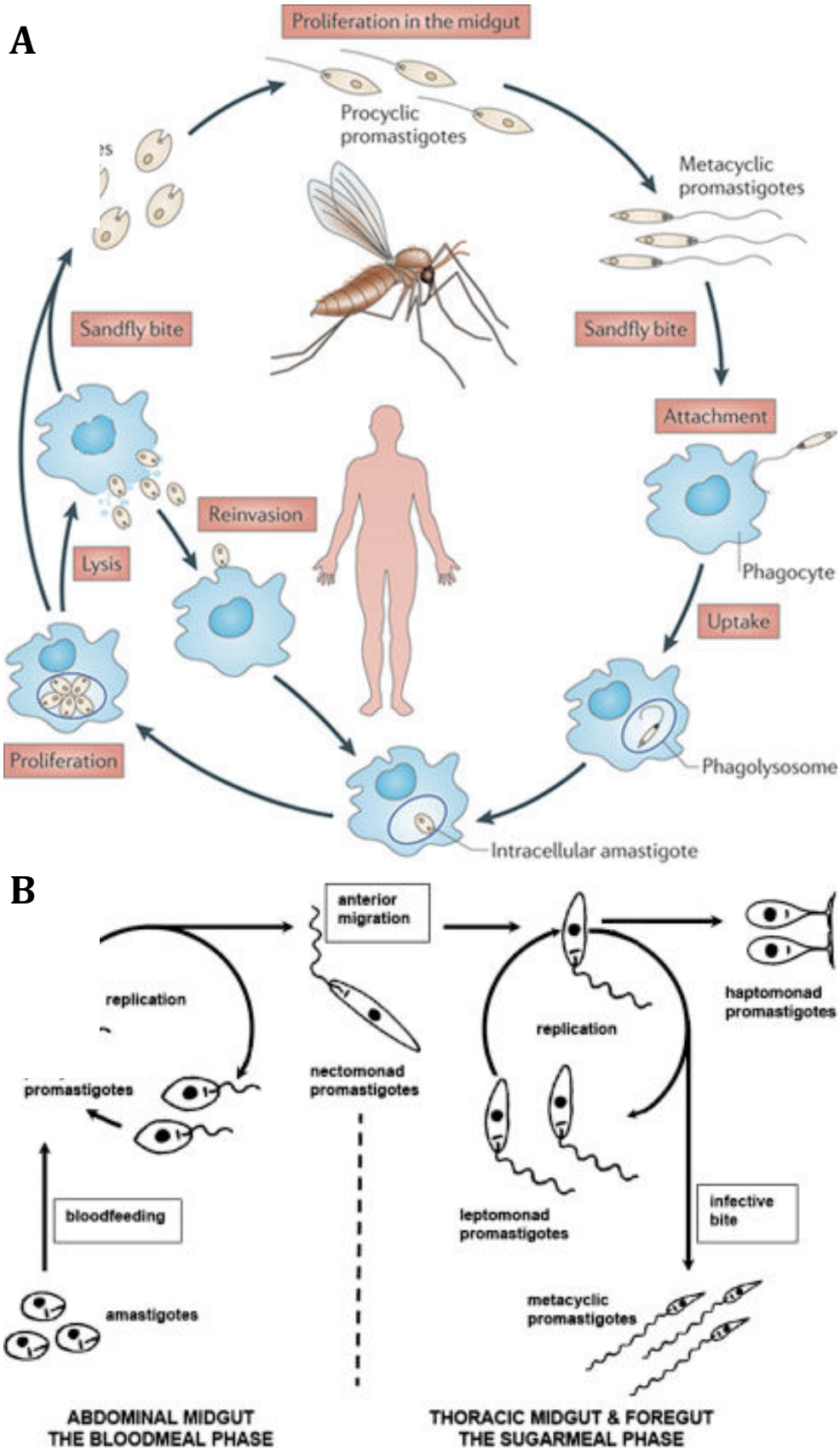


Figure 1-2: The Life cycle of *Leishmania* parasites. A) Overall life cycle, including both mammalian and insect stages (source: Kaye & Scott 2011). B) More detailed breakdown of insect life-cycle stages (source: Bates 2007).

1.1.3 The genome and transcriptional biology of *Leishmania*

Leishmania parasites possess a genome of approximately 32 Mb, with a conserved predicted proteome across multiple species of around 7,400 genes (Rogers *et al*. 2011). Species such as *L. major*, *L. donovani* and *L. infantum* possess 36 chromosomes, while other species possess fewer due to fusion events, notably *L. mexicana*, which possesses 34 chromosomes due to fusions between chromosomes eight and 29, and chromosomes 20 and 36; *L. braziliensis* also possesses 35 chromosomes due to a fusion between chromosomes 20 and 34 (Britto *et al*. 1998). One remarkable feature of *Leishmania* genomes is their apparent karyotypic plasticity. While most species appear to show basal diploidy (triploid *L. braziliensis* being an exception), frequent chromosomal copy number variations (CNVs) are observed (Downing *et al*. 2011, Rogers *et al*. 2011). Chromosome 31 (by *L. major* numbering, equivalent to chromosome 30 in *L. mexicana*) generally exhibits a copy number greater than two and is frequently tetraploid, an adaptation which has been suggested to relate to the evolution of parasitism due to an enrichment of genes upregulated in the mammalian amastigote stage on this chromosome (Akopyants *et al*. 2009, Fiebig *et al*. 2015). Variation is observed within a population of parasites, as evidenced by non-integer copy number ratios, a phenomenon known as mosaic aneuploidy (Rogers *et al*. 2011, Sterkers *et al*. 2012). Frequent stochastic amplifications, deletions and rearrangements are also found, often associated with repetitive sequences (Ubeda *et al*. 2014). Indeed, population genomics demonstrates a high background of CNV within broader populations (Imamura *et al*. 2016), complicating efforts to link specific copy number changes to phenotypic variations such as altered drug sensitivity.

The genomes of *Leishmania* and other kinetoplastid organisms are highly compact, with most genes lacking both introns and promoters. Individual genes with no evident functional association are grouped within large polycistronic arrays that are constitutively transcribed, greatly limiting the potential for transcriptional regulation (Myler *et al*. 1999, Clayton & Shapira 2007). Transcription initiation takes place at so-called ‘strand-switch regions’ between polycistrons at loci associated with histone H3 acetylation (Thomas *et al*. 2009). During maturation, splicing occurs at splice acceptor sites in intergenic regions (Figure 1-3), resulting in the addition of a 39 nucleotide spliced leader (SL)

sequence bearing a trimethyl cap to the 5' end of the newly generated monocistron (Clayton & Shapira 2007). The 3' end of this (generated as a result of a downstream splicing event) is polyadenylated.

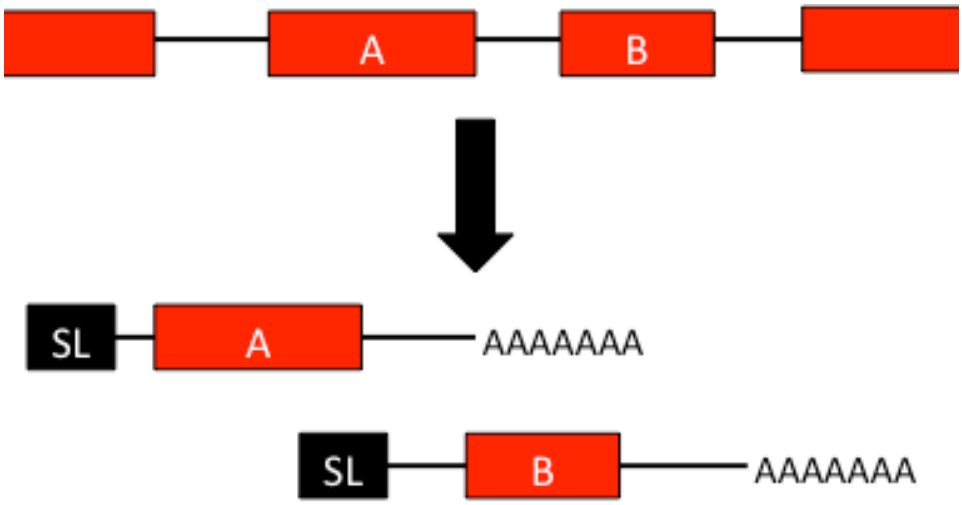


Figure 1-3: Trans-splicing in *Leishmania*. Long polycistronic units are *trans*-spliced into mature monocistronic mRNAs. Mature mRNAs are shown for genes A and B within one such unit. During splicing, a 39 nucleotide spliced leader (SL) bearing a 5'-trimethyl cap is added to the 5' end of the coding sequence, and the 3' end is polyadenylated.

In the absence of transcriptional regulation, control over RNA abundance is exerted primarily through regulation of RNA stability. Genome wide analysis of decay rates after transcriptional inhibition revealed broad variation in stability between transcripts (Fadda *et al*. 2014) and specific sequence elements associated with instability have been identified, particularly of the SIDER2 class (Azizi *et al*. 2017). However, in addition to *cis*-acting elements controlling stability, numerous RNA binding proteins have been identified in kinetoplastid genomes, and it may be that multiple *trans*-acting factors may bind sites, chiefly in the 3' untranslated regions (3'-UTRs) of transcripts, creating possibilities for complex regulation (Clayton 2014). Hence regulation of RNA abundance appears to play an important role in life cycle progression, as evidenced by RNA-seq experiments in *Leishmania*. In *L. major*, progression from metacyclic promastigotes to intracellular amastigotes showed enrichment for upregulated genes associated with responses to cellular and oxidative stress, and reduced expression of protein kinases and translation-associated genes (Dillon *et al*. 2015). A separate study compared *L. mexicana* log-phase promastigotes and intracellular amastigotes (Fiebig *et al*. 2015). This showed no enrichment of gene ontology groups amongst genes upregulated in amastigotes, whereas

promastigotes showed upregulation of genes associated with tRNA charging, various metabolic pathways (including sterol biosynthesis) and flagellar motility.

Beyond RNA abundance, translational regulation also appears to play an important role. 3'-UTR elements appear to play a dominant role in this, as in the case of regulation of *HSP83* translation in response to heat shock (Zilka *et al*. 2001). A broader understanding of translational control in developmental regulation in *Trypanosoma* species has been enabled through ribosome profiling, a technique which measures ribosome abundance on individual transcripts (Jensen *et al*. 2014, Vasquez *et al*. 2014, Smircich *et al*. 2015), revealing a greater dynamic range in gene expression than that identified at the level of RNA abundance.

1.2 *Leishmania* metabolism

Analysis of *Leishmania* genomes has suggested a high capacity for metabolic complexity (El-Sayed *et al*. 2005), likely due to adaptation to multiple host niches. Metabolite analyses of proliferating promastigotes showed rapid uptake of glucose from the medium, along with a range of amino acids (Creek *et al*. 2012), suggesting an important role of glycolysis in energy metabolism. In turn, succinate, acetate, alanine and CO_2 are secreted in high glucose medium (Creek *et al*. 2012). Analysis of amastigotes is complicated by host cell metabolism; however, use of axenic amastigotes, an extracellular model for mammalian stage parasites, has helped, although the use of extremely rich media for this stage complicates metabolic analysis. Nevertheless, available evidence suggests that amastigotes exhibit reduced glycolytic flux and increased reliance on fatty acid ß-oxidation (Creek *et al*. 2012), and in *L. braziliensis*, a broad reduction in metabolic activity has been observed on promastigote to amastigote differentiation (Jara *et al*. 2017). As parasites, *Leishmania* have limited capacity to synthesise a range of essential metabolites, including amino acids and purines, and so must rely on uptake from the extracellular environment (McConville & Naderer 2011). A number of pathways are of particular importance to the interaction of *Leishmania* parasites with drugs; these are discussed in greater detail below.

1.2.1 The pentose phosphate pathway

NADPH is an important cellular cofactor produced through the reduction of $NADP^+$. It is used as a source of cellular reducing power, required in *Leishmania* both in biosynthetic reactions and in defence against oxidative stress, which is incurred both within the macrophage phagolysosome and in response to treatment with a range of antileishmanial compounds (Moreira *et al*. 2011). The pentose phosphate pathway (PPP) is a major route of production of NADPH, as well as production of metabolites for biosynthesis, such as ribose 5-phosphate required in the synthesis of nucleic acids (Figure 1-4). Conventionally it is divided into oxidative and non-oxidative branches. The former converts glucose 6-phosphate to ribulose-5-phosphate, generating NADPH through oxidation reactions catalysed by the enzymes glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase. The latter facilitates the interconversion of

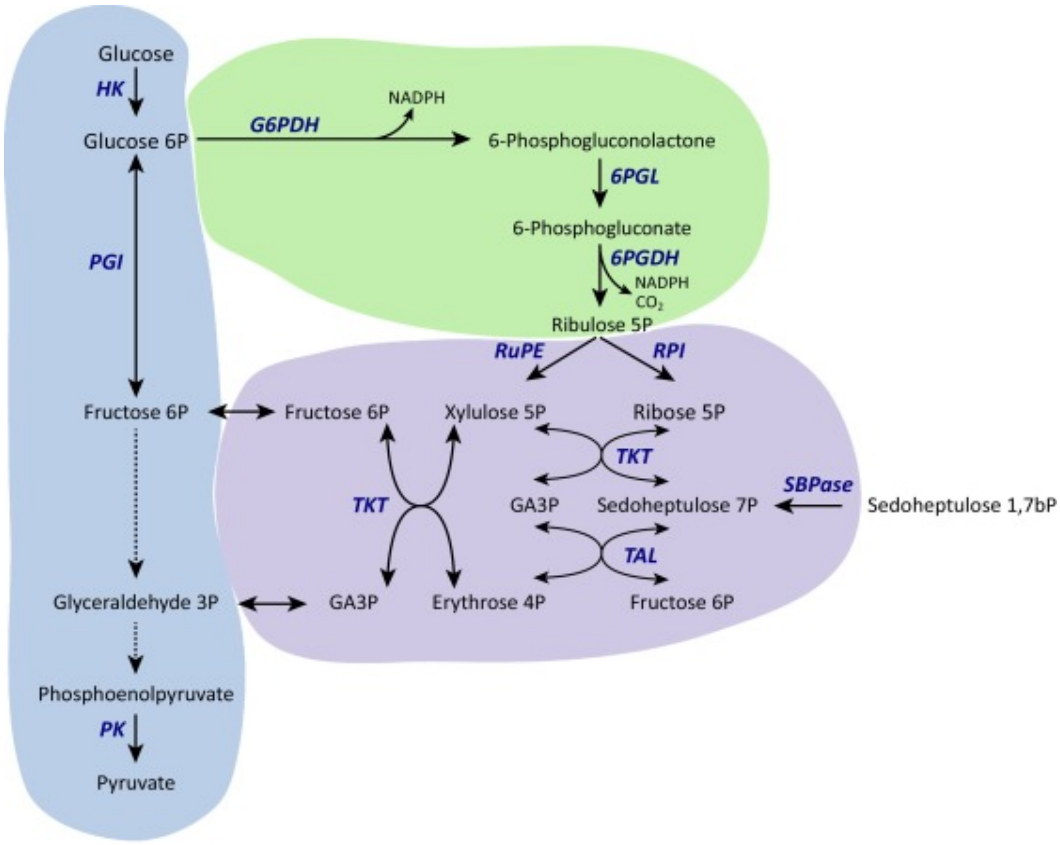


Figure 1-4: The pentose phosphate pathway. Green represents the oxidative branch, purple the non-oxidative branch, which may be more metabolically flexible than the classical pathway shown here suggests. Blue represents a simplified form of the glycolysis pathway. HK: hexokinase; PGI: phosphoglucose isomerase; PK: pyruvate kinase; G6PDH: glucose 6-phosphate dehydrogenase; 6PGL: 6-phosphogluconolactonase; 6PGDH: 6-phosphogluconate dehydrogenase; RPI: ribose 5-phosphate isomerase; RuPE: ribulose 5-phosphate epimerase; TKT: transketolase; TAL: transaldolase; SBPase: sedoheptulose-1,7-bisphosphatase (source: Kovářová & Barrett 2016)

sugar phosphates through transketolase and transaldolase, which shuffle two and three carbon units respectively between substrates. The promiscuity of these enzymes allows significant metabolic flexibility in mode of operation of this branch (Kovářová & Barrett 2016). Exposure of *L. donovani* promastigotes to oxidative stress-inducing agents leads to increased PPP flux and increase in both RNA and protein levels of numerous PPP enzymes in both oxidative and non-oxidative branches (Ghosh *et al*. 2015). *Leishmania* possesses secondary routes to NADPH generation in addition to the PPP, particularly the malic enzyme, which produces NADPH via the conversion of malate to pyruvate.

1.2.2 Trypanothione and polyamine metabolism

Polyamines are low molecular weight molecules that are cationic at cellular pH and play an essential role in cellular proliferation. Polyamine synthesis in trypanosomatids involves conversion of arginine to ornithine via arginase, and onwards to generation of the diamine putrescine by ornithine decarboxylase (Figure 1-5); from there spermidine synthase converts putrescine further to the polyamine spermidine, although both putrescine and spermidine can be acquired from the environment in *Leishmania* (Colotti & Ilari 2011). The essentiality of this pathway in *T. brucei* is highlighted by efficacy of eflornithine, an irreversible inhibitor of ornithine decarboxylase, in the treatment of African trypanosomiasis (Bacchi *et al*. 1994). Alternative routes to acquisition of polyamines from their environment by *Leishmania* and *T. cruzi* prevent this drug eliciting a therapeutic effect in these organisms.

Polyamine metabolism plays a key role in defence against oxidative stress in *Leishmania* since spermidine is a precursor to the dithiol molecule trypanothione. While many eukaryotes use the glutathione/glutathione disulfide redox couple in detoxification of oxidising agents, trypanosomatids lack many of the enzymes required. Trypanothione is used instead, synthesised through ligation of spermidine and glutathione by trypanothione synthetase. This process is dependent on glutathione synthesis, and the essential enzyme gamma-glutamylcysteine synthase, catalysing an intermediate step in this process, is rate-limiting for trypanothione biosynthesis (Mukherjee *et al*. 2009).

Trypanothione is produced from trypanothione disulfide by NADPH-dependent trypanothione reductase. Trypanothione functions as an electron donor, particularly to thioredoxin dithiol proteins called tryparedoxins (Colotti & Ilari 2011).

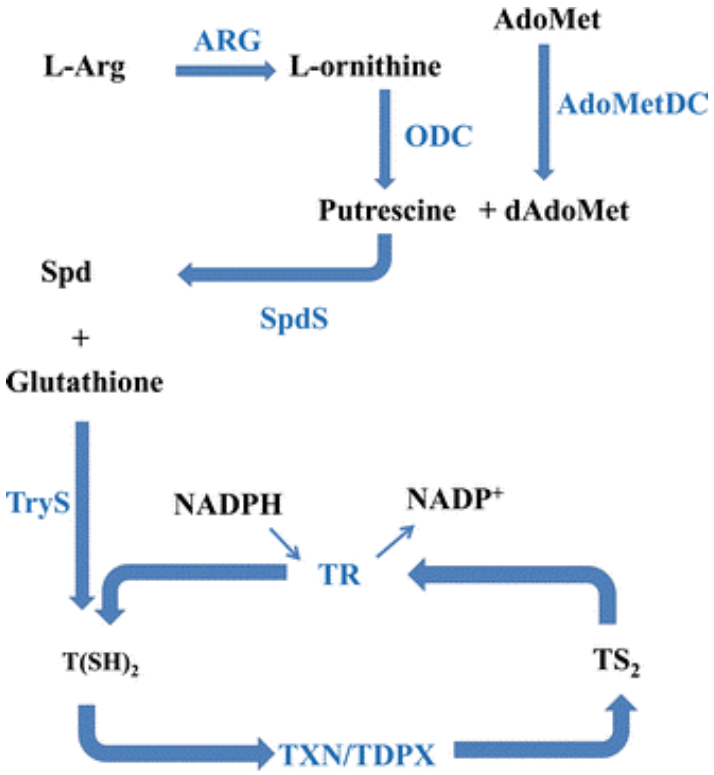


Figure 1-5: Polyamine metabolism in trypanosomatids. Polyamine synthesis and trypanothione metabolism are shown. L-Arg: L-arginine; AdoMet: *S*-adenosylmethionine; dAdoMet: *S*-adenosyl-5'-(3-methylthiopropylamine); Spd: spermidine; $T(SH)_2$: trypanothione; TS_2: trypanothione disulfide; ARG: arginase (note absent from *T. brucei*, which acquires ornithine from the environment); ODC: ornithine decarboxylase (note absent from *T. cruzi*); AdoMetDC: *S*-adenosylmethionine decarboxylase; SpdS: spermidine synthase; TryS: trypanothione synthetase; TR: trypanothione reductase; TXN: tryparedoxin; TDPX: tryparedoxin peroxidase (source: Colotti & Ilari 2011).

Tryparedoxins themselves are a group of proteins of the thioredoxin superfamily, but are highly divergent and only found in kinetoplastids (Castro & Tomás 2008). They are efficiently reduced by trypanothione but only inefficiently by glutathione. Located within both the cytosol and the mitochondrion in *Leishmania*, these highly abundant proteins are used as a source of electrons for the trypanothione peroxidases (Castro & Tomás 2008). These enzymes are important in detoxification of hydrogen peroxide and other peroxides, and are of two types. The former of these are the 2-cysteine peroxiredoxins, which use redox-active cysteines in order to reduce their substrates, and have broad specificity in detoxification of H_2O_2, organic peroxides and peroxinitrites. The other subtype is the nonselenium glutathione peroxidases, in which the

selenocysteine commonly found in such enzymes is replaced by Cys. Unusual for this class of enzymes, nonselenium glutathione peroxidases in trypanosomatids have only a poor affinity for glutathione, preferring tryparedoxin as a source of reducing equivalents (Castro & Tomás 2008). Again, these have a broad specificity for H_2O_2 and organic peroxides.

1.2.3 Oxidative stress control beyond trypanothione metabolism

While the trypanothione/trypanothione disulfide redox couple and its associated metabolism play an important role in defence against oxidative stress in trypanosomatids, a number of other mechanisms exist for ROS detoxification. A plant like peroxidase has been found in *Leishmania*, which is dependent on ascorbate for activity (Adak & Datta 2005). This represents another route to H_2O_2 degradation, although trypanothione may also play a role here in reduction of monodehydroascorbates produced during the reaction of this enzyme (Colotti & Ilari 2011). Removal of superoxide ions ($O_2^{\cdot -}$) is facilitated through the action of four iron superoxide dismutases, two of which are located in the mitochondrion, and two distributed between the cytosol and the glycosome (Dufernez *et al*. 2006). Furthermore, a second thiol-containing compound, ovothiol A, has been detected in *Leishmania*. This is present in the insect stages of several species of *Leishmania* parasite, but interestingly, while it was detected in *L. major* skin lesion-derived amastigotes, it was not found in *L. donovani* splenic amastigotes (Ariyanayagam & Fairlamb 2001). Ovothiol A itself is a histidine-derived mercaptoimidazole which has previously been observed in several marine metazoans, but its function in kinetoplastids is unclear. No ovothiol A-dependent peroxidase activity has been detected in *Leishmania* extracts, and it is not a substrate for trypanothione reductase; moreover, it is a less efficient non-enzymatic scavenger of hydrogen peroxide than trypanothione (Ariyanayagam & Fairlamb 2001). The redox potential of ovothiol A/ovothiol disulfide is also less electronegative than trypanothione/trypanothione disulfide, meaning ovothiol disulfide is likely to oxidise trypanothione (Ariyanayagam & Fairlamb 2001). Therefore, it appears not to play a major role in detoxification of peroxides, although may act in other ways, such as scavenging of free radicals.

1.2.4 Sterol metabolism

Sterols are an essential constituent of eukaryotic membranes, playing an important role in membrane stability and fluidity, as well as membrane protein function. *Leishmania* parasites differ from their mammalian hosts because, like fungi, their sterol composition consists primarily of ergostane-type sterols, rather than mammalian cholesterol. Ergostane-type sterols differ from cholesterol in possessing a methyl group at the C24 position in the side chain, as well as through the presence of a 7(8)-double bond in addition to the 5(6)-double bond present in cholesterol. These double bonds allow formation of a conjugated system, providing rigidity to the molecule, and influence membrane fluidity and lipid raft formation (Xu *et al*. 2001). Furthermore, a second double bond may protect against oxidative stress both because a more planar structure allows for greater interaction with phospholipids (Wiseman 1993) and through allowing delocalisation of radical electrons across the conjugated system (Dupont *et al*. 2012).

The presence of sterol containing a 24β-methyl group has been found to be an absolute requirement in *Saccharomyces cerevisiae*, although even minimal amounts permit viability, suggesting a possible role of this group in specific interactions with proteins or other binding partners (Pinto & Nes 1983). A similar requirement has been observed in bloodstream form *T. brucei*, in which growth inhibition mediated by prevention of C24-methylation (either by RNAi knockdown of sterol C24-methyltransferase or pharmacological inhibition by 25-azalanosterol) was rescued by provision of small amounts of ergosterol in the medium (1.2 μM) on a background of 7.8 μM cholesterol from serum lipids (Haubrich *et al*. 2015). This is in spite of the fact that *T. brucei* sterol composition in the bloodstream form largely consists of host-acquired cholesterol, but was suggested by the authors to represent a "growth-sparking" function of ergosterol, beyond a general membrane structural role.

1.2.4.1 Initial stages of sterol biosynthesis

The initial stages of ergosterol biosynthesis are similar to the cholesterol pathway in mammals, with the generation of mevalonate by 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase being followed by further steps of the

mevalonate pathway to squalene and hence to lanosterol (Figure 1-6). HMG-CoA reductase in humans is the target of statins, and on this basis its homologue in *Leishmania* has been considered as a drug target, with particularly strong inhibition of *L. donovani* growth by atorvastatin, which was rescued by supplementation of ergosterol (Dinesh *et al*. 2014). Labelling studies suggest *Leishmania* promastigotes favour leucine as a carbon source for sterol biosynthesis, in contrast to *Trypanosoma cruzi*, which exhibits a strong preference for acetate (Ginger *et al*. 2000). Unusually, leucine enters the biosynthesis pathway through a direct route, in contrast to most eukaryotes in which leucine is incorporated indirectly via acetyl CoA and acetoacetate. *Leishmania* parasites also display distinct compartmentalisation of this pathway: in *L. major*, while HMG-CoA synthesis and reduction takes place in the mitochondrion, mevalonate kinase shows a clear localisation in the glycosomes (peroxisome-related kinetoplastid organelles) and farnesyl diphosphate synthase resides in the cytosol (Ortiz-Gómez *et al*. 2006, Carrero-Lérida *et al*. 2009).

1.2.4.2 Ergostane-type sterol biosynthesis from lanosterol

Lanosterol is subjected to a series of modifications, particularly removal and addition of methyl groups and desaturation and saturation of the ring structure (Figure 1-7). This yields ergostane-type sterols as final products, including ergosterol itself, but also its isomer ergosta-5,7,24(28)-trienol (also known as 5-dehydroepisterol), which has been identified as the major sterol constituent in multiple *Leishmania* species (de Souza & Fernandes Rodrigues 2009, Andrade-Neto *et al*. 2016). In differing from mammalian cholesterol biosynthesis, this section of the pathway has been exploited as a drug target in fungi, particularly through use of the major class of antifungals, the azoles, which inhibit lanosterol 14α-demethylase (also known as CYP51, or ERG11 in yeast). However, other important targets include morpholine inhibitors of sterol C14-reductase (yeast ERG24) and fenhexamid, an agricultural inhibitor of C3-keto reductase (annotated in Figure 1-7 as short chain dehydrogenase, putative; ERG27 in yeast) (Debieu *et al*. 2001). CYP51, along with two other enzymes important in amphotericin B resistance, are discussed further below.

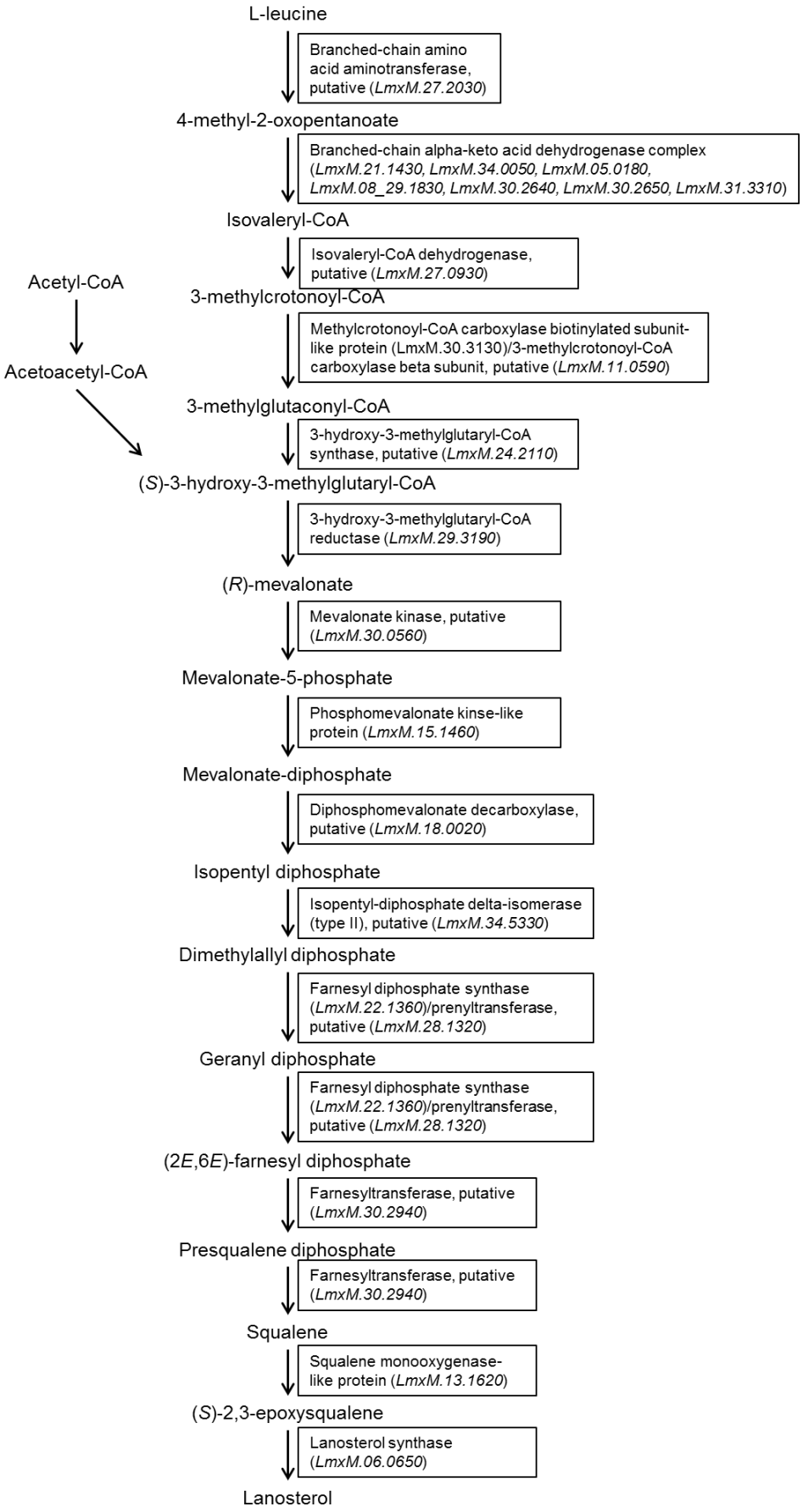


Figure 1-6: Initial stages of ergosterol biosynthesis in *Leishmania mexicana*. Lanosterol production from L-leucine. Adapted from *L.major* entry at biocyc.org (Doyle *et al*. 2009), with *L. mexicana* homologues shown here.

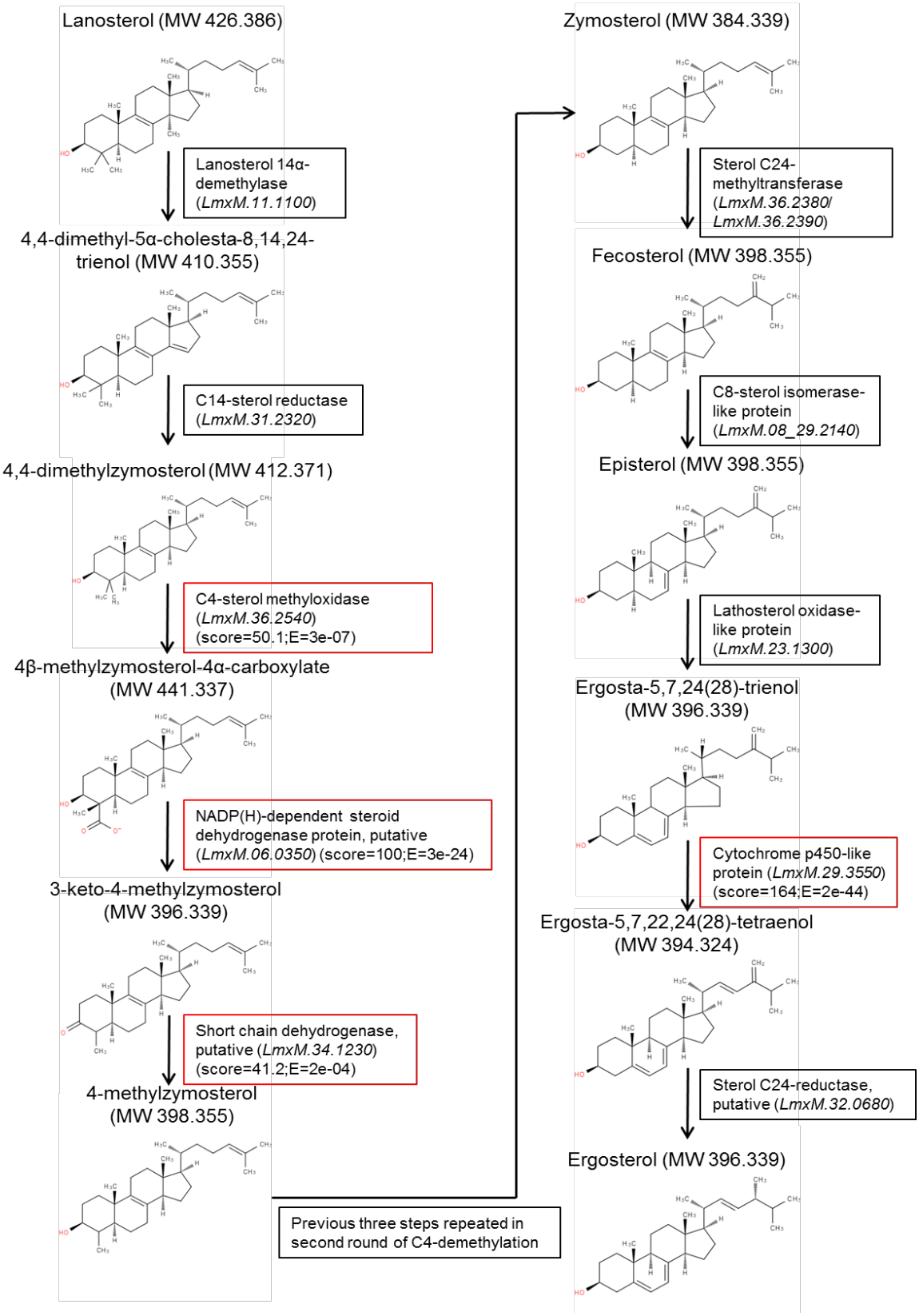


Figure 1-7: Synthesis of ergosterol from lanosterol in *Leishmania mexicana*. Simplified form of ergosterol biosynthesis from lanosterol, although side reactions and alternative pathways may occur. Where enzymes are in red, this is because of lack of annotation in biocyc.org or Kyoto encyclopaedia of genes and genomes (KEGG) databases, meaning that the top hit using the basic local alignment search tool (BLAST) is shown, along with score and probability of match (E). Oxygen-containing groups are in red; this is by convention only. Adapted from *L. major* entry at biocyc.org (Doyle *et al*. 2009) and the KEGG database (Kanehisa & Goto 2000).

1. Lanosterol 14α-demethylase (CYP51):

CYP51 is a haem-containing member of the cytochrome P450 family of monooxygenases. In a three-step, oxygen and NADPH-dependent mechanism, it catalyses the removal of a methyl group at the C14 position of the steroid backbone (Waterman & Lepesheva, 2005). CYP51 is essential for aerobic growth in yeast (Kalb *et al*. 1987). However, in *Leishmania*, this essentiality appears to be species dependent; in *L. donovani*, this gene is essential (McCall *et al*. 2015), whereas *L. major* parasites are viable but exhibit reduced heat tolerance, virulence and membrane stability, suggested by the authors to relate to accumulation of toxic C14-methyl sterols (Xu *et al*. 2014). In *Leishmania*, the role of CYP51 is complicated by the presence of a related essential enzyme, CYP5122A1, encoded by *CYP5122A1*, a single allele knockout of which exhibited reduced virulence and ergosterol production (Verma *et al*. 2011). In *L. mexicana*, CYP51, along with the subsequent enzyme in the pathway, C14-sterol reductase, localise predominantly to the endoplasmic reticulum when overexpressed (Mwenechanya *et al*. 2017).

2. Sterol C24-methyltransferase (SMT):

SMT catalyses the transfer of a methyl group from *S*-adenosylmethionine to the C24 position located in the steroid side chain, an important step which differentiates plant, fungal and leishmanial sterols from those found in animals. Because of this it has been considered as an antileishmanial drug target, inhibited by both the azasterol class of compounds (Haughan *et al*. 1995, Lorente *et al*. 2004) and the tricyclic antidepressant imipramine (Andrade-Neto *et al*. 2016). It is also under consideration as part of a polyprotein vaccine against leishmaniasis (Coler *et al*. 2015). Primary sequence analysis of *L. mexicana* SMT *LmxM.36.2380* does not predict the presence of transmembrane domains; however SMT in yeast is membrane-associated (Nes 2000), and localises to the endoplasmic reticulum in *L. major* (Jiménez-Jiménez *et al*. 2008). A ‘steric-electric plug’ mechanism has been proposed as a mode of action (Nes 2000, Figure 1-8). This begins with nucleophilic attack by the electrons in the 24(25)-double bond on the electrophilic carbon of *S*-adenosylmethionine, producing a 24,25-bridged carbenium cation intermediate that is subsequently stabilised by the deprotonation of the newly added methyl group by an active

site base. This therefore forms the 24(28) double bond of the substrate, with the remaining C24 proton transferred to the C25 position. SMT from *T. brucei* has been characterised *in vitro* and exhibits specificity for zymosterol as a substrate (Zhou *et al*. 2006); however, multiple C24-methylated products were generated, including products varying in position of the side chain double bond (fecosterol (ergosta-8,24(28)-dienol), ergosta-8,24(25)-dienol, ergosta-8,25(27)-dienol) as well as a double C24-alkylated sterol, 24,24-dimethyl ergosta-8,25(27)-dienol. The role of these side-products is unclear.

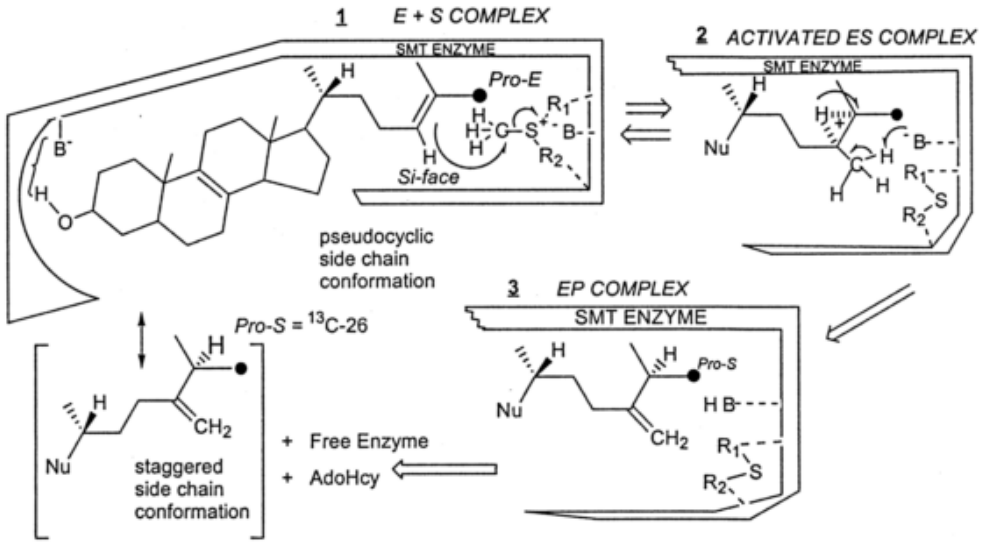


Figure 1-8: The 'steric-electric-plug' model of sterol C24-methyltransferase activity. The enzyme (E) and sterol substrate (S) form an initial covalent complex (ES) through nucleophilic attack of the 24(25)-double bond on the sulfur-bound methyl group. This forms a carbenium ion which is stabilised by bridging of the C24-hydrogen across the 24(25)bond. Subsequent proton donation from an active site base forms the enzyme-product complex (EP), and the product is subsequently released (source: Nes 2000).

The reference genomes of several *Leishmania* species (including *L. mexicana*, *L. major* and *L. infantum*) contain two tandemly arrayed copies of the *SMT* gene. In *L. mexicana*, their sequences differ by a single base at position 391 (G in *LmxM.36.2380*, A in *LmxM.36.2390*), causing a conservative amino acid substitution at position 131 (V to I). The functional significance of increased copy number is unclear, although could be a way of increasing expression in the absence of transcriptional regulation. Interestingly, however, RNA-seq data show a twofold decrease in *LmxM.36.2380* expression on differentiation from promastigote to intracellular amastigote, whereas *LmxM.36.2390* remains unchanged (Fiebig *et al*. 2015). These data suggest stage-specific variation in relative importance of these two gene copies between life cycle stages,

although it must be noted that given the similarity of the two coding sequences, accurately differentiating between effects on each copy is challenging.

3. Lathosterol oxidase-like protein:

Also known as sterol C5-desaturase (SC5D), this enzyme introduces a 5(6)-double bond in the conversion of episterol to 5,7,24(28)-ergostatrienol, creating the characteristic double bond conjugation observed in ergostane-type sterols. Disruption of *erg3*, the gene encoding SC5D in pathogenic yeast *Candida albicans*, attenuated virulence and hyphenation (Chau *et al*. 2005), *C. albicans* clinical isolates carrying *erg3* mutations have nonetheless been isolated and were resistant to azole drugs (Martel *et al*. 2010), suggesting that *erg3* mutants can still retain pathogenicity. SC5D enzymes are polytopic membrane proteins that are dependent on oxygen and electrons donated by NADPH via cytochrome-b_5. Catalytic function is dependent on a series of highly conserved His-rich motifs, which may have a role in binding catalytic iron atoms (Taton *et al*. 2000). Substitution of two of these residues in *Arabidopsis thaliana* (H203L, H222L, H222E) dramatically reduced enzyme function (Taton *et al*. 2000).

1.2.4.3 Regulation of ergosterol biosynthesis and sterol content

Little is known about the regulation of ergosterol biosynthesis in kinetoplastids. In yeast, later stages occur at the endoplasmic reticulum and protein factors interact within a multisubunit complex known as the ergosome (Mo & Bard 2005). Transcription of ergosterol biosynthesis genes is regulated in fission yeast *Schizosaccharomyces pombe* by the sterol regulatory element binding protein (SREBP), along with other oxygen-requiring pathways (Todd *et al*. 2005), and gain of function mutations in expression regulators leading to overexpression of ergosterol biosynthesis genes are often observed in fluconazole resistance in *Candida* (Popp *et al*. 2017). Disruption of *erg3* in *C. parapsilosis* is also associated with increased expression of numerous ergosterol biosynthesis genes (Rybak *et al*. 2017), raising the possibility of end product-mediated regulation.

The unusual transcriptional biology of kinetoplastids means that transcriptional regulation of ergosterol biosynthesis is unlikely to be conserved between *Leishmania* and fungi. Transcriptomics data demonstrate decreased expression

of several sterol biosynthesis genes on differentiation to amastigotes in addition to *LmxM.36.2380* described above, namely (in order of pathway appearance) *LmxM.21.1430*, *LmxM.34.0050*, *LmxM.05.0180*, *LmxM.31.3310*, *LmxM.24.2110*, *LmxM.29.3190*, *LmxM.15.1460*, *LmxM.18.0020*, *LmxM.34.5330*, *LmxM.13.1620*, *LmxM.06.0650*, *LmxM.11.1100*, *LmxM.31.2320*, *LmxM.06.0350*, *LmxM.34.1230*, *LmxM.08_29.2140* and *LmxM.29.3550* (Fiebig *et al*. 2015). Only *LmxM.32.0680*, the last enzyme in the pathway, showed increased expression (2.6-fold). Complementing what appears to be a decrease in ergosterol biosynthesis genes in amastigotes, a marked drop of around 80% in ergosterol content has been noted in *L. infantum* and *L. donovani* on differentiation from promastigotes to amastigotes released from infected macrophages, accompanied by a substantial increase in the level of cholesterol (Bouazizi-Ben Messaoud *et al*. 2017). A similar increase in cholesterol content and decrease in various ergostane-type sterols has been noted during differentiation of log-phase *L. infantum* promastigotes to stationary and metacyclic forms (Yao & Wilson 2016). By contrast, a twofold rise in ergosterol, and increases in other biosynthetic intermediates, have been noted on differentiation of *L. braziliensis* from promastigotes to axenic amastigotes (Jara *et al*. 2017). Nonetheless, this was still accompanied by a greater relative increase (21-fold) in cholesterol. As *Leishmania* does not have a cholesterol biosynthesis pathway, cholesterol content is believed to be derived from the medium, and growth in cholesterol-free medium caused a > 90% drop in cholesterol content (Yao & Wilson 2016). Within macrophages, *L. mexicana* infection causes accumulation of free cholesterol around parasitophorous vacuoles, and labelled cholesterol is incorporated into parasite membranes (Semini *et al*. 2017). Furthermore, *L. mexicana* infection causes upregulation of macrophage SREBP target genes (Semini *et al*. 2017), and *L. donovani* activates host SREBP2 to increase HMG-CoA reductase-mediated cholesterol synthesis (Mukherjee *et al*. 2014). *L. amazonensis* promastigotes can also counteract azole toxicity through uptake of exogenous low density lipoprotein as a cholesterol source (Andrade-Neto *et al*. 2011).

Stage-specific regulation of sterol metabolism has implications for development of this pathway as a source of drug targets. While azoles such as ketoconazole and posaconazole, inhibitors of CYP51, have been considered as drugs and

frequently show potent activity *in vitro*, clinical trials have been markedly more equivocal (Emami *et al*. 2017). Moreover, a clinical trial of posaconazole in Chagas disease showed initial efficacy in clearance of *T. cruzi* infection, but during follow-up a far higher rate of relapse was observed in comparison to the currently used nitroheterocyclic compound benznidazole (Molina *et al*. 2014). Lack of complete clearance of intracellular *T. cruzi* parasites even after long-term exposure to high CYP51 inhibitor concentrations has also been noted *in vitro*, as well as high strain-dependent variability (Moraes *et al*. 2014). These observations could arise due to altered metabolism *in vivo* relative to *in vitro* assays, in part due to the increased reliance on exogenous cholesterol relative to parasite-derived ergosterol, as discussed above, but also the possible presence of quiescent parasite forms less reliant on *de novo* sterol synthesis. If true, such possibilities could represent serious problems for use of ergosterol biosynthesis inhibitors, which have proved so successful in control of fungal infection. It may be, however, that if ergosterol itself possesses a "growth-sparking", or signalling, role beyond its function as a bulk membrane sterol (as discussed at the start of section 1.2.4), it could be possible to target this pathway in such a way as to produce cidal effects in even cells that are metabolically quiescent, or where sterol content is largely host-derived. An improved understanding of the broader roles of ergosterol for cellular function is therefore required.

1.3 Antileishmanial drugs: use, mode of action, and resistance mechanisms

Antileishmanial chemotherapy relies on a small range of compounds, all of which face limitations with efficacy and/or toxicity (Figure 1-9). For several decades, pentavalent antimony-based compounds have been the treatment of choice for most forms of leishmaniasis, particularly sodium stibogluconate and meglumine antimoniate. However, these compounds exhibit severe toxicity, including cardiac arrhythmias, hepatotoxicity and pancreatitis (Sundar & Chakravarty 2013). Moreover, efficacy of these compounds has been in decline, particularly in the Indian subcontinent where visceral leishmaniasis treatment failures have been associated with increased resistance amongst *L. donovani* clinical isolates *in vitro* (Lira *et al*. 1999), eventually creating a necessity to abandon use of antimonial compounds in this region (Sundar *et al*. 2000). This has led to increased reliance on secondary compounds, as well as greater need for the use

of combination therapies and adoption of novel drugs. In particular, amphotericin B (AmB) is gaining prominence as the treatment of choice for visceral leishmaniasis and other forms. The various individual antileishmanial drugs are discussed in greater detail below.

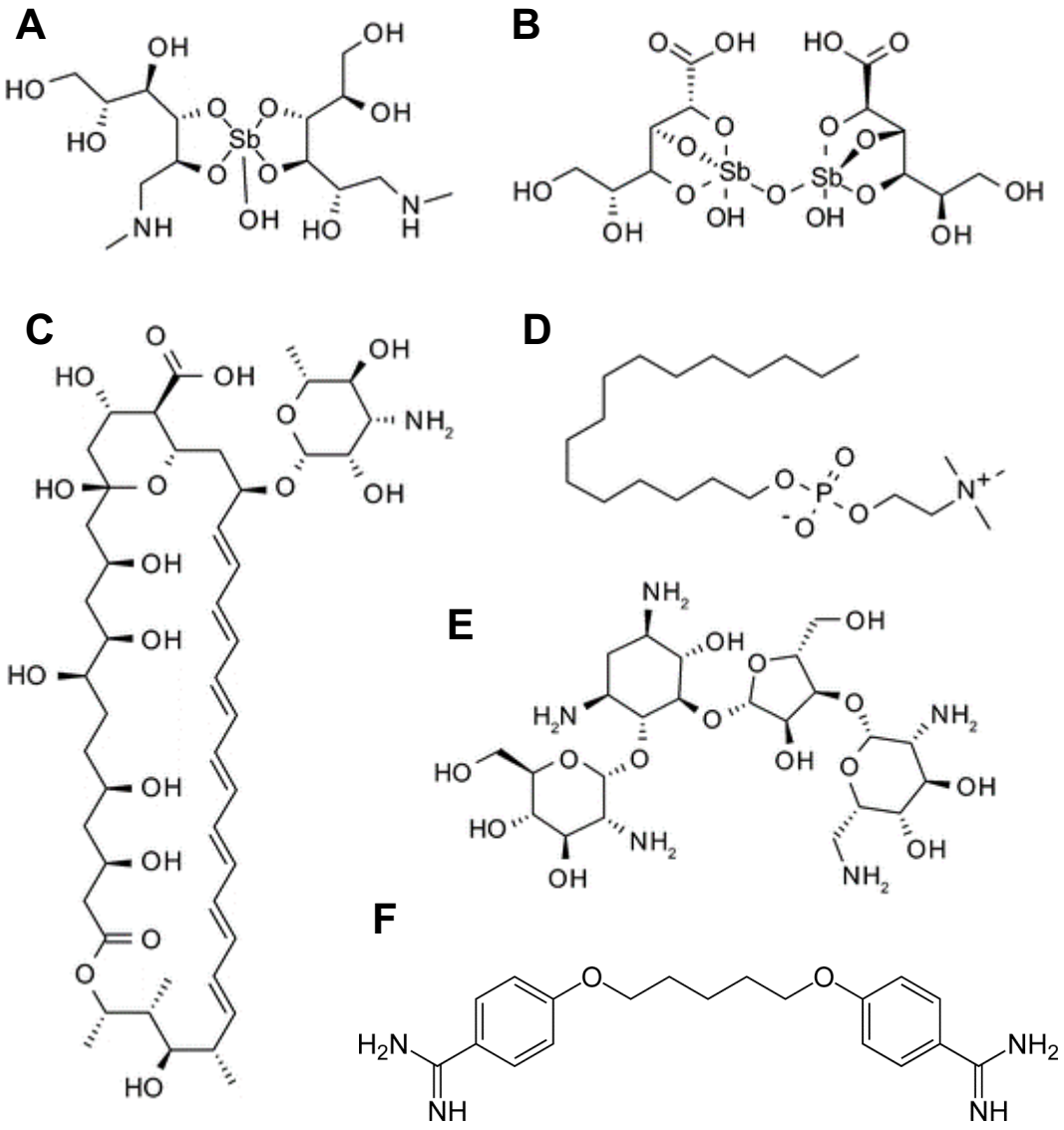


Figure 1-9: Chemical structures of antileishmanial drugs. Available treatment options for leishmaniasis. A) meglumine antimoniate; B) sodium stibogluconate; C) amphotericin B; D) miltefosine; E) paromomycin; F) pentamidine.

1.3.1 Pentavalent antimonials

Despite widespread use of pentavalent antimony-based compounds (Sb^{V}) for many years, their mechanism of action is not well understood. Activity depends on the reduction of Sb^{V} to trivalent antimony, Sb^{III}, as in the compound

potassium antimonyl tartrate (PAT). This may occur either in the amastigote (Shaked-Mishan *et al*. 2001) or in the host macrophage (Hansen *et al*. 2011), possibly due to an interaction with cellular thiols (dos Santos Ferreira *et al*. 2003). Uptake of Sb^{III} into *Leishmania* parasites is mediated by the aquaglyceroporin AQP1 (Gourbal *et al*. 2004), and once inside, Sb^{III} is known to generate oxidative stress and induce apoptosis (Moreira *et al*. 2011). This may be in part due to interference with cellular thiol metabolism, as Sb^{III} is known both to induce efflux of cellular trypanothione and glutathione in *L. donovani* and to inhibit trypanothione reductase, resulting in accumulation of the oxidised disulfide forms of these compounds (Wyllie *et al*. 2011). The nature of the interaction between Sb^{III} and trypanothione reductase has been solved by co-crystallisation in complex with reduced NADPH (Biaocco *et al*. 2009). However, immune factors may also play a role in antimonial activity *in vivo*, and Sb^{V} has been shown to activate macrophage production of reactive oxygen species (ROS) and nitric oxide production via phosphoinositide 3-kinase and mitogen-activated protein kinase signalling (Basu *et al*. 2006).

Resistance to antimonial compounds in both parasite lines selected for resistance *in vitro* and clinical isolates involves a number of mechanisms. Altered *AQP1* expression has been implicated in antimony-resistant *L. major* isolates (Gourbal *et al*. 2004), and population genomics has revealed the recombination-mediated spread of a frame shifted variant of the gene in Indian subcontinent *L. donovani* populations in which antimony resistance is prevalent (Imamura *et al*. 2016). Given the observed effects of antimonial compounds on thiol metabolism and oxidative stress, upregulation of tryparedoxin peroxidase has been noted in antimony-resistant *L. braziliensis* and *L. infantum* lines selected *in vitro* (Matrangolo *et al*. 2013), and resistant *L. donovani* isolates have shown increased expression of various thiol-metabolising enzymes, as well as increased cellular non-protein thiol content (Mukhopadhyay *et al*. 2011). Proteomics analysis of an *in vitro* selected antimony-resistant *L. infantum* line also revealed increased abundance of the ABC transporter multidrug-resistant protein (MRP) A, overexpression of which was sufficient to induce resistance in wild-type parasites (Brotherton *et al*. 2013).

Antimony-resistant clinical isolates also have enhanced capacity for immunomodulation, including through induction of macrophage interleukin (IL)-10 production and macrophage MRP1 expression (Mukherjee *et al*. 2013). This induction arose due to glycan mediated interactions with a host toll like receptor 2/3 heterodimer, which triggered signalling via extracellular signal-related kinase (ERK) and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB). Antimony-resistant *L. donovani* isolates may also be more adept at driving regulatory T cell responses (Guha *et al*. 2014). In part as a result of this immunomodulatory phenotype, as well as increased tolerance to oxidative stress, antimony-resistant *L. donovani* isolates have been suggested to possess greater fitness than their sensitive counterparts, in contrast to the standard dogma that drug resistance is associated with fitness costs (Vanaerschot *et al*. 2013). Resistant isolates display greater infectivity to macrophages (Carter *et al*. 2005) as well as higher parasite burden in mouse models (Vanaerschot *et al*. 2011), and patients with Sb^V treatment failure display higher pre-treatment parasite loads than those treated successfully (Thakur *et al*. 2003). These data suggest resistance to pentavalent antimonials is likely to persist long after the abandonment of use of these drugs.

1.3.2 Miltefosine

Miltefosine (MILT, also known as hexadecylphosphocholine) is a lysophospholipid analogue that was initially investigated for its role as an antitumour agent before its antileishmanial activity was discovered. As the only orally bioavailable antileishmanial currently in use, it has been widely administered in areas where resistance has limited the use of pentavalent antimonials, but faces limitations due to cost, frequent gastrointestinal and other adverse events, and its teratogenicity, which requires women of child-bearing age to observe contraception during and for three months after treatment (Sundar & Chakravarty 2013). Like antimonial drugs, the mechanism of MILT antileishmanial action is poorly understood, but induction of ROS production and apoptosis is known to play an important role (Moreira *et al*. 2011). However, an additional involvement of lipid metabolism has been observed: alterations to phospholipid composition are observed on MILT treatment of *L. donovani* (Imbert *et al*. 2012), and depletion of plasma membrane sterols in *L. donovani* impairs

MILT action, possibly by influencing lipid raft formation (Saint-Pierre-Chazalet *et al*. 2009).

Given the long half-life of this drug (~1 week), the requirement for lengthy treatment administration (28 days), risk of poor adherence given adverse gastrointestinal effects, and its widespread use as a monotherapy, there is a significant danger of emergence of MILT resistance (Sundar & Chakravarty 2013). In fact, since its registration for use in India in 2002, a decline in efficacy has been noted, with relapse rates of 20% 12 months after treatment observed in Nepalese VL (Rijal *et al*. 2013). While such trends can be down to a number of factors, including host immune responses and treatment compliance, evidence has emerged of decreased amastigote MILT sensitivity in VL-endemic regions of India in comparison to non-endemic regions (Prajapati *et al*. 2012), potentially as a result of higher drug pressures in endemic areas. Moreover, two unrelated Indian *L. donovani* strains (one Sb^{III}-resistant, the other sensitive) isolated from patients refractory to MILT treatment were recently found to exhibit greater than tenfold reduction in sensitivity to miltefosine relative to reference strain DD8, with resistance also observed in amastigotes (Srivastava *et al*. 2017). In both strains, mutations were found in the miltefosine transporter (MT), an ATP-dependent phospholipid translocase (*LmxM.13.1530* in *L. mexicana*). This protein forms a complex with its ß-subunit, called LdRos3 in *L. donovani* (Pérez-Victoria *et al*. 2006), and mutation or altered expression of either subunit can lead to reduced MILT sensitivity, associated with reduced uptake of the drug (Pérez-Victoria *et al*. 2006, Seifert *et al*. 2007, Laffitte *et al*. 2016, Shaw *et al*. 2016, Vacchina *et al*. 2016). In addition to effects on MILT accumulation, disruption of MT function leads to altered lipid composition, particularly with regard to phosphatidylcholines and lysophosphatidylcholines (Shaw *et al*. 2016). Resistance to MILT has also been associated with decreased ROS production in response to oxidative stress-inducing agents and upregulation of genes associated with defence against oxidative stress (Das *et al*. 2013). *L. donovani* isolates from MILT relapse patients also demonstrated reduced MILT sensitivity *in vitro*, reduced MILT accumulation and better tolerance to oxidative stress in comparison with pre-treatment isolates from the same patients (Deep *et al*. 2017). Intriguingly, these relapse-derived isolates also showed increased infectivity of mouse

peritoneal macrophages, suggesting higher fitness in drug-resistant parasites could be observed for MILT as well as for Sb^v.

1.3.3 Paromomycin

Paromomycin (PMM) is a parenterally-administered aminoglycoside antibiotic which has been in use for Indian VL since 2006. It has also shown some efficacy within a topical formulation for Old and New World CL treatment (Sundar & Chakravarty 2013). Aminoglycosides are known to inhibit translation, and PMM treatment led to decreased cytoplasmic and mitochondrial protein synthesis in *L. donovani* (Jhingran *et al*. 2009); additionally, PMM was shown to inhibit translation by *L. mexicana* ribosomes *in vitro*, and bound specifically to *Leishmania* rRNA where no binding was observed to mammalian rRNA (Fernández *et al*. 2010). A decrease in mitochondrial membrane potential after PMM treatment has also been observed (Jhingran *et al*. 2009). In contrast to many other antileishmanial compounds, however, PMM does not trigger ROS production (Moreira *et al*. 2011).

One important limitation of PMM is the apparent ease with which resistance can be selected. Experimental selection of resistance has been achieved both in promastigotes (Jhingran *et al*. 2009, Bhandari *et al*. 2011) and intracellular amastigotes (Hendrickx *et al*. 2012), as well as *in vivo* by serial passage through three Syrian golden hamsters (Hendrickx *et al*. 2015). Resistant lines exhibit reduced PMM uptake (Jhingran *et al*. 2009) and increased expression of ABC-type efflux transporters MRPA and multidrug resistance gene 1 (MDR1) (Bhandari *et al*. 2011). Greater resistance to complement-mediated lysis and increased expression of IL-10 in infected host cells have also been noted in PMM-resistant lines, as has greater tolerance to nitrosative (but not oxidative) stress (Bhandari *et al*. 2011). Mechanisms of resistance may be stage-dependent as PMM-resistant cells selected as amastigotes did not retain resistance in the promastigote stage (Hendrickx *et al*. 2012). PMM-resistant *L. infantum* selected as intracellular amastigotes have also shown evidence of enhanced infectivity relative to wild-type cells (Hendrickx *et al*. 2016). The ease of resistance selection supports the current use of PMM as a combination therapy, in which PMM has an additional advantage due to lack of evidence of cross-resistance to other drugs in a PMM-resistant line (Bhandari *et al*. 2014).

1.3.4 Pentamidine

Pentamidine (PENT) is an aromatic diamidine widely used to treat human African trypanosomiasis as well as the fungal disease pneumocystis pneumonia. It has been used to treat both VL and CL, although declining efficacy and high toxicity (with adverse effects including hypoglycaemia, nephrotoxicity and hypotension) has limited its use (Sundar & Chakravarty 2013). Uptake of PENT has been studied more extensively in *T. brucei* where it is facilitated by multiple transporters, including the P2 aminopurine transporter, TbAT1 and aquaglyceroporin 2 (Carter *et al*. 1995, De Koning 2001, Munday *et al*. 2014), although aquaglyceroporin 2 may mediate uptake by acting as a PENT receptor for endocytosis (Song *et al*. 2016). Once inside the cell, PENT accumulates inside the mitochondrion and appears to be able to bind to kinetoplast DNA through its minor groove binding activity, in turn specifically inhibiting kinetoplast DNA replication (Yang *et al*. 2016). A reduction in membrane polar lipids and intracytosolic neutral lipids has also been observed after PENT treatment in *Leishmania* (Basselin & Robert-Gero 1997), as well as reduced intracellular polyamines (Basselin *et al*. 1997), consistent with non-competitive inhibition of polyamine uptake by this drug (Basselin *et al*. 2000). Experimental resistance to PENT is associated with reduced accumulation within the mitochondrion, resulting from lower mitochondrial membrane potential (Basselin *et al*. 2002), and alteration of kinetoplastid DNA and reduction of AT-rich binding sites have been observed (Basselin *et al*. 1998). In addition, increased membrane fluidity has been noted in PENT-resistant *L. donovani* and *L. amazonensis*, and while membrane polar lipids and intracytoplasmic neutral lipids were lower than in wild-type parasites, overall phospholipid content increased in resistant lines; interestingly this comprised of a reduction in membrane phospholipid but increase in cytoplasmic phospholipid (Basselin & Robert-Gero 1997). Finally, altered polyamine metabolism and decreased ornithine decarboxylase activity have been noted in resistant lines (Basselin *et al*. 1997).

1.3.5 Amphotericin B

Amphotericin B (AmB) is a polyene macrolide antifungal drug that has been used for many years as an antileishmanial agent. Previously, when administered as AmB deoxycholate, its use has been limited by severe adverse effects, including

nephrotoxicity, hypokalemia and myocarditis, thus requiring hospitalisation during treatment (Sundar & Chakravarty 2013). However, in recent decades, novel lipid formulations have been developed which reduce toxicity; of these, the most widely used is a liposomal formulation produced by Gilead Sciences, called AmBisome. This drug has particularly been used to treat Indian VL, although efficacy against CL has also been shown in some cases (Sundar & Chakravarty 2013), and it has been recommended as the treatment of choice for the VL elimination programme in India (Singh *et al*. 2016). A limitation of AmBisome is its high cost; however, in 2009 Gilead agreed a preferential pricing agreement with the World Health Organisation to supply AmBisome to developing countries at $18 per 50 mg vial (Moon *et al*. 2011). More recently, Gilead has agreed to donate the drug, an agreement which was renewed in 2016 with Gilead's commitment to provide 380,000 vials to a number of VL-endemic countries including Bangladesh, Ethiopia, India, Nepal, South Sudan and Sudan (Gilead Sciences 2016). One development which has reduced cost further and simplified treatment regimens is the use of 10 mg/kg AmBisome as a single dose, which was shown to be non-inferior and less expensive, with fewer adverse events, than a conventional course of AmB deoxycholate therapy in Indian VL (Sundar *et al*. 2010). However, widespread use of AmBisome as a single-infusion monotherapy creates the risk of incomplete parasite clearance, leading to emergent drug resistance. Therefore, understanding the mechanisms by which resistance can arise to this drug is a priority.

1.3.5.1 Modes of action of amphotericin B

The principal basis of the antileishmanial activity of AmB is its specific affinity for ergostane-type sterols, as found in the membranes of *Leishmania* and fungi (Figure 1-10A). AmB has been shown to induce nonlethal potassium ion leakage in *L. donovani*, but at concentrations greater than 100 nM, it can self-assemble to form aqueous pores that conduct both cations and anions, leading to loss of osmotic homeostasis and cell lysis (Ramos *et al*. 1996). These aqueous pores are approximately 5 Å in radius, and can conduct water, ions, and other polar small molecules such as glycerol and urea (Andreoli 1974). This is facilitated by the amphipathic structure of AmB. Studies on synthetic liposomes indicate that introduction of ergosterol into membranes, but not cholesterol, facilitates AmB-induced permeability changes (HsuChen & Fiengold 1973). In this study,

introduction of episterol, which possesses an identical side chain to ergosterol but lacks the 5(6)-double bond, failed to facilitate AmB-induced permeability changes, whereas 5,7-cholestadienol, which shares its side chain with cholesterol but contains ergosterol-type 5(6)-7(8) double bond conjugation, permitted AmB-mediated permeabilisation. These data suggest that 5(6)-7(8) desaturation is more important for AmB activity than C24-methylation. More recently, however, the ergosterol side chain has been shown in liposomes to be important for AmB-mediated ion conductance (Nakagawa *et al*. 2014). In particular, both C24-methylation and the 22(23)-double bond were required for maximal activity. This is an interesting observation, given that 5,7,24(28)-ergostatrienol has been detected as the dominant sterol in *Leishmania* promastigotes (de Souza & Fernandes Rodrigues 2009, Andrade-Neto *et al*. 2016), and that this sterol possesses C24-methylation but lacks the 22(23)-double bond. However, it is possible that the 24(28)-double bond can act in a similar fashion, something that was not tested by the authors. While many *in vitro* studies have interrogated factors leading to aqueous channel formation, lower concentrations of AmB are fungicidal without inducing pore formation, acting simply through binding to and sequestering AmB within fungal membranes (Gray *et al*. 2012) (Figure 1-10B).

More broadly, AmB-mediated permeabilisation is dependent not only on the specific sterol present, but also phospholipid content, and interactions between sterol and phospholipid within the membrane (Bolard *et al*.1980). As a lipophilic molecule, AmB is predicted in molecular dynamics simulations to insert into the membrane independent of sterol content (Neumann *et al*. 2016). However, its ability to cause membrane permeabilisation is highly dependent on membrane fluidity, influenced by both sterol and phospholipid content (HsuChen & Fiengold 1973); hence AmB activity decreases with increased membrane fluidity.

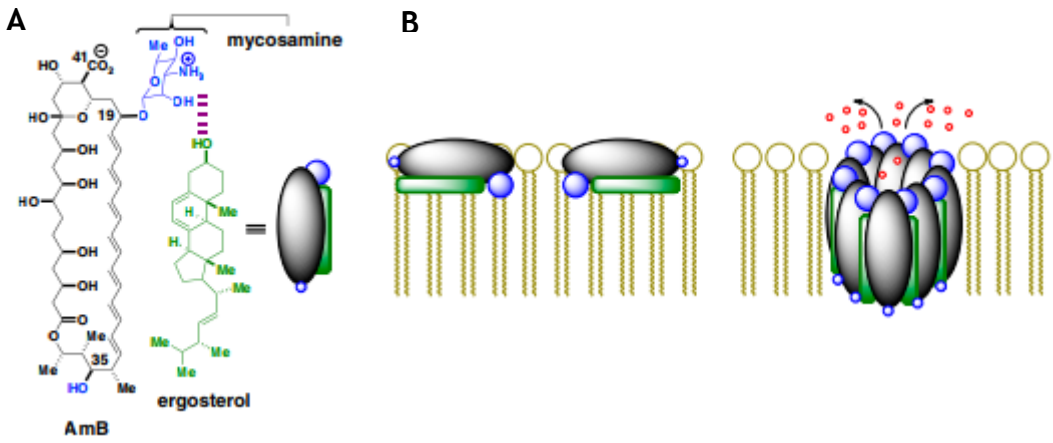


Figure 1-10: The interaction of AmB with ergosterol-containing membranes. A) Ergosterol interacts with AmB both through hydrophobic interactions and hydrogen bonding between the 3-hydroxyl group of ergosterol and the mycosamine iminosugar group. B) AmB-mediated membrane disruption may either be due to direct binding and sequestering of ergosterol within the membrane or through formation of AmB-ergosterol hydrophilic pores. Pores allow movement of ions and other small molecules (red circles) across the membrane (source: Gray *et al*. 2012).

AmB has also been associated with induction of oxidative stress, both in *Leishmania* parasites (Moreira *et al*. 2011) and in several species of fungi (Mesa-Arango *et al*. 2014). The molecular basis of ROS production has not been elucidated in *Leishmania*, and while AmB is known to generate radicals as a result of autooxidation (Lamy-Freund *et al*. 1985), the physiological relevance of this phenomenon has not been demonstrated. Rotenone-mediated inhibition of the mitochondrial respiratory chain prevented ROS generation and protected against AmB-dependent killing in yeast (Mesa-Arango *et al*. 2014) and in *Aspergillus* fungi (Shekhova *et al*. 2017), suggesting that AmB-induced ROS is of mitochondrial origin. A systems biology approach revealed activation of a signalling cascade in AmB-treated *C. albicans* and *S. cerevisiae*, involving both guanosine triphosphatases Ras1 and Ras2, as well as protein kinase A, leading to tricarboxylic acid cycle- and respiration-dependent ROS production (Belenky *et al*. 2013), consistent with a central role for the mitochondria. These data suggest that ROS generation occurs downstream of action at the membrane. However, whilst exogenous catalase (which detoxifies H_2O_2) did not protect against AmB-mediated K^+ leakage in *C. albicans*, it did prevent lethality, as well as release of [^{3}H]leucine introduced into cells by preincubation (Sokol-Anderson *et al*. 1986). While this demonstrates that permeabilisation of the membrane to K^+ ions occurs upstream or independently of ROS generation, it is unclear how leucine release relates to the formation of aqueous pores. Nevertheless, lytic

activity of AmB was far less pronounced under hypoxic conditions in comparison to air (Sokol-Anderson *et al*. 1986), providing further confirmation of the role of oxygen in AmB-mediated cell death.

As with antimony-based compounds, there is evidence that AmB also acts directly on host cells. Primary murine macrophages treated with AmB exhibited increased H_2O_2 production, and nitrite (spontaneously formed from nitric oxide) was also produced on stimulation with interferon-γ; tumour necrosis factor-α and interleukin-12 were also more abundant in treated mice (Ghosh *et al*. 2013). Immunodulation by AmB results from its activation of signalling by toll-like receptor (TLR)-1 and TLR-2, leading to the release of proinflammatory cytokines (Sau *et al*. 2003, Razonable *et al*. 2005). While TLR-2 signalling can underlie AmB toxicity, liposomal AmB diverts signalling in neutrophils from TLR-2 to TLR-4, producing fewer pro-inflammatory effects while still activating neutrophils to an antifungal state (Bellocchio *et al*. 2005). This may partially explain better tolerability of liposomal forms of the drug. Finally, it has been proposed that AmB may prevent *Leishmania* attachment to macrophages by direct interaction with macrophage cholesterol (Paila *et al*. 2010); however, while the authors showed pre-incubation of macrophages with AmB followed by washing of the drug did reduce subsequent *L. donovani* infection, the concentrations for activity (> 5 μM) were well above antileishmanial concentrations (~100 nM), and even above those known to reduce primary murine macrophage viability (> 1 μM) (Ghosh *et al*. 2013), and so both residual drug and host cell toxicity are plausible alternative explanations for the phenomena described. Mechanisms of AmB action are summarised in Figure 1-11.

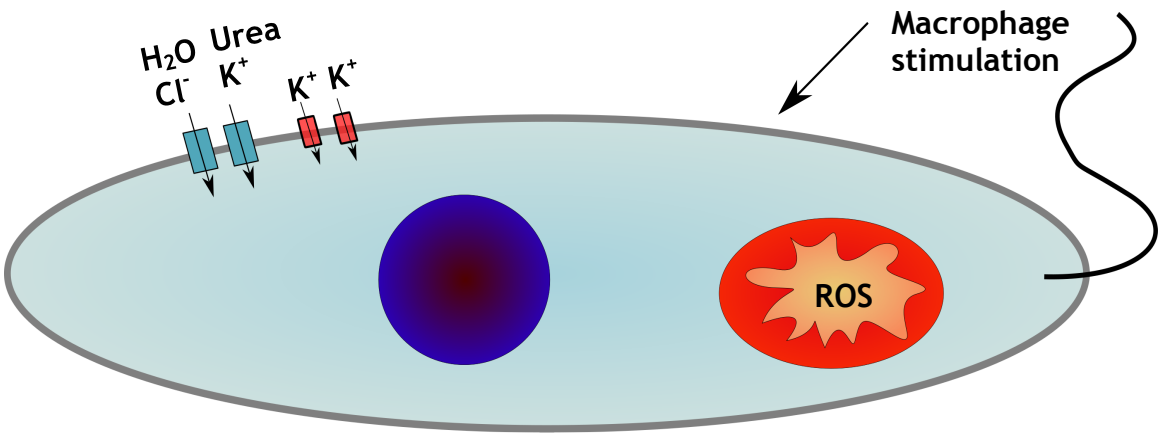


Figure 1-11: Summary of AmB mechanism of action. AmB-dependent cell killing is mediated by multiple processes, including the formation of two different types of channels (cation selective at low concentrations (in red), aqueous at > 100 nM (in turquoise)), as well as mitochondrion-dependent ROS generation and direct stimulation of host responses. The cell nucleus and mitochondrion are signified in dark blue and red, respectively.

1.3.5.2 Amphotericin B resistance

While AmB has been widely used for decades as an antifungal compound, development of resistance in fungi has been restricted to isolated events, in contrast to other antifungal classes such as the azoles. However, fungal pathogens other than the traditionally dominant *C. albicans* are emergent, and display more widespread resistance to AmB, including *C. glabrata* (Amirrajab *et al.* 2016) and *C. auris* (Lockhart *et al.* 2016), suggesting that resistance may develop more easily in different genetic backgrounds. Indeed, *L. donovani* strains derived from patients experiencing AmB treatment failure have shown reduced susceptibility to the drug after isolation (Srivastava *et al.* 2011, Purkait *et al.* 2012), signifying that the emergence of resistance in *Leishmania* is a real concern.

AmB resistance has been more comprehensively studied in yeasts. The most pronounced feature observed in multiple *Candida* species is a loss of ergostane-type sterols from the membrane, along with mutations in the ergosterol biosynthesis pathway (Sokol-Anderson *et al.* 1988, Young *et al.* 2003, Chau *et al.* 2005, Vandeputte *et al.* 2007, Forastiero *et al.* 2013, Vincent *et al.* 2013, Rybak *et al.* 2017). In cases where changes to sterol composition have been observed, typically ergosterol is replaced with alternative sterols, varying dependent on the nature of the lesion. For example, mutations in *erg11* (lanosterol 14α-demethylase, *LmxM.11.1100* in *L. mexicana*) associated with cross resistance to

both azoles and AmB in *C. tropicalis* resulted in loss of ergosterol and increased abundance of C14-methylated sterols, specifically the ergosterol biosynthetic intermediate lanosterol, and another, 14-methylfecosterol (Figure 1-12A) (Forastiero *et al*. 2013). Virulence in a caterpillar larva model of infection was similar to drug-sensitive lines, but greater susceptibility to brefeldin A and hygromycin B was noted, suggested by the authors as evidence of possible increased membrane permeability. By contrast, mutation of *erg3* (sterol C5-desaturase, *LmxM.23.1300*) in *C. albicans* is associated with loss of ergosterol and accumulation of ergosta-7,22-dienol (Figure 1-12B), in which the 5(6)-double bond introduced by this enzyme is absent (Chau *et al*. 2005). *C. albicans erg3* mutants were found to exhibit reduced virulence in mice and impaired filament formation, although fungal burden in mouse kidneys remained unaffected. Mutation in other enzymes has also been associated with loss of ergosterol in clinical isolates including *erg2* (sterol C8-isomerase, *LmxM.08_29.2140*) in *C. albicans* (Vincent *et al*. 2013) and *erg6* (sterol C24-methyltransferase, *LmxM.36.2380*/*LmxM.36.2390*) in *C. glabrata* (Vandeputte *et al*. 2007). In the former case, *erg2* mutations were associated with greatly reduced virulence and tissue invasion, increased sensitivity to oxidative stress and defective filament formation (Vincent *et al*. 2013). Similar phenotypes were observed in a line exhibiting mutation in both *erg11* and *erg3* (Vincent *et al*. 2013). Mutation in *erg6* was also associated with reduced virulence and growth rate, as well as thinning of the cell wall (Vandeputte *et al*. 2007). While these methods of sterol identification and quantification have largely relied on gas chromatography-mass spectrometry (GC-MS), the conjugation of 5(6)- and 7(8)-double bonds in ergostane-type sterols confers a distinctive ultraviolet (UV)-range spectrum (Sokol-Anderson *et al*. 1988). This allows distinction between sterols with and without this double bond conjugation (Figure 1-13). Mutation in *erg2*, as well as double inactivation of *erg3* and *erg11* results in loss of this spectrum (Vincent *et al*. 2013), whereas it is retained in *erg6* mutants (Vandeputte *et al*. 2007). Thus pathogenic fungi exhibiting altered sterol composition associated with AmB resistance have been observed in clinical isolates, but these are frequently associated with fitness costs and reduced virulence.

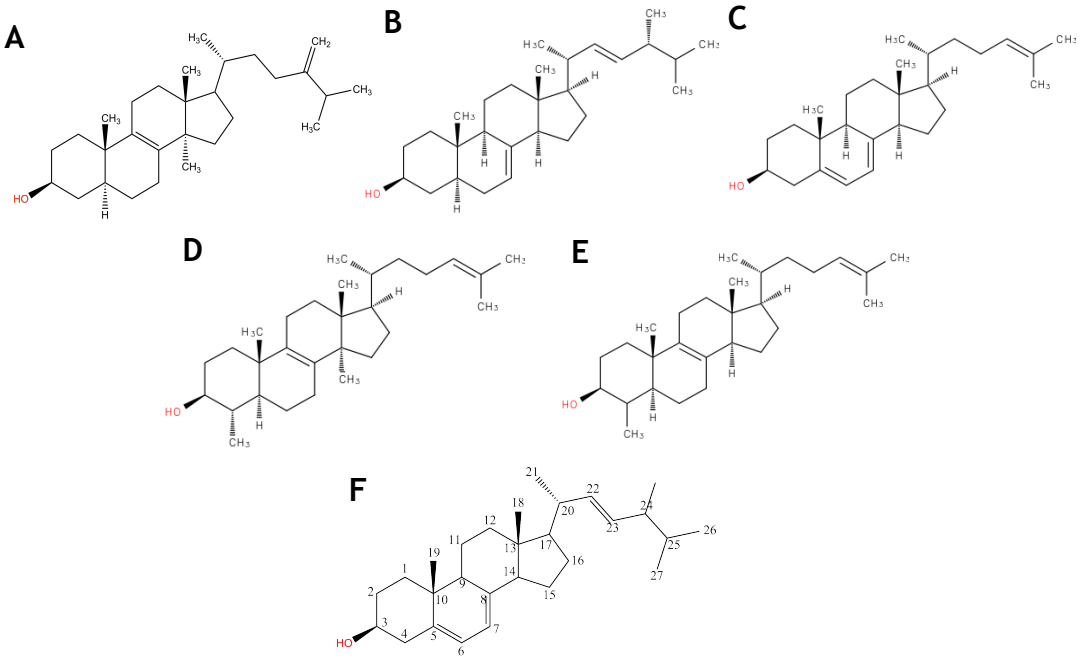


Figure 1-12: Accumulating sterols in AmB-resistant *Leishmania* and fungi. A) 14-methylfecosterol (Forastiero *et al*. 2013, Mwenechanya *et al*. 2017); B) ergosta-7,22-dienol (Chau *et al*. 2005); C) cholesta-5,7,24-trienol (Mbongo *et al*. 1998, Purkait *et al*. 2012); D) 4,14-dimethylcholesta-8,24-dienol (Al-Mohammed *et al*. 2005); E) 4-methylzymosterol (Fernandez-Prada *et al*. 2016). Note that all lack C24-methylation, 5(6)-desaturation, or both. F) The structure of ergosterol, with all carbon positions numbered as refered to in the text.

Where studies have been conducted on sterol biosynthesis in AmB-resistant *Leishmania* lines, these have mostly been on lines selected for resistance *in vitro* (Mbongo *et al*. 1998, Pourshafie *et al*. 2004, Al-Mohammed *et al*. 2005, Fernandez-Prada *et al*. 2016, Mwenechanya *et al*. 2017), although analysis has also been performed on a resistant clinical isolate (Purkait *et al*. 2012). In all cases, loss of ergostane-type sterols was described, but overall sterol composition varied between lines. In *L. donovani* promastigotes selected for AmB resistance *in vitro* (Mbongo *et al*. 1998) and derived from an AmB non-responsive patient (Purkait *et al*. 2012), cholesta-5,7,24-trienol (Figure 1-12C) was the predominant sterol. By contrast, in resistant *L. mexicana* promastigotes and amastigotes, the major sterol component was 4,14-dimethylcholesta-8,24-dienol (Figure 1-12D) (Al-Mohammed *et al*. 2005), whereas in *L. infantum* promastigotes, 4-methylzymosterol accumulated (Figure 1-12E)(Fernandez-Prada *et al*. 2016). Most recently, an AmB-resistant *L. mexicana* line was shown to accumulate primarily 14-methylfecosterol (Figure 1-12A) (Mwenechanya *et al*. 2017).

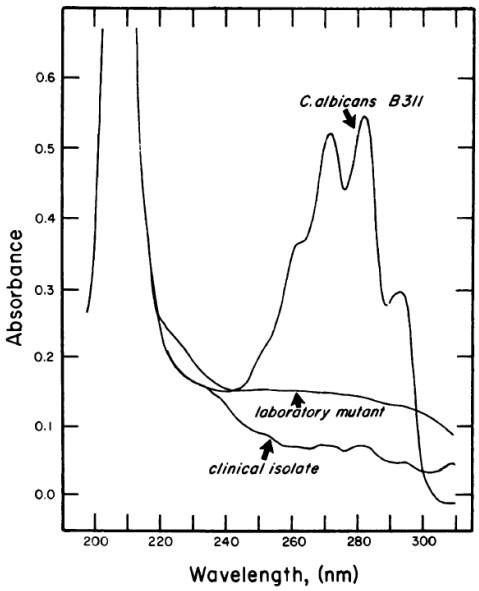


Figure 1-13: Loss of UV absorbance in AmB-resistant *Candida*. Experiment from Sokol-Anderson *et al* 1988, in which UV absorbance of *C. albicans* sterol extracts is measured. While wild-type yeast display a distinctive UV spectrum, this is lost in both resistant lines.

For the most part, the genetic basis of these changes have not been characterised in *Leishmania*. Pourshafie and colleagues identified altered *SMT* transcription in AmB-resistant *L. donovani* promastigotes (Pourshafie *et al*. 2004). They identified two transcripts, *SCMT1A* and *SCMT1B,* differing particularly in the 3'-UTR. *SCMT1A* was highly expressed in wild-type cells but absent in the resistant line, whereas *SCMT1B* was only marginally detectable in wild-type but upregulated in resistant parasites. Introduction of a spliced leader RNA was not detectable in *SMT* from AmB-resistant cells, suggesting that this transcript may also have been abnormally spliced. Purkait and colleagues found evidence of a similar change in their *L. donovani* clinical isolate, with loss of one transcript and upregulation of the other (Purkait *et al*. 2012). In neither of these cases was a role in resistance directly verified, although the accumulation of cholesta-5,7,24-trienol in this clinical isolate, which lacks C24-methylation, is consistent with loss of function in this enzyme. However, the actual genetic basis of the changes observed to *SMT* expression has not been identified, complicating the use of this as a potential marker for resistance. In the *L. mexicana* promastigote line selected by Mwenechanya and colleagues a single non-active site mutation, N176I, was identified in *CYP51* (Mwenechanya *et al*. 2017). Replacement of this enzyme by overexpression of the wild-type copy restored ergosterol production and AmB sensitivity. Interestingly, in this case, whilst increase of 14-methylergosta-8,24(28)-dienol and 4,14-dimethylergosta-

8,24(28)-dienol were clearly increased in GC-MS analysis, another sterol with a mass consistent with 4,4-dimethyl-5α-cholesta-8,14,24-trienol was also detectable in AmB-resistant cells, at an abundance of 6.7%. This sterol, which was not detectable in wild-type cells, is actually the product of the CYP51 reaction. The accumulation of a product of the mutated enzyme was unexpected and raised the possibility that disruption of the pathway due to the N176I mutation actually occurs downstream of CYP51, potentially through disruption of protein-protein interactions with other enzymes in the pathway required for function. Notably, however, ergosterol was not completely lost in this line (declining from 86% to 13%), and hence some pathway function was likely retained. An important consideration in interpretation of these data is the difficulty of correctly identifying sterols, given the abundance of highly similar isomers. This is discussed further (see section 1.4.1.2). Overall, however, in all cases the major sterol in AmB-resistant *Leishmania* lines appears to lack either C24-methylation or 5(6)-desaturation, or in some cases both, highlighting the importance of these structures for AmB activity.

As with the mechanism of action of AmB itself, AmB resistance appears to involve more than just altered ergosterol biosynthesis (Figure 1-14). In accordance with a role for phospholipid composition and overall membrane fluidity, these parameters have been observed to be altered in resistant lines. *Kluyveromyces lactis* yeast selected for reduced AmB sensitivity displayed high membrane fluidity and a decreased sterol/phospholipid ratio, along with a relative increase of monounsaturated fatty acids in comparison to saturated ones (Younsi *et al*. 2000). By contrast, while AmB-resistant *L. donovani* exhibited increased membrane fluidity, fatty acid saturation in these lines increased from 4% to 55% (Mbongo *et al*. 1998). This contrast may arise from species-specific differences, but may also emphasise the importance of sterol-phospholipid interactions in determining overall properties, as while ergosterol production was severely compromised in the *L. donovani* line, AmB-resistant *K. lactis* exhibited only a minor reduction in ergosterol content. Overexpression of plasma membrane proteolipid 3 was also noted in AmB-resistant *S. cerevisiae*, found to mediate resistance through alterations to sphingolipid metabolism (Bari *et al*. 2015). Finally, mutations in the miltefosine transporter were correlated with AmB-MILT cross-resistance in *L. infantum*, associated with extensive modulation

of phospholipid content, particularly increases in phospholipids containing cyclopropanated fatty acid substituents, as well as increased inositolphosphoceramide species (Fernandez-Prada *et al*. 2016). Interestingly, the mutations in MT in this line differed from those in a line selected for MILT resistance as they had a negligible effect on MILT uptake. As AmB acts on the plasma membrane itself, the effect on AmB sensitivity was proposed to be indirect through altered membrane properties, rather than through loss of MT-mediated AmB transport.

Another factor that has been described is altered response to oxidative stress. These changes may arise both because AmB-induced lethality is ROS-mediated, but also because it appears loss of ergosterol itself creates such stresses. AmB-resistant clinical *C. albicans* isolates carrying ergosterol biosynthesis mutations showed reduced virulence and greater susceptibility to killing by neutrophils, as well as far higher sensitivity to hypochlorous acid (an oxidising agent) and NO (Vincent *et al*. 2013). Similarly, AmB-resistant *L. mexicana* showed increased sensitivity to oxidative stress-inducing agents H_2O_2 and methylene blue (Mwenechanya *et al*. 2017). Additionally, elevation in ergosterol itself has been suggested as protective against treatment with Sb^{III} (which is known to induce ROS) in *L. donovani* promastigotes (Mathur *et al*. 2015). Hence it is not surprising that increased expression of factors associated with oxidative stress are observed in AmB-resistant organisms. AmB-resistant *C. albicans* lines, one selected *in vitro* and the other a clinical isolate, showed increased resistance to H_2O_2 and menadione, which induces superoxide ($O_2^{\cdot-}$) production (Sokol-Anderson *et al*. 1988). This was associated with increased expression of catalase. As ROS produced by AmB treatment is proposed to be derived from the mitochondrion, resistant *C. tropicalis* also showed a reduced mitochondrial respiration rate (Mesa-Arango *et al*. 2014). Proteomics analysis revealed increased expression of several enzymes associated defence against oxidative stress, including tryparedoxin peroxidase and trypanothione reductase (Brotherton *et al*. 2014); interestingly, however, neither of these induced resistance when overexpressed in wild-type cells. The same proteomics dataset showed extensive alterations in central carbon metabolism, so it may be that in order for increased expression of trypanothione metabolism genes to mediate resistance there must be a concomitant change in other pathways. Increased expression of PPP genes

glucose 6-phosphate dehydrogenase and transaldolase (Figure 1-4) was in itself able to reduce AmB susceptibility (Ghosh *et al*. 2015), presumably due to the role of the PPP as a key source of NADPH for trypanothione-mediated ROS detoxification. The AmB-resistant *L. donovani* clinical isolate described by Purkait and colleagues also showed increased expression of a range of enzymes in polyamine and trypanothione metabolism (Purkait *et al*. 2012), as well as upregulation of ascorbate peroxidase, a direct role for which was confirmed by induction of AmB-resistance in wild-type parasites overexpressing this gene (Kumar *et al*. 2014).

A range of other adaptations have been described. Proteomics of AmB-resistant *L. infantum* demonstrated increased expression of numerous chaperones including heat shock proteins Hsp60, Hsp70 and Hsp83 (Brotherton *et al*. 2014). Vincent and colleagues showed that in AmB-resistant *C. albicans*, high Hsp90 expression was required for survival, even in the absence of AmB (Vincent *et al*. 2013). Altered expression of the drug efflux transporter MDR1 has been observed in resistant *L. donovani*, with the resistance phenotype partially reversed by use of MDR1 inhibitors (Purkait *et al*. 2012). This may have been in part due to upregulation of silent information regulator 2, an NAD^+-dependent deacetylase involved in signalling, in this line (Purkait *et al*. 2015). Overexpression of this gene in wild-type cells increased MDR1 expression and AmB efflux, while reducing AmB sensitivity and AmB-mediated ROS generation. Finally, in a study specifically designed to detect mediators of AmB persistence in yeast, disruption of the target of rapamycin complex 1 (TORC1) pathway was found to increase persistence (Bojsen *et al*. 2016). Persistence can be defined differently from resistance and tolerance as follows: drug resistance explicitly refers to increased ability to grow in the presence of a given drug, whereas tolerant organisms are refractory to drug-mediated killing (i.e. they may not be able to grow during drug treatment, but show improved survival and recovery after treatment); persistence, by contrast, refers to survival of a subpopulation after high doses of drug treatment (Delarze & Sanglard 2015). Interestingly, if this subpopulation is isolated and expanded, organisms do not show increased resistance or tolerance relative to the pre-treatment population. The fact that persister subpopulations increase in populations with disrupted TORC1 signalling (either genetically or through treatment with the inhibitor rapamycin) suggests that reduced

translation rates (which are positively regulated by this pathway) may mediate increased persistence (Bojsen *et al*. 2016), even if it does not appear to mediate AmB-resistance.

Overall, therefore, data from both *Leishmania* and fungi suggest that AmB resistance is a multifaceted phenomenon (Figure 1-14). While disruption to ergosterol biosynthesis appears to be largely conserved, this can result from mutation in a range of genes in the pathway, resulting in variable sterol content in resistant lines, and in *Leishmania*, limited identification of genetic changes associated with these alterations has been achieved. While increased expression of oxidative stress response pathways has been noted repeatedly, some lines appear to show increased resistance to ROS (Sokol-Anderson *et al*. 1988) whereas others show hypersensitivity (Vincent *et al*. 2013, Mwenechanya *et al*. 2017); the factors underlying these discrepancies are unclear. What is most concerning is an apparent potential for cross-resistance: lesions in MT appear to directly mediate AmB-MILT cross-resistance (Fernandez-Prada *et al*. 2016), whereas upregulation of oxidative stress defences, chaperones and MDR1 have all been noted in antimony resistance. Selection of resistance to drug combinations, including with AmB, has been achieved in *L. donovani* promastigotes, and this resistance, retained in intracellular amastigotes, is associated with increased thiol abundance (García-Hernández *et al*. 2012). Cross-resistance to MILT and Sb^{III} has also been noted in AmB-resistant *L. mexicana* promastigotes (Mwenechanya *et al*. 2017). Clinically, a correlation has been noted in *L. donovani* field isolates between Sb^{V} sensitivity and both AmB and MILT susceptibility (Kumar *et al*. 2009). This raises the possibility that pre-selection of antimony resistance could facilitate loss of AmB susceptibility, exacerbated by the suggestion that pre-exposure of parasites to arsenic, which frequently contaminates water supplies in parts of India where VL is endemic, may facilitate the development of antimony resistance (Perry *et al*. 2013). Furthermore, while combination therapies are being explored (Sundar & Chakravarty 2013), there is a significant danger that resistance may emerge while AmB is distributed as a monotherapy (in the form of AmBisome or otherwise), prior to implementation of any combination strategies. This would render these approaches ineffective, and even risk the emergence of multidrug resistance, not only because of observed cross-resistance between AmB and

other drugs, but because treatment of AmB-resistant strains with AmB-containing combinations would leave the partner drug exposed.

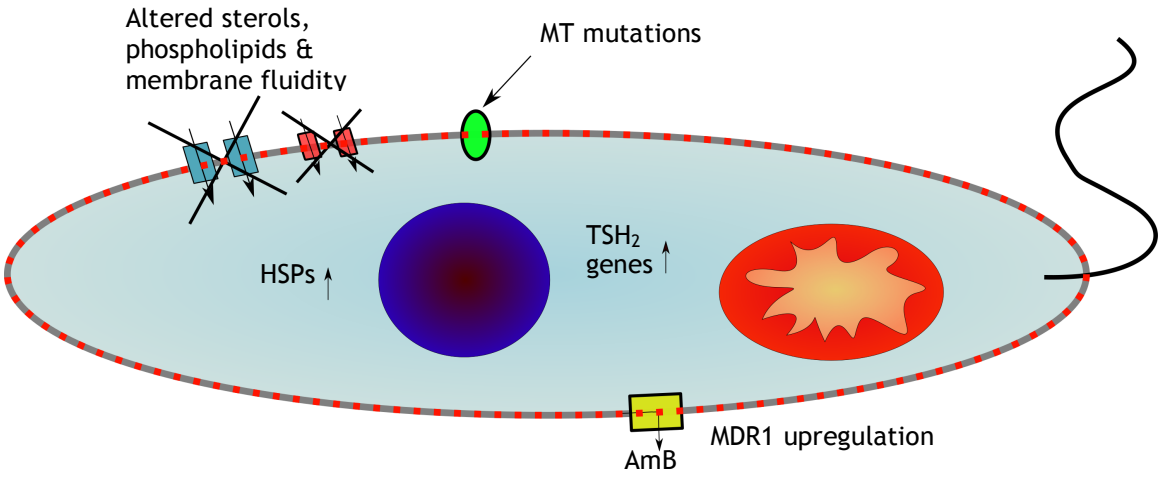


Figure 1-14: Reported mechanisms of AmB resistance in *Leishmania*. Reported changes in AmB-resistant *Leishmania* lines are shown. HSPs: heat-shock proteins; TSH_2: trypanothione disulfide; MDR1: multidrug resistance gene 1; MT: miltefosine transporter. Channels produced by AmB activity are depicted in turquoise and red, MT in green, and MDR1 in yellow. The cell nucleus and mitochondrion are signified in dark blue and red, respectively. Changes to membrane composition are indicated by dashed red lines in the membrane.

1.4 Metabolomics and next generation sequencing techniques in the investigation of parasite drug resistance

Recent years have seen an expansion in the use of so-called 'omics techniques within biology, a set of approaches which, rather than targeting a single gene or metabolite, are set up to analyse simultaneously as many as possible of a particular type of substance within a cell. Genomics, for example, analyses the whole genome of an organism whereas transcriptomics characterises and quantifies the abundance of RNA transcripts within a cell or tissue. Proteomics measures the protein complement of the cell, whereas metabolomics measures all small molecules. Each can be untargeted or targeted to varying extents. For example, genomics can focus on the whole genome (whole genome sequencing, WGS) or simply coding regions, which can be enriched in the sample, known as the exome. Metabolomics, on the other hand, can be untargeted, in which identification and relative quantification is sought for as many metabolites as

possible, or targeted, in which a class of metabolites, such as sugars, sterols or phospholipids, is analysed. In this thesis, particular use is made of metabolomics (untargeted and targeted), WGS and transcriptomics. Hence the principles underlying these and their utility and limitations are discussed.

1.4.1 Metabolomics

The term "metabolome" was first introduced in the scientific literature in 1998 (Oliver *et al*. 1998) and refers to the total cellular complement of small molecules of a molecular mass less than 1,200. While direct quantification of metabolites can be achieved using nuclear magnetic resonance-based techniques, greater sensitivity, and hence coverage of the metabolome, is offered by mass spectrometry-based detection, usually combined with chromatography. Chromatography allows resolution of metabolites prior to identification of mass, permitting relative quantification based on chromatogram peak size, resolution of isomers, and limiting of ion suppression. Ion suppression arises when the ionisation of one metabolite prevents efficient ionisation of another, thus preventing its detection (Cubbon *et al*. 2010). Both untargeted and targeted methodologies used here are described below.

1.4.1.1 Untargeted metabolomics by liquid chromatography-mass spectrometry (LC-MS)

Untargeted metabolomics aims to analyse as many metabolites as possible simultaneously; however, given the diverse properties of its analytes, this is rarely achievable by a single method. Coverage is dependent on a range of factors including quantity of sample used, extraction method, and chromatography. Coverage depends particularly on the extraction solvent: here, a 1:3:1 (v/v/v) ratio of chloroform:methanol:water is used as this has been shown to offer comprehensive metabolite dissolution after cell lysis (t'Kindt *et al*. 2010). Parameters such as temperature are also important in preventing decay of unstable metabolites, as well as perturbations to cellular metabolism. Thus, cells must be rapidly quenched to less than 10 ^{o}C, with extraction maintained at a cold temperature thereafter. Choice of column used in chromatographic separation of metabolites is also important in determining coverage. Here hydrophilic interaction liquid chromatography (HILIC) was used, as this offers favourable separation of polar metabolites (Cubbon *et al*. 2010),

which includes a large amount of central cellular metabolism, as well as metabolites of particular interest for kinetoplastids, such as trypanothione. However, there are also limitations, particularly poor resolution of hydrophobic compounds such as lipids, and conducting the procedure in an oxidising environment inevitably leads to oxidation of some compounds, including trypanothione to trypanothione disulfide (t'Kindt *et al*. 2010), meaning that quantification of these compounds is challenging or impossible. HILIC uses a hydrophilic stationary phase, along with a mobile phase containing both water and acetonitrile. Ionisation then follows, in this case by the electrospray ionisation method, which can operate in either positive ionisation (addition of protons) or negative ionisation (loss of protons) mode, and here polarity switching was used, in which positive and negative modes are alternated. Detection of masses then takes place by mass spectrometry, here using a high-resolution accurate mass ion trap mass analyser. Several studies have used untargeted metabolomics to study metabolic changes in drug-resistant *Leishmania*. These have pointed particularly towards defence against oxidative stress as an important mechanism of resistance to Sb^{III} (Berg *et al*. 2013), MILT (Canuto *et al*. 2014) and combination therapies (Berg *et al*. 2015), for example through changes to thiol metabolism, as well as alterations in other pathways such as amino acid metabolism.

1.4.1.2 Targeted metabolomics by gas chromatography-mass spectrometry (GC-MS)

GC-MS offers certain advantages over LC-MS as an approach, particularly its high resolution, lack of ion suppression and the reproducibility of fragmentation patterns and retention times (Koek *et al*. 2011). As fragmentation of metabolites can be used to define more precisely internal structures beyond the overall molecular formula (as can be defined by exact mass alone), this is beneficial for analyte identification. A limitation of GC-MS, however, is the requirement for heating and vapourisation. Certain compounds are insufficiently stable, or insufficiently volatile, for this process. Volatility can be improved by derivatisation where direct sampling is not possible, and silylation reagents (which introduce a silane group) are most universally used (Koek *et al*. 2011).

In the study of AmB resistance, GC-MS has been an important tool in sterol analysis (see section 1.3.5.2). Sterols pose particular challenges because many isoforms can exist, differing only by positions of double bonds or methyl groups, amongst other changes. These similar properties make chromatographic resolution essential, but being hydrophobic (with polarity typically found only in the C3-hydroxy group), they are poorly resolved by HILIC. Many isomers can also share the same exact mass, and their diversity makes it difficult to have separate standards for each possible metabolite. Therefore, the high resolution permitted by GC-MS, combined with the reproducibility of fragment patterns, offers major advantages. Fragmentation patterns and retention times can be compared to a database of standards (NIST) to greatly improve identification. Furthermore, silylation of sterols happens in a predictable fashion due to the presence of only one appropriate functional group (the C3-hydroxy moiety). Nevertheless, the challenges associated with identification must still not be underestimated.

As a targeted method, sterol analysis requires a sample preparation technique designed to isolate or enrich for a specific class of metabolites. For sterols, this involves initial cell lysis and metabolite extraction, followed by partitioning with a non-aqueous solvent (here heptane is used) to separate sterols from hydrophilic compounds. The method used here (adapted from Arthington-Skaggs *et al*. 1999) also involves heating in highly alkaline conditions, causing hydrolysis of saponifiable lipids (those containing an ester group). The effect of this is to convert otherwise hydrophobic compounds (such as phospholipids) into polar or even charged substituents (such as fatty acids and glycerol), thus preventing co-partitioning of these in the organic layer and leading to a purer sterol extract. Targeted methods can therefore allow for simpler analysis and more confident identification by allowing for production of a simpler analyte mixture with fewer possible constituents.

1.4.2 Next generation sequencing

The past decade has seen a revolution in sequencing technologies, with costs dropping from an estimated $2.7 billion for the original human genome project to less than $1,000 per individual genome today. This has been associated with massively parallel sequencing technologies, starting with a pyrosequencing-

based approach developed by 454 Life Sciences (later purchased by Roche). This began with emulsion polymerase chain reaction (PCR) to amplify bead-bound fragments of DNA, followed by pyrosequencing, in which bases are detected by coupling their incorporation and resulting release of pyrophosphate to enzyme-based chemiluminescence. By having the reaction take place within individual wells (which could be monitored by a charge-coupled detector), this allowed parallel sequencing of around one million individual DNA fragments, "reads", of 400-500 bases. By supplying one type of deoxynucleotide at a time, individual bases could be detected. This technique struggled, however, with homopolymeric regions, which could only be detected based on the intensity of the chemiluminescent signal as multiple bases were incorporated within a single round of sequencing. Other technologies have since arisen, but today the market has largely come to be dominated by Solexa sequencing-based technologies from Illumina. As the next generation sequencing (NGS)-based datasets within this thesis are generated using this approach, the steps involved are described in greater detail here, followed by further discussion of the specific applications used, WGS and RNA-seq (a technique for transcriptomics).

1.4.2.1 Sample and library preparation

The sample preparation method is dependent on the specific application required. WGS simply requires isolation of genomic DNA, whereas more targeted DNA-based applications, such as exome sequencing or chromatin immunoprecipitation, require enrichment of specific sequences (by use of specific sequence capture probes in the former case, or antibodies specific to a chromatin binding factor in the latter). RNA-seq typically involves isolation of total cellular RNA, although, again, more targeted applications exist. As RNA is more labile than DNA and susceptible to enzymatic degradation by environmental RNases, quality must be assessed, usually based on the integrity and relative prevalence ratio of highly abundant ribosomal RNAs (rRNAs) (although in kinetoplastids, the large subunit rRNA, equivalent to 28S in humans, is processed within the cell into multiple smaller RNAs). As the overwhelming majority of cellular RNA is rRNA, transcriptomics methods typically involve enrichment of messenger RNAs (mRNAs), either through reductive hybridisation using rRNA-specific probes or positive enrichment of polyadenylated RNA using oligo-d(T) probes. For non-model organisms, appropriate rRNA probes are rarely

available and so poly(A) selection is preferred, although selection of RNA purely from its 3'-end does require high RNA integrity to ensure that the whole length of transcripts is captured.

The next step is library preparation, which is necessary to produce DNA fragments that can be ligated to the Illumina flowcell. In the standard Illumina TruSeq DNA library preparation pipeline, the first step is fragmentation of genomic DNA to appropriate lengths (~500 bp), by mechanical forces such as sonication (Figure 1-15). Fragments of a specific range of lengths can be selected, typically using solid-phase reversible immobilisation beads to which fragments of different lengths selectively adhere under different solvent conditions (van Dijk *et al*. 2014). Fragments generated through sonication require end-repair to generate suitable DNA ends for ligation steps, in this case through generation of blunt ends, followed by addition of 5' phosphorylation and a 3' adenosine residue. Adaptors are then introduced through ligation. Adaptors contain primers for PCR amplification, as well as for sequencing. Indexes are also introduced within adaptors, which are library-specific barcodes to allow pooling of multiple libraries in a single sequencing run, referred to as "multiplexing". While PCR amplification allows use of less input material, it also introduces biases in coverage, with some sequences preferentially amplified over others (van Dijk *et al*. 2014). PCR-amplified libraries therefore require deeper sequencing to ensure adequate coverage of poorly amplified regions. For RNA-seq, library preparation follows similar steps to DNA, with the difference that reverse transcription is required to convert RNA to DNA. In the Illumina protocol this proceeds with randomly primed first strand synthesis, followed by second strand synthesis to generate double-stranded DNA. In some cases (as used in this thesis), stranded RNA-seq libraries are produced in which the original strand information is preserved (RNA itself being single-stranded, whereas the resulting DNA library is double stranded). Strand information can be preserved by a number of means, but one of the most effective is the dUTP incorporation method (Parkhomchuck *et al*. 2009) in which dUTP is incorporated into the second strand instead of dTTP. The second strand can then be degraded with uracil N-glycosylase prior to adaptor ligation, or removed simply through amplification with a polymerase that cannot use dUTP as a template. This helps ensure that only reads derived from the correct RNA strand are counted in

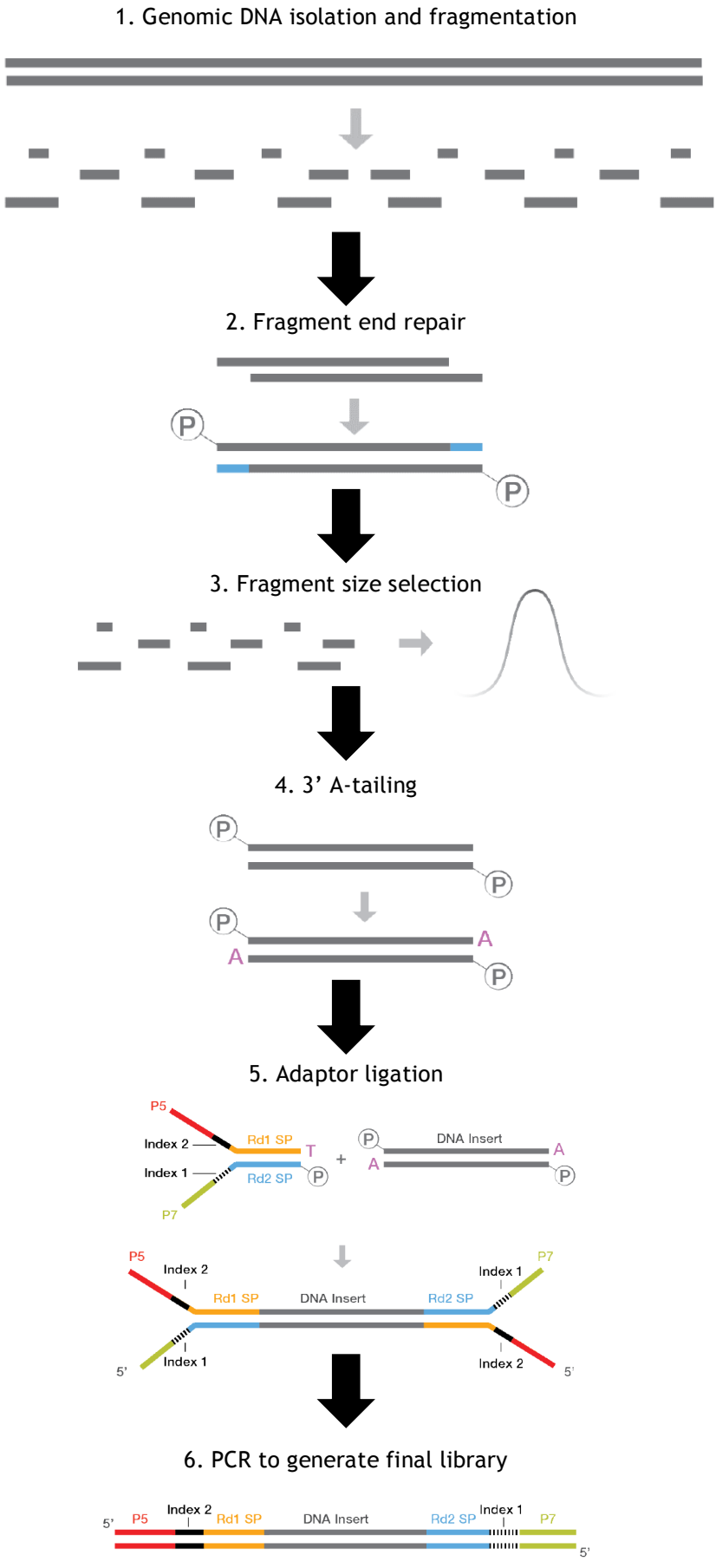

1. Genomic DNA isolation and fragmentation
2. Fragment end repair
P
P
3. Fragment size selection
4. 3’ A-tailing
P
P
P
A
A
P
5. Adaptor ligation
P5
Index 2
Rd1 SP
T
Index 1
Rd2 SP
P
P7
+
P
DNA Insert
A
A
P
P5
Index 2
Rd1 SP
DNA Insert
Rd2 SP
Index 1
P7
5’
Index 1
Index 2
5’
6. PCR to generate final library
5’
P5
Index 2
Rd1 SP
DNA Insert
Rd2 SP
Index 1
P7
5’

Figure 1-15: Generation of an Illumina library. An overview of the Illumina Truseq Nano DNA protocol for library generation, as used here. Genomic DNA is isolated and fragemented by sonication (1). Fragments are repaired (2) by filling in overhangs to form blunt ends (re-synthesised regions in blue) and generation of 5'-phosphate groups. A specific range of fragment sizes is then selected (3), followed by addition of overhanging 3' A residues (4). Fragments are ligated to adaptors (5) containing primers for amplification and hybridisation to the flow cell (P5 and P7, in red and green), sequencing primers (SPs) for reads 1 and 2 (Rd1 in yellow, Rd2 in blue), as well as indexes for multiplex sequencing. PCR is used to enrich for ligated products and generate the final library (6) (adapted from the official Illumina protocol – available at https://emea.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_truseq_nano_dna_sample_prep_kit.pdf, accessed 25/02/2018).

expression analysis.

1.4.2.2 Bridge amplification

When libraries have been accurately quantified (usually by quantitative PCR) they are applied to the Illumina flowcell, which contains a lawn of pairs of covalently bound oligonucleotides. Cluster generation occurs either in an external machine (the cBot, in the case of HiSeq sequencers) or internally within the sequencer, as for the NextSeq machine used here. This is a form of solid-state PCR used to amplify the signal from a single strand of DNA to thousands of covalently bound DNA molecules (Figure 1-16). Denatured single-stranded DNA hybridises to oligonucleotides grafted to the flowcell through adaptor sequences located at either end of the molecule (P5 or P7). First strand synthesis using the bound oligonucleotides as a primer creates a covalently bound complementary copy, allowing the original, unbound strand to be washed away after denaturation. The 3' sequence of the newly generated strand contains a sequence complementary to the other type of bound oligonucleotide, causing annealing and the formation of a bridge between the two bound oligonucleotides; this then allows the other oligonucleotide to act as a primer for synthesis of a second, complementary strand that is now also bound covalently. Cycles of this process allow generation of millions of clusters with around 2,000 molecules per cluster. The process is terminated by a final denaturation step and cleavage at the P5 primer site, allowing removal of one strand and leaving around 1,000 identical bound molecules. Blocking of the 3' end with dideoxynucleotide triphosphates (ddNTPs) prevents any interference with subsequent sequencing reactions. Accurate quantification of library input is key to ensuring that the maximal number of clusters is obtained on the flowcell surface without overlap.

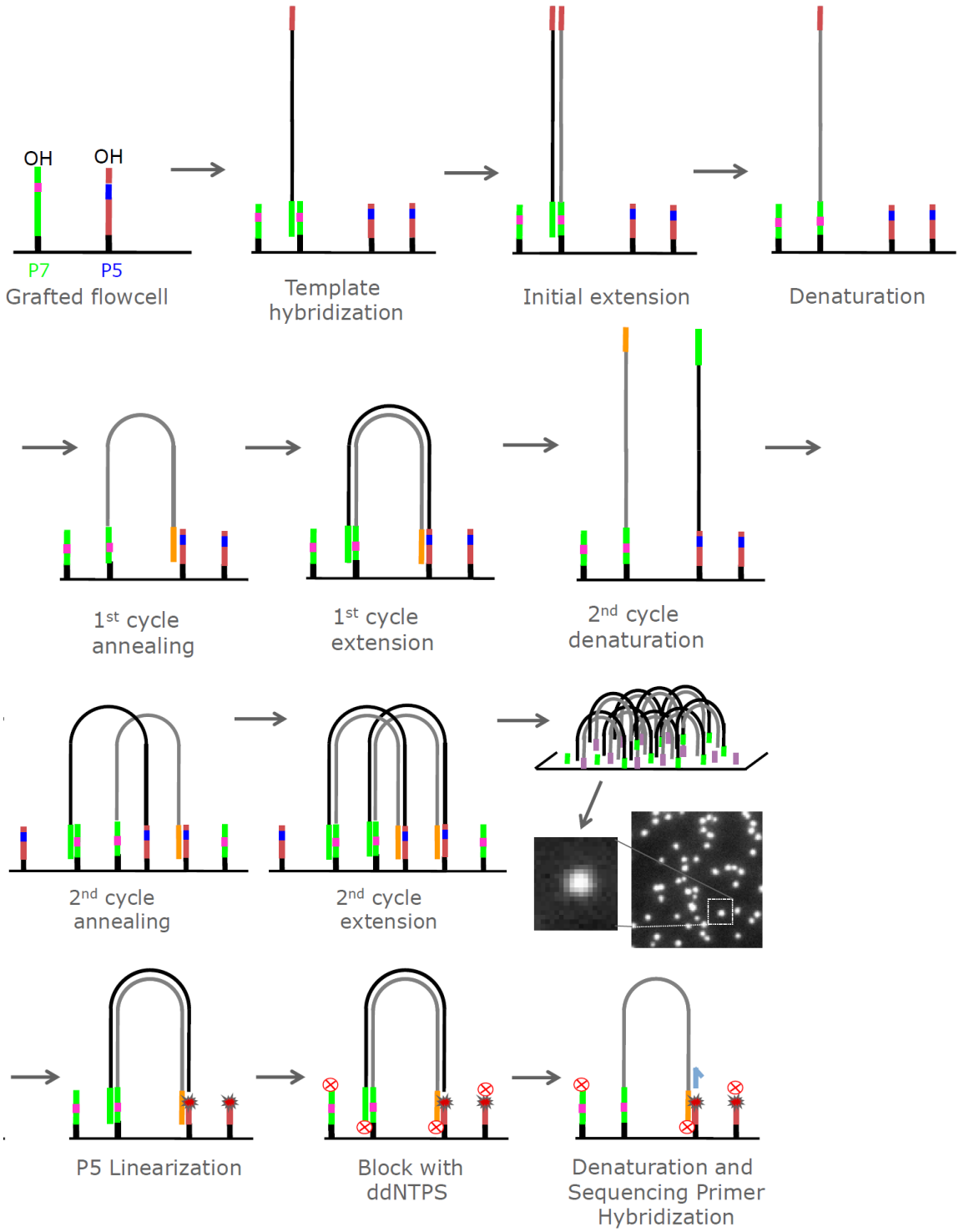


Figure 1-16: Cluster generation. Steps of bridge amplification to generate clusters of ~2,000 molecules. P5 and P7 sequences within library adaptors (green and orange, respectively, see Figure 1-15) allow binding of DNA libraries to oligonucleotides covalently bound to the Illumina flow cell. After hybridisation, these bound oligonucleotides prime sequencing of a first strand, after which denaturation leads to loss of the initial strand and retention of the newly-generated covalently bound strand. Annealing to the opposite type of oligonucleotide sequence allows extension of a new complementary strand, and repeated cycles result in bridge amplification, leading to the generation of clusters, which can be detected by fluorescence imaging during sequencing. The final row of steps shows cleavage of one strand (ligated via the P5 adaptor sequence) and removal to generate a clonal cluster of ~1,000 strands, ready for first strand sequencing (adapted from Cluster Generation slides from Genome Analysis Boot Camp, Broad Institute – available at https://www.broadinstitute.org/files/shared/illuminavids/clusterGenSlides.pdf, accessed 16/10/2016).

1.4.2.3 Sequencing by synthesis

Illumina technology uses Solexa sequencing. Like 1st generation (Sanger) sequencing, this is a sequencing-by-synthesis method in which a sequence is read based on incorporation of bases into a newly synthesised strand. Synthesis begins with binding of a primer to an internal adaptor site (Figure 1-17). Nucleotides are added that contain a fluorophore (the colour of which is dependent on the incorporated base) that acts as a 3'-blocking agent to prevent further extension. After excitation with a laser or light emitting diode source and imaging of clusters with a camera, terminal blocking reagents are removed and the next cycle can continue, until read generation is complete. Read length is limited by a number of factors, including radiation damage if lasers are used, accumulation of contaminating reagents which are inefficiently washed away, and decay of signal intensity through phasing and pre-phasing (in which some strands fail to

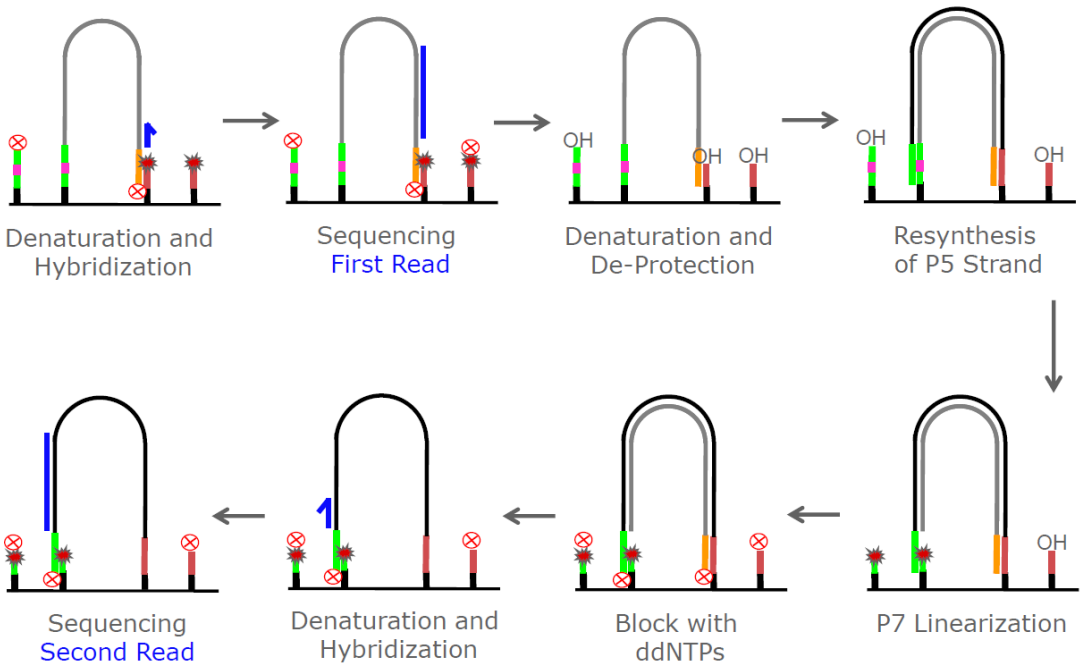


Figure 1-17: Paired end sequencing. Illumina technology allows generation of paired ends. Strands covalently bound through the P7 site (green) anneal to complementary oligonucleotides on the flow cell (red) via P5 sequences (orange) within the adaptor. This allows for primer binding and first read synthesis (blue). The newly synthesised strand is removed by denaturation, and the P5 strand previously removed (see Figure 1-16) is resynthesized. Cleavage at the P7 site allows removal of the first strand, allowing synthesis of a second read (blue) reverse-complementary to the first. (adapted from Cluster Generation slides from Genome Analysis Boot Camp, Broad Institute – available at https://www.broadinstitute.org/files/shared/illuminavids/clusterGenSlides.pdf, accessed 16/10/2016).

incorporate a base or incorporate multiple, respectively, during a single cycle thus becoming out of phase with the rest of the cluster; this is an effect which accumulates cycle-by-cycle).

A significant advantage offered by Illumina technology is the availability of paired-end sequencing. After first read synthesis is complete, denaturation to remove the newly generated strand precedes bridge-mediated synthesis, which generates a covalently bound second strand (as during cluster generation). Cleavage of the original (P7) strand and chemical 3'-blocking then occur, thus leaving clusters of single strands reverse-complementary to those used during first strand synthesis. Sequencing steps are repeated to generate a second read in a convergent orientation to the first. Thus, read pairs are produced, not only generating a greater amount of sequencing information but also aiding accurate mapping (see section 1.4.2.4). Note that one particular feature of the NextSeq is that unlike other Illumina machines, it uses two-colour chemistry in which, rather than having one colour per base, red signifies C, green T, red and green together A and unlabelled G. This simplifies the image capture and data analysis processes.

1.4.2.4 Data analysis

The generation of millions of short reads creates an analysis problem that is only soluble through considerable computational processing. Base calling is usually undertaken by an in-house algorithm and generates reads along with quality scores. These are Phred-scaled, meaning that the relationship between the quality, *Q*, and the probability of the call being inaccurate is as follows:

$$P(incorrect) = 10^{-(\frac{Q}{10})}$$

In other words, a *Q* score of 30 equates to an accuracy level of 99.9%. Initial quality processing of reads helps to remove adaptor sequences from the 5'-end of reads, as well as removing poor quality bases often found at the 3'-end for reasons explained in section 1.4.2.3. Where a reference genome is available (as is the case for *L. mexicana*), reads can be aligned to this. Fast, accurate alignment is a considerable challenge but usually involves exact matching of short subsections of the reference of length *k* (referred to as *k*-mers) to sections

of a read (known as the seed). Pre-processing of the reference genome into a set of indices allows rapid matching and extension of the seed to allow an overall, error-tolerant alignment (Reinert *et al*. 2015). In BWA (Burrows-Wheeler Alignment) (Li & Durbin 2009), the alignment tool used here, this index is created using a mathematical operation called the Burrows-Wheeler Transform. Where alignment is uncertain (e.g. due to multiple occurrences of a sequence within the reference genome), paired-end read alignment can facilitate mapping by incorporating assumptions about fragment size (and hence the presumed distance between and relative orientation of alignments of read pairs). While most alignment tools operate by these principles, specialised sequencing applications may require variants. Alignment of RNA-seq datasets, for example, often require splice-aware aligners such as HISAT (Kim *et al*. 2016), which accommodates intron splicing by facilitation of mapping across exon-exon boundaries. As introns are not prevalent in *Leishmania*, however, these are not used here.

Subsequent steps depend on the specific application required. For calling of single nucleotide polymorphisms (SNPs) and small insertions and deletions (InDels), FreeBayes (Garrison & Marth 2012) is used here. While earlier mapping-based variant calling software used models operating on a single base position at a time, FreeBayes is haplotype-based, relying on local assembly and operating on the reads directly. This is particular advantageous in regions of lower quality mapping and assembly, as well as where InDels are present (Rimmer *et al*. 2014). SNPs and InDels are detected using a Bayesian statistical framework (Garrison & Marth 2012), then subsequent custom steps are used to quality filter and annotate variants.

For RNA-seq data, numerous applications are possible including detection of transcripts and splice variants through tools such as Cufflinks (Trapnell *et al*. 2010). However, the most common application is differential expression analysis, which measures changes in gene expression through relative abundance of reads mapping to specific genomic features. Various tools exist, applying different statistical models. Technical variability in coverage of a given feature is modelled by a Poisson distribution; however, this has been found to be excessively restrictive for modelling of inter-sample variability. Largely due to

costs, replicate number is typically low for RNA-seq experiments, posing challenges for appropriate modelling of data. DESeq2 (Love *et al*. 2014), the tool employed here, uses a negative binomial model. This models variance by encompassing both a Poisson component dependent only on the mean normalised count, and an over-dispersion component derived from the observed variability of the genomic feature itself, as well as an expected dispersion estimate based on the relationship between mean normalised count and dispersion across the dataset as a whole. This overcomes the problem of low replicate number by using information of how mean coverage correlates with dispersion across all detectable genes.

1.4.2.5 WGS and RNA-seq of *Leishmania* parasites

Both WGS and RNA-seq have been applied previously to *Leishmania*. This has been facilitated by the availability of reference genomes for kinetoplastid pathogens for over a decade (El-Sayed *et al*. 2005), starting with *T. brucei*, *T. cruzi* and *L. major* but since followed by numerous other pathogenic and non-pathogenic species. Reference genome availability greatly enables the use of NGS by allowing rapid alignment of reads, and resources such as the online database TriTrypDB (part of the EuPathDB project, http://tritrypdb.org/tritrypdb/) provide a wealth of annotation for kinetoplastid genomes. Genomics has increasingly been used in the study of drug resistance in *Leishmania*, and for antimony resistance this has included sequencing of *L. donovani* clinical isolates in order to detect SNPs indicative of a resistant phenotype, as well as population-based analysis to show the spread of resistance-associated genotypes (Downing *et al*. 2011, Imamura *et al*. 2016). Regarding drugs to which resistance is not yet widespread, however, comparative genomics of resistant lines selected *in vitro* with isogenic wild-type lines has proved to be an informative approach, particularly when combined with metabolomics-based techniques (Vacchina *et al*. 2016, Fernandez-Prada *et al*. 2016, Mwenechanya *et al*. 2017). There are limitations, however, regarding Illumina-based sequencing, which must be considered when addressing the utility of this approach. As *Leishmania* genomes are known to be highly plastic (see section 1.1.3), structural variants and CNVs are common. The short reads produced by Illumina sequencing offer limited information regarding larger scale structural changes. While CNVs can be detected based on the number of reads

mapping to a given genomic feature (Rogers *et al*. 2011) (as used for RNA-seq expression analysis), the precise structural changes are difficult to characterise. Further evidence can be obtained using split reads (individual reads in which different subsections map to separate reference loci), but where sequencing is paired-end, read concordance is also informative (concordant reads are those which map to the same chromosome in a convergent fashion, with an inter-pair distance in accordance with predicted fragment size distribution), as used by structural variant predictors such as LUMPY (Layer *et al*. 2014). Nevertheless, such variants, especially where they occur in repetitive regions (making precise alignment more challenging), represent difficulties for this approach. Fortunately, long read real-time sequencing technologies, such as Pacific Biosciences' single molecule real time sequencing approach (McCarthy 2010) or pore based sequencing from Oxford Nanopore Technologies (Madoui *et al*. 2015), are gaining in accuracy and throughput. Pacific Biosciences sequencing has already been used in defining the position of the unusual trypanosomatid-specific modified base 'J' in *Leishmania* (Genest *et al*. 2015), and nanopore sequencing has been used to differentiate *Leishmania* species (Imai *et al*. 2017). However, the utility of these technologies goes well beyond such specialised applications. In the near future they should enable advancement beyond the current generation of kinetoplastid reference genomes and more powerful characterisation of large scale structural variants, by generation of multiple-kilobase reads which map across repetitive or unstable genomic regions.

By contrast, RNA-seq has been far less extensively used in *Leishmania*. This is in part because of kinetoplastids' unusual transcriptional biology (see section 1.1.3) and the fact that translational control, not captured by RNA-seq, appears to be highly important in gene regulation. Where it has been used, RNA-seq has been applied to studies of cell cycle differentiation (Dillon *et al*. 2015, Dillon *et al*. 2015b, Fiebig *et al*. 2015, Fernandes *et al*. 2016). This has helped to refine the predicted *Leishmania* transcriptome, as well as identify broad changes involved in differentiation. Indeed, the high dynamic range of RNA-seq is favourable to experiments on intracellular amastigotes, where 80-90% of reads may map to macrophages (Dillon *et al*. 2015). While other techniques, such as proteomics, may struggle to detect low abundance proteins, this problem can be overcome in RNA-seq simply by obtaining more reads (albeit at higher expense).

It also allows simultaneous profiling of the macrophage, offering an insight into host-pathogen interactions (Dillon *et al*. 2015, Fernandes *et al*. 2016). RNA-seq has rarely been used in studies of drug resistance in *Leishmania*. Profiling of MILT-resistant *L. donovani* revealed a large number of differentially expressed genes, but only a small minority of these showed fold changes greater than 1.5 (Vacchina *et al*. 2016). Differentially expressed transcripts showed a wide range of functions including enzymes and transporters, although one result of interest is a more than 100-fold downregulation of *hsp83.1*, an unexpected result given that this chaperone is known to protect from stress. *hsp70*, by contrast, showed a more modest threefold increase. Transcriptomics has been used previously to detect ascorbate peroxidase upregulation in the AmB-resistant *L. donovani* clinical isolate described by Purkait and colleagues (Kumar *et al*. 2014). However, this was done using microarray technology and the full expression changes were not reported, nor is the dataset publicly available. Hence the utility of RNA-seq for studies of AmB-resistance is currently unproven.

1.5 Aims

The risk of emergence of drug resistance is a major threat to leishmaniasis control programmes. While resistance to AmB is not yet widespread, increasing reliance on this drug and sporadic reports of drug resistance suggest this is a considerable concern. Certain features have been identified in AmB-resistant *Leishmania*, and alterations to sterol content appears to be well-established, but there is limited information regarding the genetic basis of these changes, as well as about the consequences of these changes for fitness. Furthermore, while altered sterol C24-methyltransferase transcription has been observed in two cases (Pourshafie *et al*. 2004, Purkait *et al*. 2012), genetic events leading to these changes have not been explored. The current published studies have tended to focus on individual AmB-resistant lines, making it challenging to identify which genomic, metabolic and phenotypic features arise stochastically during drug selection and which are conserved across multiple independently selected lines. Defining which changes are conserved during the selection of resistance is an important step, both in predicting the consequences of AmB resistance for fitness and sensitivity to other drugs, and in design of approaches to monitor for the emergence of AmB resistance.

Thus, the principal aims of the project were as follows:

1. To select four independent AmB-resistant *L mexicana* lines and to characterise these, defining the degree to which phenotypes are shared, such as effects on growth, infectivity, drug sensitivity and metabolism

2. To identify mutations and transcriptional changes associated with AmB-resistance, in particular in the ergosterol biosynthesis pathway but also more broadly

3. To functionally validate the role of these changes in AmB resistance and other observed phenotypic changes by gene complementation methods

4. To characterise further the specific genomic events leading to observed changes in sterol C24-methyltransferase expression

2 Materials and methods

2.1 *Leishmania* cell culture methods

2.1.1 Culture of *Leishmania* promastigotes

Leishmania mexicana strain M379 promastigotes were routinely cultured in HOMEM media (GE Healthcare, Supplementary file 2-1 for recipe) supplemented with 10% fetal bovine serum (FBS, Gibco) (herein referred to as complete HOMEM, cHOMEM) at 25 ^{o}C in non-vented flasks (Corning). Routine maintenance involved diluting parasite cultures every three to five days into fresh medium to a density of approximately 10^5 cells/ml. For use in assays (unless otherwise stated), parasites were diluted to this density three days in advance to generate a population at 5-10 x 10^6 cells/ml, defined here as mid-log phase. For quantification of parasite density, parasites were counted in a Neubauer haemocytometer; prior to this, parasites were fixed by mixing in a 1:1 ratio with 2% formaldehyde in phosphate-buffered saline (PBS), and allowed to settle for three to five minutes prior to counting. Where parasites were harvested by centrifugation, this was achieved by centrifugation at 1,250 x *g* for ten minutes, unless otherwise stated.

2.1.2 Alamar blue drug sensitivity assays

Promastigote drug sensitivity was determined by Alamar Blue assay, adapted from Räs *et al*. 1997, with modifications. Drugs were purchased from Sigma, unless otherwise stated. Solution stocks of AmB, glucose oxidase (GOX), imipramine, methylene blue, miltefosine hydrate, potassium antimonyl tartrate, pentamidine isethionate (May & Baker) and paromomycin sulfate used water as the solvent, sterilised using a 0.22 μM filter. Ketoconazole stocks were dissolved in methanol, spiroxamine and fenpropidin in ethanol, and menadione in dimethylsulfoxide (DMSO). Working stocks were made fresh on the day of the assay by dilution in cHOMEM to two times the desired maximal concentration for the drug assay. Twofold dilution series were made in opaque 96-well plates in 100 μl in cHOMEM; either 11 or 23 different concentrations were used (depending on foreknowledge of expected concentration range), and one extra

well was included containing only cHOMEM. All dilution series were performed in technical duplicates.

Mid-log phase parasites were then counted, harvested by centrifugation and resuspended in cHOMEM to a density of 2 x 10^6 cells/ml. These were then applied on top of the drug dilution series at 100 μl/well, giving a final starting cell density of 10^6 cells/ml. Plates were incubated at 25 ^oC for 72 hours, before adding 20 μl resazurin solution (0.49 mM in PBS, pH 7.4), then incubating for a further 48 hours. Resazurin dye is reduced by cellular metabolism to red fluorescent resorufin, thus acting as a marker of cell viability. Fluorescence was measured using a FLUOstar OPTIMA microplate reader (BMG Labtech) with an excitation wavelength of 530 nm and an emission wavelength of 590 nm. IC_{50}, the concentration at which parasite growth inhibition of 50% occurred, was calculated using the R package drc version 2.5 (Ritz & Streibig 2005) using a four parameter dose-response model, validated against the sigmoidal curve model in GraphPad Prism software. Statistical differences between conditions were determined using a two-tailed, unpaired Student's *t*-test.

2.1.3 Selection of drug resistance and obtaining clonal populations

Resistance to AmB was selected in four independent lines derived from one wild-type parental line by continuous passaging with stepwise increasing concentrations of drug, starting with 50 nM (AmB IC_{50} values for wild-type parasites are typically 50-70 nM). Parasites were routinely sub-passaged to a density of 10^6 cells/ml into AmB-containing medium (cHOMEM, supplemented in this case with 100 units/ml penicillin/0.1 mg/ml streptomycin (Sigma)). The concentration was increased twofold when a notable change in growth rate was observed, to a degree similar to wild-type cells growing in the absence of drug. AmB sensitivity was routinely measured by Alamar blue assay (section 2.1.2) in comparison to a wild-type line continuously cultured in parallel in the absence of drug. Selection was halted after approximately 290 days, after which stable growth was achieved in 400 nM AmB in two lines (AmBRA and AmBRB) and in 200 nM in the other two lines (AmBRC and AmBRD). Clonal populations were isolated by limiting dilution into round-bottomed 96-well plates in cHOMEM supplemented with AmB (400 nM for AmBRA and AmBRB, 200 nM for AmBRC and

AmBRD). Clonality was decided using the rule that if cells could be observed in fewer than one in six wells after a period of one to two weeks, populations in wells positive for *Leishmania* cells were considered clonal. Throughout selection, samples were taken for cryostorage at multiple stages, preserved in HOMEM with a final concentration of 15% FBS and 5% DMSO; samples were frozen gradually to -80 ^{o}C, followed by long-term storage in liquid nitrogen.

2.1.4 Promastigote growth assays

Mid-log phase promastigotes were counted as described in section 2.1.1 and diluted to an initial concentration of 10^5 cells/ml in cHOMEM. Parasite density was counted every 24 hours for seven days. For this assay, all counts were performed twice, with an average count taken between these.

2.2 Macrophage and *in vivo* methods

Mice used for primary macrophage work and *in vivo* experiments were purchased from Harlan, and kept at the Central Research Facilities, University of Glasgow, Glasgow, U.K. Mice were maintained under standard animal facility conditions and in accordance with local and home office regulations. Methods for macrophage isolation and differentiation (sections 2.2.1 and 2.2.2) are derived from the method of Weischenfeldt & Porse 2008. All mammalian cells were maintained in Roswell Park Memorial Institute (RPMI) medium (Gibco, Supplementary file 2-1 for recipe), supplemented with 10% FBS, 2 mM glutamine and 100 units/ml penicillin/0.1 mg/ml streptomycin (Sigma) (complete RPMI, cRPMI).

2.2.1 Production of L929-conditioned medium

L929 cells are mouse fibroblast cells derived from adipose tissue of a C3H/An mouse, with genotype $APRT^+$ and $HPRT^+$ (adenosine phosphoribosyltransferase and hypoxanthine ribosyltransferase). These secrete macrophage colony-stimulating factor (M-CSF), a cytokine that stimulates the differentiation of macrophages (Weischenfeldt & Porse 2008). L929 cells were maintained in cRPMI at 37 ^{o}C, 5% CO_2. For generation of L929-conditioned medium, a vented 150 cm^2 flask (Corning) of confluent cells were detached using trypsin-ethylenediaminetetraacetic acid solution (Gibco), and split into five new 150

cm^2 flasks, with 100 ml cRPMI added per flask. Four days after cells became confluent in the new flasks, the culture supernatant was obtained and filtered twice through separate 0.2 μm vacuum filters (Merck Millipore) to remove any cells. Supernatant aliquots were stored at -20 °C and not frozen again after thawing. M-CSF content was quantified using a mouse M-CSF DuoSet ELISA kit (R&D Systems) by Clément Regnault. Low batch-to-batch variation was observed, with typical values being around 500 pg/ml.

2.2.2 Isolation and differentiation of primary bone marrow-derived macrophages

Eight to 12-week old C57BL/6 mice were euthanised by cervical dislocation or CO_2 asphyxiation. Animals were sprayed with 70% ethanol and dissected to isolate the femurs and tibiae from the hind legs, with bones placed in cRPMI prior to use. The ends of each bone were removed with 70% ethanol-sterilised scissors and bone marrows were flushed out with cRPMI using a 23-guage needle (Henke Sass Wolf Fine-Ject). The same needle was used to homogenise cells, followed by straining through a 70 μm cell sieve (EASYstrainer, Grenier Bio-One). Cells were sedimented by centrifugation (300 x *g*, 5 min) and resuspended in cRPMI. Viability was checked using trypan blue viability stain (Sigma) and volume was adjusted to 12 ml with cRPMI. This was divided into 12 9 cm petri-dishes, to which were also added 7 ml cRPMI and 2 ml L929-conditioned medium (20% v/v L929-conditioned medium). Cells were incubated at 37 °C for three days, prior to adding a further 7 ml cRPMI and 2 ml L929-conditioned medium. After a further three days (six days incubation in total), plates were washed three times with serum-free RPMI to remove non-adherent cells, thus isolating the macrophage population as adherent cells. These were detached by addition of ice-cold PBS followed by incubation at 4 °C for at least 5 min and subsequent cell scraping. Detached cells were pooled and sedimented by centrifugation (300 x *g*, 5 min), and resuspended in cRPMI at 5 x 10^5 cells/ml. Macrophages were then allowed to re-adhere overnight, and infection experiments were performed immediately after.

2.2.3 *Leishmania* macrophage infection experiments

Stationary phase promastigotes were sedimented by centrifugation and resuspended in warm cRPMI. In all cases described here, stationary phase promastigotes were derived from five day-old cultures after seeding at an initial densitiy of 10^5 cells/ml, reaching a final density of 1-2 x 10^7 cells/ml. Supernatant was removed from macrophages, before parasites were added at a parasite:macrophage ratio of 10:1. After variable incubation times (see below), extracellular parasites were removed by washing five times with serum-free RPMI, followed by replacement with cRPMI. During infections, wells were washed daily three times with serum-free RPMI, followed by replacement of the medium, to prevent build-up of proliferating extracellular parasites. Incubation of infected macrophages was at 32 ^oC, with 5% CO_2.

2.2.3.1 Intracellular growth and AmB sensitivity assays

For these assays, macrophages were seeded into 24-well plates with each well containing one UV-sterilised circular glass cover slip, at 2.5 x 10^5 cells/well. Thus on incubation overnight, macrophages adhered to the cover slips. Parasites, resuspended in warm cRPMI at a density of 5 x 10^6 cells/ml, were added after removal of supernatant. For intracellular growth assays, infection proceeded for four hours, prior to removal of extracellular parasites. Infected macrophages were then incubated for variable lengths of time (seven, 24, 48, 96 and 144 hours) prior to staining (see below). For AmB sensitivity assays, infection proceeded for 24 hours, prior to removal of extracellular parasites. Medium containing drugs was prepared as a serial twofold dilution in cRPMI, which was then added to wells to produce 11 AmB concentrations (with a maximal concentration of 4 μM), along with one drug-free well. After 24 hours drug treatment, wells were washed three times and replaced with further medium containing drugs at the same concentrations, and staining occurred 24 hours later (48 hours drug treatment in total).

Before staining, wells were washed three times with warm PBS. All supernatant was removed and cells were fixed with 100% methanol for 15 seconds. Methanol was removed and once dry, cells were stained with 12.5% Giemsa stain (Sigma) in Sörensen buffer (3.7 mM KH_2PO_4, 0.7 mM Na_2HPO_4, pH 7.2) for 10 min.

Infection was quantified by light microscopy, counting at least 100 macrophages per well and recording the number of infected macrophages as well as the total number of parasites infecting those macrophages. For determination of AmB sensitivity, the ratio of total parasites to total macrophages was used to calculate an IC_{50}, making use of the drc package in R (as in section 2.1.2).

2.2.3.2 Infection of macrophages for RNA extraction

For this assay, macrophages were seeded into 6-well plates at 2 x 10^6 cells/well. After adherence overnight, macrophages were infected with 2 x 10^7 stationary phase promastigotes, resuspended in warm cRPMI to a density of 10^7 cells/ml, for four hours. After removal of extracellular parasites, infected cells were incubated in cRPMI for a further 72 hours, washing daily as described above. Finally, cells were washed three times with warm PBS. RNA was extracted using the Nucleospin RNA kit (Macherey-Nagel) according to manufacturer's instructions except that the initial lysis was carried out by applying the lysis buffer supplemented with ß-mercaptoethanol directly to the wells, followed by recovery of the cell lysate. Uninfected macrophages were included as a control.

2.2.4 *Leishmania* infection *in vivo*

Prior to mouse infections, stationary phase promastigotes were resuspended in PBS at a density of 5 x 10^7 parasites/ml. Infections of 8-12 week old female BALB/c mice were carried out by subcutaneous footpad injection (performed by Anne-Marie Donachie), with regular monitoring for footpad swelling thereafter, measuring three times with footpad callipers and taking the mean value. Parasites were recovered from the popliteal lymph node draining from the site of infection. Lymph nodes were physically disrupted and forced through a 100 μm cell sieve (EASYstrainer, Grenier Bio-One), and homogenates were then added to cHOMEM supplemented with 200 units/ml penicillin/0.2 mg/ml streptomycin (Sigma) and 50 μg/ml gentamycin. Cultures were then incubated at 25 ^oC for several days in order to allow differentiation to and proliferation of promastigotes.

2.3 Metabolite methods

2.3.1 Untargeted metabolomics

2.3.1.1 Sample extraction

For each sample, 10^8 mid-log phase promastigotes (double-counted as described in section 2.1.1) were obtained. For AmB-resistant cell lines, these were grown in the absence of AmB, but not for more than one passage since growth in AmB. Cells were rapidly cooled to < 10 °C with use of a dry ice-ethanol bath, rapidly agitating during cooling to avoid local freezing. With metabolism quenched, all subsequent stages were conducted on ice or at 4 °C. Cells were sedimented by centrifugation and resuspended in ice cold PBS in 1.5 ml Eppendorf tubes, before further sedimentation at 1,900 x *g* for 10 min. All supernatant was removed and pellets were resuspended in 1:3:1 v/v chloroform:methanol:water. After agitation by rapid shaking at 1,500 rpm for one hour, cell debris was sedimented by centrifugation at 17,000 x *g* for 10 min, and supernatant was retained. A blank sample was prepared by adding extraction solvent directly to empty Eppendorf tubes at the same time as to the samples, and carrying out subsequent shaking and centrifugation steps in parallel. A quality control (QC) sample was produced by pooling extracts from each sample in equal amounts, and topping this up with further extraction solvent. All samples were stored at -80 °C.

2.3.1.2 Liquid chromatography-mass spectrometry

All steps in LC-MS analysis were carried out by Suzanne McGill at Glasgow Polyomics. Metabolites were separated by HILIC using the Dionex UltiMate RSLC system (Thermo Fischer Scientific) with a ZIC-pHILIC column (Merck Sequant). The column was maintained at 30 °C and samples were eluted using a linear gradient between two mobile phase solvents, 20 mM ammonium carbonate in water (A), and acetonitrile (B). The A:B ratio of solvents was initially 20:80, followed by a linear change to 80:20, over a period of 15 min. The ratio was then changed to 95:5, as rapidly as possible, then held for 2 min. After this, the ratio was changed back to 20:80 as rapidly as possible and held for 7 min. The flow rate for this was 0.3 ml/min. Samples were maintained at 4 °C prior to injection, and an injection volume of 10 µl was used. Samples were run in a

randomised order and pooled QC samples were run after every fourth sample to allow for assessment of reproducibility of the instrument over time.

For MS analysis, masses were detected using a Thermo Orbitrap Exactive (Thermo Fisher Scientific) operating in polarity switching mode. MS settings were as follows: resolution was 50,000 with a mass-to-charge ratio (m/z) range of 70-1,400, the automatic gain control target was 10^6, the probe temperature was 150 ^oC and the capillary temperature 275 ^oC. Flow rates were 40 for sheath gas, 5 for auxiliary gas and 1 for sweep gas (arbitrary units). Positive mode ionisation used a source voltage of +4.5 kV, a capillary voltage of +50 V, a tube voltage of +70 kV and a skimmer voltage of +20 V. Negative mode ionisation used a source voltage of -3.5 kV, a capillary voltage of -50 V, a tube voltage of -70 V and a skimmer voltage of -20 V. Immediately prior to each analysis batch, mass calibration was performed with each polarity. Calibration used Thermo calmix (Pierce calibration solutions, Thermo Scientific) standards for m/z < 1,400, and to cover small metabolites, low-mass contaminants were also included: $C_2H_6NO_2$ (m/z 76.0393) for positive ion electrospray ionisation and $C_3H_5O_3$ (m/z 89.0244) for negative ion electrospray ionisation. Lock-mass correction was applied to each analytical run using these ubiquitous low-mass contaminants, to enhance calibration stability.

2.3.1.3 Data analysis

Data were analysed using the mzMatch package in R (Scheltema *et al*. 2011) and IDEOM software (Creek *et al*. 2012b), as a pipeline for identification of metabolites from raw data and relative quantification based on peak area. A pooled sample root squared deviation cutoff value of 50% was used, and metabolites were quality filtered according to the following criteria: firstly, all peptides were removed as these are rarely reproducible or properly identifiable. All peaks with a retention time of less than four minutes were excluded as these form part of a bolus which cannot be accurately quantified due to ion suppression effects resulting from poor resolution. Remaining peaks were subject first to log transformation followed by a one-way analysis of variance (ANOVA) with the Benjamini-Hochberg multiple comparisons correction of P-values. A Tukey's honest significant difference post hoc test was then applied to determine significant differences between individual AmB-resistant lines and

wild-type. Statistical analysis was conducted using R (R core team 2014). Principal component analysis (PCA) plots were produced from log-transformed data using Metaboanalyst 3.0 (Xia *et al*. 2015), an online tool for metabolomics data analysis.

2.3.2 Sterol analysis

2.3.2.1 Sample extraction

Mid-log phase promastigotes were harvested by centrifugation and washed once with PBS. Pellets were resuspended in alcoholic KOH (25% m/v KOH dissolved in 3:2 v/v ethanol:water), followed by heating at 85 ^{o}C for 1 hour to hydrolyse all saponifiable lipids. Samples were allowed to cool and were then partitioned by vortexing for 30 s with *n*-heptane in a 1:1 v/v ratio. Samples were then left for approximately 20 min for aqueous and organic layers to separate, after which the upper organic layer was removed and retained as the sterol extract. Extracts were stored at -80 ^{o}C.

2.3.2.2 Gas chromatography-mass spectrometry

Two experiments were conducted, one as described in chapter 3, the other as described in chapter 5. For the chapter 3 experiment, samples were normalised by pellet mass, with 1.5 ml alcoholic KOH added per 100 mg pellet (equivalent to around 5 x 10^8 cells). This produced high variability in absolute signal intensity (due particularly to residual medium in tubes when weighing pellets), so for the chapter 5 experiment, a constant cell number of 3 x 10^8 cells was maintained instead, extracting this in 500 μl alcoholic KOH. Blanks were prepared using alcoholic KOH (heated as above for the chapter 5 experiment), vortexing the sample with an equal volume of *n*-heptane before allowing of layers to separate and retention of the organic layer. A pooled QC sample was produced by pooling equal volumes of extracts from all samples (for the chapter 5 experiment, separate pools were produced for each replicate).

GC-MS was conducted by Dr Stefan Weidt at Glasgow Polyomics, including data processing and primary analysis. Silylation of compounds was performed as follows: each sample, as well as a standard mix of nine sterols, was dried directly into vials. 50 μl N-methyl-n-trimethylsilyltrifluoroacetamide with 1%

2,2,2-trifluoro-N-methyl-N-(trimethylsilyl)-acetamide, chlorotrimethylsilane (Thermo Scientific) was then added and samples were vortexed for 10 s before incubation at 80 ^{o}C for 15 min. 50 μl pyridine was then added, along with 1 μl retention index solution (an *n*-alkane mixture containing C12, C15, C19, C22, C25 and C29 alkanes).

GC was performed using a TraceGOLD TG-5SILMS column (30 m length, 0.25 mm inner diameter, 0.25 μm film thickness, Thermo Scientific) installed in a Trace Ultra gas chromatograph (Thermo Scientific). Helium was used as the carrier gas with a flow rate of 1.0 ml/min, and 1 μl of trimethylsilane-derivatised sample was injected into a split/splitless injector using a surged splitless injection, with a splitless time of 30 s and a surge pressure of 167 kPa. An initial oven temperature of 70 ^{o}C was used, which was increased to 250 ^{o}C at a ramp rate of 50 ^{o}C/min. The ramp rate was then reduced to 10 ^{o}C/min, with a final temperature of 330 ^{o}C held for 3.5 min.

Eluting peaks were then transferred at an auxiliary transfer temperature of 250 ^{o}C to a ITQ900-GC mass spectrometer (Thermo Scientific), with a filament delay of 5 min. 70 V electron ionisation was used, with an emission current of 50 μA. The ion source was held at 230 ^{o}C, the full scan mass range was 50-700 m/z with an automatic gain control of 50%, and maximum ion time of 50 ms.

Comparison to compound standards and identification of unknown sterol peak areas was performed using TraceFinder v3.3 (Thermo Scientific). Identification of peaks to which standards did not match was conducted by comparison to the NIST library, also using TraceFinder.

2.3.2.3 UV-Vis spectroscopy

The UV spectra of sterol extracts was obtained using a Shimadzu UV-2550 UV-Vis spectrophotometer and UVProbe v2.34 software. A wavelength range of 190 nm-350 nm was scanned, with *n*-heptane used as a blank. Samples were measured in a quartz cuvette, and 8 x 10^{7} cells were used per sample. The spectrum of ergosterol itself (Sigma), at 0.05 mg/ml dissolved in *n*-heptane, was also measured for comparison.

2.4 Molecular biology methods

2.4.1 General molecular biology methods

All genomic DNA preparations were performed with mid-log phase promastigotes using the Nucleospin Tissue kit, RNA extractions were performed from mid-log phase promastigotes with the Nucleospin RNA kit, and agarose gel extractions and PCR cleanup were performed with the Nucleospin gel and PCR cleanup kit (all Nucleospin kits were purchased from Macherey-Nagel). All primers were purchased from and Sanger sequencing reactions performed by Eurofins Genomics.

2.4.2 Polymerase chain reaction of genes of interest

Genes of interest (*LmxM.23.1300*, *LmxM.13.1530*, *LmxM.36.2380*) were amplified by PCR using primers containing regions of homology to the plasmid pGL1132 (see section 2.4.3). Amplification was performed with Phusion polymerase (New England Biolabs), according to manufacturer's protocols, with the addition of 3% DMSO. The thermocycling program was as follows: 98 ^{o}C for 30 s followed by 30 cycles of denaturation at 98 ^{o}C for 10 s, annealing at variable temperatures for 30 s and elongation at 72 ^{o}C for variable lengths of time; the last stage was 10 min further elongation at 72 ^{o}C. Specificity of amplification was assessed using a 1% agarose gel, and PCR products were cleaned up. For *LmxM.23.1300* amplified from both wild-type cells and AmBRA/cl1, an annealing temperature of 63 ^{o}C was used, with 30 s extension (forward primer sequence 5'-GTCAAGCGAATTCCATATGAATGGACTTCGCCTTTAACC-3', reverse 5'-AGTGGTGGTGGTGGTGGTGCTTATGCGCCGACCTTCTTTG-3'). For *LmxM.13.1530*, an annealing temperature of 67.3 ^{o}C was used, with an extension time of 90 s (forward primer sequence 5'-GTCAAGCGAATTCCATATGAATGCCCAATCAACCGGCG-3', reverse 5'-AGTGGTGGTGGTGGTGGTGCTTATGCACTATCGAGCGATCTCC-3'). For *LmxM.36.2380*, both annealing and extension were at 72 ^{o}C, giving a total time of 60 s per cycle at this temperature (forward primer sequence 5'-GTCAAGCGAATTCCATATGAATGTCCGCCGGTGGCCGT-3', reverse 5'AGTGGTGGTGGTGGTGGTGCCTACACCTCCTTGGACGGCTTGC-3').

2.4.3 Construction of overexpression plasmids

All vectors constructed were derived from the vector pGL1132, based on the pNUS backbone (Figure 2-1) (Tetaud *et al*. 2002). This vector allows for episomal overexpression of inserted genes (green fluorescent protein, *GFP*, is the insert in pGL1132). Efficient *trans*-splicing of genes is mediated by the 5'-UTR from phosphoglycerate kinase B and the 3'-UTR from phosphoglycerate kinase A. Downstream of this, a neomycin resistance gene is encoded, flanked by the 5'-UTR from phosphoglycerate kinase B and the *Crithidia* glutathionylspermidine synthetase gene to its 3'. The vector was linearised and *GFP* excised using the restriction enzymes BglII and XhoI in NEBuffer 3.1 (all from New England Biolabs), with linearised vector purified by 1% agarose gel electrophoresis followed by gel excision. Plasmid construction and transformation were achieved using the NEBuilder ligation-free cloning system (New England Biosciences) according to the manufacturer's protocol. Briefly, PCR amplicons containing *Leishmania* genes flanked by pGL1132 homology regions (see section 2.4.2) were combined with linearised vector in a 2:1 molar ratio and incubated with NEBuilder master mix for 20 min at 50 ^{o}C, before cooling. This was used to transform kit-supplied competent cells, which were subsequently plated onto Luria-Bertani (LB) agar containing 100 μg/ml ampicillin.

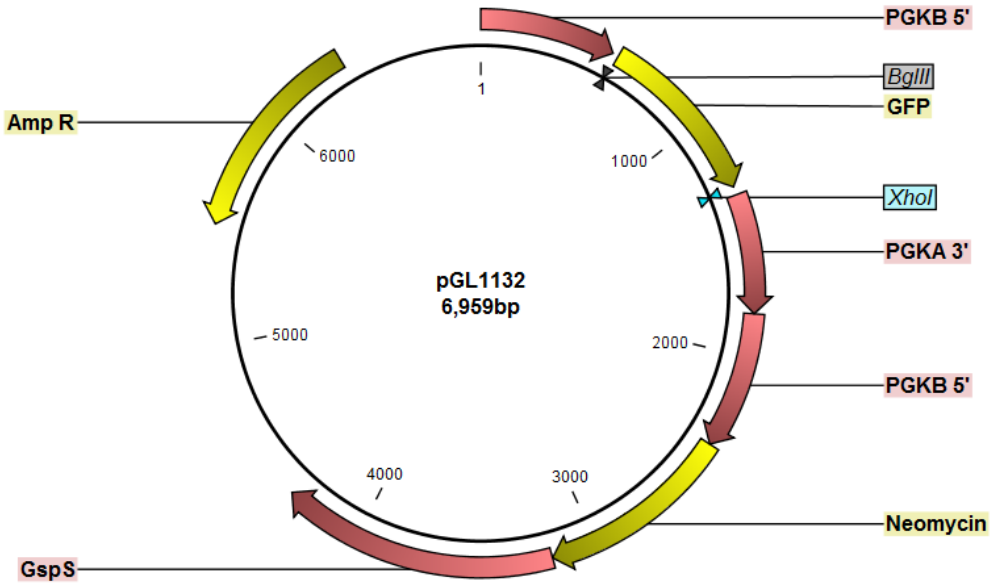


Figure 2-1: pGL1132 vector map. Vector map showing genes expressed, along with flanking UTRs to drive episomal expression in *Leishmania*. Restriction sites (BglII and XhoI) used here are shown. Yellow arrows represent genes: ampicillin resistance gene (Amp R), green fluorescent protein (*GFP*) and neomycin resistance gene (Neomycin). Red arrows represent 3'- and 5'-UTRs derived from phosphoglycerate kinases A (PGKA) and B (PGKB). (Image produced using CLC Genomics Workbench v7.6.4).

After incubation overnight at 37 °C, colonies were screened by PCR using a forward primer located upstream of the insert within the plasmid (5'-TCACATTCCGCTCTGTCCAC-3') and a gene-specific reverse primer (as used for initial gene amplification, see section 2.4.2), with an annealing temperature of 60 °C and an extension time of 30 s for *LmxM.36.2380* and 90 s for *LmxM.23.1300* and *LmxM.13.1530*. Positive colonies were amplified in LB broth with 100 μg/ml ampicillin and plasmids were isolated using the Nucleospin plasmid kit (Macherey Nagel). Inserts were amplified using vector specific primers 5'-TCACATTCCGCTCTGTCCAC-3' (forward) and 5'-GACAAAAAGGCGTCGAAGGAG-3' (reverse) with an annealing temperature of 60 °C and an extension time of 30 s. Correct sequences were verified by Sanger sequencing using the same primers. Due to the size of *LmxM.13.1530* (3315 bp), additional sequencing primers were used in the forward orientation (5'-GACGTCGTACGTATCAAGAAC-3', 5'-GCAGTGTGGTGGAACAACAAG-3', 5'-GGACGACACGGATGGCCTCAG-3', 5'-GAGATGGAGCAAGACATGCG-3').

2.4.4 Generation of *Leishmania* overexpression lines

Vectors were concentrated by ethanol precipitation prior to transfection. Briefly, 0.3 M sodium acetate (pH 5.2) was added to the vector in a 1:10 (v/v) sodium acetate:vector ratio, followed by ice cold ethanol in a 2:1 (v/v) ethanol:vector ratio. Samples were retained at -80 ^{o}C for several hours, prior to sedimentation by centrifugation at 17,000 x *g*, 10 min. Samples were washed once with 70% ethanol, followed by centrifugation again at 17,000 x *g*, 10 min and removal of supernatant. Samples were allowed to dry before resuspension in sterile water. DNA concentration was measured using a Nanodrop spectrophotometer (Thermo Fisher).

Leishmania lines were prepared for transfection as follows. Five x 10^7 late-log phase (1-1.5 x 10^7 cells/ml) were sedimented by centrifugation and resuspended in 200 μl ice cold transfection buffer (90 mM Na_2HPO_4, 90 mM $NaHPO_4$, 5 mM KCl, 50 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), 0.15 mM $CaCl_2$). This was combined with 10 μg plasmid and quickly transferred to a Gene Pulser electroporation cuvette (Bio-Rad). Transfection by electroporation was performed using an Amaxa Nucleofector II (Lonza) on program U-033. Cells were immediately transferred into cHOMEM and incubated at 25 ^{o}C overnight, before 100 x dilution into cHOMEM supplemented with 25 μg/ml G418 (Sigma). G418 is a drug that allows selection for cells expressing the Neomycin resistance gene. Cells were passaged at least twice prior to use, to eliminate non-transfected cells from the population (note that as a non-integrated, episomal vector, cloning post-transfection is unnecessary). Parasites were transfected with water as a control, and no growth was observed in the presence of G418, thus demonstrating efficient selection. All transfected parasite lines were routinely cultured in the presence of 25 μg/ml G418. Note that for Alamar blue drug sensitivity assays (section 2.1.2), parasites were incubated at a final concentration of 12.5 μg/ml to ensure that selection was maintained throughout the assay.

2.4.5 Quantitative PCR

Quantitative PCR (qPCR) was performed using normalisation to a housekeeping gene, *LmxM.36.2350*, encoding cytosolic glyceraldehyde 3-phosphate

dehydrogenase (GAPDH). The choice of housekeeping gene was made for a number of reasons: *LmxM.36.2350* is a single-copy gene which did not show altered expression either in the RNA-seq dataset generated here (see chapter 4) or a transcriptomics dataset comparing *L. mexicana* promastigotes and intracellular amastigotes (Fiebig *et al*. 2015). GAPDH has also previously been used as a reference gene for expression analysis (Bhandari *et al*. 2014, Vacchina *et al*. 2016). For CNV analysis, it has the added advantage of being on the same chromosome (chromosome 20) as *SMT* genes *LmxM.36.2380* and *LmxM.36.2390*, meaning that all changes seen are due to local effects, rather than whole chromosomal copy number changes.

2.4.5.1 Primer design

Primers for qPCR were designed using Primer Express v3 software (Applied Biosystems). Primer sequences are listed in Table 2-1. Note that *SMT* refers to primers equally specific to both *SMT* genes *LmxM.36.2380* and *LmxM.36.2390*. All primers were tested for efficiency using a serially diluted genomic DNA template (Supplementary file 2-2).

Table 2-1: Primer sequences for qPCR.

Target feature	Forward primer	Reverse primer
LmxM.13.1530	5'-ACGCAGGTAATGCGAAAAGG-3'	5'-GACGTCACCAGGGTGGATGT-3'
LmxM.23.1300	5'-TCCACCACGAGCTGTCAAAC-3'	5'-ACCCAGGCGGTCCCATATA-3'
LmxM.36.2350	5'-TCAACGACCTGCTGGATGTC-3'	5'-GCCATGCGTGGAGTCGTACT-3'
SMT (both)	5'-TGCCACGCGAAGGACAA-3'	5'-CGGGCTTGATGACACGAAA-3'
SMT 3' UTR A	5'-TGGAGTGTTTTTTGTTCTTGGTATG-3'	5'-CAGCAAGACATCAGCTAAGAAGAATAA-3'
SMT 3' UTR B	5'-TCCCTCCCTCCAAAGACATG-3'	5'-CTCTGTTCACGAAAGTTGTCATATTCT-3'
LmxM.36.2400	5'-GGTGGCTGTGGCCTTTCA-3'	5'-TCGCTTAGGGCAACCATACC-3'
LmxM.36.2540	5'-TGAGCGGTCACAACCTGAAG-3'	5'-GGAAGTGCGACGCGATGT-3'
LmxM.36.2550	5'-GGCGTACCCGGGTTTAGTTAA-3'	5'-CATTGACATCGGTGTTGTGCTT-3'

2.4.5.2 Reverse transcription

Reverse transcription of RNA, extracted as described in section 2.2.3.2 for intracellular amastigotes and 2.4.1 for promastigotes (albeit with a longer DNase treatment step of 30 min than the recommended period of 15 min), was performed using Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's protocol with a reaction time of 60 min. For each sample, a

reverse transcriptase (RT)-free reaction was performed as a control for genomic DNA contamination. The single-stranded cDNA product was diluted with nuclease-free water (Ambion) in order to produce signal within an appropriate range in subsequent qPCR reactions.

2.4.5.3 Measurement of gene expression and CNV

Gene expression was measured by quantitative reverse transcription PCR (qRT-PCR) using the Power SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer's protocol, with 2.5 μl diluted cDNA sample added per reaction, to a final volume of 25 μl. Primers were used at 300 nM final concentration, and reactions were prepared in MicroAmp Optical 96-well reaction plates (Applied Biosystems). Reactions were performed using a 7500 real-time PCR system (Applied Biosystems) according to the following protocol: samples were initially held for 2 min at 50 ^{o}C, followed by 10 min at 95 ^{o}C, then 40 cycles of 95 ^{o}C for 15 s followed by 60 ^{o}C for 60 s (during which step fluorescence was measured). ROX dye was used as a passive reference, and for initial optimisation, the dissociation curve option was used to assure specific amplification. All reactions were performed in triplicate, as was a blank water-only control, and RT-free reactions were measured in duplicate for each sample. A threshold δCt value of 6 was used as a difference between RT-treated and RT-free samples (i.e. that detected DNA was at least 64 times higher in the RT-treated than the RT-free samples), meaning that contaminating genomic DNA was deemed to make a negligible contribution to signal intensity. Note that for two samples, the first replicates of macrophages infected with AmBRB/cl2 and AmBRD/cl2, this threshold was not met, likely due to low signal (resulting from low infection rates). However, the δCt for these samples was still at least 5 (equivalent to a 32-fold difference in DNA abundance), meaning that this was tolerated.

For genomic DNA qPCR for CNV detection, reactions were carried out as described above, in technical triplicate (including for water controls), with the exception of RT-free controls, which were not relevant. Five μg DNA was used per reaction, extracted as described in 2.4.1 and quanitified using a Qubit (Thermo Fisher).

2.4.6 Sequencing of genomic DNA regions

In order to amplify the intergenic region between the two sterol C24-methyltransferase gene copies, primers were designed with the forward primer binding to the 3' end of *SMT* (5'-ACCGAAGGGCACGTATAAGG-3') and the reverse primer binding to the 5' end (5'-TTCACCATCGTCGTGGTAGC-3'). Amplification was performed using Q5 polymerase (New England Biolabs) according to the manufacturer's protocol. The thermocycling program was as follows: 98 ^{o}C for 30 s followed by 30 cycles of denaturation at 98 ^{o}C for 10 s, annealing at 60 ^{o}C for 30 s and elongation at 72 ^{o}C for 2 min; the last stage was 10 min further elongation at 72 ^{o}C. Specificity of amplification was assessed using a 1% agarose gel, and PCR products were cleaned up. The amplicon was cloned using the pGEM-T Easy vector system (Promega) according to the manufacturer's protocol, as follows: the amplicon was subjected to A-tailing using Gotaq G2 polymerase (Promega), followed by ligation into the pGEM-T Easy vector in a 1:1 vector:insert ratio with T4 DNA ligase (Promega) overnight at 4 ^{o}C. The plasmid was then used to transform competent DH5α *E. coli* on LB agar that had been overlaid with X-Gal and isopropyl ß-D-1-thiogalactopyranoside for blue-white selection, with 100 μg/ml ampicillin as a selection marker. White colonies were screened by colony PCR using the same PCR program described above, and colonies containing the insert were amplified, with vector being purified using the Nucleospin plasmid kit (Macherey-Nagel). Samples were then subjected to Sanger sequencing (performed by Eurofins Genomics), with four positive clones sequenced to account for potential PCR errors. For the first round of sequencing, M13 plasmid-specific universal sequencing primers were used, specifically M13 uni (-21) (5'-TGTAAAACGACGGCCAGT-3') and M13 rev (-29) (5'-CAGGAAACAGCTATGACC-3'). After this, insert-specific primers were designed: two in the same sense as M13 uni (-21) (5'-GAGAGAAAGAGCGGGGGAC-3', 5'-TGCGTCAAACAGGTGTGTGG-3') and two in the same sense as M13 rev (-21) (5'-GCTGTTGAGCTGCTCGTCGTTC-3', 5'-CCCTTCGCCCCATTCTTGCC-3'). Pairwise alignment using Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html) was used to align sequences from different clones and the pairwise alignment tool EMBOSS Matcher (https://www.ebi.ac.uk/Tools/psa/emboss_matcher/nucleotide.html) was

employed in order to find overlaps between consensus sequences, thus allowing the construction of the overall region into one contiguous sequence.

In order to amplify the *SMT* coding sequences in a transcript specific fashion (as described in section 5.2.2.1), Q5 polymerase was also used. In each case, the forward primer was the one used for amplification of *SMT* in section 2.4.2 and the reverse primer was that used for qPCR of transcript specific genes (*LmxM.36.2380* and *LmxM.36.2390*) as described in section 2.4.5.1. In the case of both amplicons, the PCR program was as described for the intergenic region above, with the exception that an annealing temperature of 58 ^{o}C was used along with an elongation time of 60 s. PCR products were not cloned but were sent as amplicons for Sanger sequencing.

2.5 Next generation sequencing methods

2.5.1 Library generation and sequencing

Genomic DNA and total RNA were extracted from mid-log phase parasites as described in section 2.4.1. For AmB resistant lines, samples were grown in cHOMEM in the absence of AmB, but no more than one passage since growth in AmB. Where both WGS and RNA-seq were performed for the same clone (AmBRA/cl1, AmBRB/cl2, AmBRC/cl3, AmBRD/cl2), extractions were performed from the same flask of parasites (with wild-type genomic DNA extracted from the same flask as wild-type RNA-seq replicate 1). Library preparation was performed at Glasgow Polyomics by Julie Galbraith (WGS libraries) and Dr David McGuinness (RNA-seq libraries). WGS libraries were produced using the TruSeq Nano DNA library preparation kit (Illumina) according to manufacturer's protocols except that initial fragmentation by sonication was performed using a Bioruptor Pico (Diagenode). RNA-seq libraries were prepared using the TruSeq stranded mRNA library preparation kit, including poly(A) mRNA selection (Illumina). For all samples, paired-end 2 x 75 bp reads were obtained using a NextSeq 500 sequencer (Illumina).

2.5.2 Data analysis

Output files were provided in fastq format by Glasgow Polyomics. Initial quality control was performed using FastQC v0.10.1 (Babraham Bioinformatics,

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptor removal and quality trimming were performed using Trim Galore! (Babraham Bioinformatics, https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), a wrapper script which makes use of FastQC for adaptor detection, as well as Cutadapt v1.5 for quality trimming (Martin 2011). Reads were subject to 3' trimming with a quality filter of 20, and reads less than 20 bp after trimming were excluded. Alignment was then performed to the *L. mexicana* MHOM/GT/2001/U1103 reverence genome release 9.0 (available on TriTrypDB, http://tritrypdb.org/tritrypdb/) using BWA v2.2.1 (Li & Durbin 2009) with default settings for paired-end data, and reads were then sorted according to alignment position using Samtools v1.2 (Li *et al*. 2009).

For WGS data, optical and PCR duplicates were then marked using Picard Tools v1.138 (http://broadinstitute.github.io/picard/). Per-gene coverage was calculated using HTSeq-count v0.6.1 (Anders *et al*. 2015) using default settings for unstranded, paired-end data. Chromosomal ploidy was calculated using a custom Python script (Supplementary file 2-3) as follows. First, per-gene coverage was normalised to gene length. The median normalised per-gene mapped fragment count was then calculated for each chromosome, and ploidy for a specific chromosome was calculated as double the ratio between the median normalised per-gene fragment count for that chromosome, and the median value for this across all chromosomes (double due to the assumption of basal diploidy). Displayed as an equation, this can be shown as follows for calculation of the ploidy a chromosome, with *M* representing the median length-normalised per-gene coverage for chromosome *i*:

$$Ploidy_i = 2 \times (\frac{M_i}{Median(M_{1:n})})$$

Haploid ratios were calculated as the length-normalised mapped fragment count for an individual gene divided by the median value for its parent chromosome, using a custom Python script (Supplementary file 2-4).

Identification of SNPs and InDels was then performed using Freebayes v1.1.0 (Garrison & Marth 2012), supplying per-chromosome ploidy as calculated above

(rounded to the nearest integer). Variants were annotated using Snpeff v3.6b (Cingolani *et al*. 2012), then filtered using a threshold quality score of 30, and requiring that alternative alleles appear in at least five reads across all samples. In order to filter for genotypes which differed in between strains with respect to types of alleles present (as explained in section 4.2.1.2), a custom Python script was used (Supplementary file 2-5).

For RNA-seq data, per-gene mapped fragment counting was performed without duplicate marking using HTSeq-count v0.6.1 (Anders *et al*. 2015) with default settings for reverse-sense stranded, paired-end data. Differential expression analysis was then performed using DESeq2 v1.14.0 (Love *et al*. 2014).

2.6 ^{35}S-methionine uptake assays

Uptake of ^{35}S-methionine was measured as an indicator of global translation rates. Methionine-free HOMEM was made up according to the recipe described in Supplementary file 2-1, without L-methionine. This was used to make up a stock solution of EasyTag [^{35}S]-methionine (Perkin Elmer) at 0.2 µCi/µl. Mid-log phase promastigotes were sedimented by centrifugation and resuspended at a final density of 5 x 10^6 cells/ml in methionine-free HOMEM with 10% FBS (note that methionine is likely present in FBS, but this did not appear sufficient to interfere with radiolabel uptake). For each parasite line used, cells were split into two tubes each containing 6 x 10^6 parasites. To one of these, cycloheximide (Sigma) was added from a 50 mg/ml stock in DMSO to a final concentration of 100 µg/ml. To each tube, 2.5 µl (0.5 µCi) [^{35}S]-methionine stock was added. Samples were incubated at 25 ^{o}C for variable lengths of time, before centrifugation at 1,900 x g for 5 min, washing twice in 1 ml PBS (repeating centrifugation as before between washes) and then resuspension in 200 µl 2% sodium dodecyl sulfate solution to lyse cells. Lysates were mixed with Optiphase HiSafe 2 scintillation fluid (Perkin Elmer) and scintillation counting was performed using a Hidex 300 SL scintillation counter. A count time of 60 s was used, with default settings for ^{35}S ß-emission but with the energy window adjusted to 5-700 keV.

3 Phenotypic and metabolic characterisation of four amphotericin B-resistant *Leishmania* lines

3.1 Introduction

The limited range of chemotherapeutic options for leishmaniasis, paired with increasing reliance on AmB as a treatment, makes the risk of emergent resistance to this drug an area of particular concern. In this respect, however, there has been scepticism in the field regarding the feasibility of emergence of AmB resistance, based on its long use in fungi with limited evidence of widespread resistance, as well as reports of high fitness costs that may prevent further transmission (Vincent *et al*. 2013). Establishing the phenotypic effects of selection of AmB resistance in *Leishmania in vitro* may provide insight into effects of similar selection in clinical populations. However, previous studies have tended to focus on only one (Mbongo *et al*. 1998, Pourshafie *et al*. 2004, Purkait *et al*. 2012, Brotherton *et al*. 2014, Fernandez-Prada *et al*. 2016, Mwenechanya *et al*. 2017) or two (Al-Mohammed *et al*. 2015) resistant lines. While taken overall these studies provide a useful body of evidence about the sorts of changes that may be expected, variation in both the species and the questions addressed makes direct comparison more challenging.

Here I sought to address this through independent selection of four resistant lines from a single *L. mexicana* parental line. This was in order to assess which changes arise stochastically during the selection process, and which are necessary consequences of selection of AmB resistance. One of the most important questions to address in terms of its relevance to policymaking is whether cross-resistance to other antileishmanial compounds is to be expected. Mbongo and colleagues previously revealed altered sensitivity in an AmB-resistant *L. donovani* line to a number of compounds, including hypersensitivity to PENT and ketoconazole but resistance to sinefungin (a methyltransferase inhibitor) and no change in sensitivity to PMM (Mbongo *et al*. 1998). More recently, another AmB-resistant *L. donovani* promastigote line revealed cross-resistance to PMM, mild hypersensitivity to PAT and unaltered MILT sensitivity (García-Hernández *et al*. 2012), whereas an AmB-resistant *L. mexicana* line revealed hypersensitivity to PENT but cross-resistance to ketoconazole, MILT and PAT (Mwenechanya *et al*. 2017). MILT cross-resistance was also reported in an

AmB-resistant *L. infantum* line (Fernandez-Prada *et al*. 2016), as has PAT cross-resistance (Moreira *et al*. 2011). However, the fragmented nature of this information means that variations in species and methodology cannot be ruled out as an underlying cause of this heterogeneous outlook. In this chapter, therefore, I directly compare independently selected lines to assess which changes, if any, are conserved.

Regarding broader effects on fitness, similar difficulties arise when interpreting current data. One study has identified hypersensitivity to ROS-inducing agents in AmB-resistant *L. mexicana* (Mwenechanya *et al*. 2017). Other studies, however, have reported upregulation of enzymes involved in oxidative stress (Purkait *et al*. 2012, Brotherton *et al*. 2014), and reduced accumulation of ROS in response to treatment with both AmB and other drugs (PAT and MILT) was observed in an AmB-resistant *L. infantum* line (Moreira *et al*. 2011). In these latter cases, however, direct sensitivity to ROS-inducing agents was not tested. Contradictions are also apparent in the fungal literature, with both ROS hypersensitivity (Vincent *et al*. 2013) and resistance (Sokol-Anderson *et al*. 1988) reported. Factors underlying these discrepancies have not been explored. Regarding the virulence of AmB-resistant *Leishmania*, little is known. Fernandez-Prada and colleagues noted that while an AmB-resistant *L. infantum* promastigote line was infectious to human THP-1 macrophage-like cells, infection rates were around half those for wild-type cells (Fernandez-Prada *et al*. 2016). Impaired cutaneous lesion formation has been noted in AmB-resistant *L. mexicana*; however this line was also observed to retain the resistance phenotype *in vivo* (Al-Mohammed *et al*. 2005). Nevertheless, such studies are complicated by mutations which arise during serial passaging as promastigotes irrespective of drug pressure (as noted, for example, in Mwenechanya *et al*. 2017), which may in itself contribute to loss of infectivity.

In order to understand the molecular basis of AmB resistance, as well as the sorts of phenotypes discussed above, it is important to understand metabolic changes in resistant lines. Identifying metabolic signatures of AmB resistance is useful both in that it helps to predict possible phenotypic consequences of resistance for the parasite as a whole, and may offer a source of markers of resistance for screening purposes. Of the changes described in the literature,

loss of ergostane-type sterols is by far the most consistently reported feature in AmB-resistant *Leishmania* (Mbongo *et al*. 1999, Al-Mohammed *et al*. 2005, Fernandez-Prada *et al*. 2016, Mwenechanya *et al*. 2017). The nature of these changes was discussed in greater detail in section 1.3.5.2. While loss of wild-type sterol is universal, the alternative sterols that increase in resistance are more variable. Because these sterols have different properties, it is likely they will result in different phenotypic consequences; however, once again the fragmented nature of prior studies, which focus on single resistant lines, hampers direct comparison in this respect. Thus by selecting multiple independent lines and characterising them in parallel, it should provide insight as to the sort of sterol changes which are possible, and potential consequences of those changes, which may even help to predict which sort of changes are actually likely to occur in the field.

Regarding the broader metabolic changes observed in AmB resistance, little work has been done and it is unclear whether universal metabolic signatures can be seen. One study (Berg *et al*. 2015) involved metabolic profiling of one AmB-resistant *L. donovani* line in the context of comparison with lines resistant to other drugs, as well as to drug combinations. The authors particularly noted an increase in proline and aromatic amino acids in this line, as well as in a line selected for cross-resistance to AmB and Sb^{III}. However, supplementation of wild-type cells with proline did not cause increase resistance to AmB (although it did result in decreased Sb^{III} sensitivity). Changes in lipid metabolism, including phospholipids and sphingolipids, were also observed in this line, as indeed such changes have been described by others (Fernandez-Prada *et al*. 2016). Mevalonate, a precursor to both sphingolipid and sterol biosynthesis, was also decreased.

Beyond this study, other work has been done on metabolomic characterisation of *Leishmania* resistant to MILT and Sb^{III} (Canuto *et al*. 2014, Berg *et al*. 2015). A multiple-platform analysis (LC-MS, GC-MS and capillary electrophoresis-MS) revealed decreases in metabolites of the arginine-polyamine pathway, including of arginine, ornithine and citrulline, in MILT-treated *L. donovani* promastigotes, but increases in these metabolites, along with the polyamine spermidine, in MILT-resistant parasites (Canuto *et al*. 2014). Trypanothione disulfide and its

biosynthetic precursor, glutathionyl-sperimidine, increased in MILT-resistant parasites, whereas trypanothione disulfide decreased after MILT treatment. Broader changes in amino acid metabolism were observed, including an increase in all but two (glycine and cysteine) out of the 20 proteinogenic amino acids. Many similar changes were also observed in Sb^{III}-resistant *L. donovani* (Berg *et al*. 2013), including increases in numerous free amino acids and polyamine and thiol metabolism-related metabolites.

I aimed, therefore, to characterise newly selected AmB-resistant *L. mexicana* to determine whether similar metabolic alterations are seen to those previously described, and how consistent these changes are between independent lines. This approach allows differentiation between changes that are universal and those which arise stochastically during selection, as well as potentially allowing association of specific metabolic changes with other phenotypes.

3.2 Results

3.2.1 Selection of AmB-resistant lines

AmB resistance was selected in four independent *L. mexicana* M379 lines from a single parental progenitor (termed AmBRA, AmBRB, AmBRC and AmBRD). AmB was increased in a stepwise manner from a starting concentration of 50 nM (being approximately equal to the wild-type starting IC_{50}), doubling each time stable growth was achieved. Wild-type cells were also cultured continuously in parallel without AmB, and sensitivity was routinely monitored (Figure 3-1). On ceasing selection, three clones were obtained from each line by limiting dilution, and sensitivity was determined (Figure 3-2). As IC_{50} values were very similar for individual clones obtained from the same line, the majority of analysis was carried out with only one clone per line, the one with the highest mean IC_{50} (AmBRA/cl1, AmBRB/cl2, AmBRC/cl3, AmBRD/cl2).

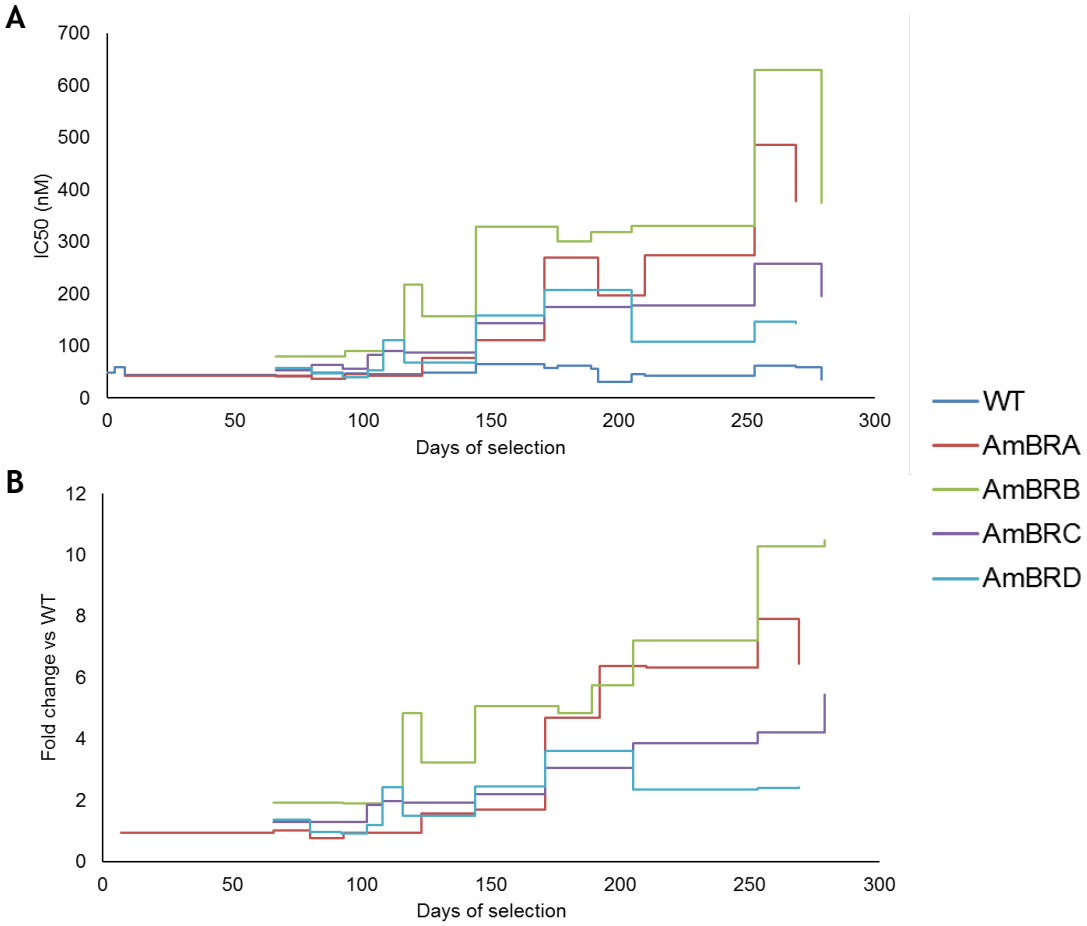


Figure 3-1: Selection of AmB resistance. AmB resistance was selected over around 280 days and changes to AmB sensitivity were routinely monitored. A) AmB sensitivity expressed as IC_{50} in nM, as determined by Alamar blue assay. B) Fold change in IC_{50} in comparison to a wild-type line cultured in parallel in the absence of drug.

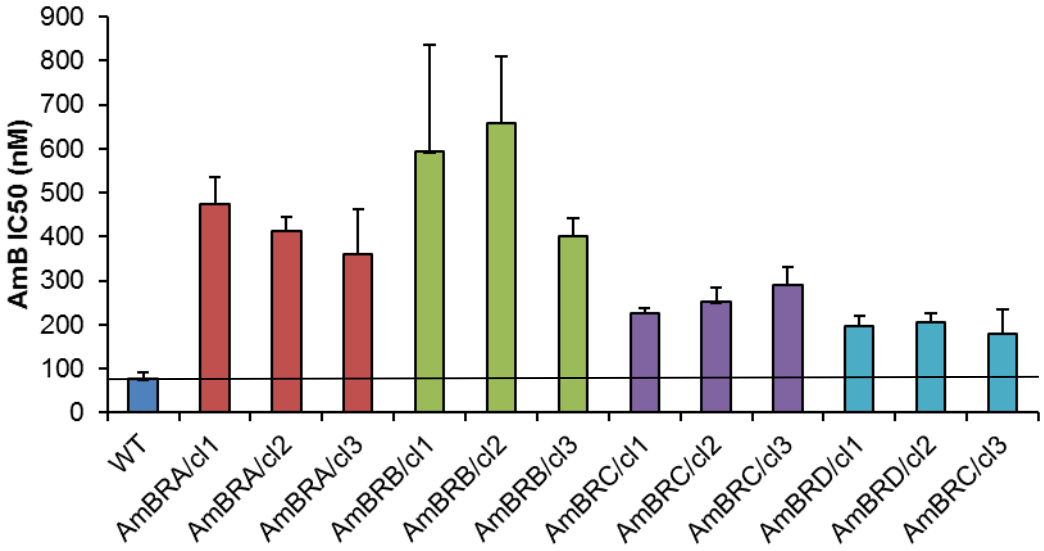


Figure 3-2: AmB sensitivity of individual clones. Individual clones from each independent line were obtained by limiting dilution and assayed for AmB sensitivity. Values are the mean of three biological replicates and error bars denote standard deviation. The horizontal black line represents wild-type IC_{50}.

Resistance was found to be variable between different lines, despite selection in parallel (Table 3-1). In comparison to wild-type cells, lines showed increases in IC_{50} of 7.3-fold, 8.4-fold, 3.9-fold and 2.6-fold for AmBRA/cl1, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 respectively. The stability of the AmB resistance phenotype was assessed by serially passaging resistant lines at least 15 times in the absence of drug pressure. In all cases IC_{50} values remained higher than for wild-type parasites, and similar to the levels in selected cells, confirming that the resistance phenotype was stable (Table 3-1).

3.2.2 Sensitivity of AmB-resistant lines to antileishmanial agents

3.2.2.1 Sensitivity to antileishmanial drugs

The presence of cross-resistance or hypersensitivity was determined for five antileishmanial drugs (Table 3-1). While all lines were resistant to AmB, resistance to other drugs was more heterogeneous. For PAT, significant hypersensitivity was observed for AmBRB/cl2, with no change in other lines. For MILT, two lines were found to be hypersensitive, AmBRA/cl1 and AmBRC/cl3; AmBRB/cl2, by contrast, showed a 2.3-fold increase in MILT resistance. PMM also exhibited variable efficacy, with hypersensitivity noted in AmBRA/cl1 and cross-resistance in AmBRC/cl3. The only change consistent across all lines was hypersensitivity to PENT. Interestingly, lines that were most AmB-resistant appeared to show the greatest degree of PENT hypersensitivity.

3.2.2.2 Sensitivity to ergosterol biosynthesis inhibitors

Sensitivity to three ergosterol biosynthesis inhibitors was assessed (Table 3-1). Ketoconazole targets the enzyme CYP51 (*LmxM.11.1100*), whereas both fenpropidin and spiroxamine inhibit sterol C14-reductase (*LmxM.31.2320*). No changes in sensitivity to ketoconazole or fenpropidin were noted, but AmBRB/cl2 and AmBRD/cl2 showed evidence of cross-resistance and hypersensitivity, respectively, to spiroxamine. I also included imipramine, a tricyclic antidepressant previously shown to inhibit C24-methylation in *Leishmania* (Andrade-Neto *et al*. 2016). Interestingly, all four lines showed mild but significant hypersensitivity to this compound.

Chapter 3

Table 3-1: Sensitivity of AmB-resistant lines to other antileishmanial agents.

Sensitivity to a range of compounds is given in μM, with the exception of GOX which is in mU/ml. Mean IC_{50} values are shown, ± standard deviation. Statistically different values from wild-type ($P < 0.05$) are shown in bold. Assays for each compound were performed as four biological replicates, with the exception of AmB after 15 passages without drug, for which three biological replicates were performed. Abbreviations used are as follows: AmB – amphotericin B, PAT – potassium antimonyl tartrate, MILT – miltefosine, PMM – paromomycin, PENT – pentamidine, GOX – glucose oxidase

	Wild-type	AmBRA/cl1	AmBRB/cl2	AmBRC/cl3	AmBRD/cl2
AmB	0.0585 ± 0.00314	**0.427 ± 0.0225 ($P = 5.62 \times 10^{-8}$)**	**0.492 ± 0.0496 ($P = 2.28 \times 10^{-6}$)**	**0.231 ± 0.0282 ($P = 1.88 \times 10^{-5}$)**	**0.153 ± 0.0370 ($P = 0.00223$)**
AmB (15 passages)	0.0541 ± 0.0135	**0.437 ± 0.0456 ($P = 1.53 \times 10^{-4}$)**	**0.574 ± 0.101 ($P = 9.06 \times 10^{-4}$)**	**0.228 ± 0.0136 ($P = 9.67 \times 10^{-5}$)**	**0.124 ± 0.0242 ($P = 0.0121$)**
PAT	105 ± 14.1	94.4 ± 14.2 ($P = 0.323$)	**65.9 ± 2.51 ($P = 0.00152$)**	85.2 ± 17.4 ($P = 0.123$)	146 ± 48.4 ($P = 0.155$)
MILT	41.6 ± 9.87	**10.7 ± 1.01 ($P = 7.87 \times 10^{-4}$)**	**96.1 ± 2.35 ($P = 2.6 \times 10^{-4}$)**	**22.0 ± 2.57 ($P = 0.0106$)**	29.5 ± 5.83 ($P = 0.0783$)
PMM	57.5 ± 17.6	**13.9 ± 1.45 ($P = 0.00262$)**	60.3 ± 8.30 ($P = 0.785$)	**132 ± 45.1 ($P = 0.0219$)**	87.9 ± 33.5 ($P = 0.161$)
PENT	2.87 ± 0.0648	**0.613 ± 0.0191 ($P = 7.57 \times 10^{-10}$)**	**0.313 ± 0.0284 ($P = 4.73 \times 10^{-10}$)**	**0.702 ± 0.0304 ($P = 1.36 \times 10^{-9}$)**	**0.694 ± 0.0292 ($P = 1.28 \times 10^{-9}$)**
Ketoconazole	10.8 ± 4.45	8.27 ± 1.52 ($P = 0.331$)	7.58 ± 3.22 ($P = 0.291$)	12.5 ± 2.12 ($P = 0.513$)	6.22 ± 1.70 ($P = 0.105$)
Fenpropidin	64.8 ± 21.0	37.4 ± 16.2 ($P = 0.084$)	74.5 ± 10.3 ($P = 0.438$)	71.9 ± 13.7 ($P = 0.591$)	39.4 ± 17.8 ($P = 0.114$)
Spiroxamine	14.0 ± 0.645	12.7 ± 4.37 ($P = 0.596$)	**39.4 ± 12.0 ($P = 0.00541$)**	12.4 ± 6.86 ($P = 0.671$)	**3.94 ± 1.80 ($P = 4.51 \times 10^{-5}$)**
Imipramine	55.0 ± 2.65	**42.3 ± 2.52 ($P = 4.51 \times 10^{-4}$)**	**48.8 ± 3.01 ($P = 0.0217$)**	**45.6 ± 2.13 ($P = 0.00153$)**	**44.0 ± 0.74 ($P = 2.12 \times 10^{-4}$)**
GOX	2.39 ± 0.0320	**1.78 ± 0.0877 (1.3×10^{-5})**	2.84 ± 1.06 ($P = 0.423$)	2.89 ± 1.11 ($P = 0.402$)	2.19 ± 0.180 ($P = 0.0783$)
Menadione	5.68 ± 0.685	4.81 ± 0.789 ($P = 0.148$)	**4.08 ± 0.224 ($P = 0.00443$)**	4.94 ± 0.698 ($P = 0.181$)	6.34 ± 0.710 ($P = 0.229$)
Methylene blue	5.67 ± 1.31	**0.186 ± 0.0425 ($P = 1.62 \times 10^{-4}$)**	**0.301 ± 0.0592 ($P = 1.82 \times 10^{-4}$)**	**1.57 ± 1.29 ($P = 0.00437$)**	**1.18 ± 0.243 ($P = 5.26 \times 10^{-4}$)**

3.2.2.3 Sensitivity to oxidative stress-inducing agents

GOX is an enzyme which catalyses the oxygen-dependent oxidation of glucose to glucono-1,5-lactone, generating H_2O_2. H_2O_2 is itself highly unstable and therefore not suited for the Alamar blue assay used here (which takes place over five days). GOX was therefore used as a stable source of H_2O_2 in the medium. Of the four AmB-resistant lines, AmBRA/cl1 exhibited mild hypersensitivity, whilst the other three lines exhibited no change (Table 3-1). Menadione has been used previously as a source of intracellular superoxide ions (Ghosh *et al*. 2015). Again, little change in sensitivity was observed, although AmBRB/cl2 demonstrated mild hypersensitivity. Sensitivity to another agent, methylene blue, was also determined. Methylene blue induces intracellular oxidative stress through oxidation of NADPH to $NADP^+$, thus depleting an important source of reducing capacity in the cell (Maugeri *et al*. 2003). All lines were found to be hypersensitive to this compound. AmBRA/cl1 in particular was found to be 30-fold hypersensitive.

3.2.3 Fitness phenotype characterisation of AmB-resistant lines

3.2.3.1 Promastigote growth curves

Growth rate in the promastigote stage was assessed. Overall no large changes were observed, however AmBRA/cl1 appeared to exhibit a minor growth defect during log-phase (the first 72 hours), and there was evidence that AmBRD/cl2 reached a higher density than wild-type during stationary phase. To test this, linear models were generated comparing wild-type and each resistant line individually (cell line as one factor, time as the other). This revealed a significant effect of cell line only for AmBRD/cl2 ($P = 0.00249$). However, when only the first 96 hours were included, before stationary phase was reached, AmBRA/cl1 revealed moderate significance ($P = 0.0374$). Time was a significant factor under all conditions tested. Therefore, selection of AmB resistance does not appear to have major effects on growth in the promastigote stage.

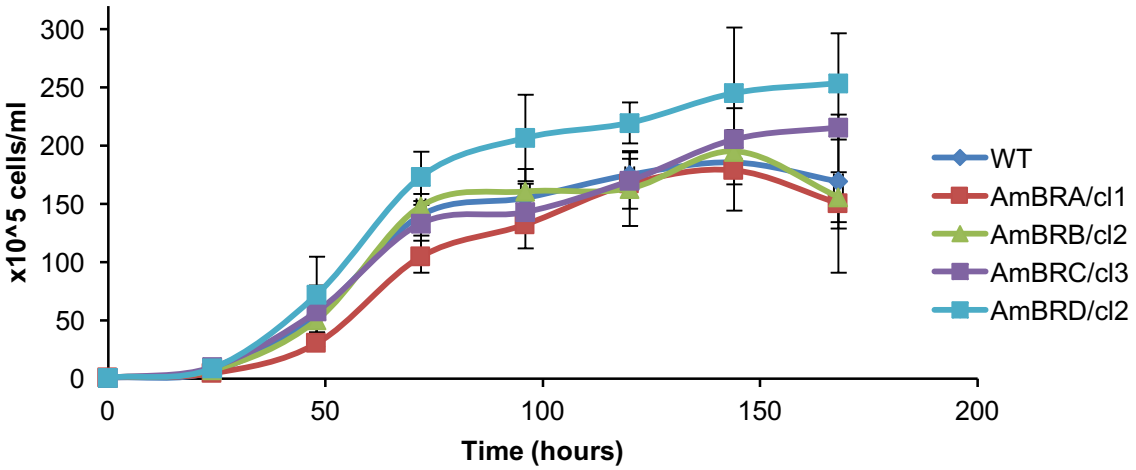


Figure 3-3: Growth of *Leishmania* promastigotes. Increase of cell density over time in the promastigote stage was measured by counting using a haemocytometer, starting at a density of 10^5 cells/ml.

3.2.3.2 Infectivity and AmB sensitivity in macrophages

The ability of resistant lines to infect and replicate in primary murine bone marrow-derived macrophages (BMDMs) was assessed. Infection levels were determined after infection for four hours, with time points at seven, 24, 48, 96 and 144 hours after initial infection (including the four hour initial infection period). Because long-term passaging as promastigotes (as during resistance selection) can reduce virulence, a wild-type line which had been cultured in parallel with resistant lines in the absence of drug for over 50 passages was included, as well as the parental line which had been subject to fewer than 15 passages since recovery from a mouse. Percentage infection (the percentage of total macrophages which were infected with parasites) was calculated (Figure 3-4A). This revealed that AmBRA/cl1, AmBRB/cl2 and AmBRD/c2 all had much lower infection rates than low passage wild-type cells, although this was comparable to highly passaged wild-type cells, suggesting that poor infection rates could result simply from long-term growth as promastigotes. AmBRC/cl3, by contrast, showed infection rates at least as high as low passage wild-type parasites, with 75% of macrophages infected after 144 hours in comparison to only 55% for low passage wild-type cells. In all cases, an initial drop in infection was observed between 24 and 48 hours. This is likely due to macrophage-mediated killing of some parasites in the initial phase of infection.

As uninfected macrophages become more likely to detach over time, percentage infection can in itself be a misleading value as it can artificially increase due to loss of uninfected macrophages alone. Therefore, the number of parasites per infected macrophage was also determined, and showed similar dynamics to percentage infection (Figure 3-4B). While initially, after seven hours, there were an average of 2.7 and 3.4 parasites per infected macrophage for low passage wild-type and AmBRC/cl3 parasites, respectively, this increased by 144 hours to 5.9 and 7.9, respectively, thus providing clear evidence of intracellular replication. By contrast, other lines showed only marginal evidence of growth, if any. In order to determine the statistical significance of these changes, I generated linear models comparing each cell line with low passage wild-type parasites, with cell line and time as explanatory variables and parasites per infected macrophage as the response variable. This revealed no effect of cell line for AmBRC/cl3 ($P = 0.406$) whereas a significant effect was observed for high passage wild-type ($P = 1.65 \times 10^{-6}$), AmBRA/cl1 ($P = 2.72 \times 10^{-4}$), AmBRB/cl2 ($P = 2.22 \times 10^{-7}$) and AmBRD/cl2 ($P = 4.83 \times 10^{-5}$). Time was a significant factor in all cases. By contrast, out of all resistant lines, only AmBRC/cl3 showed a significant effect of cell line in comparison to high passage wild-type cells ($P = 3.08 \times 10^{-4}$), with time again a significant factor in all cases. Therefore, while infectivity rates for AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2 are low, this may be attributable to effects of long term passaging.

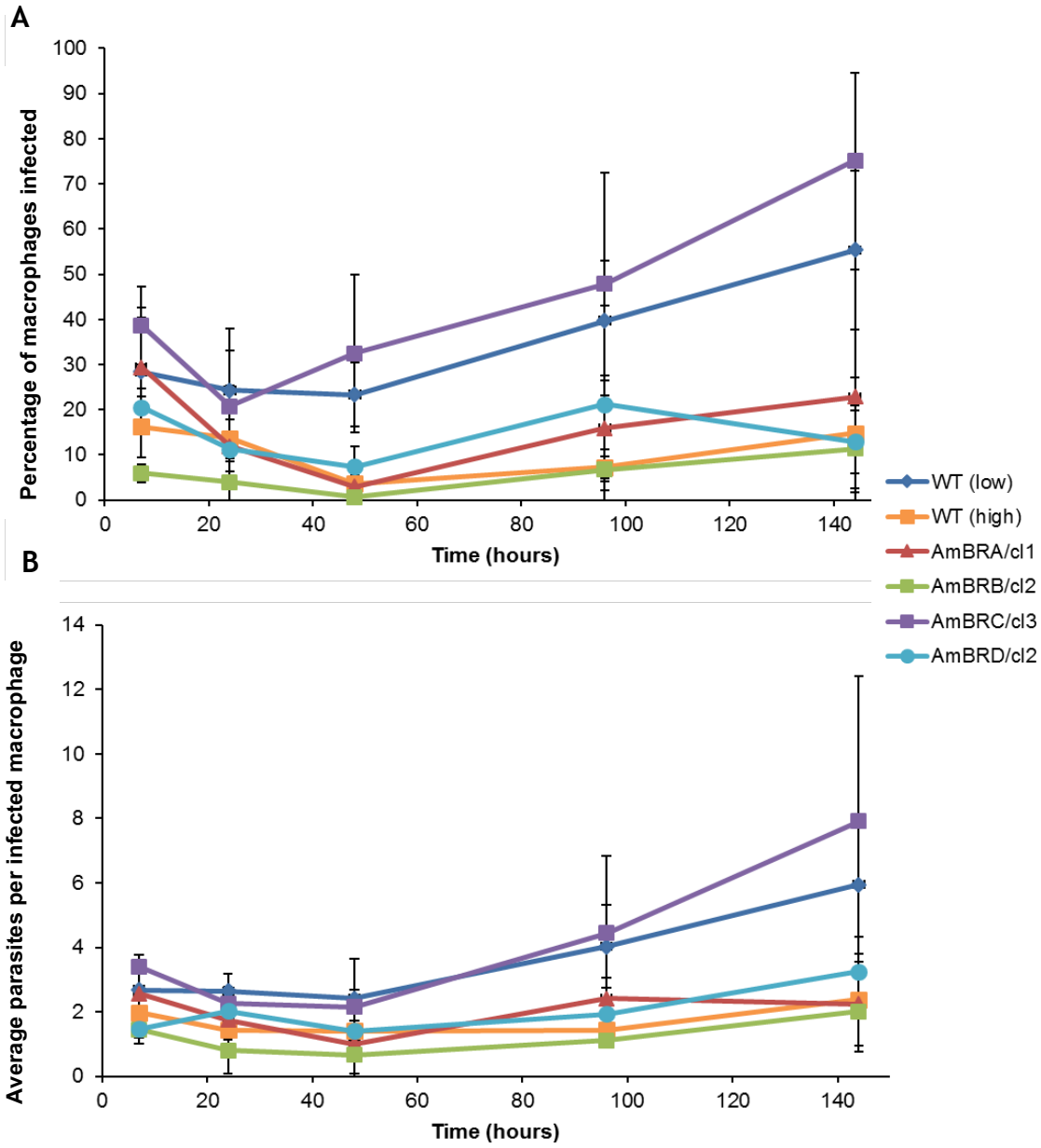


Figure 3-4: Infectivity of AmB-resistant lines in primary murine macrophages. Macrophages were infected for four hours and then incubated for variable periods of time before quantification of infection load by Giemsa staining. WT (low) and (high) relate to wild-type parasites subjected to fewer than 15 or greater than 50 passages, respectively, since isolation from a mouse. At least 100 total macrophages were counted for each sample. A) Percentage of macrophages infected. B) Average number of parasites per infected macrophage, calculated as the total number of intracellular parasites counted divided by the number of infected macrophages. Three biological replicates (with macrophages derived from three separate mice) were used, with mean values displayed and error bars representing standard deviation.

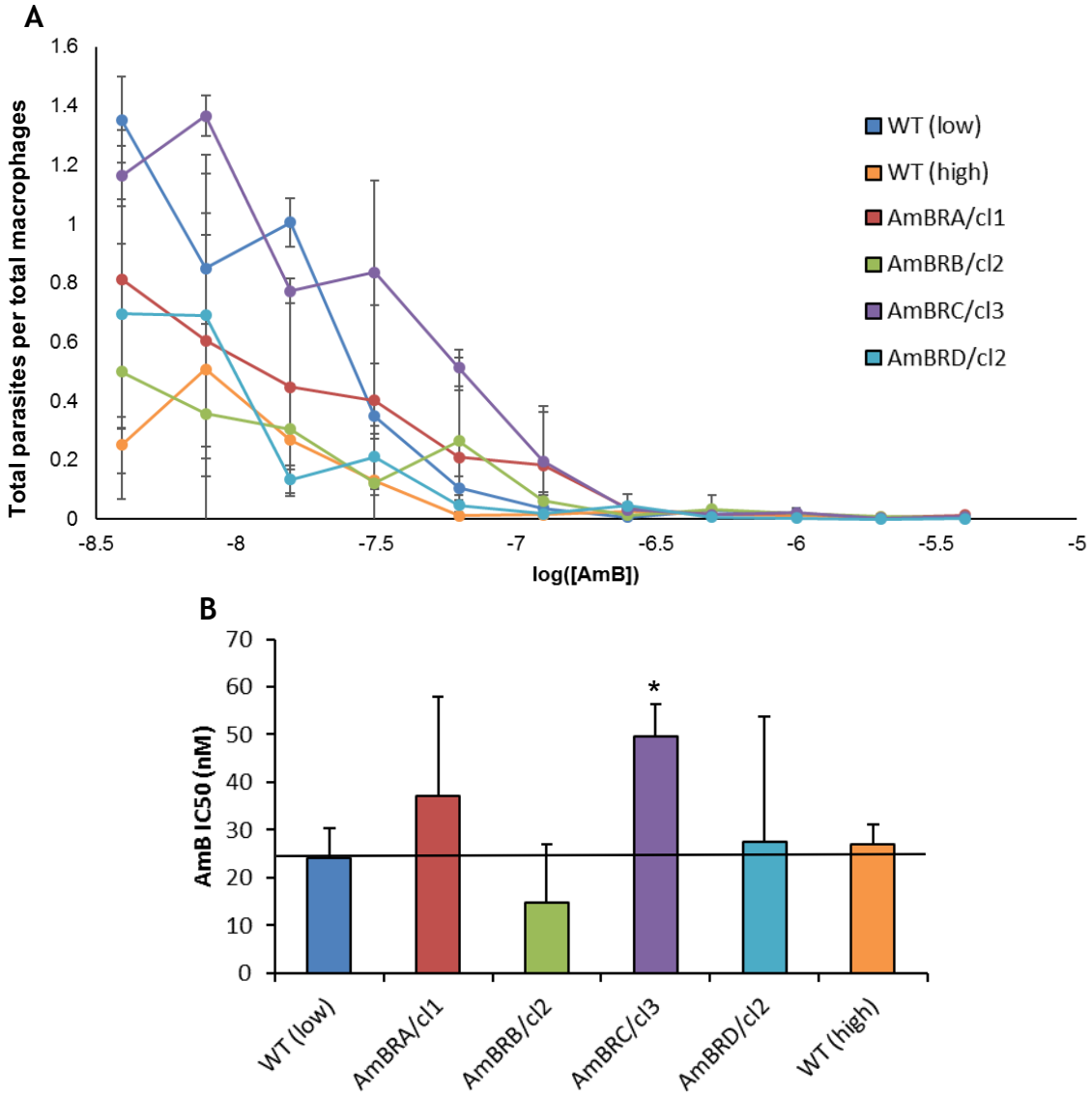


Figure 3-5: AmB sensitivity of intracellular amastigotes. Macrophages were infected for 24 hours prior to treatment with variable concentrations of AmB for a further 48 hours and quantification by Giemsa stain. WT (low) and (high) relate to wild-type parasites subjected to fewer than 15 or greater than 50 passages, respectively, since isolation from a mouse. Infection burden was quantified as the average ratio of total parasites counted against total macrophages counted, and three biological replicates (with macrophages derived from three separate mice) were obtained. A) Mean numbers of parasites per macrophage for all parasite lines tested, plotted against the log-transformed molar AmB concentration. B) IC_{50} values for intracellular amastigotes. Error bars represent standard deviation. Asterisks denote statistical difference from wild-type low passage cells ($P < 0.05$). The horizontal black line represents wild-type IC_{50}.

Sensitivity to AmB as intracellular amastigotes was also determined. BMDMs were infected for 24 hours before removal of any extracellular parasites and exposure to serial twofold dilutions of drug. After a further 48 hours infection burden was quantified and used to calculate IC_{50}. In this case, all samples were harvested at the same time, and hence time-dependent effects on percentage infection due to preferential loss of uninfected macrophages were not expected. Therefore,

the indicator of infection burden used was parasites per macrophage, including both infected and uninfected macrophages (which therefore takes into account both the total percentage infection and the number of parasites per individual infected macrophage) (Figure 3-5A). Comparison of IC_{50} values (Figure 3-5B) revealed that only one line, AmBRC/cl3, retained a significantly higher IC_{50} to AmB than low passage wild-type cells (49.7 nM compared to 24.1 nM, P = 0.00813). It must be noted, however, that variability was very high for some of the other resistant lines. This may be due to biological factors, such as variation between macrophages derived from different individual mice, or technical factors such as the difficulty of reliably quantifying infection levels when infection rates are low (and therefore small changes in absolute counts can have a large influence on the calculated IC_{50}). Nevertheless, it is noteworthy that similar levels of variation were not observed for the high passage wild-type cells, despite having similarly low infection rates.

3.2.3.3 Infectivity of resistant lines *in vivo*

In order to determine whether any of the AmB-resistant lines were capable of survival *in vivo*, BALB/c mice were infected with stationary phase promastigotes. Since AmBRC/cl3 was the only line to show high infection levels in macrophages, two mice each were initially infected with wild-type and AmBRC/cl3 parasites. However, one mouse each was also infected with AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. For wild-type and AmBRC/cl3 mice, footpad swelling was evident by four weeks after infection, at which point mice were sacrificed and in both cases viable promastigotes could be derived from lymph node homogenates. For wild-type, one mouse showed no change in footpad swelling, the other showed swelling from 1.6 mm up to 2.2 mm. For AmBRC/cl3, both mice showed swelling from 1.6 mm to 1.8 mm. For AmBRD/cl2, swelling from 1.6 mm to 2.1 mm was evident by nine weeks after infection and viable promastigotes were recovered from lymph nodes. No swelling was ever observed for AmBRA/cl1 and AmBRB/cl2, hence lymph nodes were not taken. Recovered promastigotes were tested to see whether the resistance phenotype was retained. Wild-type parasites obtained from lymph nodes had an IC_{50} for AmB of 73.7 nM (± 6.7 nM), whereas AmBRC/cl3 and AmBRD/cl2 had IC_{50} values of 229 nM (± 2.4 nM) and 173 nM (± 36.1 nM) respectively, both of which were statistically different from wild-type.

Hence both parasite lines showed stability of the resistance phenotype even after infection for several weeks in mice.

3.2.4 Characterisation of sterol changes in AmB resistance

Having thus established a range of phenotypic characteristics of the AmB-resistant lines, I sought to determine the basis of these phenotypes, starting with effects on cellular metabolism. Due to the abundance of evidence linking changes in sterol composition to AmB resistance, I aimed to characterise the sterols present in these lines.

3.2.4.1 UV spectroscopy

The conjugated double bond system present in ergosterol (across C5, C6, C7 and C8) gives it a distinctive UV absorbance spectrum that is absent in cholesterol and other sterols (Figure 3-6) (Sokol-Anderson *et al*. 1988). Previously, this has been used in fungi to detect changes to sterol composition (Sokol-Anderson *et al*. 1988, Vincent *et al*. 2013). I therefore sought to determine whether a similar spectrum could be detected in *Leishmania* sterol extracts and whether this was altered in the AmB-resistant lines. Another AmB-resistant line, known to exhibit disrupted ergosterol biosynthesis due to a mutation in *CYP51* (Mwenechanya *et al*. 2017), was included for comparison.

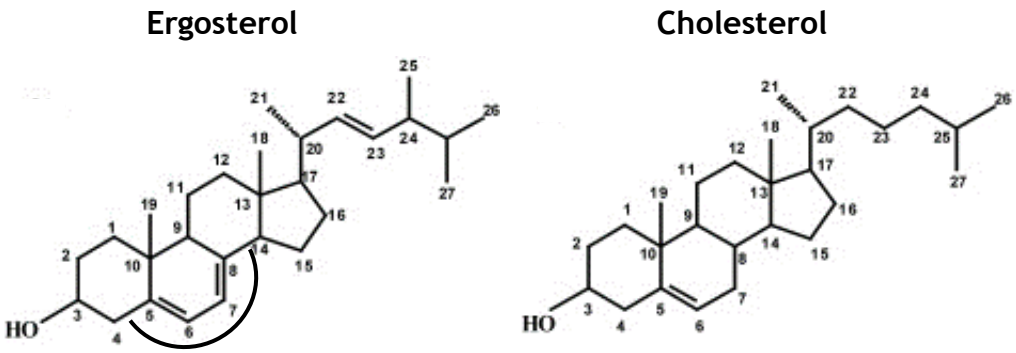


Figure 3-6: Comparison of double bond conjugation in ergosterol and cholesterol. The presence of 5(6)- and 7(8)-double bonds in ergosterol (left) allows conjugation and delocalisation of electrons across this four-carbon system. The presence of this system allows absorption within the UV range. By contrast, cholesterol (right) has only a 5(6)-double bond, which does not permit UV absorption. Hence only sterols with ergostane-type double bond conjugation have this sepctrum.

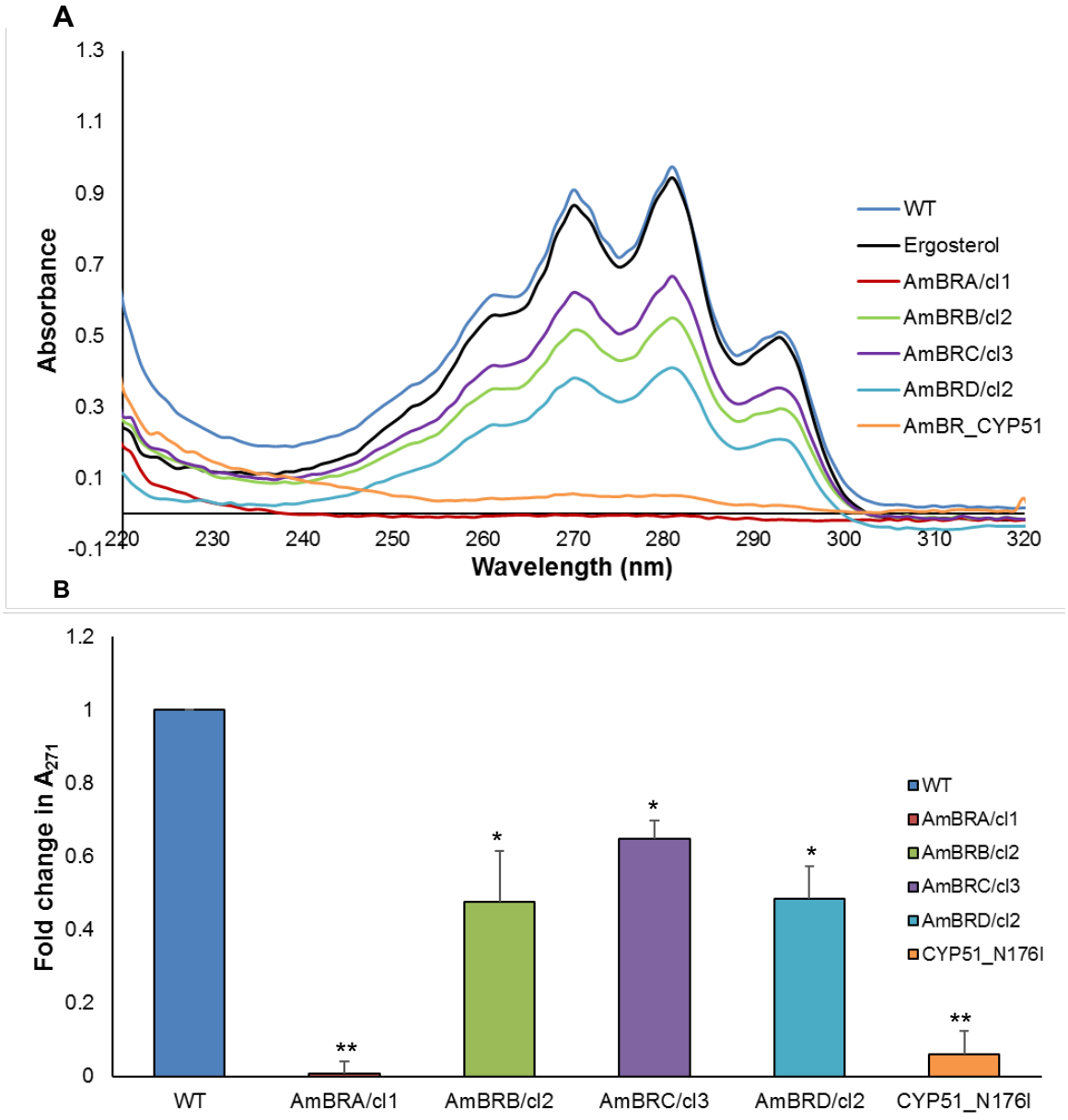


Figure 3-7: UV absorbance of *Leishmania* sterol extracts. A) UV spectra of sterol extracts from wild-type and AmB-resistant lines. Representative of three independent biological replicates, with 0.05 mg/ml ergosterol dissolved in *n*-heptane included as a standard. B) Absorbance at 271 nm, expressed as a fold change in comparison to wild-type cells. Values are the mean of three independent biological replicates, with error bars representing standard deviation. Asterisks denote statistical difference from wild-type sterol extracts (*P < 0.05, **P < 0.005).

The UV spectrum of wild-type *Leishmania* sterol extracts shows a strong similarity with 0.05 mg/ml ergosterol (Figure 3-7A), with high peaks at 271 and 281 nm, and a lower intensity peak at 293 nm. Similar spectra, albeit with reduced magnitude, were visible for AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. By contrast, little evidence could be observed of any UV spectra for AmBRA/cl1 or the *CYP51* mutant line (although a slightly higher basal level of absorbance was observed in the *CYP51* mutant). In order to quantify changes more precisely,

absorption at 271 nm was examined. Because of drift in absolute absorbance values between replicates that appeared to be a batch effect, this was expressed as fold change relative to wild-type cells. AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 all exhibited moderate but significant decreases in average A_{271} relative to wild-type (0.48-fold with P = 0.0232, 0.65-fold with P = 0.00631, and 0.48-fold with P = 0.00956, respectively), whereas far bigger decreases were noted for AmBRA/cl1 and the *CYP51* mutant (0.006-fold with P = 3.68×10^{-4} and 0.06-fold with P = 0.00149, respectively). This suggests that sterol composition varied between resistant lines and that double bond conjugation was largely absent in the sterols of AmBRA/cl1 and the *CYP51* mutant, but not AmBRB/cl2, AmBRC/cl3 or AmBRD/cl2.

3.2.4.2 GC-MS

In order more precisely to determine the nature of alterations to sterol composition, a sterol-focused GC-MS approach was used (Table 3-2, Figure 3-8). Peak areas were used to derive total sterol composition. P-values adjusted for multiple comparisons are given in full in Supplementary file 3-1. Of the panel of control sterols that were included, only one of them, cholesterol, matched to peaks found in any samples at a frequency of greater than 1%. Mean relative abundance values of less than 1% were excluded on the basis that it was unclear how close these were to the detection limit and hence how reliably detected they were over background. Cholesterol, which is taken up from the medium, was present at a relative abundance of 1-7% in all samples, with a significant decrease in AmBRA/cl1 (P = 0.0213). Interestingly, ergosterol was not detected in any samples. Instead, the most abundant sterol in wild-type parasites (82.6% of total sterol) was an isomer of ergosterol ($C_{28}H_{44}O$) that exhibited a different retention time (11.52 min) from the ergosterol standard (11.20 min). Based on previous sterol analyses in other *Leishmania* species (de Souza & Fernandes Rodrigues 2009, Andrade-Neto et al. 2016), this was suggested to represent ergosta-5,7,24(28)-trienol. Further evidence is provided by the UV spectra of sterol extracts (Figure 3-7), in which the wild-type spectrum matches almost perfectly with that of pure ergosterol, thus suggesting that the major wild-type sterol possesses the same desaturation in the core ring structure as ergosterol itself. Based on matching to NIST libraries, a secondary sterol, ergosta-7,22-dienol, was also abundant (11.7%).

Table 3-2: GC-MS analysis of sterol composition in *Leishmania* lines.

Percentage composition is derived from the peak area divided by the total sum peak area for all detected sterol peaks, and shown for each line on the right hand columns of the table. Peaks representing less than 1% mean composition are excluded from this table. Only cholesterol was included as a standard, with other peaks not matching to any standard used. Other peaks were identified based on comparison to NIST libraries, with the exception of ergosta-5,7,24(28)-trienol which was identified on the basis of comparison to literature (de Souza & Fernandes Rodrigues 2009, Andrade-Neto et al. 2016). Mean percentage composition values for three biological replicates are shown, ± standard deviation. Abbreviations: RT – retention time, TMS – trimethyl silane (the derivatising agent used to improve compound volatility).

RT (min)	Major fragment mass (*m/z*)	Molecular ion (*m/z*)	-TMS (*m/z*)	Formula -TMS	Putative identification	Evidence	WT	AmBRA/ cl1	AmBRB/ cl2	AmBRC/ cl3	AmBRD/ cl2
10.70	368	459	386	$C_{27}H_{46}O$	Cholesterol	Match to standard	3.00 ± 0.750	1.23 ± 0.151	3.71 ± 0.637	6.98 ± 3.63	5.54 ± 4.29
11.19	349	455	382	$C_{27}H_{42}O$	Cholesta-5,7,22-trienol	NIST score 678	1.30 ± 1.40		79.8 ± 0.537	85.8 ± 4.87	85.9 ± 6.53
11.30	343	456	383	$C_{27}H_{44}O$	Cholesta-5,7,24-trienol	NIST match 699 to desmosterol, wrong RT			5.85 ± 0.830	5.79 ± 0.909	5.65 ± 1.46
11.52	363	469	396	$C_{28}H_{44}O$	Ergosta-5,7,24(28)-trienol	NIST match 804 to ergosterol, wrong RT	82.6 ± 0.308		9.53 ± 0.714		1.83 ± 3.17
11.64	343	470	397	$C_{28}H_{46}O$	Ergosta-7,22-dienol	NIST score 701	11.7 ± 1.30	97.3 ± 0.245			

Large shifts in sterol composition were detected in all resistant lines. In AmBRA/cl1, peaks relating to cholesta-5,7,22-trienol and ergosta-5,7,24(28)-trienol were undetected. By contrast, 97.3% of cellular sterol was ergosta-7,22-dienol, hence fitting with the loss of UV spectrum as this sterol lacks 5(6)-desaturation. However, in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, a separate set of changes were seen. Loss of both dominant sterols in wild-type cells (ergosta-5,7,24(28)-trienol and ergosta-7,22-dienol) was observed, combined with a rise in cholesta-5,7,22-trienol from 1.30% in wild-type parasites to 79.8%, 85.8% and 85.9% in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, respectively. A small increase in cholesta-5,7,24-trienol was also noted, which was statistically significant ($P < 0.05$) in all three lines.

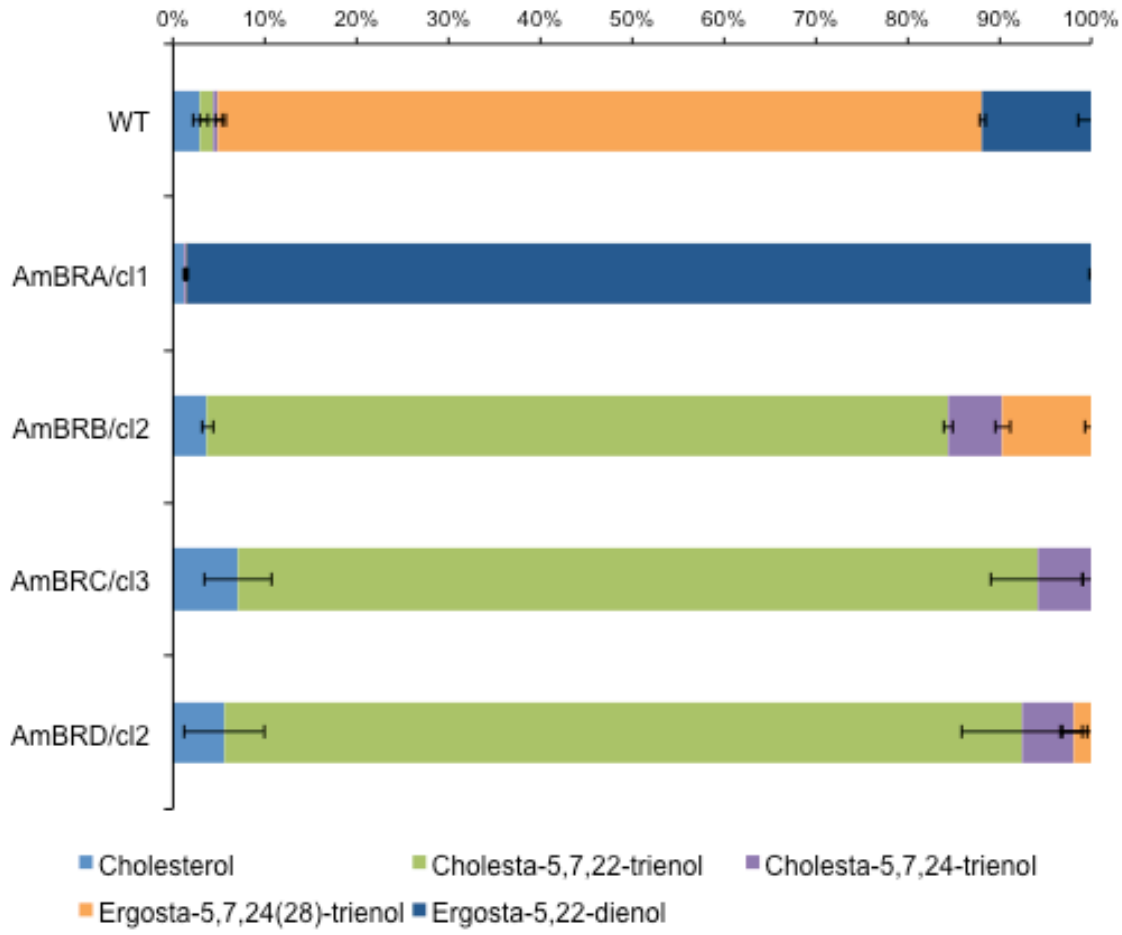
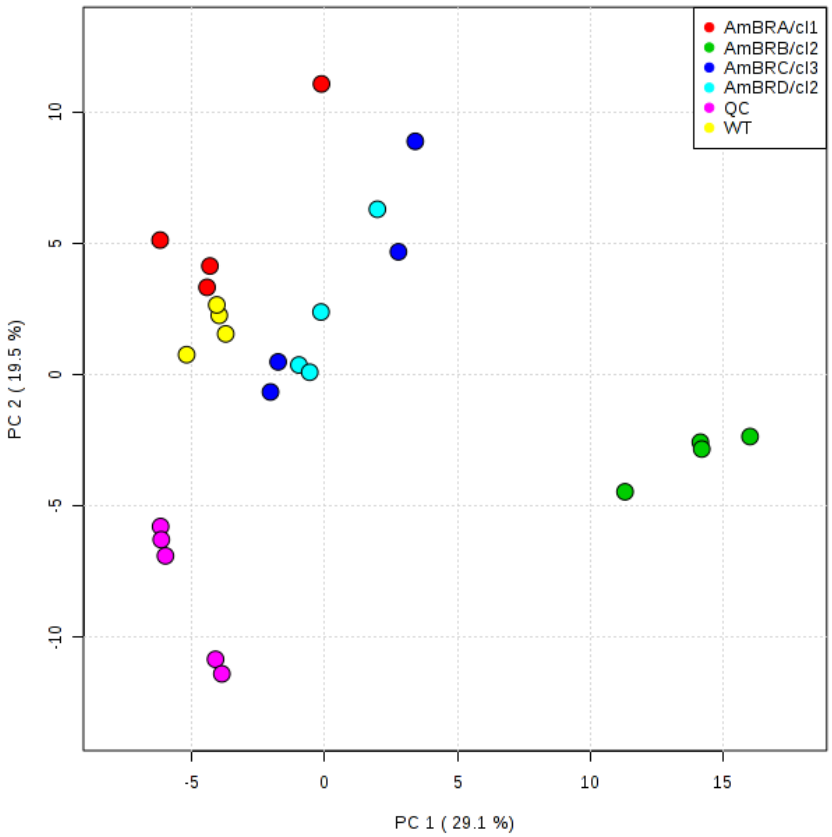


Figure 3-8: Sterol composition of wild-type and AmB-resistant *Leishmania* parasites. Sterol composition was determined by GC-MS. Stacked blocks represent the mean percentage of each sterol as a proportion of total sterol detected (based on GC peak area). Error bars represent the standard deviation across three biological replicates.

3.2.5 Untargeted metabolomics

While sterol changes in AmB resistance are well documented, I aimed to take a wider view of metabolism with regard to resistance-associated changes. Hence an untargeted LC-MS-based approach was used, with four biological replicates for each sample. Initial targeting detected a total of 1156 putative metabolites. The IDEOM file generated for all data is included as Supplementary file 3-2. Exclusion of unidentified metabolite peaks and peptides from analysis and filtering on the basis of retention time and manual inspection of peaks (see section 2.3.1.3) resulted in 233 metabolites remaining (Supplementary file 3-3 for the filtered list with statistical analysis). Principal components analysis (PCA) revealed a strong separation of AmBRB/cl2 samples, as well as technical replicates of a quality control derived from pooled metabolites across all samples (Figure 3-9A). However, limited resolution was observed amongst other resistant lines. By contrast, when pooled samples were excluded (Figure 3-9B), good resolution was achieved between all lines, particularly along principal component 2 (*y*-axis). Interestingly, principal component 1 (*x*-axis), which accounted for the most variation, separated AmBRB/cl2 samples from other lines. In order to detect significantly varying metabolite peaks I initially performed an analysis of variance (ANOVA) test, which revealed 159 metabolites with a multiple comparisons-corrected P-value of less than 0.05. On these, I applied Tukey's HSD post-hoc test to determine significant changes between wild-type and resistant lines. Of 142 metabolite peaks differing between wild type and resistant lines ($P < 0.05$), 64, 99, 32 and 86 were different from wild-type in AmBRA/cl1, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, respectively.

A



B

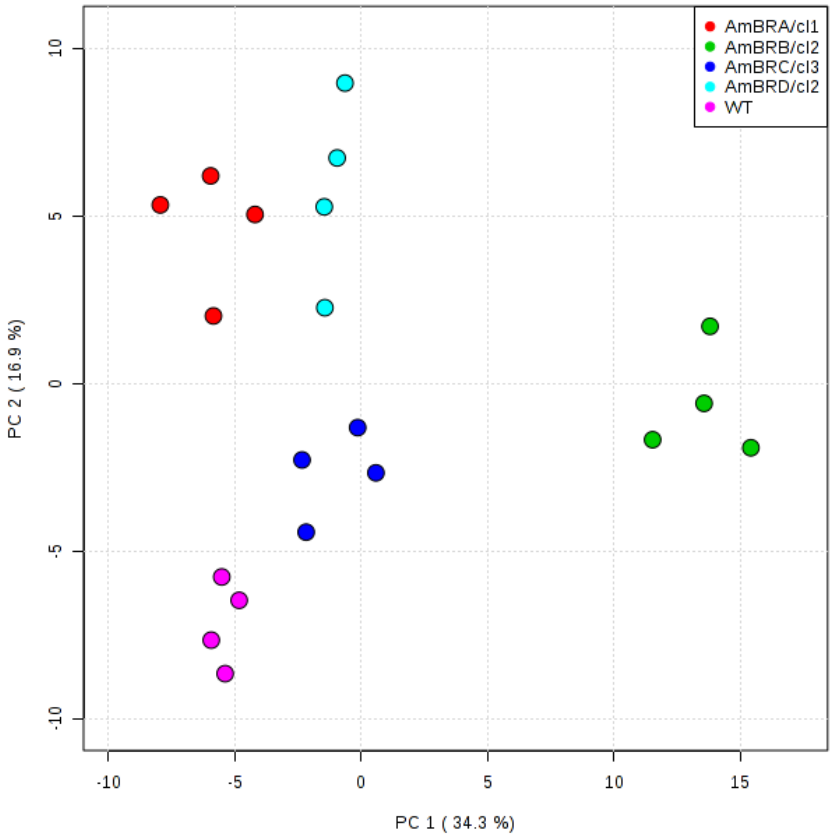


Figure 3-9: PCA plots of untargeted LC-MS metabolomics data. PCA plots were generated using Metaboanalyst 3.0 software (Xia *et al*. 2015). A) Pooled samples are included in analysis. B) Pooled samples are excluded from analysis. Note the colour scales are different in each case.

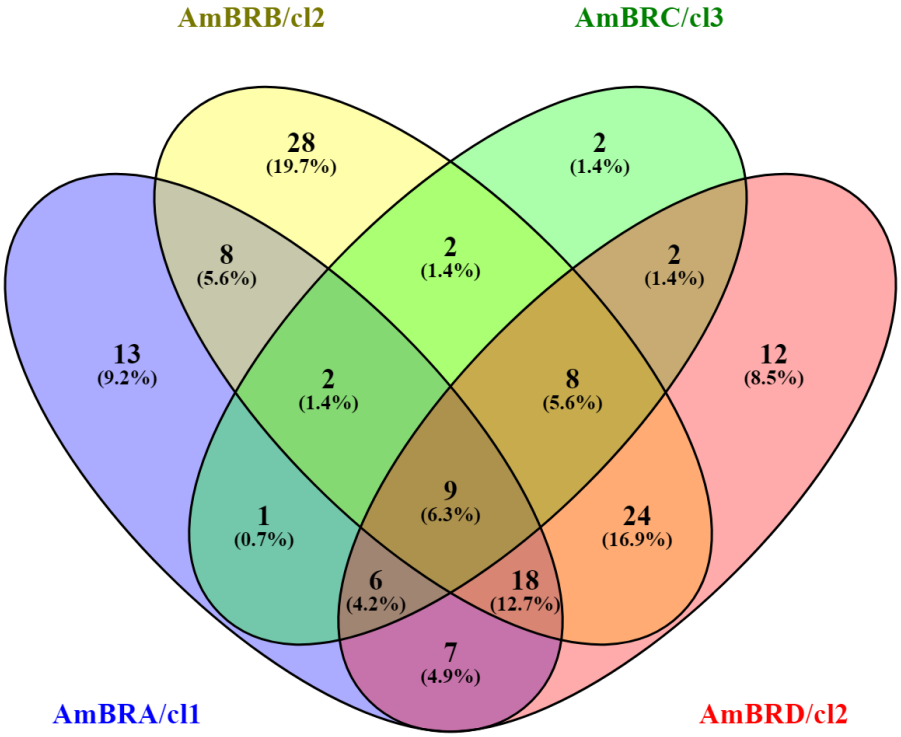


Figure 3-10: Overlap of significantly altered metabolites between resistant lines. Metabolites which were significantly altered between wild-type and individual AmB-resistant lines ($P < 0.05$, Tukey's HSD) were included in construction of a Venn diagram using online tool Venny v2.1 (Oliveros 2015).

Comparison of significantly altered metabolite peaks between resistant lines revealed only limited overlap. 55 metabolite peaks showed differential intensity from wild-type in only one resistant line, with 44 shared between two, 34 between three, and only nine across all four (Figure 3-10). This implies that metabolic changes are highly heterogeneous across individual resistant lines, thus complicating analysis. Of the nine metabolite peaks which were significantly altered across all four lines (Figure 3-11), eight of these showed declines in abundance, and only one, identified as succinate, showed moderate increases (1.3 to 1.6-fold). Of the other metabolites identified, three were associated with amino acid metabolism (O-phospho-L-serine, (R)-1-aminopropan-2-ol, and D-alanyl-D-alanine), and three ([FA (16:2)] N-hexadecyl-ethanolamine, [SP dimethyl,amino(18:0/2:0)] 2S-dimethylaminooctadecane-1,3R-diol and [FA methyl,hydroxy(5:0)] 3R-methyl-3,5-dihydroxy-pentanoic acid) were associated with lipid metabolism. Only succinate and N-acetyl-D-glucosamine showed positive identification via standards. Individual pathways are considered in greater detail below.

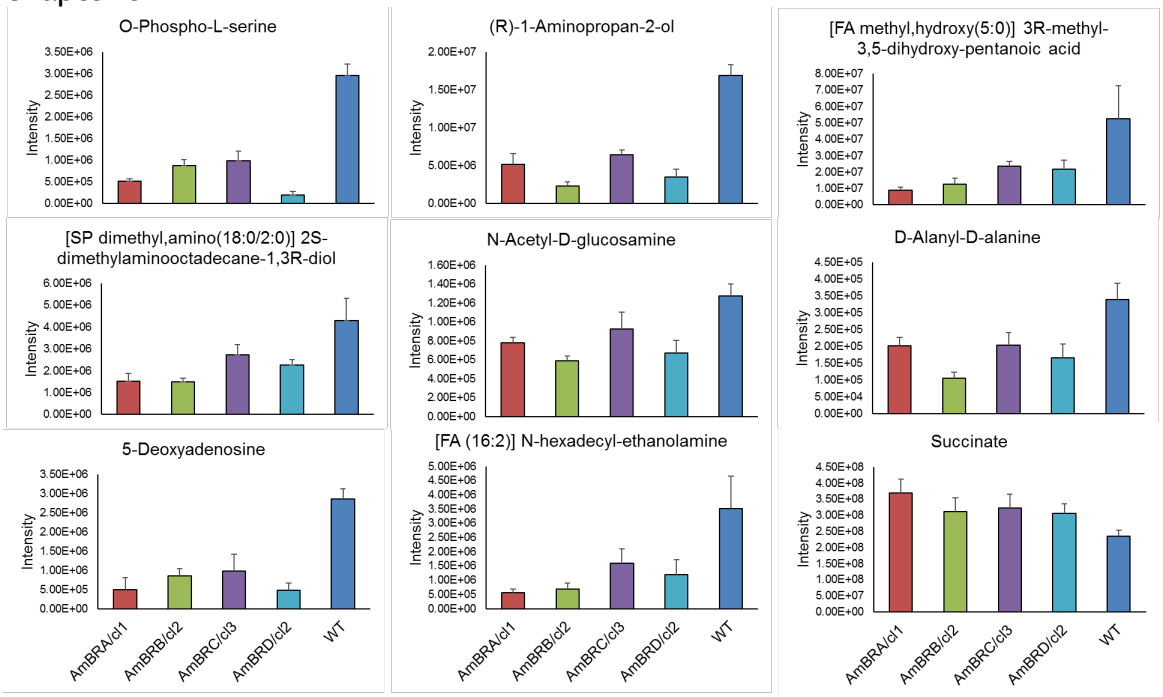


Figure 3-11: Metabolites significantly altered in all four resistant lines. Metabolites which revealed significant differences from wild-type ($P < 0.05$, Tukey's HSD) across all four AmB-resistant lines are shown. Mean values of four biological replicates are plotted, with error bars representing standard deviation. Intensity values are in arbitrary units, derived from liquid chromatograph peak area.

With respect to confidence of identification, it should be noted here that most metabolites were identified on the basis of mass alone, as matching to retention times through use of standards was not employed (either because a standard was not included, because the standard did not give an acceptable peak, or because the metabolite did not match the known standard). In this case, identifications should be regarded as putative annotations, equivalent to the metabolomics standards initiative level two (Salek *et al*. 2013). On the other hand, where peaks are stated to have matched to a standard, these are equivalent to MSI level one, and therefore greater confidence can be assumed regarding identification. For MSI level two identification, therefore, it is important to consider that isomers cannot be distinguished.

3.2.5.1 Lipid metabolism

Lipid metabolism includes fatty acids, saponifiable lipids (such as sphingolipids and glycerophospholipids) and non-saponifiable lipids (sterols). It should be noted that HILIC-based LC-MS is not the ideal tool, given the hydrophobicity of many of these compounds, which leads to many being eluted early, before the retention time cutoff of four minutes employed here. Furthermore, due to the

complexity of lipid molecular structures and the abundance of possible isomers, identification in the absence of standards should be treated with caution. For this reason, the sterol-focused GC-MS offers a more reliable picture of sterol changes (section 3.2.4.2). Given these caveats, however, 52 peaks were identified as related to lipid metabolism after filtering for peak quality and retention time, of which 36 showed significant alterations in the ANOVA and 33 showed significant differences between at least one resistant line and wild-type parasites. In AmBRA/cl1, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, 16, 25, four and 20 significant changes were observed respectively. Two peaks were identified as sterols after filtering (Figure 3-12), [ST (3:0)] cholest-5Z,7Z,24-trien-3beta-ol (mass 382.32) and [ST methoxy,dimethyl(3:0/3:0/3:0)] (5Z,7E)-(1S,3R,22R)-22-methoxy-26,27-dimethyl-24a-homo-9,10-seco-5,7,10(19)-cholestatrien-23-yne-1,3,25-triol (mass 484.35). With regards to the former, this was only detected in lines AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. It is worth noting that both cholestatrienols identified in GC-MS (Table 3-2) increase in these lines, particularly cholesta-5,7,22-trienol. Thus, this peak likely corresponds to one of the sterols accumulating in these lines. The other sterol identified shows a decline only in AmBRA/cl1 and AmBRB/cl2. This is identified as a secosteroid, but it is unlikely to be correctly identified, and does not correspond to any sterol previously found in *Leishmania*, having a far higher molecular weight than any found in the ergosterol biosynthesis pathway. Indeed as identification is based on mass alone, it is not known whether this is a sterol, or an adduct or fragment of another molecule.

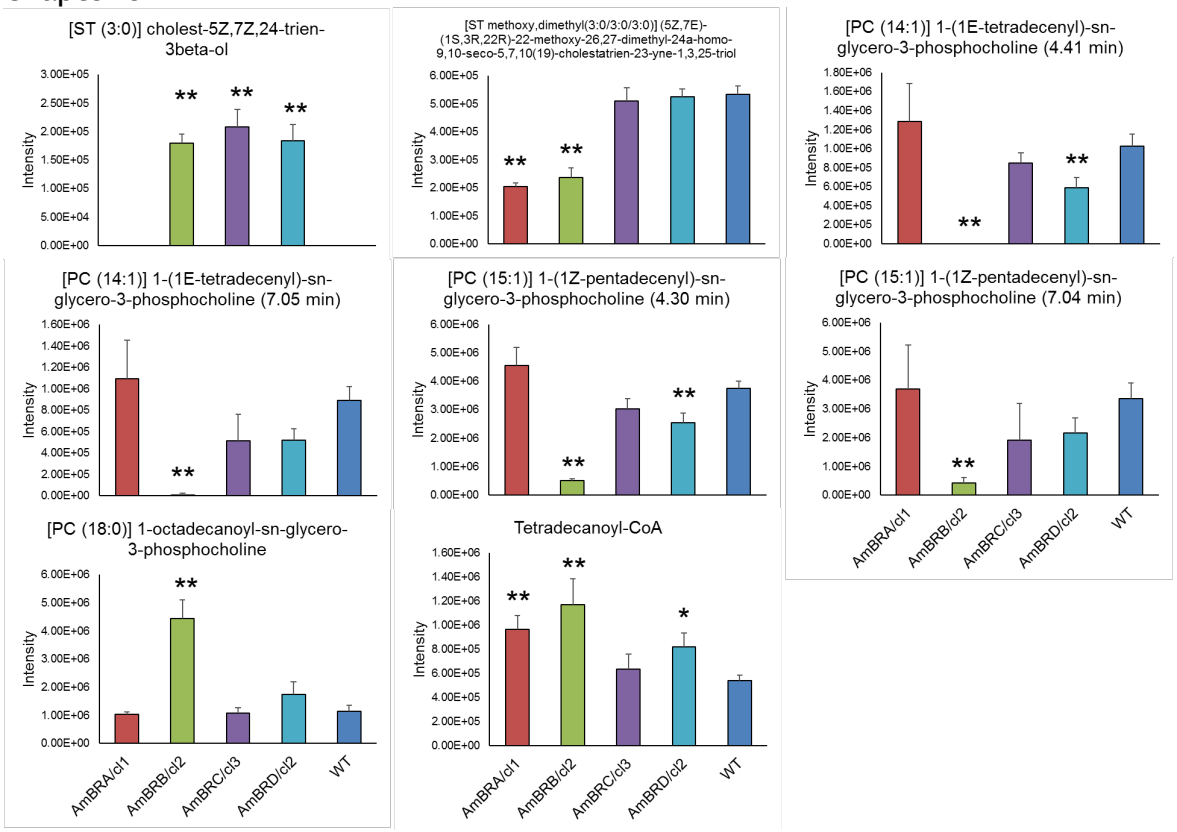


Figure 3-12: Selected changes in lipid metabolites. Changes in peak area for a number of lipid metabolites as discussed in the text. Mean values of four biological replicates are plotted, with error bars representing standard deviation. Asterisks represent significant differences from wild-type (*P < 0.05, **P < 0.005). Intensity values are in arbitrary units, derived from liquid chromatograph peak area.

With regard to other lipids, as noted the uncertainty of metabolite identification means that focus on specific metabolites is problematic. However, overall, certain patterns are evident from the data (Figure 3-13), particularly for AmBRB/cl2. Of 25 lipids with differential abundance in AmBRB/cl2, 22 showed decreased expression relative to wild-type parasites. These include particularly sphingolipids (eight peaks) and glycerophospholipids (nine peaks). Of the peaks showing the greatest decline, two are identified as [PC (14:1)] 1-(1E-tetradecenyl)-sn-glycero-3-phosphocholine (mass 451.31, retention times 4.41 min and 7.05 min, fold changes 0 and 0.008) and [PC (15:1)] 1-(1Z-pentadecenyl)-sn-glycero-3-phosphocholine (mass 465.32, retention times 4.30 min and 7.04 min, fold changes 0.13 and 0.14). Besides the abovementioned sterol changes, two significant increases are noted relative to wild type, one in [PC (18:0)] 1-octadecanoyl-sn-glycero-3-phosphocholine (mass 323.36) and the other in tetradecanoyl-CoA (mass 488.66) of 3.9-fold and 2.2-fold, respectively. Overall, these changes suggest extensive remodelling of glycerophospholipid and

sphingolipid metabolism particularly in AmBRB/cl2, with some decreases also present in other lines (Figure 3-13).

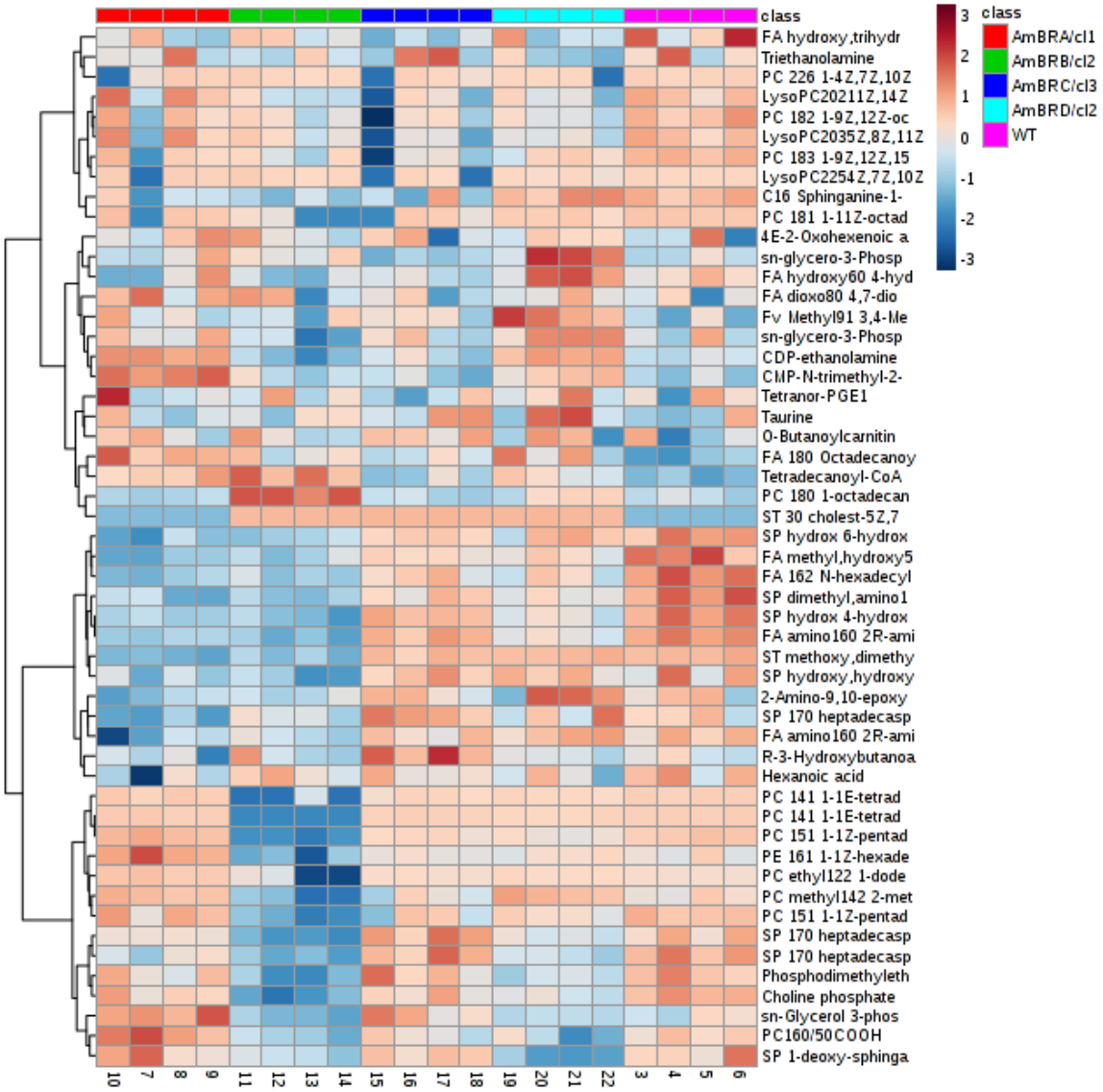


Figure 3-13: Heatmap of changes to lipid metabolism. A heatmap containing all 52 identified metabolites associated with lipid metabolism, as generated by Metaboanalyst v3.0 (Xia *et al*. 2015). Columns represent individual biological replicates, and colours represent log-transformed intensity normalised by row. Metabolite names are cut off and altered automatically by the analysis software; however, two-letter acronyms signify the following: FA – fatty acid; PC – phosphatidylcholine; SP – sphingolipid; ST: sterol.

3.2.5.2 Carbohydrate and energy metabolism

After filtering, a total of 53 peaks were identified as metabolites related to carbohydrate and energy metabolism. Of these, 31 differed significantly from wild-type parasites in at least one resistant line (nine in AmBRA/cl1, 21 in AmBRB/cl2, six in AmBRC/cl3 and 17 in AmBRD/cl2.

The PPP has been implicated in defence against oxidative stress caused by various antileishmanial agents (Ghosh *et al*. 2015). A number of PPP metabolites were detected here, namely 6-phospho-D-gluconate, D-ribose-5-phosphate and D-sedoheptulose-7-phosphate (Figure 3-14). D-ribose and D-glucono-1,5-lactone, whose phosphorylated forms are pathway members, and deoxyribose and 2-deoxy-D-ribose-5-phosphate, derived from D-ribose, were also detected (section 1.2.1). D-sedoheptulose-1,7-bisphosphate and D-ribose-1,5-bisphosphate, the bis-phosphorylated forms of PPP metabolites D-sedoheptulose-7-phosphate and D-ribose-5-phosphate, respectively, were also detected. None of these were validated by a standard. Seven of these showed significant alteration compared to wild-type parasites. AmBRA/cl1 demonstrated reduced abundance of 6-phospho-D-gluconate and D-ribose. AmBRB/cl2 showed decreases in D-ribose-1,5-bisphosphate and D-glucono-1,5-lactone, but a twofold increase in deoxyribose. AmBRC/cl3 showed a twofold rise in D-sedoheptulose-1,7-bisphosphate and AmBRD/cl2 showed a reduction in 6-phospho-D-gluconate, D-ribose-1,5-bisphosphate and D-sedoheptulose-1,7-bisphosphate. Overall, therefore, the changes appear to be heterogeneous, although broad PPP changes are not apparent in any individual line. Moreover, NADPH itself, a PPP product of major importance with regard to drug resistance, was unchanged in all lines.

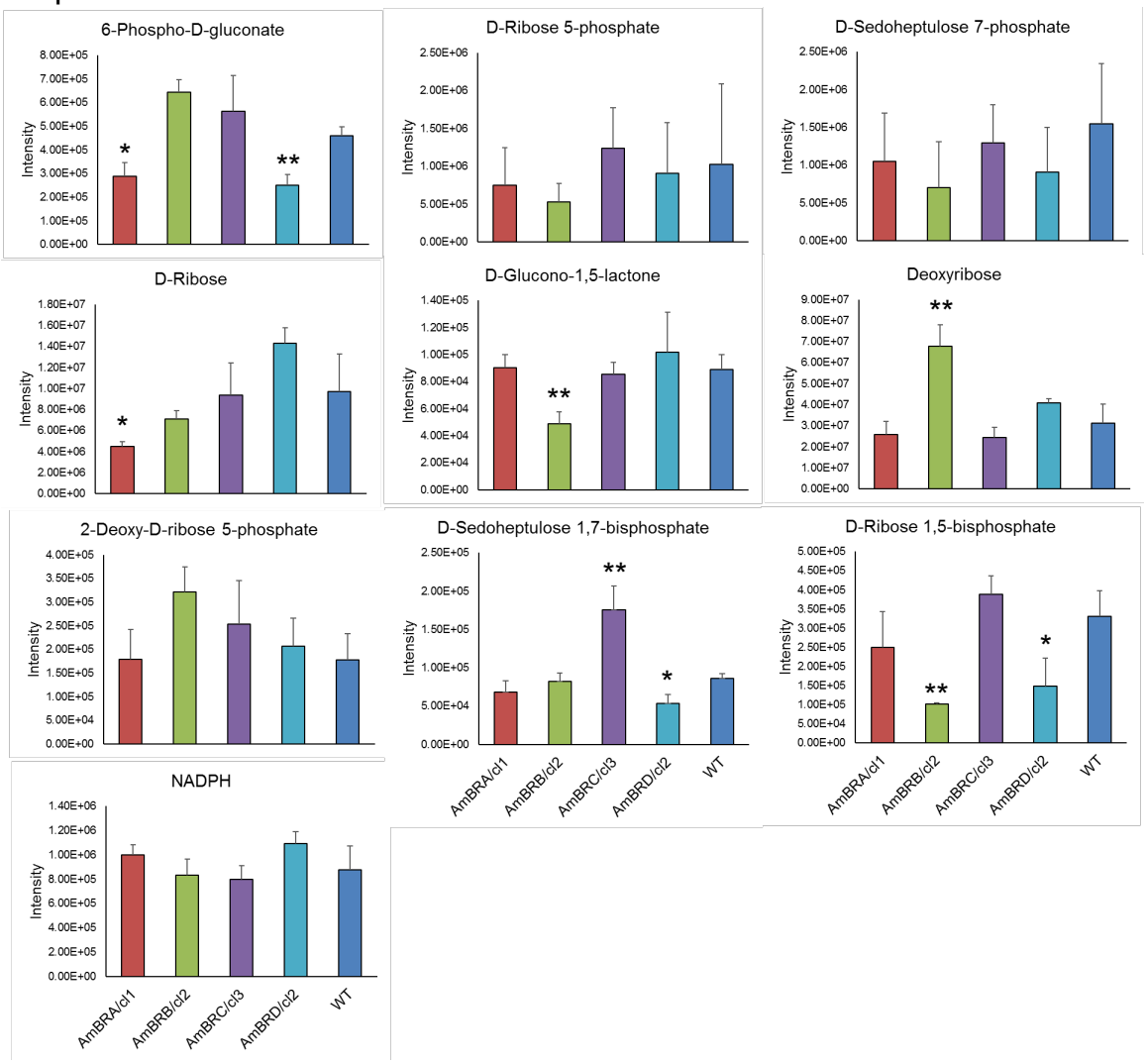


Figure 3-14: Pentose phosphate pathway metabolites. Changes for pentose phosphate pathway metabolites detected by LC-MS. Mean values of four biological replicates are plotted, with error bars representing standard deviation. Asterisks represent significant differences from wild-type (*P < 0.05, **P < 0.005). Intensity values are in arbitrary units, derived from liquid chromatograph peak area.

A number of metabolites in the glycolysis pathway were detected, namely D-glucose, D-fructose-1,6-bisphosphate, 3-phospho-D-glycerate, phosphoenolpyruvate and pyruvate (Figure 3-15). Two peaks associated with D-glucose were detected (mass 180.06, retention times 12.08 min and 13.52 min); these presumably refer to two sugar isomers, but as neither match to available standards, differentiation is not possible. A D-glucose standard was included in the run, but excluded due to poor peak quality; this had a retention time of 12.77 min, matching to neither of the above peaks. 3-phospho-D-glycerate, phosphoenolpyruvate and pyruvate were all validated by matching to standards. Of these metabolites, three showed significant differences between wild-type and resistant lines: D-glucose (retention time 13.52 min) showed an almost 50%

decrease in AmBRB/cl2, whereas 3-phospho-D-glycerate showed a 1.8-fold increase in the same line. D-fructose-1,6-bisphosphate showed significant decreases in AmBRA/cl1 and AmBRD/cl2, with a significant increase in AmBRC/cl3. Notably, AmBRB/cl2 also exhibited increases in the levels of phosphoenolpyruvate, pyruvate and the hexose with retention time 12.08 min, but these were not significant; nevertheless, it suggests the possibility of greater glycolytic flux in this line. D-lactate, a product of the anaerobic mode of this pathway, interestingly showed increases in all four lines, which were significant in AmBRA/cl1, AmBRB/cl2 and AmBRC/cl3 (note that this cannot be distinguished from its enantiomer, L-lactate, but as the product of glycolysis it is the more likely identification).

Tricarboxylic acid (TCA) cycle metabolites were detected, namely succinate, CoA (not a TCA metabolite but a cofactor involved), acetyl-CoA, citrate and cis-aconitate (Figure 3-15). Two peaks for CoA were detected (mass 255.71, retention times 4.68 min and 10.61 min), but neither was validated by a standard. Succinate, as described above, is increased in all four lines. Acetyl-CoA showed moderate decreases in all four lines, which were statistically significant in AmBRB/cl2 and AmBRD/cl2. No changes were observed in citrate or cis-aconitate, however ATP levels were significantly increased in AmBRA/cl1 and AmBRB/cl2. By contrast, AMP levels (validated by a standard) were unchanged. NADH, also validated by a standard, showed a significant increase in AmBRA/cl1.

Finally, changes to a number of polysaccharide type molecules were observed Figure 3-15). Precise identification of these from molecular mass alone is challenging, but these were identified in IDEOM as follows: maltose (mass 342.12), maltotriose (mass 504.17), glycogen (mass 666.22), cellopentaose (mass 828.28) and cellohexaose (mass 990.32). With the exception of maltose (which showed significant increases in AmBRA/cl1 and AmBRD/cl2, with no decreases), all showed decreases in AmBRB/cl2 and AmBRD/cl2, with a more moderate decrease in AmBRA/cl1 and no change in AmBRC/cl3. The molecular formulae of these, respectively, are $C_{12}H_{22}O_{11}$ (di-hexose), $C_{18}H_{32}O_{16}$ (tri-hexose), $C_{24}H_{42}O_{21}$ (tetra-hexose), $C_{30}H_{52}O_{26}$ (penta-hexose) and $C_{36}H_{62}O_{31}$ (hexa-hexose). Given that *Leishmania* normally stores carbohydrate as polysaccharides of D-mannose (an

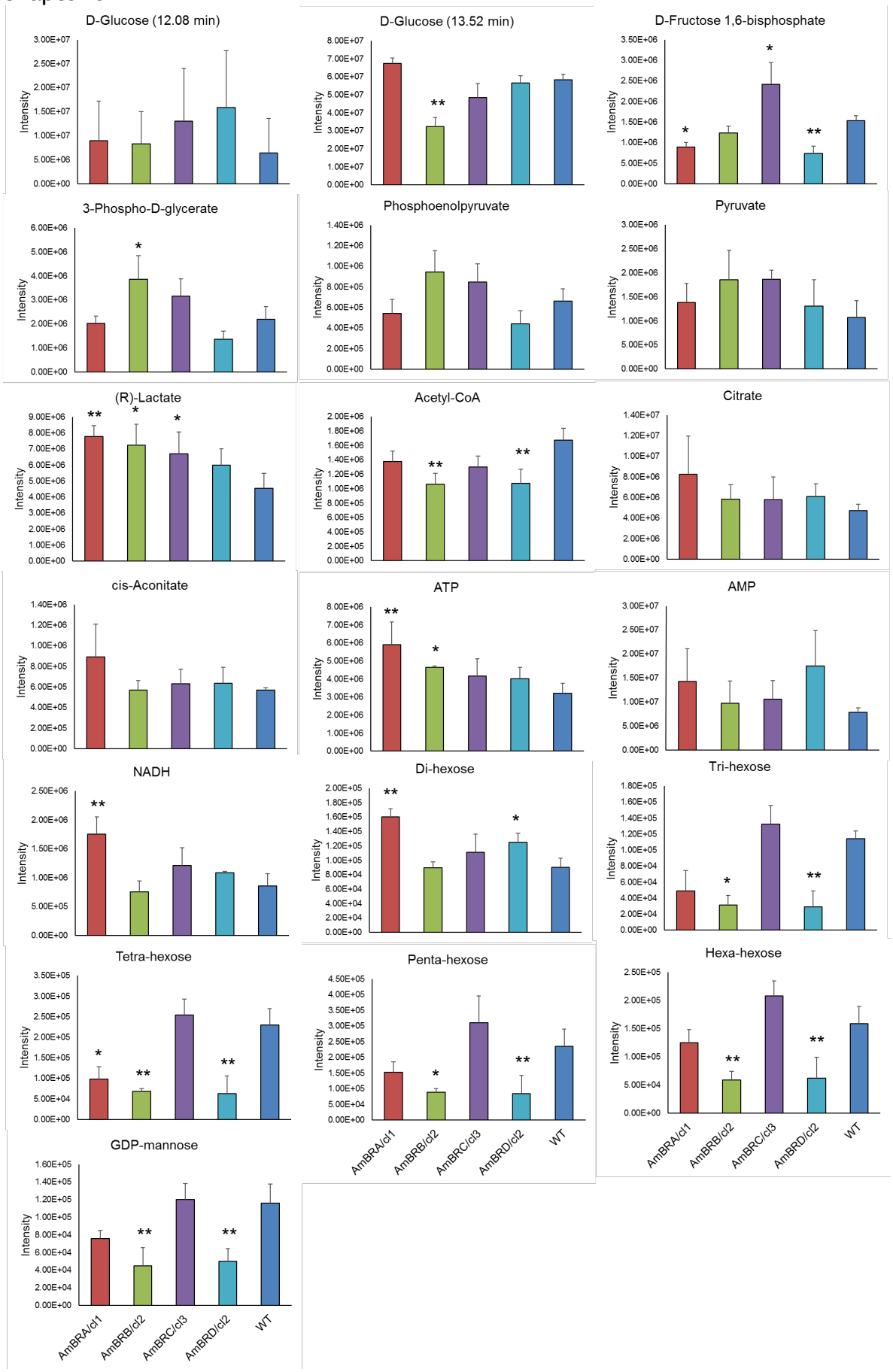

D-Glucose (12.08 min)
D-Glucose (13.52 min)
D-Fructose 1,6-bisphosphate
3-Phospho-D-glycerate
Phosphoenolpyruvate
Pyruvate
(R)-Lactate
Acetyl-CoA
Citrate
cis-Aconitate
ATP
AMP
NADH
Di-hexose
Tri-hexose
Tetra-hexose
Penta-hexose
Hexa-hexose
GDP-mannose
Intensity
AmBRA/cl1
AmBRB/cl2
AmBRC/cl3
AmBRD/cl2
WT

Figure 3-15: Metabolites in glycolysis, the TCA cycle and polysaccharide metabolism. Changes in carbohydrate and energy metabolism, as detected by LC-MS. Mean values of four biological replicates are plotted, with error bars representing standard deviation. Asterisks represent significant differences from wild-type (*P < 0.05, **P < 0.005). Intensity values are in arbitrary units, derived from liquid chromatograph peak area.

isomer of glucose and fructose) known as mannogen, it is possible that these represent two, three, four, five and six-sugar units of this polysaccharide, and indeed the precursor to mannogen biosynthesis, GDP-mannose, shows a very similar pattern of changes to the three to six-hexose units.

3.2.5.3 Amino acid and polyamine metabolism

A total of 91 peaks were detected after filtering that were associated with amino acid metabolism, 56 of which demonstrated significant differences in abundance between wild-type parasites and at least one resistant line. The number of changes per line were 25 for AmBRA/cl1, 39 for AmBRB/cl2, 18 for AmBRC/cl3 and 38 for AmBRD/cl2. Of the amino acids found in proteins, 11 were differentially abundant in comparison to wild-type in at least one line: arginine, glutamate, glutamine, glycine, histidine, lysine, serine, phenylalanine, proline, threonine and tryptophan.

Of particular interest to drug resistance are polyamine and trypanothione metabolism (Figure 3-16). While polyamines themselves are poorly detected by the LC-MS system used here, other metabolites may be more easily resolved and quantified. In particular, the two stages before putrescine in its biosynthesis, L-arginine and L-ornithine (both validated by standards), are both significantly decreased in AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2, with a slight but non-significant increase in AmBRC/cl3. S-adenosyl-L-methionine, which is converted into S-adenosyl-5'-(3-methylthiopropylamine), the aminopropyl donor used to create spermidine from putrescine, is significantly decreased in AmBRB/cl2 and AmBRD/cl2. N-acetylputrescine, the acetylated version of putrescine, was detected and may indicate changes to putrescine abundance; this exhibited significant increases in both AmBRB/cl2 and AmBRC/cl3. The redox-active thiol and precursor of trypanothione biosynthesis, glutathione (also validated by a standard), shows significant changes in all but AmBRC/cl3; however, while AmBRB/cl2 decreases, substantial increases are observed for both AmBRA/cl1 and AmBRD/cl2. Trypanothione was also observed to greatly increase,

particularly in AmBRA/cl1 but also significantly in AmBRC/cl3 and AmBRD/cl2. Note that two peaks were initially identified as trypanothione, one with a mass of 361.65 and retention time of 12.41 min, the other with a mass of 723.30 and a retention time of 19.34 min. However, the 19.34 min peak was excluded because it was of much poorer quality, and the 723.30 mass could also be found at the earlier retention time. The 361.65 *m/z* is predicted to arise due to ionisation by double protonation of trypanothione (mass 723.86). Trypanothione disulfide was also detected (mass 360.64, retention time 14.53 min), but no change in abundance is evident. Although both trypanothione and trypanothione disulfide were included with the standards, trypanothione was not detected and trypanothione disulfide was initially excluded due to poor peak quality. However, the retention time of the trypanothione disulfide standard, 14.46, is very similar to that of the peak identified as trypanothione disulfide in the samples, helping to validate further its identity.

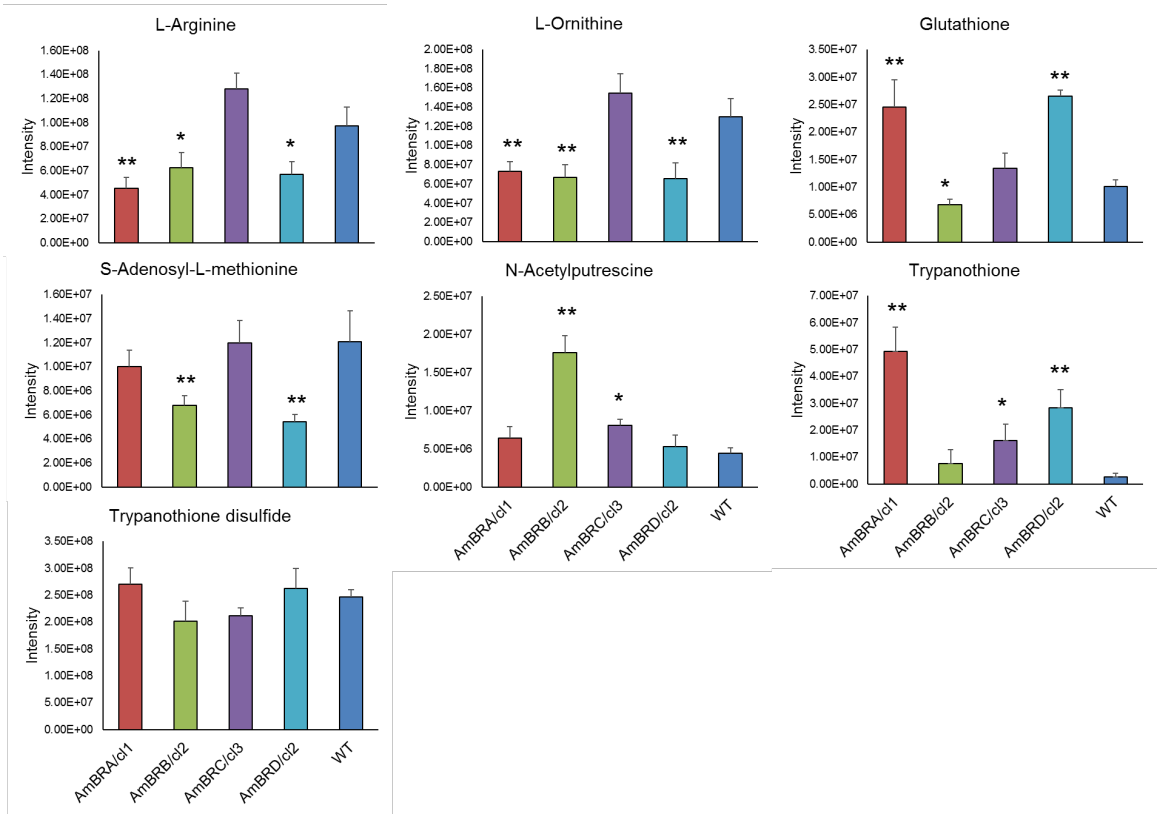


Figure 3-16: Trypanothione metabolism. Changes in trypanothione metabolism, as detected by LC-MS. Mean values of four biological replicates are plotted, with error bars representing standard deviation. Asterisks represent significant differences from wild-type (*P < 0.05, **P < 0.005). Intensity values are in arbitrary units, derived from liquid chromatograph peak area.

In other areas of amino acid metabolism, observed numerous changes to synthesis and degradation pathways were observed (Figure 3-17). With regard to arginine metabolism, a similar pattern of downregulation in AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2 to L-arginine itself was noted for its degradation intermediates, 4-guanidinobutanal and 4-guanidinobutanamide (Figure 3-18). Alterations to proline metabolism were seen. L-proline itself decreased in the same three lines (though only significant in AmBRB/cl2 and AmBRD/cl2). Two peaks were identified as its degradation product, L-1-pyrroline-3-hydroxy-5-carboxylate (mass 129.04). One peak, with a retention time of 11.96 min, decreased in all AmB-resistant lines, significantly in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2; no change was observed in the other, with a retention time of 9.15 min. As no standards are available, distinction between the two is not possible but possible isomers include 1-pyrroline-4-hydroxy-2-carboxylate, another degradation product of L-proline. Two peaks were observed for the intermediate in the conversion of L-arginine to L-proline, L-glutamate-5-semialdehyde (mass 131.06, retention times 8.27 min and 11.97 min). Again, neither was validated by a standard. The 8.27 min peak was unchanged in AmBRB/cl2 and AmBRD/cl2, but showed significant increases in AmBRA/cl1 and AmBRC/cl3. Conversely, the 11.97 min peak showed significant decreases in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. Interpretation without a standard is challenging, since isomers include hydroxyprolines, which are themselves degradation products of L-proline. However, another proline degradation product, L-4-hydroxyglutamate semialdehyde, significantly increased in AmBRC/cl3, but substantially albeit non-significantly decreased in AmBRA/cl1.

Other amino acids which showed abundance of multiple synthesis and degradation intermediates include L-lysine, L-histidine, L-serine and L-glycine. While it is not possible to comment on all of these changes, it should be noted that two metabolites are annotated as involved in glycine and serine metabolism were universally decreased, as described above: O-phospho-L-serine, the phosphorylated form of serine, and (R)-1-aminopropan-2-ol, a compound derived from glycine. L-serine itself showed a generalised decrease, which was significant in all lines except AmBRC/cl3, whereas glycine only showed a significant decline in AmBRD/cl2. The source of changes to O-phospho-L-serine is unclear, since its immediate precursor, 3-phospho-hydroxypyruvate, was not

detected and similar changes are not detected in the metabolite preceding this, 3-phospho-D-glycerate (section 3.2.5.2). As phosphorylation of serine residues is a frequently observed post-translational modification, O-phospho-L-serine can arise from phosphoprotein degradation. In this context, reduced levels of this metabolite could result from a reduced rate of this degradation.

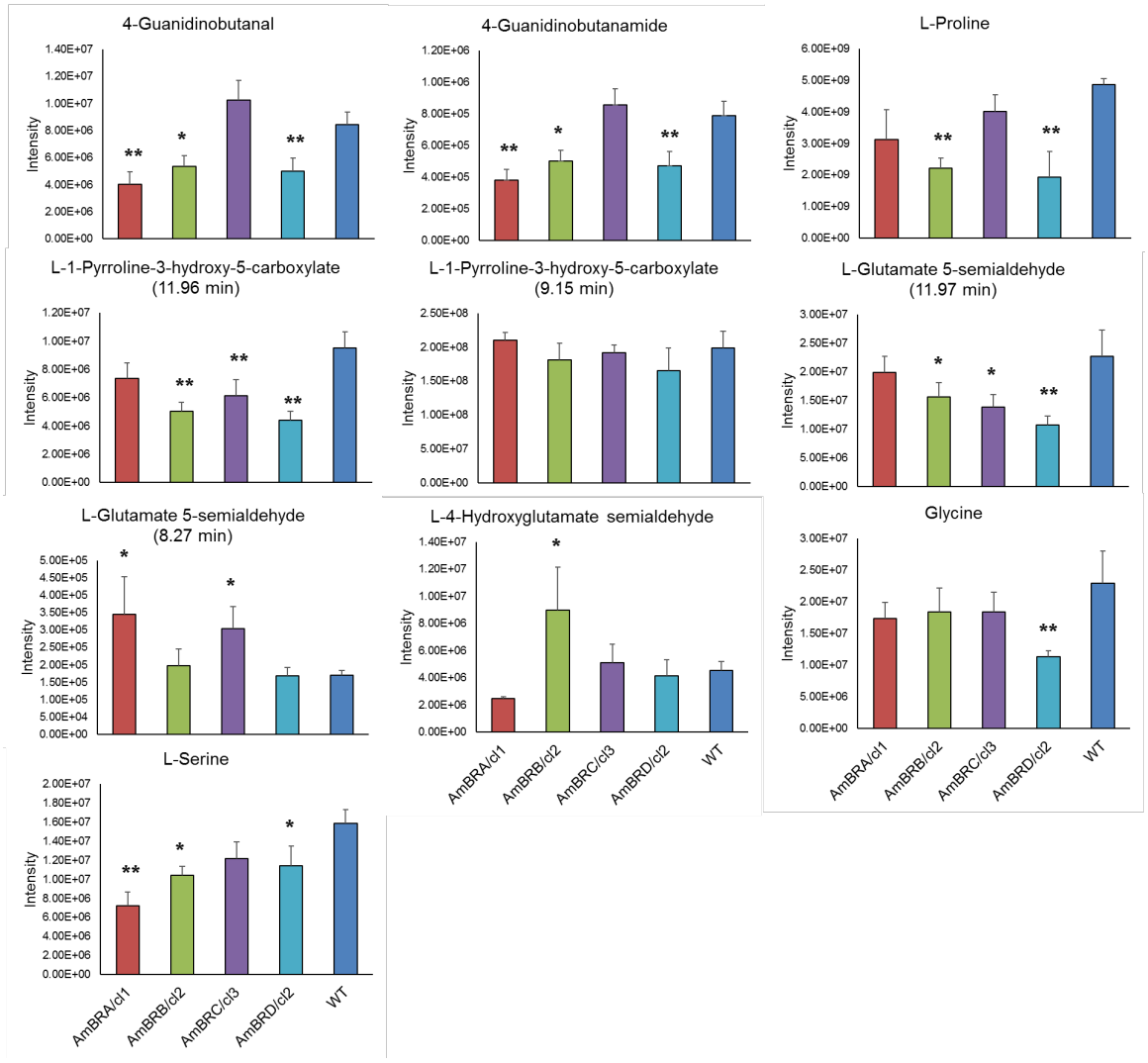


Figure 3-17: Arginine and proline metabolism. Changes in arginine and proline metabolism, as detected by LC-MS. Mean values of four biological replicates are plotted, with error bars representing standard deviation. Asterisks represent significant differences from wild-type (*P < 0.05, **P < 0.005). Intensity values are in arbitrary units, derived from liquid chromatograph peak area.

Chapter 3

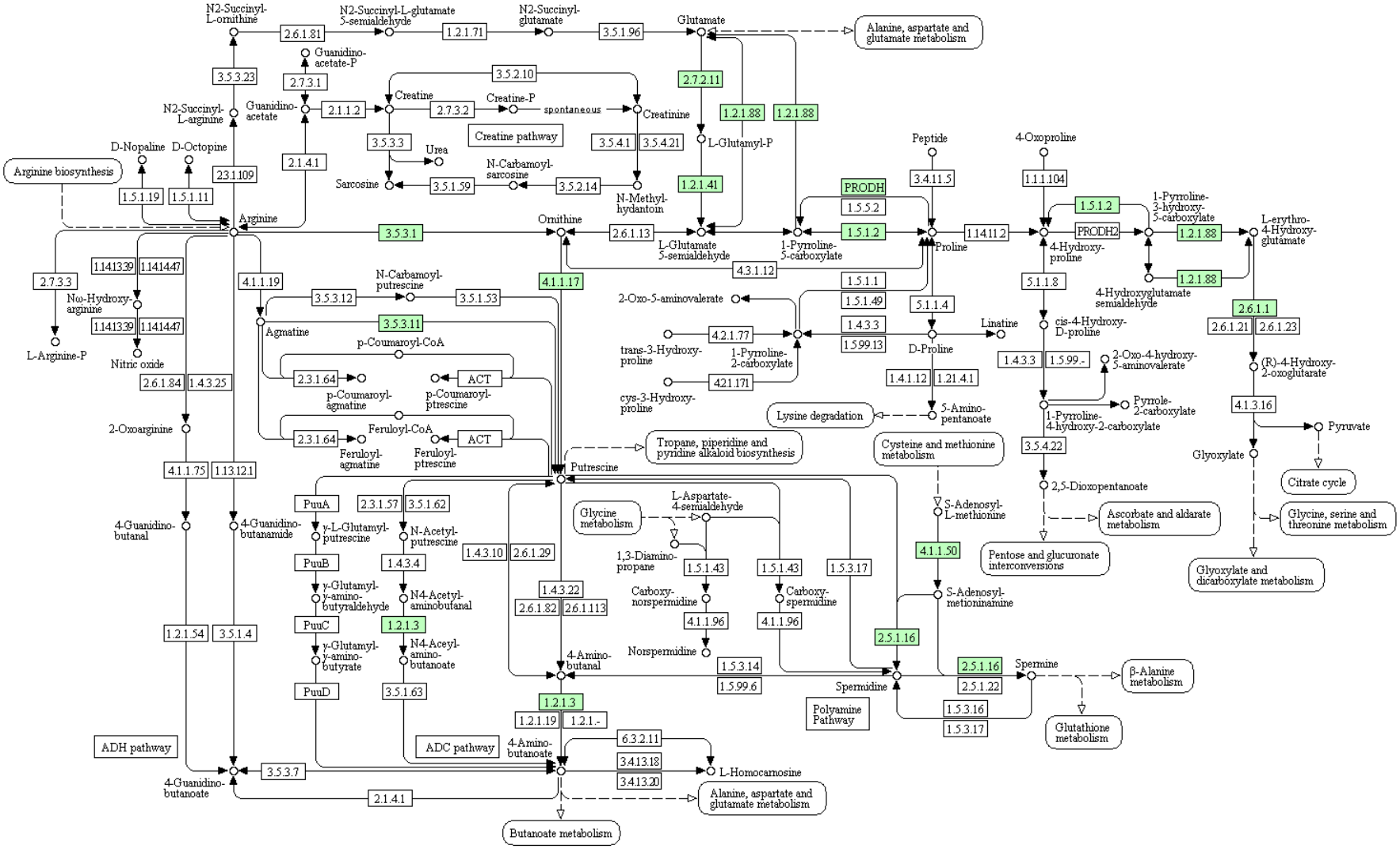

N2-Succinyl-L-ornithine
N2-Succinyl-L-glutamate 5-semialdehyde
N2-Succinyl-glutamate
Glutamate
Alanine, aspartate and glutamate metabolism
Guanidino-acetate-P
Creatine
Creatine-P
spontaneous
Creatinine
Creatine pathway
N2-Succinyl-L-arginine
Guanidino-acetate
D-Nopaline
D-Octopine
Arginine biosynthesis
Urea
Sarcosine
N-Carbamoyl-sarcosine
N-Methyl-hydantoin
L-Glutamyl-P
Peptide
4-Oxoproline
Arginine
Ornithine
L-Glutamate 5-semialdehyde
1-Pyrroline-5-carboxylate
Proline
4-Hydroxy-proline
1-Pyrroline-3-hydroxy-5-carboxylate
L-erythro-4-Hydroxy-glutamate
4-Hydroxyglutamate semialdehyde
Nω-Hydroxy-arginine
L-Arginine-P
Nitric oxide
N-Carbamoyl-putrescine
Agmatine
p-Coumaroyl-CoA
p-Coumaroyl-agmatine
p-Coumaroyl-ptrescine
ACT
2-Oxo-5-aminovalerate
trans-3-Hydroxy-proline
cys-3-Hydroxy-proline
1-Pyrroline-2-carboxylate
D-Proline
Linatine
cis-4-Hydroxy-D-proline
2-Oxo-4-hydroxy-5-aminovalerate
(R)-4-Hydroxy-2-oxoglutarate
Pyrrole-2-carboxylate
1-Pyrroline-4-hydroxy-2-carboxylate
2,5-Dioxopentanoate
Pyruvate
Glyoxylate
Citrate cycle
Glycine, serine and threonine metabolism
Ascorbate and aldarate metabolism
Pentose and glucuronate interconversions
Glyoxylate and dicarboxylate metabolism
Lysine degradation
5-Amino-pentanoate
2-Oxoarginine
Feruloyl-CoA
Feruloyl-agmatine
Feruloyl-ptrescine
Tropane, piperidine and pyridine alkaloid biosynthesis
Cysteine and methionine metabolism
Putrescine
S-Adenosyl-L-methionine
4-Guanidino-butanal
4-Guanidino-butanamide
PuuA
γ-L-Glutamyl-putrescine
N-Acetyl-putrescine
PuuB
γ-Glutamyl-γ-amino-butyraldehyde
N4-Acetyl-aminobutanal
PuuC
γ-Glutamyl-γ-amino-butyrate
N4-Aceyl-amino-butanoate
PuuD
Glycine metabolism
L-Aspartate-4-semialdehyde
1,3-Diamino-propane
Carboxy-norspermidine
Carboxy-spermidine
S-Adenosyl-metioninamine
Norspermidine
4-Amino-butanal
Spermidine
Spermine
β-Alanine metabolism
Polyamine Pathway
Glutathione metabolism
ADH pathway
ADC pathway
4-Amino-butanoate
4-Guanidino-butanoate
L-Homocarnosine
Alanine, aspartate and glutamate metabolism
Butanoate metabolism

Figure 3-18: Map of arginine and proline metabolism. Numbers represent enzymes as per their enzyme commission (EC) number, with green boxes representing enzymes encoded by genes known to be present in the *L. mexicana* genome. Small circles represent positions of individual metabolites (labelled) and arrows between these show the direction of metabolic conversion between metabolites. Dotted arrows represent metabolic links to and from other areas of metabolism not depicted here. Source: KEGG map for *L. mexicana* arginine and proline metabolism (Kanehisa & Goto 2000).

3.3 Discussion

3.3.1 Summary of changes in individual AmB-resistant lines

In order to distinguish features that arise during selection for AmB resistance in a deterministic fashion, in comparison to those that only sometimes arise, I selected four independently derived AmB-resistant lines from the same *L. mexicana* parent. However, because extensive variation was observed between resistant lines, it is important before further discussion to briefly summarise changes in each line in turn, and the changes that have been detected.

3.3.1.1 Changes in AmBRA

This line exhibited around 7.3-fold resistance to AmB relative to wild-type parasites. By contrast, it was hypersensitive to MILT, PMM and PENT, but without altered sensitivity to ergosterol biosynthesis inhibitors. It also showed mild hypersensitivity to GOX, but the greatest hypersensitivity to methylene blue of any line. An increased sensitivity to oxidative stress therefore appears to accompany AmB resistance in this line. A mild growth defect was observed in the promastigote stage, with poor infectivity in macrophages and no evidence of conservation of AmB resistance as intracellular amastigotes. Total loss of ergosterol was accompanied by accumulation of a sterol lacking double bond conjugation, identified as ergosta-7,22-dienol. In terms of broader metabolism, 61 significant changes were observed including moderate changes in lipid metabolism (Figure 3-13). The PPP was unaltered, however other changes in carbohydrate metabolism were seen, notably increases in organic acids succinate and lactate. While a disaccharide identified as maltose (but quite possibly di-mannose) was substantially increased, polysaccharides as a whole were moderately, although not always significantly, decreased. Finally, multiple alterations were noted in amino acid metabolism, particularly decreases in arginine metabolism and large increases in the thiols trypanothione and glutathione.

3.3.1.2 Changes in AmBRB

Resistance in this line was the most pronounced, 8.4-fold increased over wild-type parasites. Hypersensitivity to PAT and PENT was observed, but by contrast, this line demonstrated cross-resistance to MILT, as well as the ergosterol biosynthesis inhibitor spiroxamine. GOX sensitivity was unchanged from wild-type parasites, but mild hypersensitivity to menadione and more pronounced methylene blue hypersensitivity were exhibited. This indicates that there is some change in susceptibility to oxidative stress, but that this differs from the type seen in AmBRA/cl1. AmBRB/cl2 exhibited normal growth in the promastigote stage, but was poorly infective to macrophages and did not retain the resistance phenotype as intracellular amastigotes. A large decrease in wild-type sterol was associated with accumulation of cholestatrienols, particularly cholesta-5,7,22-trienol. Extensive changes to phospholipid and sphingolipid metabolism were also observed, with particularly strong increases in several phosphatidylcholine species. PPP members themselves were unchanged, although some species related to PPP metabolites displayed differential abundance from wild-type. A trend towards increases in late glycolytic intermediates, as well as lactate and succinate, was noted but polysaccharide metabolism appeared broadly to decrease. Multiple decreases in arginine and proline metabolism are observed, but there was little change in trypanothione and even a slight decrease in glutathione.

3.3.1.3 Changes in AmBRC

Moderate AmB resistance was obtained, 3.9-fold over wild-type parasites. AmBRC/cl3 was cross-resistant to PMM, but hypersensitive to MILT, PENT and methylene blue. This line exhibited similar growth to parental wild-type cells both as promastigotes and intracellular amastigotes, and retained the resistance phenotype as intracellular parasites. Sterol changes were similar to AmBRB/cl2, although loss of wild-type sterol was even more pronounced. This line exhibited the fewest metabolite changes relative to wild-type parasites. Some changes to carbohydrate metabolism were observed, most notably increases in fructose-1,6-bisphosphate, lactate and succinate. Arginine metabolism was unaffected, although increases in trypanothione were observed as well as altered proline metabolism.

3.3.1.4 Changes in AmBRD

This was the least resistant line, with 2.6-fold resistance over wild-type parasites. No cross-resistance was observed, but hypersensitivity to PENT, spiroxamine and methylene blue. AmBRD/cl2 exhibited no growth defect in the promastigote stage, with this line possibly reaching stationary phase at a higher density than wild-type parasites; however, loss of infectivity and of the AmB resistance phenotype in intracellular amastigotes were exhibited. Sterol changes bore strong similarities to AmBRB/cl2 and AmBRC/cl3. Changes to lipid metabolism were less extensive than in AmBRA/cl1 and AmBRB/cl2, but some changes to carbohydrate metabolism were observed, including declines in 6-phospho-D-gluconate, fructose-1,6-bisphosphate, acetyl-CoA and numerous polysaccharides. Multiple metabolites in arginine and proline metabolism were decreased, as was glycine, but increases in both glutathione and trypanothione were observed.

3.3.2 Selection of AmB resistance results in variable sensitivity to a range of agents

Four AmB-resistant lines were independently selected. Whilst cells were passaged continuously in the presence of drug for around 180 days, it should be noted that the degree of resistance remained moderate, particularly for AmBRC and AmBRD. This is less than in previous attempts to select AmB-resistance in *L. mexicana*, which yielded 16-fold (Al-Mohammed *et al*. 2005) and 23-fold (Mwenechanya *et al*. 2017) resistance over wild-type parasites. In the latter case, selection was actually over a shorter period of time (210 days); therefore, it is possible that these differences in magnitude reflect mainly stochastic occurrence of resistance mutations, some of which lead to stronger resistance phenotypes than others.

Cross-sensitivity to other antileishmanial drugs was highly heterogeneous (Table 3-1), particularly for PAT, MILT and PMM, with both cross-resistance and hypersensitivity noted for the latter two, and moderate hypersensitivity to PAT observed in only one line (AmBRB/cl2), with no change in the others. Previous studies have suggested that selection for AmB resistance has variable effects on sensitivity to other compounds, yet it has not been possible to separate effects

of the *Leishmania* species used and the method of selection from these observations. For example, PAT hypersensitivity has been observed in AmB-resistant *L. donovani* (García-Hernández *et al*. 2012), whereas AmB-resistant *L. infantum* and *L. mexicana* promastigotes appear to be PAT cross-resistant (Moreira *et al*. 2011, Mwenechanya *et al*. 2017). This study clearly shows that parasites of the same species, selected by the same method for AmB resistance, can vary in their sensitivity profile to other antileishmanial drugs. This is important in demonstrating that none of these features is an essential accompanying phenotype of resistance, and therefore they may or may not be expected to occur should resistance arise in the field. This is in contrast to the arguments made by Moreira and colleagues that defence against oxidative stress is a universal feature of resistance to AmB, MILT and PAT, with the implication that resistance to any one of these compounds should be accompanied by resistance to the other two (Moreira *et al*. 2011).

By contrast, this study, in conjunction with others (Mbongo *et al*. 1998, Mwenechanya *et al*. 2017), portrays PENT hypersensitivity as a consistent feature of AmB resistance. Whilst PENT hypersensitivity in an AmB-resistant *L. donovani* line was relatively moderate (approximately twofold), a previously selected *L. mexicana* line was 13-fold hypersensitive (Mwenechanya *et al*. 2017). In this study, PENT hypersensitivity varies from 4.1-fold to 9.2-fold, with the most AmB-resistant line (AmBRB/cl2) also being the most PENT-hypersensitive. The mechanism of action of PENT in kinetoplastids is not fully understood but appears to involve accumulation within the mitochondrion and binding to kinetoplast DNA (Yang *et al*. 2016). Intriguingly, however, this is accompanied by effects on membrane lipids and reduced intracellular polyamines (Basselin & Robert-Gero 1997, Basselin *et al*. 2000). PENT-resistant *Leishmania* exhibits altered phospholipid content, along with a reduction in membrane phospholipid content and increased membrane fluidity. Reduced mitochondrial membrane potential has also been noted (Basselin *et al*. 2002). Metabolic changes in our AmB-resistant lines are discussed in greater detail below. However, for the time being it is important to note that altered phospholipid content has been previously observed in AmB resistance (Mbongo *et al*. 1998, Fernandez-Prada *et al*. 2016). Strangely, AmB-mediated disruption of membranes is known to decrease with increasing membrane fluidity (HsuChen & Fiengold 1973) and

increased membrane fluidity is associated with AmB resistance in *Leishmania* (Mbongo *et al*. 1998). Hence, there is a contradiction that increased membrane fluidity is observed in both PENT resistance and AmB resistance (which correlates with PENT hypersensitivity). It is likely, therefore, that sensitivity to these drugs is influenced by not only bulk changes to membrane properties, but also other factors such as the specific lipid changes involved in each case. Nevertheless, since PENT is a clinically available antileishmanial drug, albeit used in only limited circumstances at present, the reproducibility of PENT hypersensitivity as an associated phenotype suggests that it may be a useful tool in case of emergent AmB resistance.

Another area of particular interest to drug resistance is sensitivity to oxidative stress. One complication in this field is the fact that different agents can contribute to different types of oxidative stress. GOX, which when added to the medium produces H_2O_2, therefore exposes cells to this form of ROS from the extracellular environment. Little change was observed in sensitivity to this enzyme, although AmBRA/cl1 was mildly but significantly hypersensitive. Neither were there pronounced changes in sensitivity to menadione, although AmBRB/cl2 exhibited mild but significant hypersensitivity. By contrast, methylene blue, which acts by depleting cellular NADPH (which is of central importance in defence against oxidative stress), revealed universal hypersensitivity across all AmB-resistant lines, up to 30-fold in AmBRA/cl1. With respect to the literature, some uncertainty exists as to the relationship between sensitivity to AmB and oxidative stress. In AmB-resistant fungi, both resistance (Sokol-Anderson *et al*. 1988) and hypersensitivity (Vincent *et al*. 2013) to ROS have been observed. In *Leishmania*, hypersensitivity to both H_2O_2 and methylene blue have been noted in one line (Mwenechanya *et al*. 2017). However, others, whilst not directly testing sensitivity to inducers of oxidative stress, have noted increased expression of factors associated with oxidative stress resistance (Purkait *et al*. 2012, Brotherton *et al*. 2014). The results of this study suggest that hypersensitivity to oxidative stress-inducing agents may depend on the specific source of ROS used, but certainly the results using GOX and menadione do not support the hypothesis that severe hypersensitivity to oxidative stress is an essential accompanying feature of AmB-resistance. While hypersensitivity is much more pronounced to methylene blue, the exact mode of action of this

compound is relatively unclear, and the question of whether other factors are influential, such as uptake of the compound, or subcellular compartment-specific induction of stress, has not been addressed.

However, whilst hypersensitivity to oxidative stress does not appear to be a necessary feature of AmB resistance, neither does the converse appear to be true, that AmB-resistant parasites are more resistant to ROS. It is, however, possible to resolve observations from the literature of oxidative stress hypersensitivity (or in this case, simple lack of resistance) with evidence of increased expression of oxidative stress-protective factors (including increased abundance of trypanothione shown here). It may be that some of the changes directly associated with AmB resistance (particularly to sterols, discussed in greater detail below) would in themselves result in ROS hypersensitivity, were it not for the compensatory changes observed in oxidative stress metabolism. In this context, it is important to consider that such changes to oxidative stress metabolism have already become widespread as a result of resistance to pentavalent antimonials (section 1.3.1), and therefore with respect particularly to Indian *L. donovani* strains, many of these putatively compensatory adaptations may already be present in the genetic background.

3.3.3 Sterol changes in AmB-resistant *Leishmania* and their relationship with parasite fitness

One feature of AmB resistance that has emerged repeatedly is loss of wild-type ergostane-like sterols. Unsurprisingly, this phenotype was observed in all four AmB-resistant lines. Wild-type parasites were found by GC-MS (Table 3-2) to contain principally two sterols, ergosta-5,7,24(28)-trienol (82.6%) and ergosta-7,22-dienol (11.7%). While both sterols possess C24-methylation, the former also possesses 5(6)-7(8) double bond conjugation. These are the two main features that have been implicated as conferring specificity for AmB towards ergosterol-containing membranes (HsuChen & Fiengold 1973, Nakagawa *et al*. 2014). Ergosta-5,7,24(28)-trienol is an isomer of ergosterol in which the double bond in the side chain is shifted from the 22(23) position to the 24(28) position. The major peak in wild-type cells was identified as this, rather than ergosterol itself, because it failed to match the retention time of an ergosterol standard run in parallel, and it has previously been detected in other analyses of *Leishmania*

sterols (de Souza & Fernandes Rodrigues 2009, Andrade-Neto et al. 2016). However, it differs from a previous study of *L. mexicana* by this group (Mwenechanya *et al*. 2017) in which ergosterol itself made up 86% of wild-type sterol, and ergosta-5,7,24(28)-trienol was barely detected. The reasons for this discrepancy are unclear; while they could reflect methodological differences (either in detection or in identification), it should be noted that ergosta-5,7,24(28)-trienol is itself a late intermediate in ergosterol biosynthesis, and changes may reflect biological variations in activity of the two enzymes downstream (Figure 1-7). In the earlier study by Mwenechanya and colleagues, the close similarity of ergosterol and ergosta-5,7,24(28)-trienol may have led to misidentification as ergosterol, due to lack of standards and of prior knowledge. Nevertheless, the identical spectrum of wild-type sterol extracts to pure ergosterol shows the close relationship of the two sterols (Figure 3-7).

In all resistant lines, this sterol was lost. The effect of this appears to have been to result in near total or total loss of either C24-methylation or 5(6)-desaturation. The four resistant lines fall into two clear groups. AmBRA/cl1 loses 5(6)-7(8)-conjugated desaturation, with near total loss of wild-type ergosta-5,7,24(28)-trienol and accumulation of the minor wild-type sterol, ergosta-7,22-dienol. The latter differs from ergosterol biosynthesis intermediate episterol only by the position of the side-chain double bond. As expected from this loss of 5(6)-7(8) double bond conjugation, no UV absorbance was observed in sterol extracts, similar to the line previously selected by our group that contained a *CYP51* mutation (Figure 3-7). Ergosta-7,22-dienol has not previously been observed to accumulate in AmB-resistant *Leishmania*; however, it became the dominant sterol in an AmB-resistant *C. albicans* isolate which possessed a mutation in *erg3* (sterol C5-desaturase).

By contrast, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 exhibit very different sterol compositions to AmBRA/cl1, but very similar to each other. Both wild-type ergosta-type (C24-methylated) sterols are almost, albeit not totally, lost, and cholesta-type sterols, particularly cholesta-5,7,22-trienol, are accumulated. Cholesta-5,7,22-trienol differs from ergosterol, only through its lack of C24-methylation. A similar sterol, cholesta-5,7,24-trienol, was previously identified as the major sterol in two AmB-resistant *L. donovani* lines (Mbongo *et al*. 1998,

Purkait *et al*. 2012). This was present at a lower frequency (5-6%) in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. All three lines retained a UV spectrum, although UV absorption at 271 nm was reduced by 40-50%. That UV absorption remains less in these lines could either relate to differences between extinction coefficients for these sterols (which seems unlikely as differences are found only in the side chains, which are distant from the site of UV absorption) or to overall lower sterol composition in these lines. One cautionary note in this respect, however, is that parasites were normalised only by cell number and not cell mass. Nevertheless, visual inspection did not reveal any evidence of size changes and so it is highly unlikely that cell mass accounts for these differences (indeed, total ion chromatograph values, representative of total metabolite loading in LC-MS analysis, suggested that control of cell number alone was adequate for ensuring metabolite yield was not dependent on cell line).

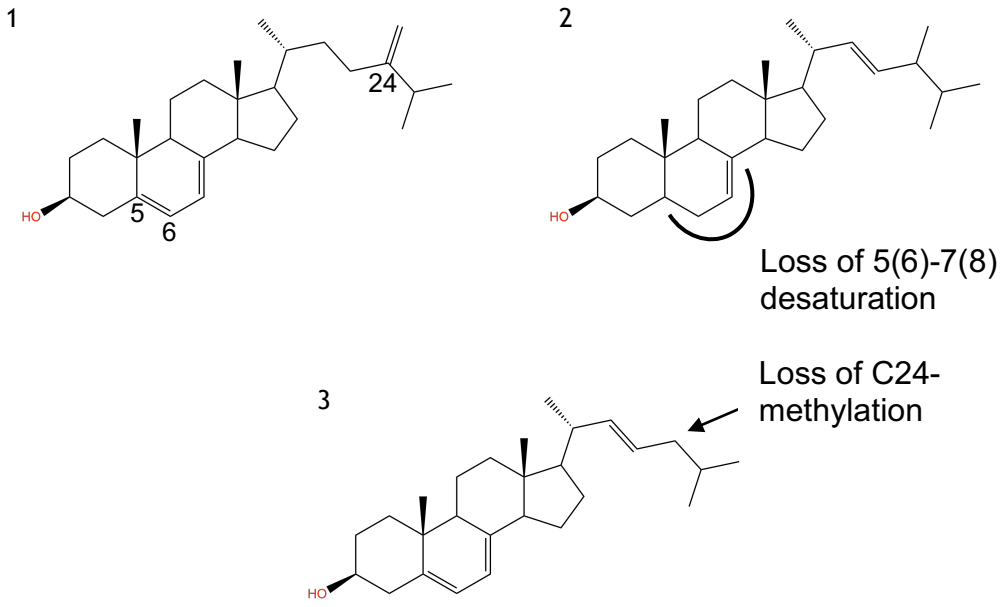


Figure 3-19: Major sterols in wild-type and resistant *Leishmania*.1) Ergosta-5,7,24(28)-trienol, the major sterol in wild-type parasites. The position of carbons 5,6 and 24 are labelled. 2) Ergosta-7,22-dienol, the major sterol in AmBRA/cl1. 3) Cholesta-5,7,22-trienol, the major sterol in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. The key points of structural difference as described in the text are indicated.

3.3.3.1 Sterol changes and sensitivity to oxidative stress

As described, there are two modes of sterol changes amongst the four lines, one of which has been seen previously in *L. donovani*, and the other that has only hitherto been observed in fungi. As sterols play a major role in many cellular processes, it is expected that different changes will have different phenotypic effects on broader parasite physiology, including in the context of consequences for parasite fitness. This is of particular importance with regard to the feasibility of emergence of such changes in clinical populations. While our ability to interpret these data in this respect is clearly limited by an expected background of other changes acquired during selection of resistance, it is nevertheless useful to compare which phenotypes are shared across AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, but different from AmBRA/cl1. Notably, only AmBRA/cl1 demonstrates evidence of a log-phase growth defect in the promastigote stage. Furthermore, only AmBRA/cl1 exhibited hypersensitivity (albeit mild) to GOX, and while no evidence was found of menadione hypersensitivity in this line, AmBRA/cl1 was the most sensitive line to methylene blue. The only line which demonstrated significant hypersensitivity to menadione was AmBRB/cl2, although this was also very mild.

Tentatively, these data might suggest that the fitness costs of the sterol change in AmBRA/cl1 were more pronounced than the one present in the other three resistant lines. Indeed, hypersensitivity to oxidative stress is consistent with a putative role of 5(6)-7(8) double bond conjugation both in supporting membrane function by maintaining sterol ring rigidity, and possible stabilisation of radical electrons (Xu *et al*. 2001, Dupont *et al*. 2012). Notably, hypersensitivity to H_2O_2 was described in a previously selected *L. mexicana* line that contained a mutation in *CYP51* (Mwenechanya *et al*. 2017). This line was shown here to exhibit almost total loss of the ergosterol-related UV spectrum, helping to strengthen the association between loss of double bond conjugation and hypersensitivity to certain types of oxidative stress. Furthermore, the H_2O_2 used in this study, along with GOX-produced H_2O_2 described here, represent an extracellular source of ROS in contrast to menadione and methylene blue. This further supports a role of sterol ring double bond conjugation in direct protection of the membrane from oxidative stress, either through stabilisation of radical electrons as described, or through allowing closer packing and reduced

exposure of other lipids. By contrast, AmBRB/cl2 exhibited mild hypersensitivity to menadione, an intracellular source of superoxide ions. However, given that the sterol profile in this line is similar to AmBRC/cl3 and AmBRD/cl2, this would appear to rule out sterol changes alone as a cause of this phenotype.

3.3.3.2 Sterol changes and retention of infectivity

Another key fitness phenotype to consider is the ability to infect macrophages and to differentiate into amastigotes. Due to the ease of culture of *Leishmania* in the promastigote stage, and the simplification that results from working with an extracellular parasite rather than an intracellular one (as for the amastigote stage), promastigotes have frequently been used as a model for *in vitro* experimental work, including selection and characterisation of drug resistance. There are, however, numerous differences, metabolic and otherwise, between insect and mammalian stages of the *Leishmania* life cycle. Assessment of the relevance of phenotypes observed in *Leishmania* promastigotes to the mammalian stage is therefore essential, since the mammalian stage is the one subject to drug treatment in the clinical context. I therefore assessed the ability of each AmB-resistant line to infect primary murine macrophages, macrophages being a key host cell *in vivo*. A wild-type line highly passaged in parallel in the absence of drug during resistance selection was used, to control for the well-established effects of serial passaging in the promastigote stage on parasite infectivity. Three out of four resistant lines (AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2) showed dramatically reduced infectivity and ability to replicate inside macrophages, with infection levels similar to those observed in the highly passaged wild type control. In contrast, AmBRC/cl3 retained infectivity and replication within macrophages similar to the parental line, in spite of being passaged for close to a year in the promastigote stage. AmBRC/cl3 and AmBRD/cl2 both demonstrated the ability to survive *in vivo* for several weeks, although quantitative comparison of virulence with wild-type parasites was not performed. Nevertheless, in both cases the resistance phenotype was stable in the absence of drug both in culture as promastigotes and within a mammalian host.

In the context of changes to sterol composition, this clearly shows that loss of wild-type sterols and replacement with cholesta-5,7,22-trienol is not in itself

enough to prevent macrophage infectivity (and indeed of the three lines accumulating this sterol, loss of ergostane-type sterols was most profound in AmBRC/cl3). Instead it suggests that loss of virulence in AmBRB/cl2 and AmBRD/cl2 are due to changes other than those to sterol composition. Nor does it appear that these sterol changes prevent viability *in vivo*. By contrast, it is not possible to claim whether or not the changes observed in AmBRA/cl1 are compatible with infectivity, as it is impossible to discount a role for other changes. Furthermore, while no swelling was observed in footpads for AmBRA/cl1 and AmBRB/cl2, it is possible that a small number of viable parasites may be present, and with only one mouse each infected, one cannot rule out that these experiments were anomalous. Whilst recovery of viable promastigotes is proof that parasites are viable *in vivo*, even with a very small sample size (one or two mice), the converse, that lack of swelling means lack of infectivity, is not necessarily true.

Given evidence of stage-specific changes to sterol composition (Yao & Wilson 2016, Bouazizi-Ben Messaoud *et al*. 2017) it is also important to consider whether these changes are associated with resistance in the intracellular amastigote stage. Out of the four lines resistant to AmB as promastigotes, only one, AmBRC/cl3, retained resistance in the intracellular amastigote stage. There are a number of possible explanations for this. First of all, it is noteworthy that this was also the only line that retained evidence of the ability to replicate within macrophages. Continuous passaging in increasing concentrations of drug selects for drug resistance in the strict sense (that is, the ability to grow in the presence of drug concentrations that inhibit the growth of drug-sensitive lines), and the Alamar blue assay itself explicitly measures ability of the drug to inhibit growth. The fact that growth was not witnessed in amastigote forms of AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2, therefore, means that, strictly speaking, any changes in the number of parasites per macrophage dependent on drug dose most likely relate to tolerance rather than resistance, which is a separate phenomenon (the ability of organisms to survive drug-mediated killing, even if growth itself is inhibited). Thus the lack of growth may prevent detection of resistance. Furthermore, low infection rates bring technical challenges, particularly the fact that since counts are in all cases very low, the difference between maximal and minimal infection rates is also lower in absolute terms.

Hence minor variations in absolute counts between replicates can lead to far greater variation in calculated IC_{50} values, potentially preventing detection of significant changes. It should be noted, however, that while infection rates were similarly low in highly passaged wild-type cells, both mean IC_{50} and variability were similar to low passage wild-type, despite higher rates of infection and replication in the latter case. A final possibility is that there is a biological basis for the lack of resistance in AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2, in that the resistance mechanisms acquired by these lines are not applicable in the amastigote stage. However, based on sterol composition alone, one cannot say that this is the case, as sterol changes are so similar between AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, despite the fact that resistance is conserved in the amastigote stage in AmBRC/cl3 but not the other two lines. In lines with poor infectivity and intracellular growth it is also unclear whether differentiation-associated physiological changes actually occur.

A final observation, with respect to possible effects of sterol changes on drug sensitivity, is the lack of an obvious sterol-associated signature in sensitivity to ergosterol biosynthesis inhibitors. All four lines show hypersensitivity to imipramine; however, this changes is so small that the biological significance is unclear, a surprising result given the almost complete loss of sterol C24-methylation in lines AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. Furthermore, imipramine has other effects on *Leishmania* physiology, including effects on proline transport and proton motive force (Zilberstein *et al*. 1990), which may account for this phenotype. While no changes were observed in ketoconazole or fenpropidin sensitivity, AmBRB/cl2 and AmBRD/cl2 were cross-resistant and hypersensitive, respectively, to spiroxamine, despite the near identical effects on ergosterol biosynthesis in these lines. It may be that other biological differences are responsible; however, it should also be considered that this compound might have off-target effects responsible for toxicity, irrespective of changes to sterol metabolism. This seems likely since the relative sensitivities to spiroxamine and fenpropidin differ, despite nominally both targeting sterol C14-reductase.

3.3.4 Broader changes in metabolism

Multiple mechanisms have been implicated in AmB resistance beyond changes to sterol biosynthesis, including increased expression of oxidative stress-related genes and changes to phospholipid composition. Therefore, LC-MS metabolomics was conducted in order to determine whether broader metabolic changes were observable in AmB resistance, and whether these changes were conserved across multiple independent resistant lines.

Initial data analysis using the IDEOM pipeline (Creek *et al*. 2012b) identified a large number of peaks, some of which were not of sufficient quality for reliable identification and relative quantitation. I therefore employed both initial criterion-based filtering based on retention time and metabolite category, and manual removal of poor quality peaks from the dataset. This helps to reduce the number of false positives but also helps to increase the chance of detection of significant changes in higher quality data by reducing the overall number of hypotheses tested. PCA analysis showed good separation of samples only after pools were excluded; this also showed that AmBRB/cl2 was divergent from other lines (including wild-type). Despite normalisation of data through log-transformation, it is possible that certain metabolites were dominant in PC1, resulting in the separation observed, and indeed loadings plots suggested that certain lipid species played a particularly strong role in this.

I employed statistical analysis using an initial ANOVA with P-value correction to reduce the number of false positives arising through multiple comparison, followed by Tukey's HSD to detect significant changes in comparison to wild-type, as direct differences between resistant lines were not specifically explored. After this analysis, the total number of significantly changing metabolites represented a high percentage of all metabolites detected after filtering (68%), a figure which varied in individual lines from 14% to 42%, with only limited overlap. This suggests the scope of metabolic variation is considerable. One specific limitation of this dataset is the lack of an explicit control for high passaging, as the wild-type control used was the parental line. A previous study by our group into metabolomics of AmB resistance (Mwenechanya *et al*. 2017) used a highly passaged wild-type as a control. This creates the complication that sporadic changes occurring during passaging of the wild-type

line are likely to result in changes in comparison to AmB-resistant parasites that will be interpreted as alterations arising during selection of resistance. Whilst the experimental design used here avoids this problem, it also raises the question of whether metabolite differences arise as a result of passaging or due to resistance selection. To some extent, the use of multiple independent resistant lines circumnavigates this issue, as changes that universally occur during high passaging *in vitro* (and hence would be spotted through use of a highly passaged wild-type control) should occur in all four lines, which only accounts for a small minority (6%) of significant changes. Nevertheless, it creates the problem that those metabolite changes could either reflect universal changes in AmB resistance or universal consequences of serial passaging. These data should therefore be interpreted with caution. An ideal experimental design would therefore include both the parental line and a derived highly passaged wild-type line as controls; however, this also adds to experimental complexity and expense.

Changes in specific pathways are discussed in greater detail below:

3.3.4.1 Lipid metabolism

The limitations of HILIC with respect to resolution of hydrophobic molecules, and the difficulties of identification of these large metabolites, are discussed in greater detail above (section 3.2.5.1). Therefore, it is necessary to remain cautious when discussing lipids, but it may still be valuable to examine patterns in the data. These overall patterns, as shown in Figure 3-13, clearly suggest that while in AmBRC/cl3 and AmBRD/cl2, with some exceptions, lipid levels remain relatively similar to wild-type, in AmBRA/cl1 there is a trend in some towards decreased levels, and most of all there is a clear decline in many lipids in AmBRB/cl2. Several species, particularly 14:1 and 15:1 phosphatidylcholines (monounsaturated), are greatly reduced specifically in AmBRB/cl2 whereas an 18:0 phosophatidylcholine is greatly increased. Several other unsaturated phospholipid species are observed to decrease in this line, as well as several sphingolipid species (which in fact include 17:0 saturated sphingolipid species).

Mbongo and colleagues previously described a large decrease in unsaturated fatty acid species linoleate and oleate and a large increase in stearate (18:0),

with an overall increase in percentage saturated fatty acid from 4% to 57% in AmB-resistant *L. donovani* promastigotes (Mbongo *et al*. 1998). While it is harder from this untargeted analysis (where many lipids may be undetected, or excluded on peak quality or retention time grounds) to provide as complete a picture of fatty acid residue saturation, nonetheless it suggests that similar changes are observed in AmBRB/cl2, with far less evidence in other lines. It is interesting to note, however, that tetradecanoyl-CoA, the CoA derivative of the 14:0 fatty acid tetradecanoic acid, is increased in three out of four lines (Figure 3-12). This in itself may indicate that a relative increase in saturated fatty acid biosynthesis occurs beyond AmBRB/cl2.

Fernandez-Prada and colleagues previously described phospholipid changes in an AmB-resistant *L. infantum* line that contained mutations in the miltefosine transporter (Fernandez-Prada *et al*. 2016). This study particularly identified increases in various inositol-phosphoceramides, as well as decreases in phosphatidylinositol species and increases in some phosphatidylethanolamines. With respect to individual fatty acids, overall composition remains relatively unchanged, albeit with some increases in the cyclopropyl fatty acid species C19Δ and 24:0. Overall, however, no change in overall fatty acid saturation was observed. The lipid species identified are quite different from those described here, as the methodology itself (electrospray ionisation mass spectrometry, with a lipid focused extraction protocol and Nanomate interface for total lipids, GC-MS of derivatives for fatty acids) differs considerably. This makes the results from this study hard to compare directly with our own. Nevertheless, the clear contradiction with the study by Mbongo with regard to fatty acid saturation shows that very different lipid changes can arise during AmB resistance selection. It may be that very different combinations of lipid species (including both phospholipids and sterols) can result in similar overall changes to membrane properties, such as fluidity and AmB binding affinity.

3.3.4.2 Carbohydrate and energy metabolism

Multiple lines of evidence implicate central carbon metabolism as having the potential for involvement in AmB resistance in *Leishmania*. The PPP is one such area of metabolism that appears to have an important role in responses to antileishmanial drugs, through production of NADPH as a source of reducing

power for ROS detoxification pathways. Overexpression of glucose-6-phosphate dehydrogenase and transaldolase both lead to protection against various oxidative stress-inducing agents such as H_2O_2 and menadione, as well as antileishmanial drugs AmB, sodium antimonyl gluconate and MILT (Ghosh *et al*. 2015), which are known to act via generation of oxidative stress (Moreira *et al*. 2011). Given these observations, it is perhaps surprising that there is no evidence of increased PPP metabolism in any of the four AmB-resistant lines described here. While sporadic significant changes are observed (Figure 3-14) these are not conserved between resistant lines, and moreover are generally speaking decreases rather than increases. Consistent with this, NADPH is not increased in any line. Thus, it does not appear that PPP adaptations are an observed phenotype of selection of AmB resistance.

By contrast, there is evidence that the PPP is involved in response to AmB treatment. Several PPP enzymes, including glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase and transaldolase, show increased expression in response to oxidative stress (Ghosh *et al*. 2015). Recent untargeted LC-MS experiments examining initial metabolic changes after 15 minutes AmB treatment (prior to parasite lysis) indicate large increases in PPP metabolites, including D-ribose-5-phosphate, D-sedoheptulose-7-phosphate, D-glucose-6-phosphate and D-erythrose-4-phosphate (Reilly, Alpizar Sosa & Binti Ithnin, unpublished observations). Therefore, it is feasible that adaptations do exist, but are observed only as a response to drug-induced stress, with greater capacity for increase in pathway flux when needed, as opposed to altered basal PPP metabolism. This has yet to be tested, however.

With respect to broader central carbon metabolism, specifically glycolysis and the TCA cycle, little is known about a role in AmB resistance in *Leishmania*. In fungi, one observation of relevance is that, in keeping with a role of mitochondrion-generated ROS in AmB-mediated killing, mitochondrial respiration rates are observed to decrease in AmB-resistant *Candida* species (Mesa-Arango *et al*. 2014). Both the TCA cycle and respiration have been implicated in cell death caused by AmB and other fungicidal drugs (Belenky *et al*. 2013). Therefore, one might expect reduced TCA flux in AmB-resistant *Leishmania*. There is little evidence of this, however (Figure 3-15); while acetyl-CoA was reduced in

AmBRB/cl2 and AmBRD/cl2, citrate, cis-aconitate and malate were unchanged and succinate actually increased slightly in all lines. There was, however, some evidence of increased glycolysis, particularly in AmBRB/cl2, where 3-phospho-D-glycerate increased and both pyruvate and phosphoenolpyruvate moderately but not significantly increased. Moreover, lactate was increased in all lines, significantly so in three. The universal increase in succinate, with no evidence of increases in other TCA metabolites, may indicate increased glycosomal production of this metabolite from phosphoenolpyruvate, which occurs in *Leishmania* as a route of NAD^+ regeneration during glycolysis, in a similar fashion to lactate production (Saunders *et al*. 2011). A general trend towards increased ATP, significant in AmBRA/cl1 and AmBRB/cl2, was also observed. No significant changes in AMP were observed, suggesting an increased ATP/AMP ratio. It is possible that increased ATP results from higher fermentative metabolism, resulting in accumulation of lactate and succinate as waste products. While it cannot be ruled out that this simply results from long-term passaging (particularly as the environment is glucose-rich, and may therefore permit increased reliance on energy-inefficient fermentative pathways), it is also possible that increased fermentation may allow reduced dependency on mitochondrial oxidative phosphorylation, therefore reducing vulnerability to drug-induced mitochondrial ROS. The relevance of this to the more nutrient-limited environment of the macrophage phagolysosome, however, is unclear.

The most striking phenotype observed in carbohydrate metabolism is, however, the decreased abundance of several polysaccharides equivalent to three, four, five and six hexose units in AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2 (Figure 3-15). In *Leishmania*, mannogen is believed to be the major polymer used for carbohydrate storage, suggesting that these species reflect mannose polymers. In keeping with this, the precursor of mannogen biosynthesis, GDP-mannose, shows a similar pattern of decline in these three lines, suggesting that mannogen biosynthesis is reduced. Mannogen appears to be an important reserve in nutrient-limited environments, and is established as a virulence factor, as developmental stages or mutant *L. mexicana* lines with low mannogen content do not survive heat shock, nor are they able to differentiate into amastigotes or infect macrophages (Ralton *et al*. 2003). As mannogen is not required for growth as promastigotes, it is feasible for mutations in mannogen biosynthesis to arise

during serial passaging in this stage. It is very possible, therefore, that these changes have arisen stochastically during selection of resistance in AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2. Indeed, the fact that infectivity (which is retained only in AmBRC/cl3) correlates with polysaccharide abundance but not sterol content (as AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 share very similar sterol profiles), means that loss of mannogen biosynthesis is a more likely cause of loss of infectivity than disruption to sterol biosynthesis. This is less clear for AmBRA/cl1, however, since the sterol changes observed in this line have not been observed elsewhere. Nevertheless, caution must be used in identifying altered mannogen biosynthesis in loss of virulence since overall fewer changes are observed in AmBRC/cl3 across a broad range of areas of metabolism; therefore, other factors may also play a role.

3.3.4.3 Amino acid and thiol metabolism

Metabolites related to amino acid metabolism represent the largest portion of identified metabolites (39%), with 56 metabolites differing significantly from wild-type parasites across all lines. Many of these changes, however, are not conserved between resistant lines. It is possible, therefore, that these changes have little to do with selection of drug resistance but represent stochastic changes acquired during growth in a nutrient rich environment. The changes that are most widespread are in arginine and proline metabolism. Arginine metabolism is of particular interest because of its connection to polyamine biosynthesis and hence to trypanothione production, which is key to protection against oxidative stress. Polyamines themselves, such as putrescine and spermidine, are poorly detected by the HILIC LC-MS platform used here. However, N-acetyl-putrescine was detected, and showed increases in AmBRB/cl2 and AmBRC/cl3, potentially serving as an indicator of similar changes in putrescine. Trypanothione levels were increased in all lines, although not significantly in AmBRB/cl2, suggesting that its biosynthesis is increased. Interestingly, the oxidised form, trypanothione disulfide, remains unchanged. This may suggest that the ratio of reduced to oxidised trypanothione is increased, thereby providing greater capacity for protection against ROS. However, as trypanothione can be oxidised during extraction (t'Kindt *et al.* 2010), this ratio may be unreliable.

Despite increases in trypanothione, as well as its precursor, glutathione, in some lines, L-arginine and L-ornithine themselves were decreased in AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2, as were the arginine degradation products, 4-guanidinobutanal and 4-guanidinobutanamide. S-adenosyl-L-methionine, processed to the aminopropyl donor required for spermidine production, S-adenosyl-5'-(3-methylthiopropylamine), is also decreased in AmBRB/cl2 and AmBRD/cl2. As putrescine and spermidine were not detected, it is more difficult to interpret how trypanothione increases in spite of these upstream decreases, but it may be that increased glutathione in AmBRA/cl1 and AmBRD/cl2 is a contributing factor. Changes to N-acetylputrescine did not correlate with trypanothione, with the biggest increase in the former observed in AmBRB/cl2, the only line in which trypanothione did not significantly increase. Alternatively, it is possible that decreases in arginine metabolism arise as a result of increased trypanothione biosynthesis, depleting the substrate pool without adequately replenishing it. What is noteworthy, however, is that AmBRB/cl2, the line with the highest degree of AmB resistance, exhibits little evidence of increased trypanothione and a significant decrease in glutathione levels. Therefore, high AmB resistance does not necessarily correlate with high trypanothione levels.

Proline, along with various proline-related metabolites, was observed to decrease significantly in AmBRB/cl2 and AmBRD/cl2, with small but non-significant decreases in other lines. Other metabolites in proline metabolism were detected and found to be altered in abundance; however, the presence of multiple peaks with the same mass (and no standard) means that it is difficult to interpret the biological significance of these changes. For example, L-1-pyrroline-3-hydroxy-5-carboxylate is produced from proline by proline dehydrogenase. However, whilst one peak shows declines very similar to proline itself, the other shows no significant changes. A more detailed analysis of the exact changes to proline metabolism would therefore require greater use of standards. Nevertheless, the decrease in proline is a surprising result. Previous metabolomics analysis has found increases in proline to be universal across a range of single- and multi-drug resistant *L. donovani* lines, including an AmB-resistant line (Berg *et al*. 2015), and proline has independently been observed to increase in MILT-resistant *L. donovani* (Canuto *et al*. 2014). Proline itself is known to be protective against environmental stresses, including oxidative

stress, and can even act as a ROS scavenger *in vitro* (Kaul *et al*. 2005), although supplementation with proline does not in itself reduce drug sensitivity to AmB (although it does for Sb^{III}) (Berg *et al*. 2015). Therefore, it is unclear why decreased proline abundance might be observed here, but could suggest species-specific differences between this study and the two *L. donovani* studies described above. *L. mexicana* may simply not use proline to protect against stress in the same way as for *L. donovani*.

3.4 Conclusions

The potential for emergent AmB resistance is a considerable threat to leishmaniasis control programmes, but whilst loss of ergostane-type sterols appears to be a consistently observed phenotype, other conserved signatures are unclear. Here evidence is presented of extensive heterogeneity between AmB-resistant lines in terms of drug sensitivity, fitness and metabolotype. These data question a number of previous observations derived from previous individual analyses of AmB resistance, namely that selection of resistance leads to cross-resistance to other antileishmanial drugs that act through generation of ROS (Moreira *et al*. 2011, Mwenechanya *et al*. 2017), and that AmB resistance is associated with sensitivity to oxidative stress, both in *Leishmania* (Mwenechanya *et al*. 2017) and in yeast (Vincent *et al*. 2013). Rather, while altered sensitivity to ROS and other antileishmanial compounds may arise, they are not essential features of selection of AmB resistance, nor is loss of infectivity. Sterol changes are also heterogeneous, with AmBRA/cl1 displaying a different sterol profile from AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, and other sterol changes have been observed in the literature. Similar heterogeneity is observed in broader metabolism, complicated here by phenotypes such as reduction of polysaccharide biosynthesis that may arise stochastically during the process of resistance selection.

This poses a challenge in terms of identification of resistance-associated markers, an essential tool for surveillance of resistance in the field and prediction of drug response. While disruption of ergosterol biosynthesis is a conserved phenotype, GC-MS-based sterol analysis is not a practical tool for surveillance. Whilst UV spectroscopy represents a more feasible solution in terms of potential for the development of portable tests, it is evident here that loss of

the ergosterol-associated UV spectrum is only observed in certain lines. Thus, the utility of such a test would depend on the specific changes present in an individual AmB-resistant isolate.

An important limitation of this study has been the focus on fitness and infectivity *in vitro*. While I have shown that at least two lines (AmBRC/cl3 and AmBRD/cl2) are infectious *in vivo*, I have not quantified that infectivity in terms of rates of lesion formation, nor determined whether resistance is truly exhibited within the mammalian host. Nevertheless, AmBRC/cl3 showed no evidence of loss of fitness in a primary macrophage model, suggesting that loss of C24-methylation (as observed in this line) is consistent with maintenance of virulence, and loss of virulence in other lines with similar sterol changes is likely due to other factors such as reduced mannogen biosynthesis. Indeed, it may well be that much of the phenotypic heterogeneity described here only arises because of lack of purifying selection during serial passage of axenic promastigotes in a growth rich medium. It will be interesting to determine whether a narrower range of phenotypes are found in AmB-resistant parasites compatible with pathogenicity *in vivo*.

There is one phenotype, however, which may be of particular importance in terms of its applicability. All AmB-resistant lines described here, in addition to those previously tested (Mbongo *et al*. 1998, Mwenechanya *et al*. 2017), exhibit hypersensitivity to PENT. It will be important to determine which specific changes underpin this relationship. Clearly PENT is a drug that faces considerable limitations in terms of its efficacy and safety profile. However, in parasite populations where resistance to both pentavalent antimonials and MILT is emerging or prevalent, and rapid selection of PMM resistance is largely expected, hypersensitivity to PENT could represent an important tool for management and suppression of any emergent AmB resistance.

4 Characterisation of AmB resistance-associated genetic and expression changes through next generation sequencing

4.1 Introduction

The previous chapter describes the identification of a range of phenotypic and metabolic changes observed in AmB resistance. Thus far, however, I have not described the genetic basis of any of these changes. The stability of resistance in the absence of drug pressure would suggest that genetic changes, which are not easily reversible, underlie the phenotype observed. Identification of the genetic basis of such changes offers a number of advantages. First of all, from a practical perspective, identification of mutated genes facilitates the development of screening assays to detect such changes, which may be more straightforward than detection of the metabolic effects themselves (such as altered sterol composition). Secondly, experimentally speaking, putative resistance-associated mutations can be tested directly using molecular biology methods, such as reintroduction of wild-type versions of mutated genes to determine whether wild-type phenotypes are restored.

Next generation sequencing (NGS)-based methods offer significant advantages over traditional sequencing methods in that they offer a genome-wide perspective rather than one limited to targeting specific candidate genes. First of all, NGS methods allow identification of changes at genomic loci that may not previously have been suspected. However, they also provide a broader range of types of information about changes, besides local changes such as SNPs and InDels. On the basis of the total number of reads mapping to specific genomic features, one can infer copy number variations, both at the level of individual genes or genomic loci, and of whole chromosomes. This is particularly pertinent to *Leishmania*, which is known to have high levels of CNV at both of these levels (Rogers *et al*. 2011).

Whole genome sequencing (WGS) has been highly informative in determination of resistance-associated changes in drug-resistant parasites, as discussed in section 1.4.2.5. However, because access to NGS technology has only recently become widespread, few studies have applied it in the context of AmB-resistant

Leishmania. Currently, the two mutations identified are the N176I mutation in sterol biosynthesis gene *CYP51* and the G433S mutation in the miltefosine transporter. Interestingly, while both mutations are homozygous, neither resulted in total loss of gene function, with C14-methylated sterols still observed in the *CYP51* mutant (*CYP51* encodes for sterol C14-demethylase) (Mwenechanya *et al*. 2017), and MILT still accumulating in the MT mutant (Fernandez-Prada *et al*. 2016). Changes in sterol C24-methyltransferase transcript abundance have also been identified (Pourshafie *et al*. 2004, Purkait *et al*. 2012), but no evidence has been produced linking *SMT* expression changes to genomic alterations, nor has a functional role for such changes been directly validated.

Beyond AmB, WGS has been used in a number of studies to identify changes associated with antileishmanial drug resistance. These outline some of the diversity of changes that can be detected through this technique. For example, both SNPs and frameshift mutations have been detected in the *MT* gene in MILT-resistant *Leishmania* (Coelho *et al*. 2012, Laffitte *et al*. 2016), and a two-nucleotide frameshift mutation in aquaglyceroporin-1 is widespread in Indian antimony-resistant *L. donovani* isolates (Imamura *et al*. 2016). However, besides point mutations, copy number changes have also been implicated in resistance, particularly local gene amplification through generation of circular amplicons. For example, in an antimony-resistant *L. infantum* line, a circular amplicon containing the MRPA ABC transporter gene (associated with drug efflux) was linked with increased expression of that gene (Leprohon *et al*. 2008). Clinical isolates displaying antimony resistance have also been found to have amplifications in genes including those encoding MRPA and tryparedoxin peroxidase (Jeddi *et al*. 2014). Interestingly, two AmB-resistant lines derived from the lizard parasite *L. tarentolae* displayed circular amplicons, whose copy number was linked circumstantially to AmB resistance levels; however, the specific gene amplifications responsible were not identified (Singh *et al*. 2001). Thus, CNV appears to be a frequent mode of mutation during drug resistance, although it must be noted that the limitations of short read sequencing (section 1.4.2.5) can place constraints on characterisation of the specific structural variations giving rise to such phenomena.

While the functional effects of mutations can be obvious where they involve changes in coding sequences, many mutations may have effects on gene function that are harder to predict from sequence changes alone. RNA-seq provides genome-wide information on changes in gene expression, which may in turn be influenced by changes at the genome level. In contrast to WGS, RNA-seq has rarely been used to study drug resistance in kinetoplastids. This is in part due to the unusual transcriptional biology of these organisms, in which most of the genome is constitutively transcribed and RNA levels are moderated on the basis of stability, with translational regulation (which is not detected by RNA-seq) being a major source of expression changes. Examples of where RNA-seq has been used in this context were described in section 1.4.2.5. However, it is worth noting at this stage that expression changes have been described at the *protein* level in AmB resistance in *L. infantum* (Brotherton *et al*. 2014), including increases in both trypanothione reductase and tryparedoxin peroxidase. In a separate study, increased expression of various enzymes involved in trypanothione metabolism, including trypanothione synthetase, trypanothione reductase, tryparedoxins and tryparedoxin peroxidases, were revealed to increase at the RNA level using a semiquantitative PCR assay in an AmB-resistant *L. donovani* clinical isolate (Purkait *et al*. 2012), and increased ascorbate peroxidase expression was also detected in the same line (Kumar *et al*. 2014). Therefore, it is feasible that similar expression changes may underlie the increased abundance of trypanothione in many of these lines, as well as some of the other metabolic phenotypes discussed in the previous chapter. Furthermore, a number of other changes previously associated with AmB resistance, including in drug transporter protein MDR1 (Purkait *et al*. 2012) and various chaperone proteins (Brotherton *et al*. 2014), manifest as altered expression. In cases of changes detected by the study of Brotherton and colleagues, it is unclear firstly how conserved such changes are in other resistant lines, and secondly whether these changes can be detected on the level of transcript abundance, or arise solely due to altered translational or post-translational regulation.

In order to derive an overall picture of changes in gene function in AmB resistance, WGS and RNA-seq were therefore performed in parallel. This allowed us to identify changes at both genomic and transcriptomic levels, as well as

permitting analysis of how the former might influence the latter in determining the AmB resistance phenotype.

4.2 Results

4.2.1 Whole genome sequencing of AmB-resistant *Leishmania*

In order to identify changes associated with AmB resistance at the genome level, one clone was sequenced from each resistant line, AmBRA/cl1, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 (those used for metabolomics and sterol analysis), as well as the wild-type parental line. 34-40 million reads (17-20 million read pairs) were obtained for each sample (Supplementary file 4-1), and high percentage alignments to the *L. mexicana* reference genome of approximately 98% were achieved. With a genome size of 32 Mb, the minimum coverage per sample was 80-fold, meaning that good average read-depth was achieved across all samples.

4.2.1.1 Changes in overall chromosome ploidy

Before examining changes at the gene sequence level, I examined whole chromosome ploidy changes. These were calculated as the median length-normalised per-gene coverage for each chromosome divided by the median value across all chromosomes to provide a ploidy ratio (section 2.5.2) (Figure 4-1A). This revealed extensive evidence of ploidy shifts between lines. However, there were no universal shifts in ploidy in comparison to wild-type shared across all AmB-resistant lines (except chromosome four, see below), suggesting that ploidy changes may have happened spontaneously irrespective of drug pressure. In order to convert ploidy ratios into absolute ploidy values, as required during variant calling (section 4.2.1.2) ploidy ratios were multiplied by a default ploidy of two and rounded to the nearest integer (Figure 4-1B). This revealed similar heterogeneity. Chromosome 30 was polyploid in all lines (three to four copies) as previously described for multiple *Leishmania* species (Rogers *et al*. 2011). Most changes in copy number were increases, with decreases only observed in chromosomes two and 12 (both in AmBRD/cl2), and chromosome four (in all four lines). However, as chromosome four decreases from triploidy in wild-type cells to diploidy in AmB-resistant parasites, the expected default ploidy of four in this chromosome (Rogers *et al*. 2011) means that this may simply represent a spontaneous ploidy increase in wild-type cells.

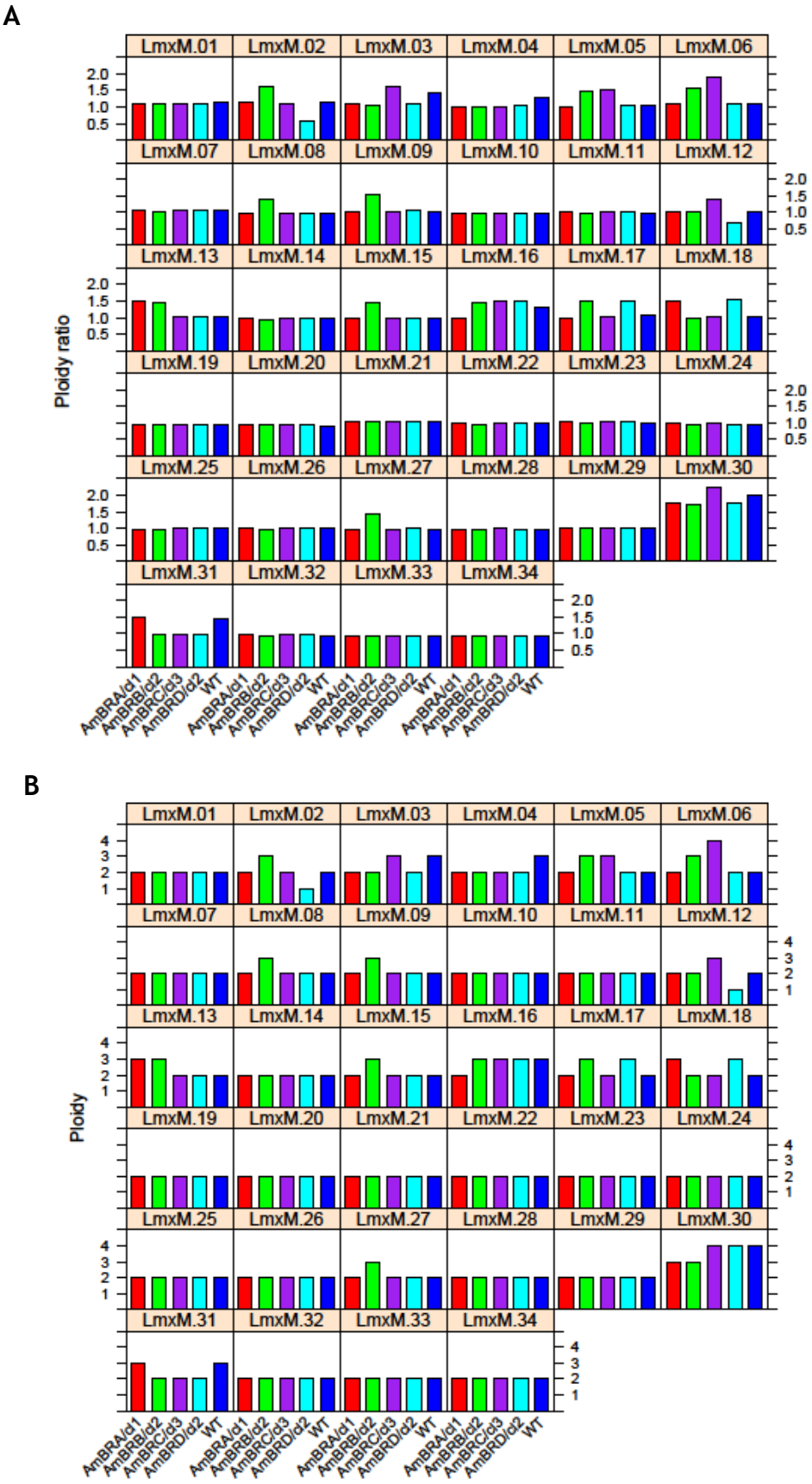


Figure 4-1: Ploidy changes derived from WGS data. A) Ploidy ratios were calculated for each chromosome as the ratio between median length-normalised per-gene coverage for an individual chromosome over the median across all chromosomes. B) Ratios were converted into absolute ploidy values by multiplying these values by two (assuming this to be the default ploidy) and rounding to the nearest integer.

4.2.1.2 Mutations leading to coding changes

I calculated local variants, including SNPs and InDels, using Freebayes (Garrison & Marth 2012). This allowed us to input calculated per-chromosome ploidy values as shown in Figure 4-1B into genotype estimates, allowing greater sensitivity of variant detection for non-diploid chromosomes (for example, expected allele frequencies for a diploid chromosome would be 0%, 50% or 100%, whereas on a tetraploid chromosome these would be 0%, 25%, 50%, 75% or 100%). In total, 44,336 variants were detected (Supplementary figure 4-2) after quality filtering. Of these, 6,744 showed differences from wild type (variants which changed in allele frequency between two heterozygous genotypes, e.g. reference/alternative (diploid) to reference/reference/alternative (triploid) were not included in this number) (Supplementary file 4-3). In terms of variants detected in individual resistant lines that were different from wild-type, in AmBRA/cl1, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 these numbered 1,363, 1,528, 1,386 and 1,769, respectively.

I focused on genes that led to changes in coding sequences, such as non-synonymous substitutions and frame shift mutations. The numbers of coding changes for each line sequenced were 77, 113, 101 and 151 for AmBRA/cl1, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, with nine, 12, 16 and 16 of these being homozygous in individual lines but not detectable in wild-type. However, many of these displayed poor read depth. Of those with mean read depth greater than or equal to 10 across all samples, these numbers decreased further to four, three, eight and four changes. A number of these lacked annotation, being described only as “hypothetical proteins”. Of those which had an annotated putative function (five genes), *LmxM.02.0040* (aminopeptidase P1, variant called in AmBRD/cl2), *LmxM.05.0380* (microtubule associated protein, putative, variant called in AmBRA/cl1 and AmBRC/c3) and *LmxM.14.0665* (fatty acid elongase, putative, heterozygous in AmBRB/cl2 and AmBRD/cl2, homozygous in AmBRC/cl3), the alternative allele was detected at low frequency in wild-type samples and there was possible evidence of miscalled genotypes in each case. The two remaining genes were both in the ergosterol biosynthesis pathway, *LmxM.23.1300* and *LmxM.36.2380*. Mutations in these two genes are discussed in greater detail below.

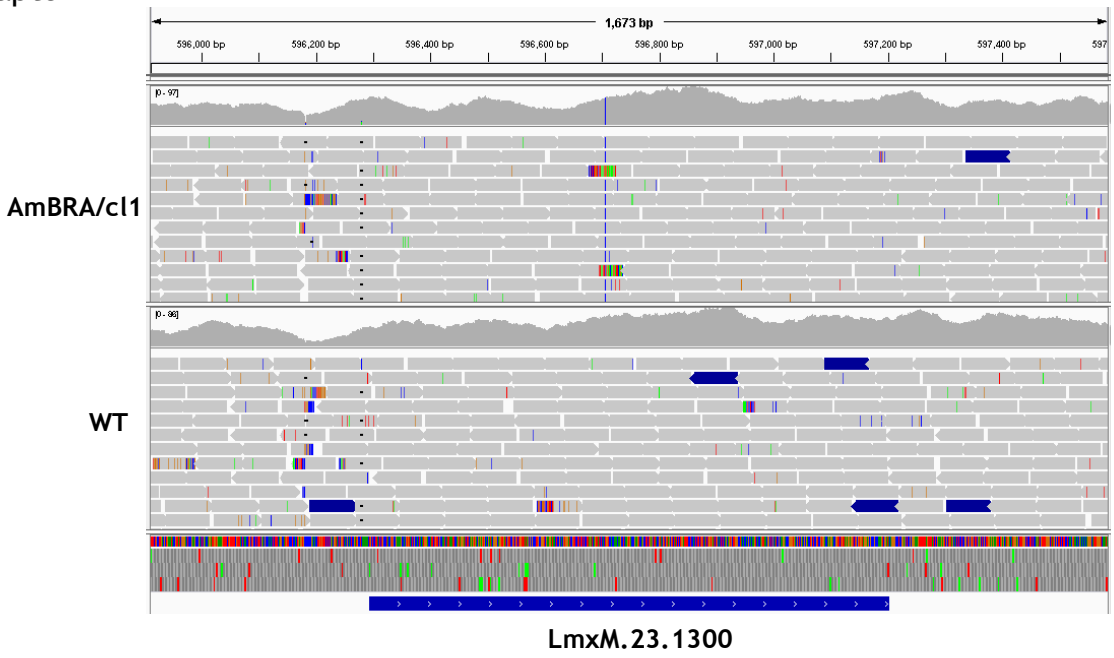


Figure 4-2: Mutation in *LmxM.23.1300*. The G415C mutation in AmBRA/cl1, as visualised using the Integrative Genomics Viewer (Robinson *et al*. 2011). Visualisation through this method can be understood as follows. At the top of the figure, the gene coordinates (here, on chromosome 23) are shown. For each line (AmBRA/cl1 top, wild-type bottom), there is one panel, which has two parts. The top part (the grey continuous segment which rises and falls) shows total read depth by genomic position. The part below this displays individual mapped reads. Concordantly aligned reads (i.e. those belonging to pairs that show alignment in the correct orientation relative to each other and with the expected distance between them, based on library fragment size) are shown in grey, with an arrow shape in one direction (where space is available) indicating orientation of alignment. Coloured reads indicate those that do not, for varying reasons, meet the expectations of read concordance. All bases that do not match the reference genome are represented as coloured vertical lines within individual reads. Note that at the center of the panel displaying the alignment for AmBRA/cl1, there is a blue line representing the same mismatch in all individual reads at the same point, indicating the position of the G415C variant, which is absent in the alignment for wild-type parasites. Below the alignment track for wild-type, the sequence of the reference genome is represented in colour-coded format, along with information about individual reading frames. At the bottom of the figure, the blue box indicates the position of the gene *LmxM.23.1300*. Please note that whilst the genome reference strain to which reads are aligned is *L. mexicana* MHOM/GT/2001/U1103, the wild-type strain here is *L. mexicana* MNYC/BZ/62/M379, hence some differences from the reference genome are expected in the wild-type strain.

LmxM.23.1300 encodes the ergosterol biosynthesis enzyme, sterol C5-desaturase (SC5D). A homozygous non-synonymous substitution (G415C) was detected in AmBRA/cl1 alone. The resulting codon change, GGG to CGG, caused an amino acid change, G139R. In order to predict the functional impact of this change, I aligned *LmxM.23.1300* against SC5D homologues from other species (Figure 4-3). This revealed G139 to be conserved across all kinetoplastid species, including *Leishmania*, *Crithidia* and *Trypanosoma* species, as well as in humans. However, tryptophan was present in *S. cerevisiae*, and glutamate in *A. thaliana*. Therefore, this residue is not universally conserved. However, it is located within a conserved His-rich motif (as annotated in the Uniprot database for the

yeast enzyme, Erg3p), disruption of which through mutation of His residues is known to greatly reduce enzyme activity (Taton *et al*. 2000). Furthermore, a glycine to arginine substitution, from a residue with a single hydrogen atom as a side chain to a much larger, positively charged side chain, represents a large functional change at this position. No structural information for sterol C5-desaturases appears to be available, however, so determining the precise impact of this substitution is challenging.

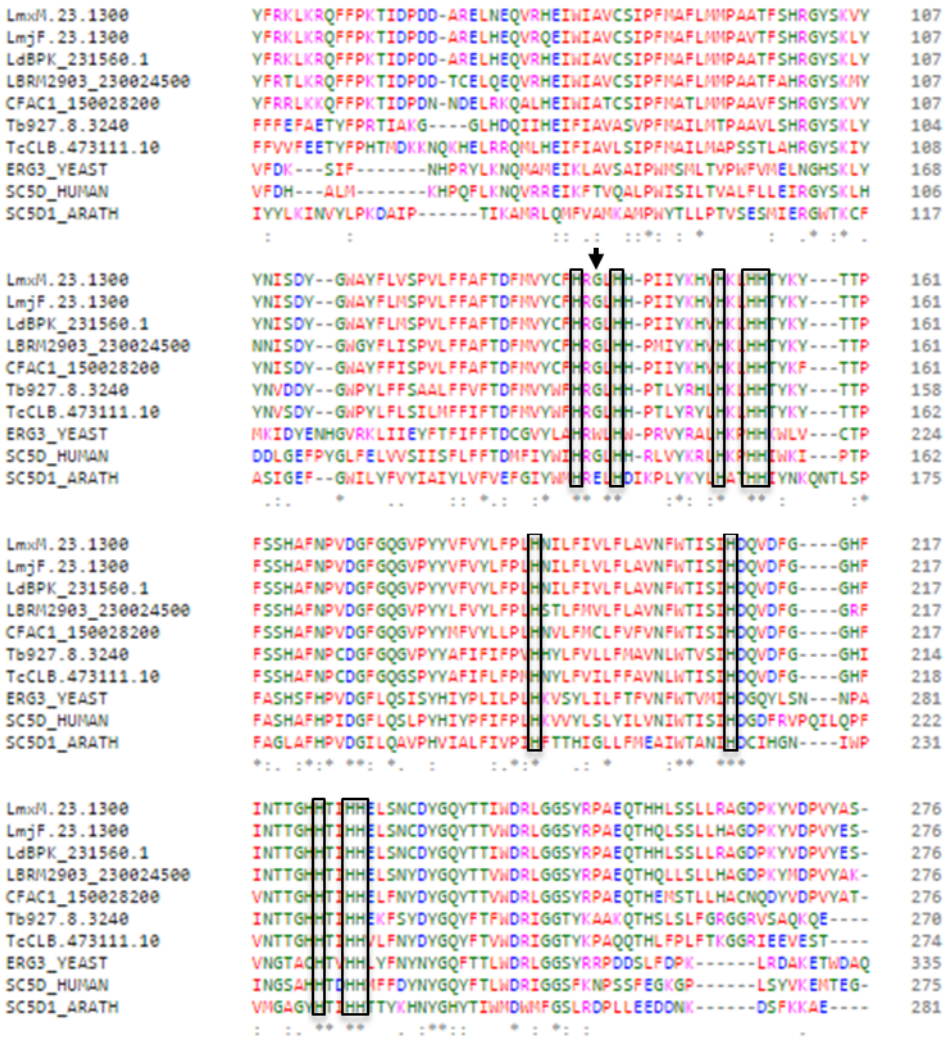


Figure 4-3: Alignment of sterol C5-desaturase sequences from kinetoplastids and other organisms. Alignments were performed using Clustal Ω (Sievers *et al*. 2011). The positions of conserved His residues are marked by boxes. The position of *LmxM.23.1300* G139 is marked with an arrow. Source species are, from top to bottom, *L. mexicana*, *L. major*, *L. donovani*, *L. braziliensis*, *Crithidia fasciculata*, *T. brucei*, *T. cruzi*, *S. cerevisiae*, *Homo sapiens* and *Arabidopsis thaliana*. Kinetoplastid sequences are derived from TriTrypDB, others from Uniprot.

The other sterol biosynthesis gene exhibiting mutations was sterol C24-methyltransferase (*SMT*). The situation here is more complex because two tandemly arranged genes encode this enzyme, *LmxM.36.2380* and

LmxM.36.2390. These differ only by one base change at position 391, G in *LmxM.36.2380* and A in *LmxM.36.2390*. This is a coding substitution, with the codon change GTC to ATC resulting in a V131I substitution. A similar variant, G961A, was also noted to be homozygous in all lines (including wild-type) in *LmxM.36.2390*. This resulted in the codon change GTT to ATT, causing a V321I substitution. Substitution between these two alkyl, branched-chain amino acids is highly conservative, so may not result in major changes to enzyme function.

In AmBRA/cl1, no changes in this gene relative to wild-type were observed. However, changes were observed in the two positions described above in the other three lines. A homozygous G391A substitution was observed in *LmxM.36.2380* in AmBRB/cl2 and AmBRC/cl3, meaning that there was an A at position 391 in both *SMT* gene copies in these lines. A homozygous G961A substitution was also observed in *LmxM.36.2380* in AmBRD/cl2, meaning that an A was found at position 961 in both gene copies in this line. No genotype was called at this position in *LmxM.36.2380* for AmBRB/cl2 and AmBRC/cl3, due to lack of coverage. Therefore, only A961 was detected in these lines.

In addition to these changes, another homozygous substitution, T215G, was observed only in AmBRC/cl3, in *LmxM.36.2380*. This results in the codon substitution, TTC to TGC, and the coding change, F72C. As phenylalanine and cysteine have dramatically different properties (a bulky, hydrophobic aromatic residue in comparison to a small, polar thiol), this might be predicted to affect enzyme function. I therefore aligned sequences from multiple different species to compare to annotated features in other organisms (Figure 4-4). F72 falls within a motif of highly conserved residues, particularly aromatic residues, which comprise a fragment derived from the *S. cerevisiae* protein that is protected from tryptic digestion by binding of an irreversible sterol analogue inhibitor (Marshall & Nes 1999). Thus, these residues represent a putative sterol binding site, suggesting that loss of a conserved hydrophobic residue in favour of a polar one is likely to affect sterol binding. Inspection of the other two residues within the alignment, V131 and V321, shows that while the position of a branched-chain aliphatic residue at these positions is clearly important, other similar residues (leucine or isoleucine) may also be found. In the case of SNPs in

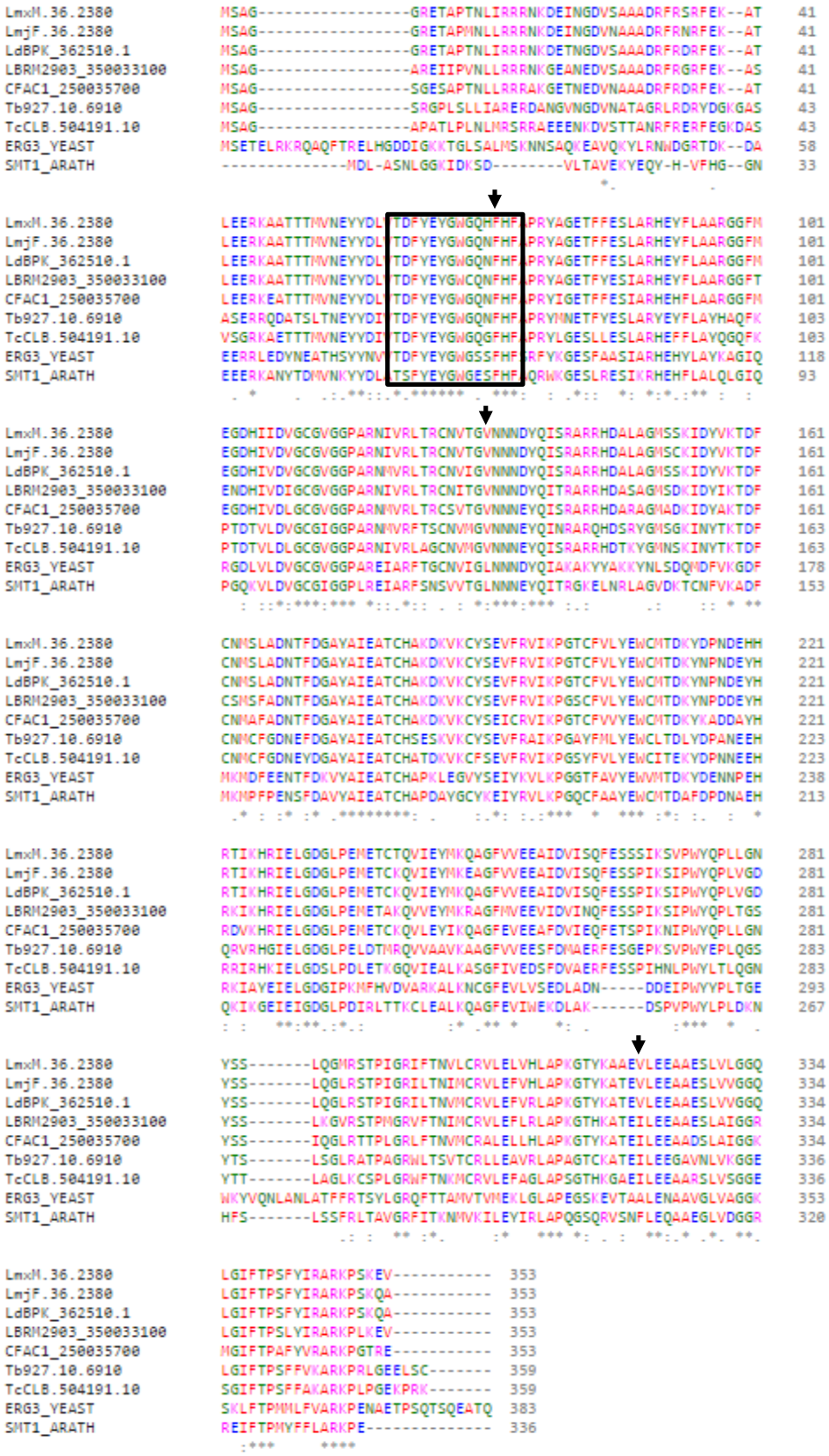


Figure 4-4: Alignment of sterol C24-methyltransferase genes from kinetoplastids and other species. Alignments were performed using Clustal Ω (Sievers *et al*. 2011). The position of the putative sterol binding site in the Erg6p enzyme from *S. cerevisiae* is marked with a box. Sites of variants described in the text are marked with arrows. Source species are, from top to bottom, *L. mexicana*, *L. major*, *L. donovani*, *L. braziliensis*, *C. fasciculata*, *T. brucei*, *T. cruzi*, *S. cerevisiae* and *A. thaliana* (the *A. thaliana* enzyme is cycloartenol C24-methyltransferase). Kinetoplastid sequences are derived from TriTrypDB, others from Uniprot.

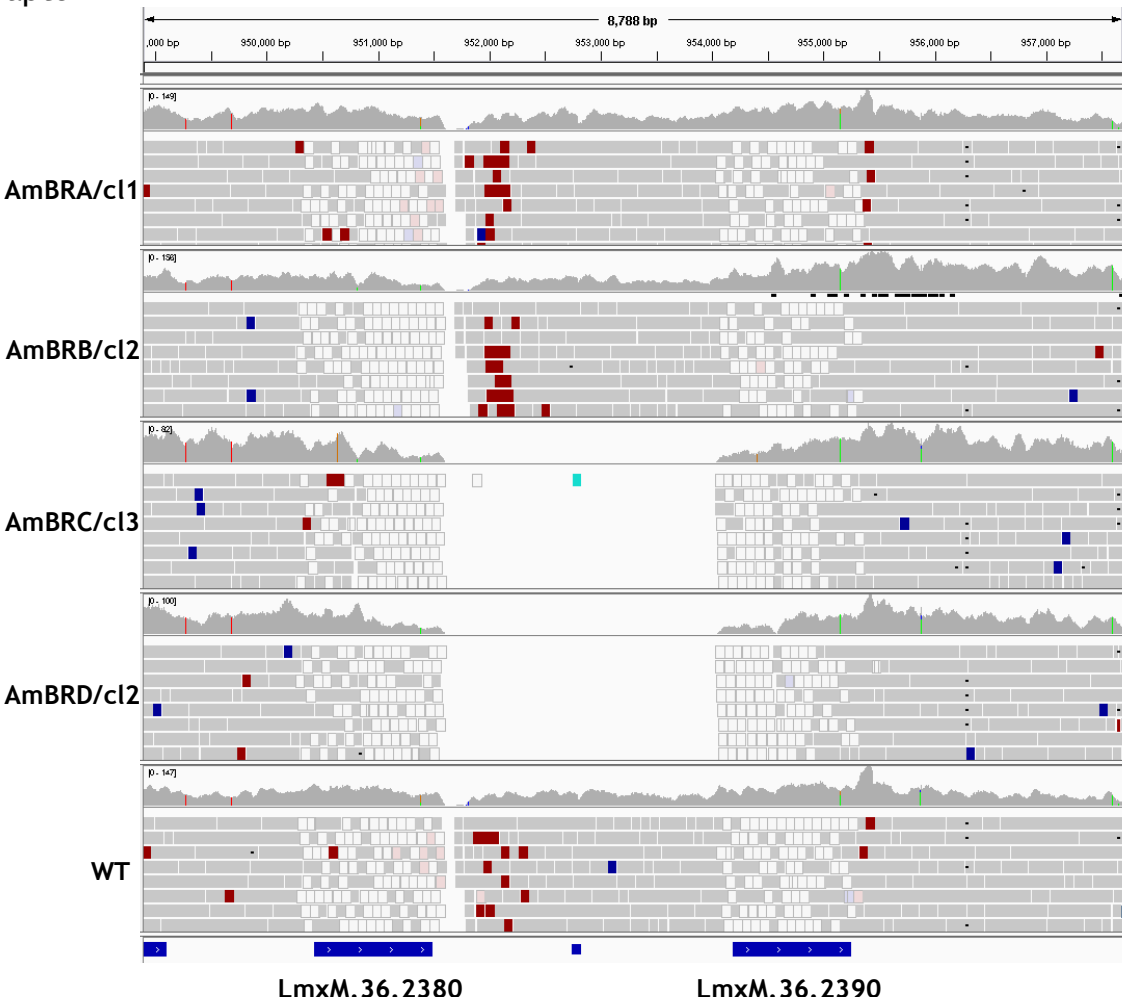


Figure 4-5: Visualisation of the alignment at the sterol C24-methyltransferase locus. The genomic region of *SMT* genes *LmxM.36.2380* and *LmxM.36.2390*, as visualised using the Integrative Genomics Viewer (Robinson *et al*. 2011). Grey blocks represent aligned reads with mapping quality greater than 0, white blocks with mapping quality of 0. Green and red lines on the coverage depth traces (top part of each panel) represent positions of variants. See Figure 4-2 for a full description of the visualisation.

both *LmxM.23.1300* and *SMT* genes, mutations were confirmed by gene amplification and Sanger sequencing.

Visualisation of the *SMT* genomic locus reveals a far more complex situation than for *LmxM.23.1300* (

Figure 4-5). In addition to the presence of SNPs, there is a total loss of coverage in the intergenic region between *SMT* copies for AmBRC/cl3 and AmBRD/cl2, indicative of possible structural variation. This is complicated by the presence of large numbers of reads with mapping quality of 0, as shown by the white boxes, within the coding sequences themselves. Because a mapping quality of 0 often relates to the presence of more than one equally suitable alignment sites, this is therefore a result of the close sequence homology between *LmxM.36.2380* and *LmxM.36.2390*. Another notable feature is the fact that there is a gap in the

alignment for all other lines just downstream of *LmxM.36.2380* within the intergenic region. The fact that reads do not align across this region even in wild-type cells is indicative of structural differences from the reference genome, either due to strain-strain differences or misassembly during reference genome construction. Therefore, whilst there is clear evidence of alterations in this region during the development of drug resistance, the presence of sequence repetition and potential problems with the reference genome make the exact nature of these changes difficult to interpret.

Coding changes identified are summarised in Table 4-1. The assignment of particular variants in *SMT* to particular gene copies is discussed further in chapter 5.

Table 4-1: Summary of genotypes at mutation sites in wild-type and resistant parasites. Nucleotide positions and genotypes are shown, with equivalent amino acid positions and substitutions given in brackets below.

	SC5D	***SMT***		
	415 (139)	215 (72)	391 (131)	961 (321)
WT	G/G (G/G)	T/T (F/F)	G/A (V/I)	G/A (V/I)
AmBRA/cl1	C/C (R/R)	T/T (F/F)	G/A (V/I)	G/A (V/I)
AmBRB/cl2	G/G (G/G)	T/T (F/F)	A/A (I/I)	A/A (I/I)
AmBRC/cl3	G/G (G/G)	G/G (C/C)	A/A (I/I)	A/A (I/I)
AmBRD/cl2	G/G (G/G)	T/T (F/F)	G/G (V/V)	A/A (I/I)

4.2.1.3 Analysis of gene copy number

Gene copy number variation is well established as a source of variability in *Leishmania* species, including in the context of drug resistance. I therefore investigated evidence of local per-gene copy number variation. This was done by calculating an initial per-gene haploid ratio for each strain (dividing the length-normalised coverage for each gene by the median normalised coverage for its chromosome), and then determining the fold change for this in comparison to

wild-type (Figure 4-6, Supplementary file 4-4). In order to eliminate false positives arising due to poor coverage of particular genes (which may arise due to incorrect assembly and can lead to stochastic variations being interpreted as large fold changes), I eliminated those genes which had a haploid ratio in wild-type cells of less than an arbitrary cutoff of 0.3 (i.e. length-normalised coverage of these genes was less than 30% of the median for that chromosome). This threshold eliminates very poorly covered genes, but retains low coverage due to, for example, heterozygous deletions in wild-type parasites (which would give a 50% decline in diploid chromosomes). I then defined genes with copy number variation as those which showed a fold change in haploid ratio of less than 0.7 or greater than 1.3. This revealed decreases in the haploid ratios of 18 genes, and increases in 66 genes.

Amongst genes showing reduced haploid ratios, most were distributed randomly, although six consecutive genes (*LmxM.08.0720-LmxM.08.0770*) all showed around half the haploid ratio of wild-type parasites in AmBRB/cl2. However, these were all annotated as “amastin-like protein, putative”), suggesting that they are a multicopy gene family, which inspection of the alignment appeared to show evidence of poor local assembly. Two loci were found to be of particular interest with regard to reduced haploid ratio. The first of these encompasses the two genes, *LmxM.13.1530* and *LmxM.13.1540*, and the second *SMT* genes *LmxM.36.2380* and *LmxM.36.2390*.

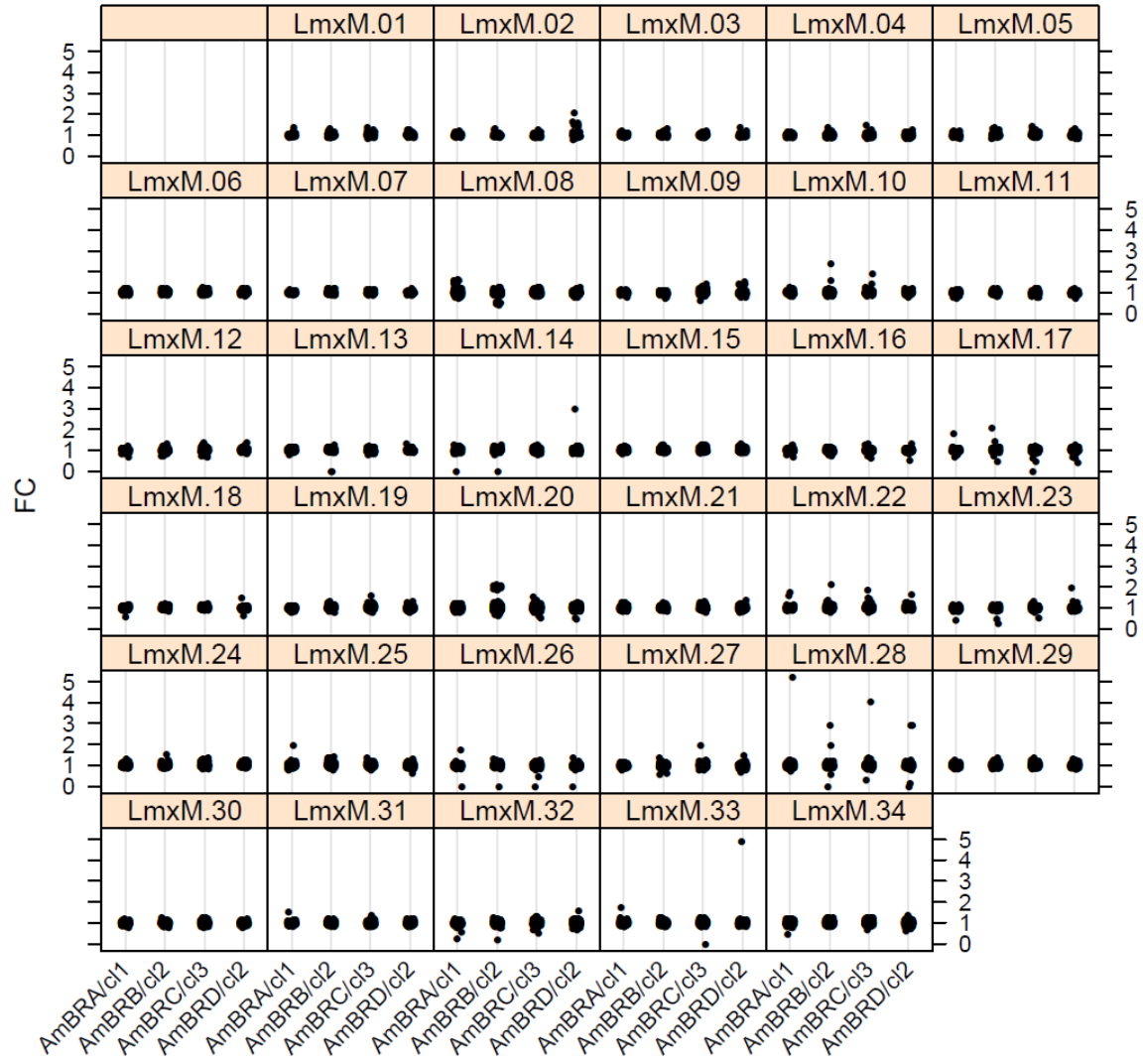


Figure 4-6: Dotplot of fold change (FC) in per-gene haploid ratio in AmB-resistant lines. Haploid ratio was calculated for each gene as the length-normalised coverage for that gene divided by the median length-normalised coverage across the whole chromosome (thus omitting effects of chromosomal CNV). Fold change was then calculated in comparison to wild-type, with poorly covered genes (wild-type haploid ratio < 0.3) omitted. Most genes cluster strongly around 1 (no change from wild-type – this will be most genes on the individual chromosome, generally hundreds of genes) with some genes, often in clusters, showing increases or decreases in individual lines.

Inspection of the former of these loci on chromosome 13 revealed clear evidence of a large deletion of approximately 8 kb encompassing both *LmxM.13.1530* and *LmxM.13.1540* (Figure 4-7). Lack of coverage across this region, as well as read pairs mapping across the large gap, provided clear evidence of this. Whilst *LmxM.13.1540* was annotated as a hypothetical protein without known function, *LmxM.13.1530* encodes the miltefosine transporter, which has previously been implicated in cross-resistance between MILT and AmB.

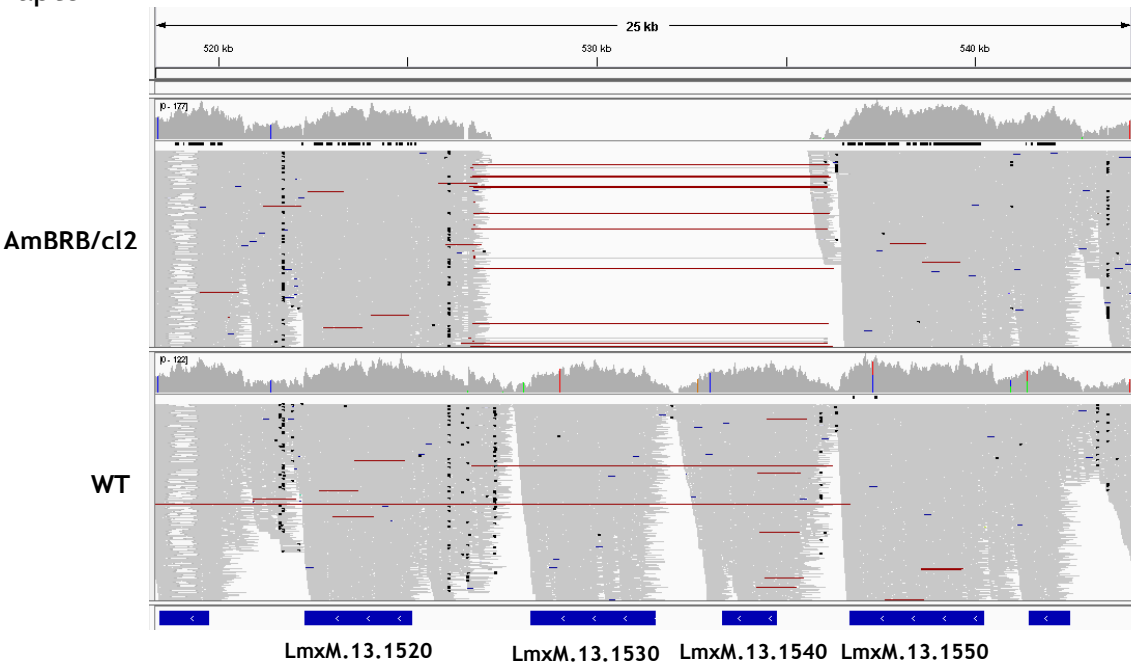


Figure 4-7: Visualisation of the deletion in chromosome 13. The genomic region of *LmxM.13.1530* and *LmxM.13.1540*, as visualised using the Integrative Genomics Viewer (Robinson *et al*. 2011). In this image, the reads are shown in a collapsed arrangement (more compact than in previous figures), and arranged as read pairs. Red lines indicate read pairs which map further apart than expected (based on expected fragment size). Note the large number of red lines mapping across the region of low coverage, indicating the presence of read pairs mapping across the putative deletion. See Figure 4-2 for a full description of the visualisation.

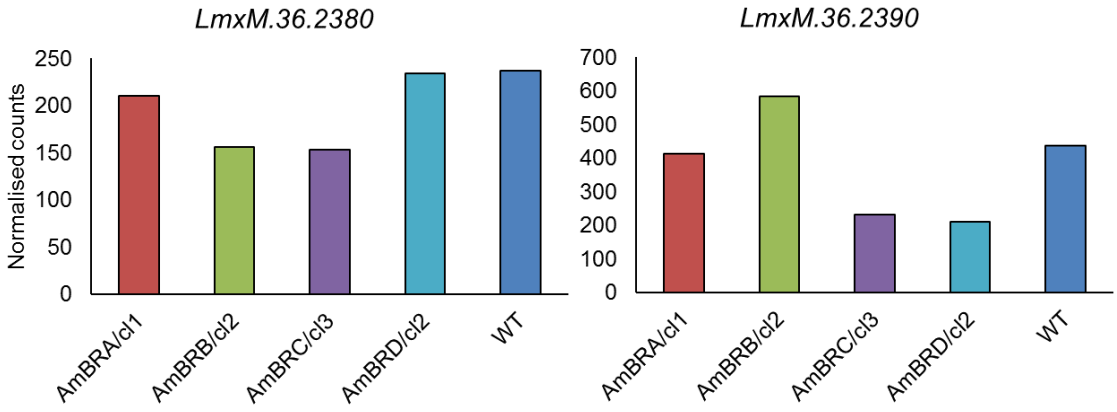


Figure 4-8: Coverage of *SMT* genes in WGS data. Per-gene coverage of *SMT* genes *LmxM.36.2380* and *LmxM.36.2390* was normalised to library size using size-factor normalisation in DESeq2 (Love *et al*. 2014).

By contrast, decreases in the *SMT* locus on chromosome 20 were less pronounced, but observed in a greater number of lines (Figure 4-8). While little change was observed in AmBRA/cl1, mixed changes were observed in the other three lines: in AmBRB/cl2, coverage was decreased in *LmxM.36.2380* but increased in *LmxM.36.2390*; AmBRC/cl3 saw decreased coverage in both genes whereas AmBRD/cl2 exhibited decreased coverage in *LmxM.36.2390* only.

In addition to genes where decreased haploid ratios were observed, a number of genes showed evidence of amplification. While some were isolated, most showed evidence of clustering, with multiple genes expanded in tandem. These putative multi-gene amplification events were detected on chromosomes two, eight, nine and 20 (Figure 4-9). The amplifications on chromosomes two, eight and nine, in AmBRD/cl2, AmBRA/cl1 and AmBRD/cl2, respectively, involved a 50% increase in haploid ratio, suggesting that the amplification occurred only on one out of two chromosome copies (although interestingly, chromosome two was predicted haploid in AmBRB/cl2 – therefore, this situation may be more complex, and mosaic ploidy changes may be involved). It is unclear exactly whether these changes occurred sporadically or were selected by drug pressure, and this is complicated by the likelihood that even if changes are selected, this is likely only due to selection for amplification of one gene, with the other genes simply increasing in copy as a secondary effect. It is notable, however, that the amplification in chromosome two encompassed *LmxM.02.0730*, which is annotated as a 'dehydrogenase/oxidoreductase-like protein'. This could feasibly have an influence either in metabolism or in defence against oxidative stress; however, no further information is available and the predicted protein is only 57 amino acid residues in length. Also of interest is the glycerophospholipid metabolism gene *LmxM.09.1040*, which is amplified on chromosome nine in AmBRD/cl2. This encodes a putative phospholipid:diacylglycerol acyltransferase.

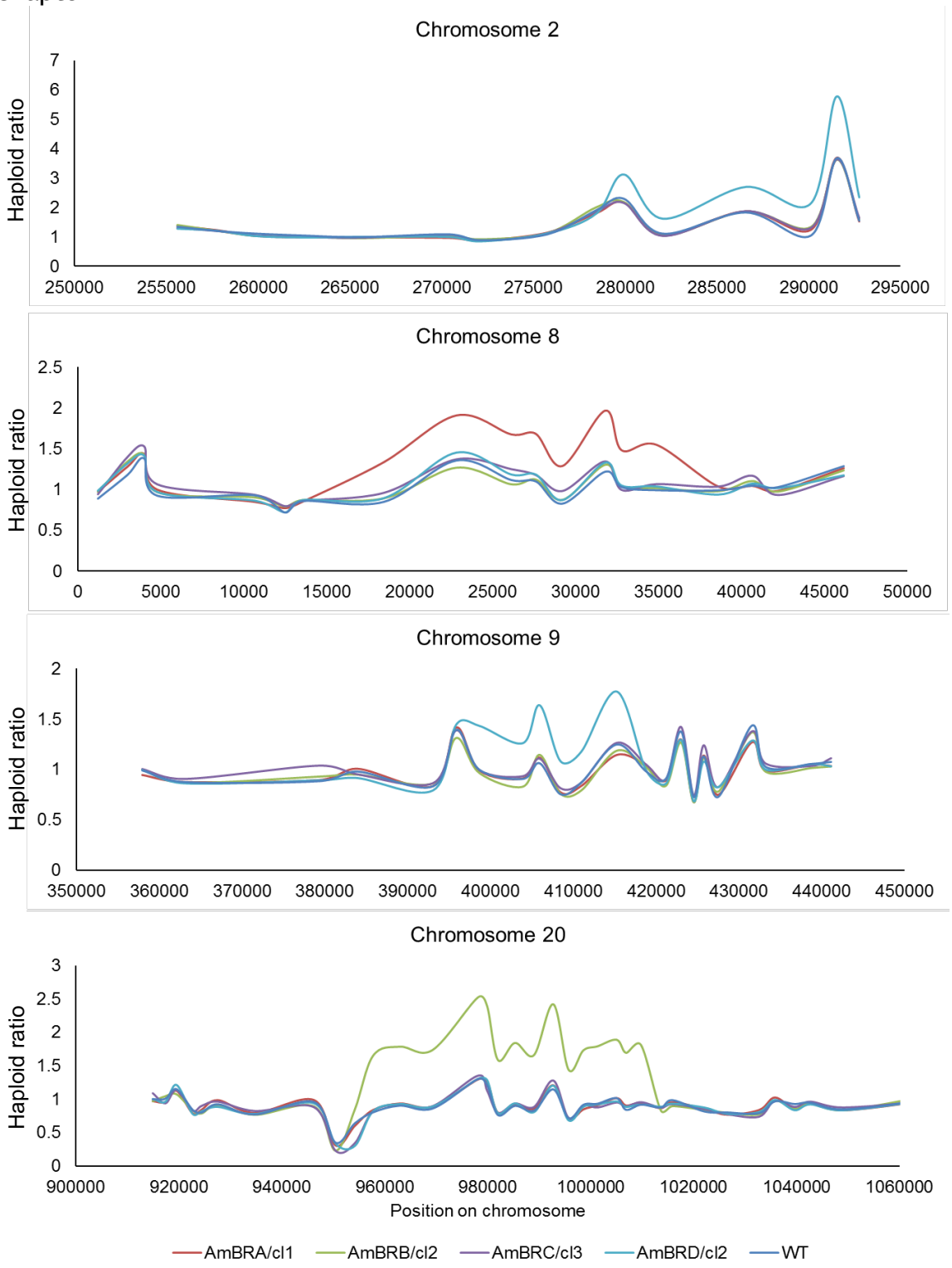


Figure 4-9: Multi-gene amplifications detected in WGS data. Per-gene haploid ratios are plotted against gene start position for four genomic regions where evidence of amplification is observed.

The largest degree of amplification, however, was observed in AmBRB/cl2 on chromosome 20, in which a region spanning 16 genes showed a twofold increase in haploid ratio. Genes within this putative amplified region are annotated with a number of functions, including a chromatin assembly factor (*LmxM.36.2500*)

and an RNA helicase (*LmxM.36.2530*), but most remarkably, the boundary for the amplified region falls in between the two *SMT* genes, *LmxM.36.2380* and *LmxM.2390*, with the former showing decreased coverage in AmBRB/cl2, and the latter, increased (Figure 4-8). Furthermore, the last gene in this amplicon, *LmxM.36.2540*, encodes the ergosterol biosynthesis enzyme sterol C4-methyloxidase, the protein which catalyses the first stage in C4-methyl removal from the sterol ring structure.

4.2.1.4 Identification of loss of heterozygosity events

During serial passage *in vitro*, *Leishmania* parasites are believed to divide ameiotically, in other words without any form of sexual reproduction. Therefore, there is a mechanistic question of how homozygous mutations arise in the absence of this form of recombination. One possible mechanism is through large-scale loss of heterozygosity (LOH) events, in which homologous recombination-mediated repair pathways lead to the replacement of one haplotype with another. I compared the position of heterozygous sites between wild-type parasites and the lines carrying these mutations, AmBRA/cl1 and AmBRB/cl2, for chromosomes 23 and 13, respectively (Figure 4-10). This revealed that the majority of heterozygous sites appeared to be shared between AmB-resistant and wild-type parasites. However, in both cases, regional depletion of heterozygous sites present in wild-type parasites was evident in resistant lines. Intriguingly, these regions in both cases extended from one end of the chromosome inwards, to around the point where *LmxM.23.1300* and *LmxM.13.1530*, respectively, were located. These data imply that homozygous changes to both genes were associated with extensive LOH events. As a result of these events, homozygous non-synonymous changes in the region were also associated. In AmBRA/cl1, these were in *LmxM.23.1310* (dynein heavy chain, putative), *LmxM.23.1330* (DNA polymerase zeta catalytic subunit) and *LmxM.23.1540* (hypothetical protein). In AmBRB/cl2, these were *LmxM.13.1570* (phosphoprotein phosphatase-like protein), *LmxM.13.1640* (mitogen-activated protein kinase 7) and *LmxM.13.1680* (pyrroline 5-carboxylate reductase). All of these variants were heterozygous in all other lines.

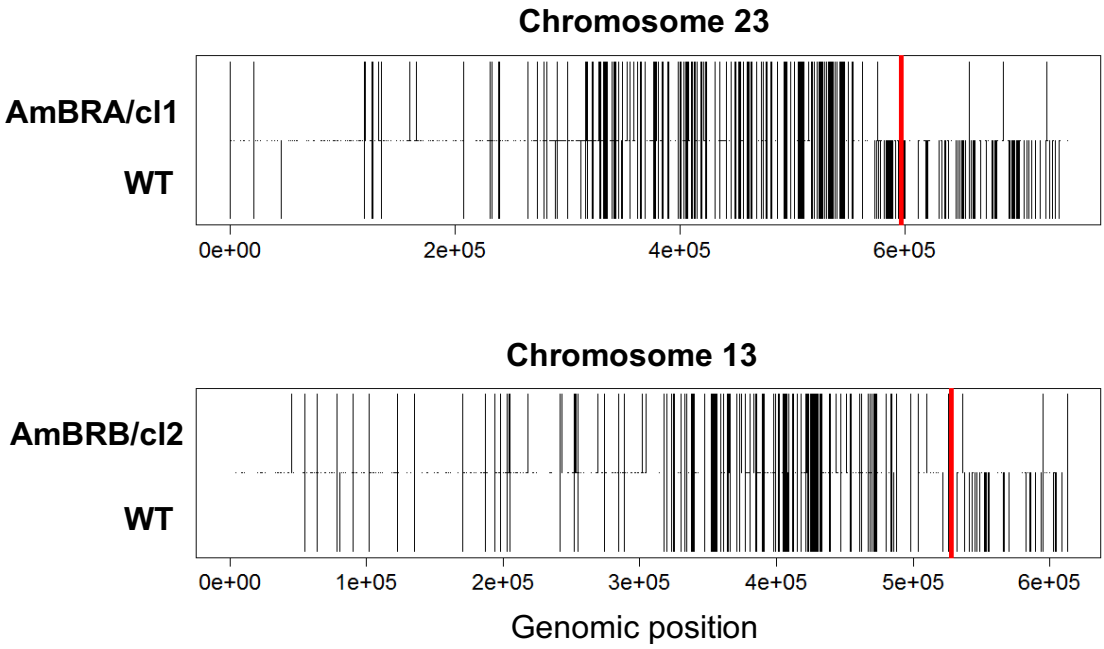


Figure 4-10: Loss of heterozygosity events associated with homozygous variants. Putative LOH events are visualised by plotting all heterozygous sites (black vertical lines) throughout chromosomes 23 and 13 for wild-type (bottom part of plot) and AmB-resistant lines (top part of plot, AmBRA/cl1 for chromosome 23, bottom part AmBRB/cl2 for chromosome 13). The positions of *LmxM.23.1300* (top) and *LmxM.13.1530* (bottom) are marked with vertical red lines.

4.2.2 Determination of transcriptome changes by RNA-seq

While WGS is highly useful in detecting alterations to gene function through mutations or deletions in coding sequences, it is challenging to interpret the consequences of such changes for gene expression. Therefore, I aimed to assess whether altered RNA abundance levels play a role in acquisition of AmB resistance, as well as associated phenotypic and metabolic changes, through RNA-seq transcriptomics (alignment statistics are provided in Supplementary file 4-1). In order to capture maximum biological variability three clones from each resistant line were sequenced, with RNA from AmBRA/cl1, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 prepared from the same flask as the DNA used for WGS. Three biological replicates of wild-type parental parasites were included for comparison. PCA plotting of data showed clear grouping of clones from independently selected lines (Figure 4-11), suggesting strong similarity between them, as indicated by their similar AmB sensitivities (see section 3.2.1).

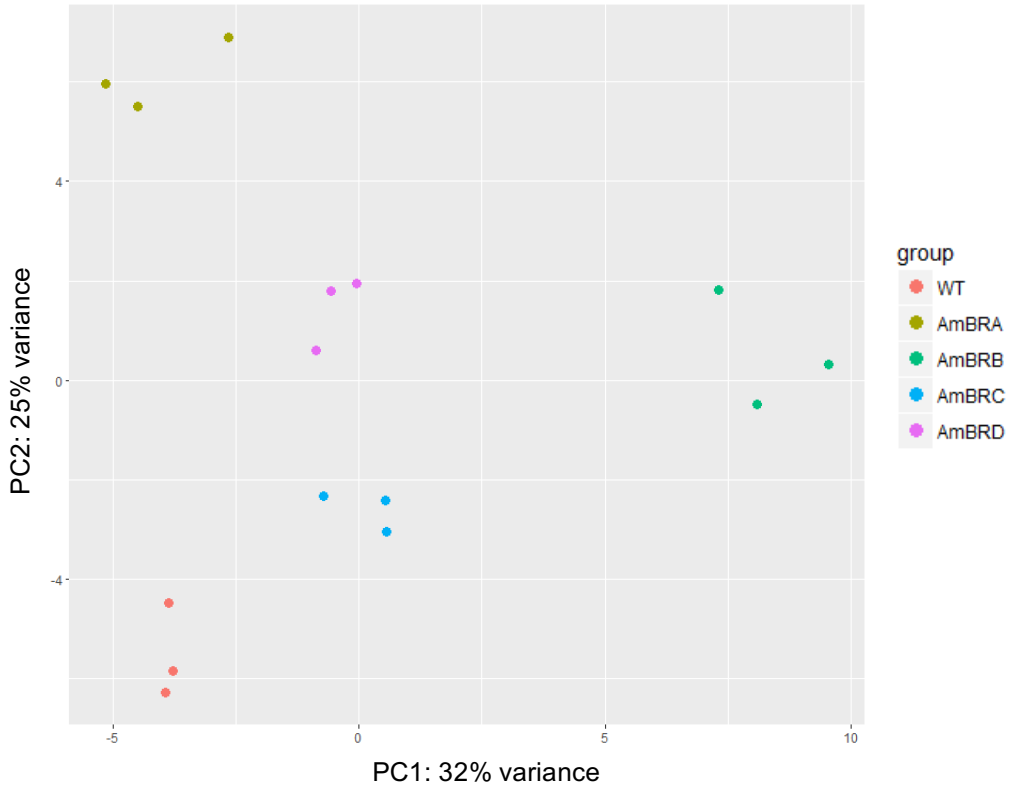


Figure 4-11: PCA plot of RNA-seq data. Generated using the R package DESeq2 (Love *et al.* 2014).

In order to detect variations in RNA abundance, I performed differential expression analysis using DESeq2 (Love *et al*. 2014). This revealed a large number of statistically significant changes ($P < 0.05$), with 2,016 in AmBRA, 3,411 in AmBRB, 1,066 in AmBRC and 1,570 in AmBRD (Supplementary file 4-5 for normalised counts, Supplementary file 4-6 for full expression analysis). Plots of mean normalised counts against fold change indicate a fairly tight distribution across both axes, with normalised counts per gene clustering tightly around the 10^3 mark (Figure 4-12). Despite the large number of differentially expressed genes (ranging from 13% up to 41% of all detected coding sequences in AmBRC and AmBRB, respectively), the vast majority of these are less than twofold ($-1 < \log_2 FC < 1$).

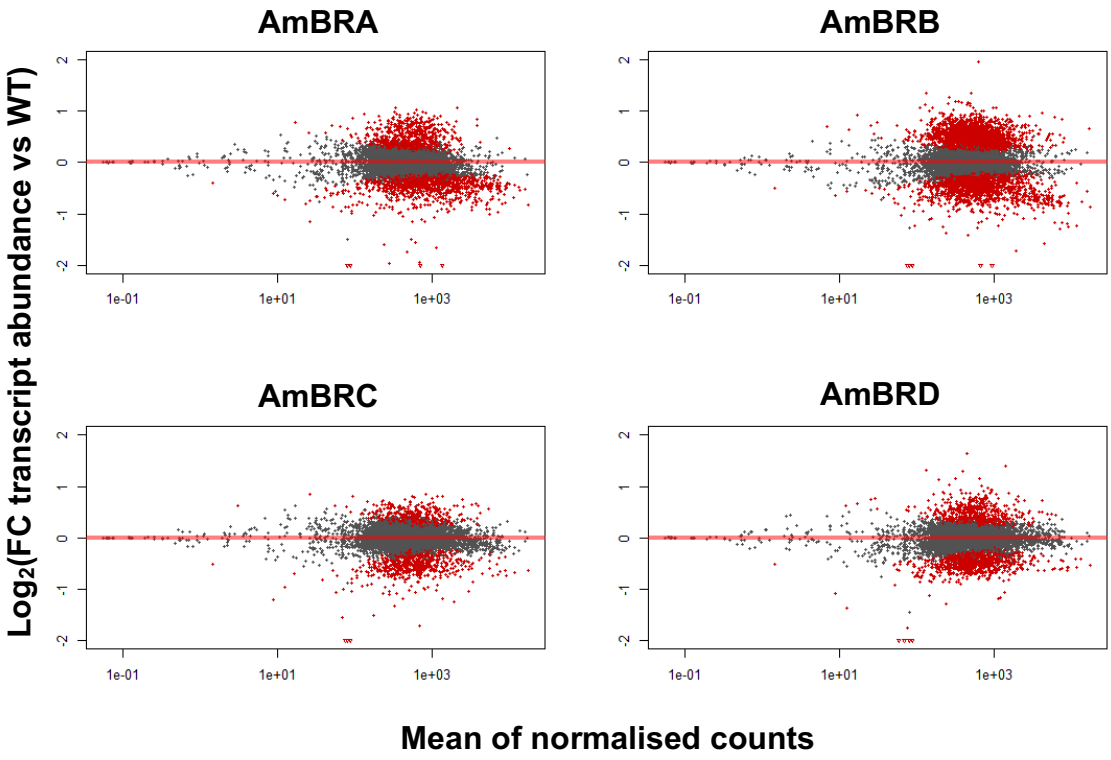


Figure 4-12: Distribution of fold change in transcript abundance in resistant lines compared to wild-type parasites. Log_2(fold change) is plotted against the mean normalised counts for each gene (shown here on a log scale). Fold change is between resistant and wild-type lines. Red dots indicate statistical significance ($P < 0.05$). Plots were produced using DESeq2 (Love *et al*. 2014).

4.2.2.1 Correlation between RNA expression and chromosome copy number

Given the large number of significant differences with low fold-changes, I investigated whether these arose due to changes in chromosomal ploidy. Figure 4-13 shows the result of plotting fold-changes in RNA abundance in comparison to ploidy changes derived from the WGS data. This reveals evidence of a relationship between changes in chromosomal ploidy state and changes in RNA abundance. In each case I performed a Kruskal-Wallis test to determine whether ploidy influenced RNA-seq data, as well as a pairwise Mann-Whitney-Wilcoxon test comparing RNA fold-changes where ploidy was unaltered compared to wild-type and those where ploidy was either increased or decreased. In all cases, there were highly significant differences ($P < 10^{-150}$). The distributions clearly show that other factors play an important role, and coercing ploidy changes to categorical variables may not properly represent cases of mosaic or intermediate ploidy changes. Nevertheless, ploidy clearly has a strong influence over RNA abundance in all lines.

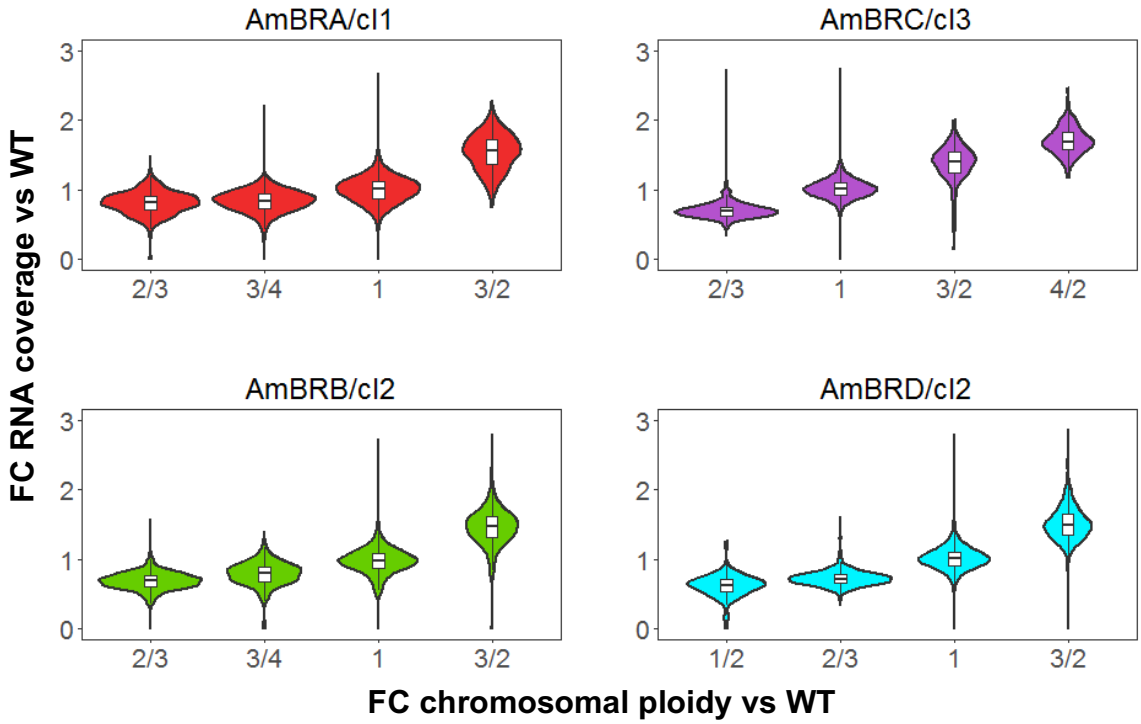


Figure 4-13: Correlation between chromosomal ploidy changes and fold-change in RNA expression in resistant lines compared to wild-type parasites. Changes in integer ploidy values (Figure 4-1B) and fold-change in RNA abundance were compared to wild-type for each resistant line. Violin plots indicate the distribution of RNA fold-changes for each type of ploidy change, with box plots overlaid showing the median and lower and upper quartiles in each case. RNA fold-change values are calculated as the normalised fragment counts for individual clones divided by the mean normalised fragment count for wild-type parasites across three biological replicates for this line. Plots were produced using the R graphics package, ggplot2 (Wickham 2009).

4.2.2.2 Differentially expressed genes with the highest fold changes

Given the very large number of genes whose differential expression was statistically significant, I decided to focus initially just on those which had a greater than twofold difference in abundance ($\log_2(FC) < -1$ or $\log_2(FC) > 1$). Note that "fold change" as expressed here refers to the expression in the resistant line divided by the expression in wild-type parasites. Therefore, a twofold increase is described as a fold-change of two, whereas a twofold decrease is described as a fold-change of 0.5. This cutoff greatly reduced numbers: 22 genes in AmBRA (19 decreases, three increases), 55 in AmBRB (33 decreases, 22 increases), 22 in AmBRC (all decreases) and 21 in AmBRD (12 decreases, nine increases). Many of these genes included hypothetical proteins, and interestingly two genes encoding the *Leishmania* virulence factor GP63 (*LmxM.10.0390* and *LmxM.10.0470*) exhibited decreased expression in all four

lines. In AmBRA, an ABC transporter-like protein, *LmxM.33.0670*, showed a 0.45-fold change in expression. While increased ABC transporter expression has been previously implicated in AmB resistance (Purkait *et al*. 2012), the transporter described in that case, MDR1 (*LmxM.33.0990*), was unchanged in this dataset, and the change in *LmxM.33.0670* was a decrease. Also of interest was a twofold increase in expression of the signalling factor calmodulin (*LmxM.09.0910* and *LmxM.09.0920*) in AmBRB. However, the most interesting changes were in genes associated with lipid metabolism and translation, and these are discussed in the following sections.

4.2.2.3 Expression changes in sterol metabolism genes

I identified 11 genes involved in ergosterol biosynthesis (section 1.2.4) that showed significant changes in expression between wild-type and at least one resistant line, with a $\log_2$(FC) cutoff of ± 0.5 (less than 0.7-fold or greater than 1.4-fold). Amongst these genes, a mixture of increases and decreases were observed (Figure 4-14). Within the early part of the pathway (L-leucine to lanosterol), both increases and decreases were observed in components of the branched-chain alpha-keto acid dehydrogenase complex (*LmxM.31.3310*, *LmxM.08_29.1830* and *LmxM.21.1430*). Increases were observed in *LmxM.15.1460* (phosphomevalonate kinase-like protein) and *LmxM.18.0020* (diphosphomevalonate decarboxylase, putative) in AmBRA and AmBRB, and AmBRC and AmBRD, respectively. *LmxM.13.1620* (squalene monooxygenase-like protein) also revealed increases in AmBRA and AmBRB. With respect to later stages in the pathway, however, decreases were more dominant. In particular, *LmxM.31.2320* (C14-sterol reductase) showed decreased expression in AmRB, AmBRC and AmBRD, and *LmxM.34.1230* (short chain dehydrogenase, putative) was decreased in AmBRA and AmBRB. In AmBRB, by contrast, a strong (3.9-fold) increase was observed in *LmxM.36.2540* (C4-sterol methyloxidase). Overall, therefore, some changes are present but are spread out between different resistant lines, limiting the conclusions that can be drawn about pathway regulation.

Most notable of all, however, were changes to the two *SMT* genes, *LmxM.36.2380* and *LmxM.36.2390* (Figure 4-15A). In the former case, a strong decrease in expression was observed in AmBRB, AmBRC and AmBRD, with

expression only 0.14-fold, 0.15-fold and 0.30-fold that of wild-type, respectively. In the case of the latter, decreases were only observed in AmBRC and AmBRD, with expression 0.60-fold and 0.24-fold that of wild-type. It should be noted, however, that almost identical coding sequences and potential problems with respect to the reference genome in this region (Figure 4-5) mean that it is difficult to differentiate adequately between the two gene copies. Crucially, there is an evident decrease in *SMT* transcript abundance across AmBRB, AmBRC and AmBRD, although it appears that, particularly for AmBRB, one gene copy may be more affected than the other.

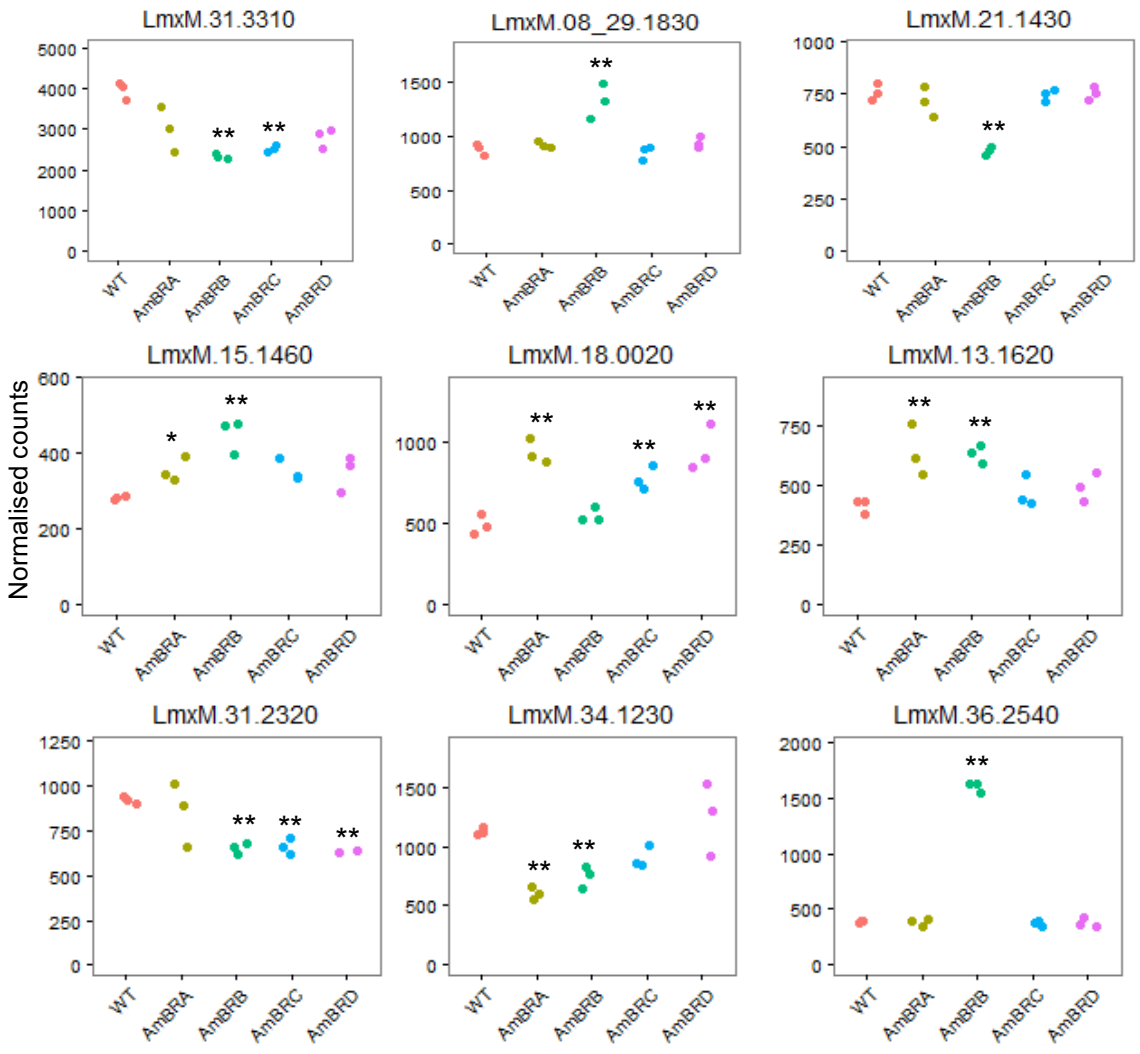


Figure 4-14: Differentially expressed genes in the sterol biosynthesis pathway. Normalised counts are plotted, with each dot representing an individual clone from each resistant line (and independent biological replicates for wild-type cells). Genes are plotted in the order of pathway appearance. Asterisks represent changes that are statistically significant (*P < 0.05, **P < 0.005). Genes are as follows: *LmxM.31.3310*, *LmxM.08_29.1830*, *LmxM.21.1430*: subunits of branched-chain alpha-keto acid dehydrogenase complex; *LmxM.15.1460*: phosphomevalonate kinase-like protein; *LmxM.18.0020*: diphosphomevalonate decarboxylase, putative; *LmxM.13.1620*: squalene monooxygenase-like protein; *LmxM.31.2320*: C14-sterol reductase; *LmxM.34.1230*: short chain dehydrogenase, putative; *LmxM.36.2540*: C4-sterol methyloxidase.

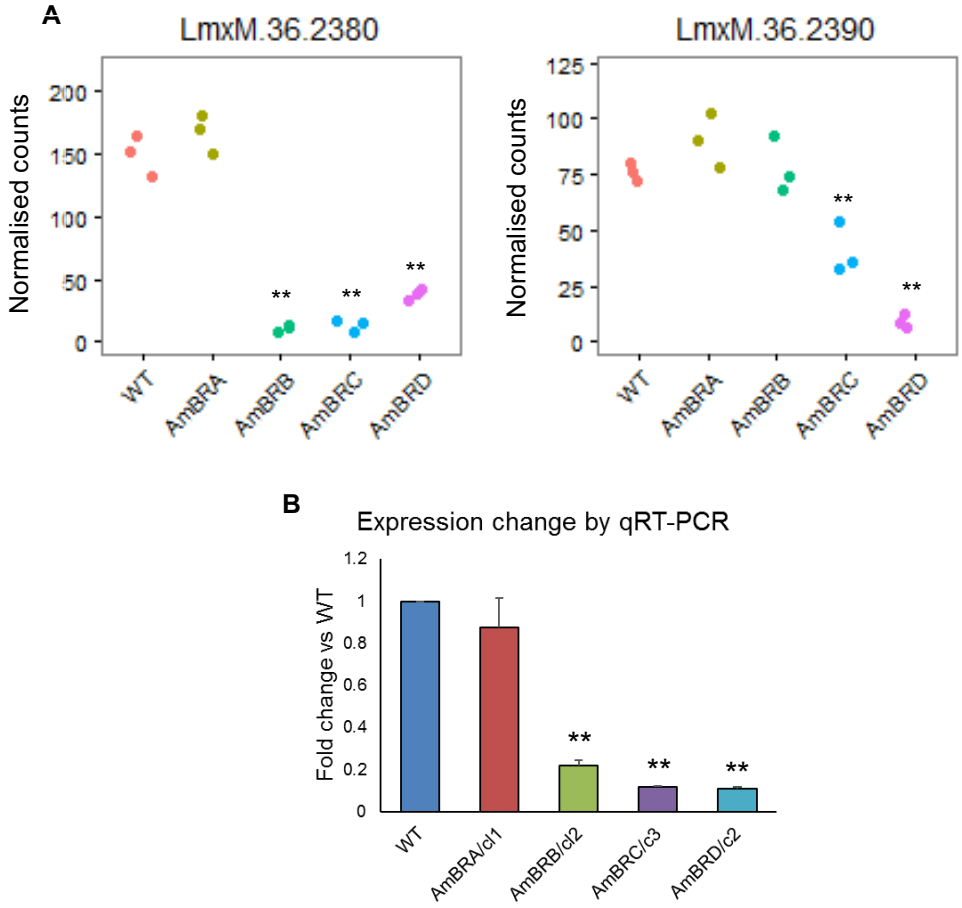


Figure 4-15: Expression changes in sterol C24-methyltransferase. A) Expression changes in *SMT* genes *LmxM.36.2380* and *LmxM.36.2390*, as determined by RNA-seq. Asterisks represent statistically significant differences in expression from wild-type (*P < 0.05, **P < 0.05). B) Fold change in expression as determined by qRT-PCR using primers that equally bind to each *SMT* coding sequence. Fold changes were calculated as $2^{(-\delta\delta Ct)}$, with mean values across three independent biological replicates (separate RNA extractions) shown. Error bars represent standard deviation. Asterisks indicate statistically significant differences in δCt from wild-type (*P < 0.05, **P < 0.005).

In order to confirm the decreased expression observed in the RNA-seq data, I used qRT-PCR (with cytosolic GAPDH, *LmxM.36.2350*, as a loading control – see section 2.4.5 for justification). As the coding sequences are so similar, I did not attempt at this stage to differentiate between *SMT* genes but simply measured overall expression, using primers that targeted sites identical between copies (Figure 4-15B). This confirmed that expression was unchanged in AmBRA/cl1 (P = 0.302), 0.88-fold that of wild-type. By contrast, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 all showed large decreases in expression, with fold-changes of 0.22 ($P = 4.68 \times 10^{-6}$), 0.12 ($P = 9.70 \times 10^{-6}$) and 0.11 ($P = 1.58 \times 10^{-6}$) that of wild type. I aimed to confirm this expression change at the protein level through western blot (Supplementary file 4-7). However, whilst a band was identified of the appropriate molecular weight that was lost in AmBRB/cl2, AmBRC/cl3 and

AmBRD/cl2, the signal was very poor with high amounts of noise, making it not possible to quantify.

4.2.2.4 Expression changes in other areas of lipid metabolism

The metabolomics data presented in the previous chapter reveal extensive changes in lipid metabolism beyond sterol biosynthesis, including in phospholipid and sphingolipid metabolism. In order to determine whether these changes had a transcriptional basis, I examined expression differences in genes that were annotated with the gene ontology (GO) terms GO:0006629 (lipid metabolic process) GO:0006633 (fatty acid biosynthetic process) and GO:0008654 (phospholipid biosynthetic process). After removing two sterol biosynthesis genes included in this list (*LmxM.23.1300* and *LmxM.36.2540*), this yielded 47 genes, of which 14 showed significant expression changes with a $\log_2$(FC) of ± 0.5 (less than 0.7-fold or greater than 1.4-fold) in at least one resistant line. Interestingly, of these genes, four were annotated as lipases (*LmxM.08_29.1260*, *LmxM.13.0200, LmxM.13.0210*, *LmxM.18.0160*, *LmxM.33.1070*). Changes were mixed in these genes, with increases reported in *LmxM.08_29.1260*, *LmxM.13.0200, LmxM.13.0210* and *LmxM.18.0160*, and *LmxM.33.1070* decreased in all four lines. Changes were also reported in desaturase genes: expression of *LmxM.10.1320* (fatty acid desaturase, putative) decreased in all four lines (although with a $\log_2$(FC) of only -0.4 in AmBRC), whereas there was a moderate but significant increase in *LmxM.26.1670* (sphingolipid delta 4 desaturase, putative) only in AmBRD. Other changes include decreased expression in *LmxM.03.0080* (glycerol-3-phosphate acyl transferase), *LmxM.03.0821* (choline/ethanolamine phosphotransferase, putative) and *LmxM.14.1360* (myo-inositol-1-phosphate synthase, decreased in all four lines, greater than twofold in AmBRD), as well as increases in *LmxM.09.1040* (phospholipid:diacylglycerol acyltransferase, putative), *LmxM.18.0810* (ethanolamine phosphotransferase, putative; this also showed a significant decrease in AmBRB, $\log_2$(FC) of -0.42), *LmxM.17.1100* (3-oxo-5-alpha-steroid 4-dehydrogenase-like protein) and *LmxM.28.2100* (glycerophosphoryl diester phosphodiesterase, putative). Most of these changes were only observed in one or two lines, with general trends only observed in *LmxM.10.1320* and *LmxM.33.1070*. One gene that was not annotated by GO term as involved in lipid metabolism but appears to have a related function, was *LmxM.24.1810*. This was one of 22 genes in AmBRC which showed

a greater than twofold decrease, and additionally showed 0.65-fold wild-type expression in AmBRB. This encoded a short chain dehydrogenase/reductase, whose orthologue in *T. brucei*, *Tb927.8.6420*, encodes beta-ketoacyl-ACP reductase 2, a fatty acid synthesis component in that organism.

I also sought to determine whether alterations in expression of phospholipid transporters were observed, using the GO term GO:0015914. The only gene that showed significant expression changes was *LmxM.13.1530*, the miltefosine transporter. Interestingly, not only did this show loss of expression in AmBRB (as suggested by the deletion described in section 4.2.1.3), but expression was also increased (1.64-fold) in AmBRA. Given that AmBRA/cl1 was observed to be MILT-hypersensitive, I performed qRT-PCR to confirm these changes. Whilst this confirmed loss of expression in AmBRB/cl2 ($P = 7.32 \times 10^{-6}$), it did not confirm an increase in AmBRA/cl1 ($P = 0.555$). By contrast, increased expression was observed in AmBRC/cl3 and AmBRD/cl2 (1.3-fold with $P = 0.0165$ and 1.4-fold with $P = 0.00105$, respectively). Therefore, whilst loss of LmxM.13.1530 expression is clearly observable for AmBRB across both methods, subtler changes in the other AmB-resistant lines were not consistent.

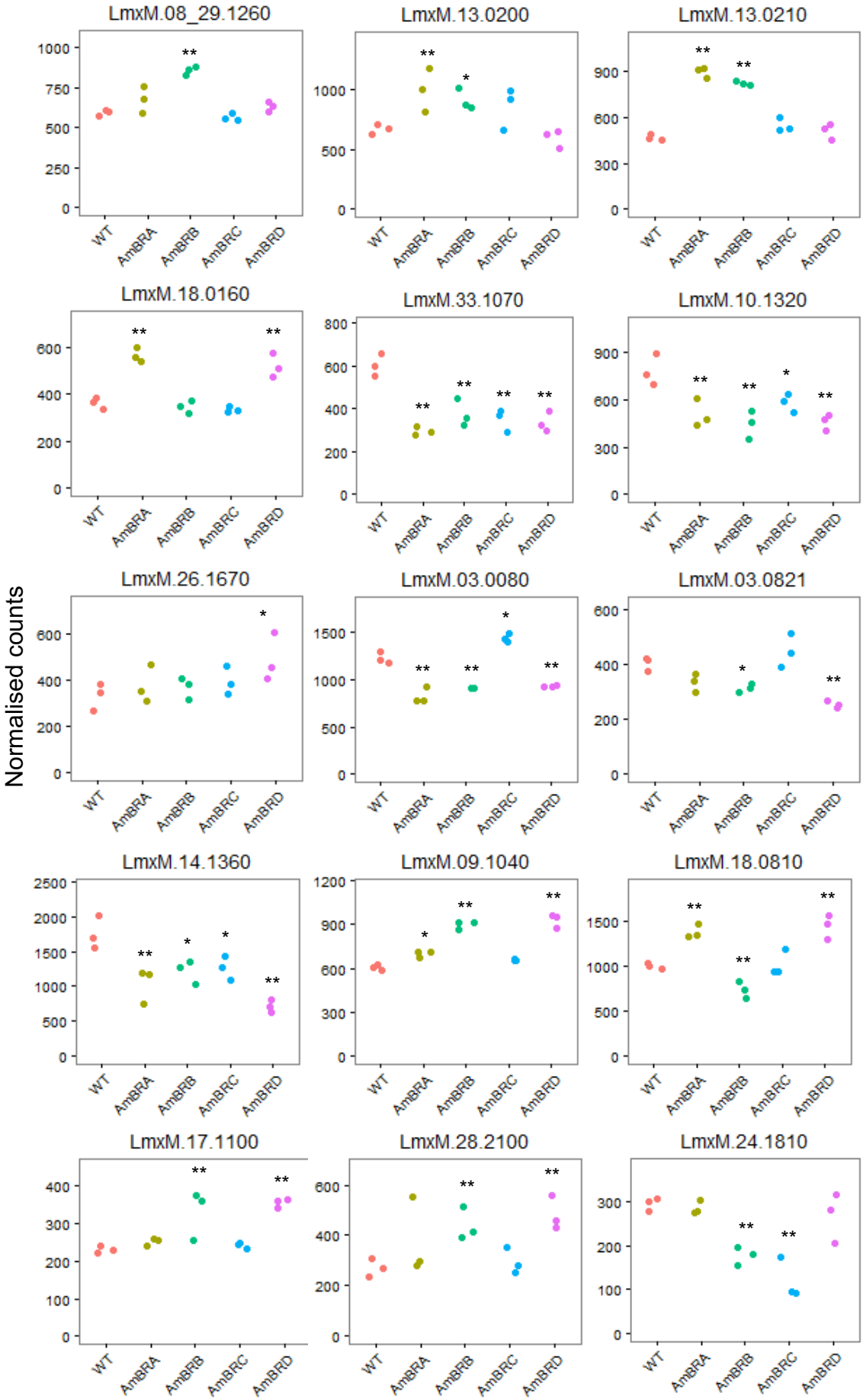

LmxM.08_29.1260
LmxM.13.0200
LmxM.13.0210
LmxM.18.0160
LmxM.33.1070
LmxM.10.1320
LmxM.26.1670
LmxM.03.0080
LmxM.03.0821
LmxM.14.1360
LmxM.09.1040
LmxM.18.0810
LmxM.17.1100
LmxM.28.2100
LmxM.24.1810
Normalised counts
WT
AmBRA
AmBRB
AmBRC
AmBRD

Figure 4-16: Other expression changes in lipid metabolism. Normalised counts are plotted, with each dot representing an individual clone from each resistant line (and independent biological replicates for wild-type cells). Genes are plotted in the order of mention within the text. Asterisks represent statistically significant differences in expression from wild-type (*P < 0.05, **P < 0.005). Genes are as follows: *LmxM.08_29.1260*: Wnt-binding factor required for Wnt secretion/Lipase (class 3), putative; *LmxM.13.0200*: lipase (class 3), putative; *LmxM.13.0210*: triglyceride lipase, putative; *LmxM.18.0160*: lipase (class 3), putative; *LmxM.33.1070*: lipase (class 3), putative; *LmxM.10.1320*: fatty acid desaturase, putative; *LmxM.26.1670*: sphingolipid delta 4 desaturase, putative; *LmxM.03.0080*: glycerol-3-phosphate acyl transferase; *LmxM.03.0821*: choline/ethanolamine phosphotransferase, putative; *LmxM.14.1360*: myo-inositol-1-phosphate synthase; *LmxM.09.1040*: phospholipid:diacylglycerol acyltransferase, putative; *LmxM.18.0810*: ethanolamine phosphotransferase, putative; *LmxM.17.1100*: 3-oxo-5-alpha-steroid 4-dehydrogenase-like protein; *LmxM.28.2100*: glycerophosphoryl diester phosphodiesterase, putative; *LmxM.24.1810*: short chain dehydrogenase/reductase, putative.

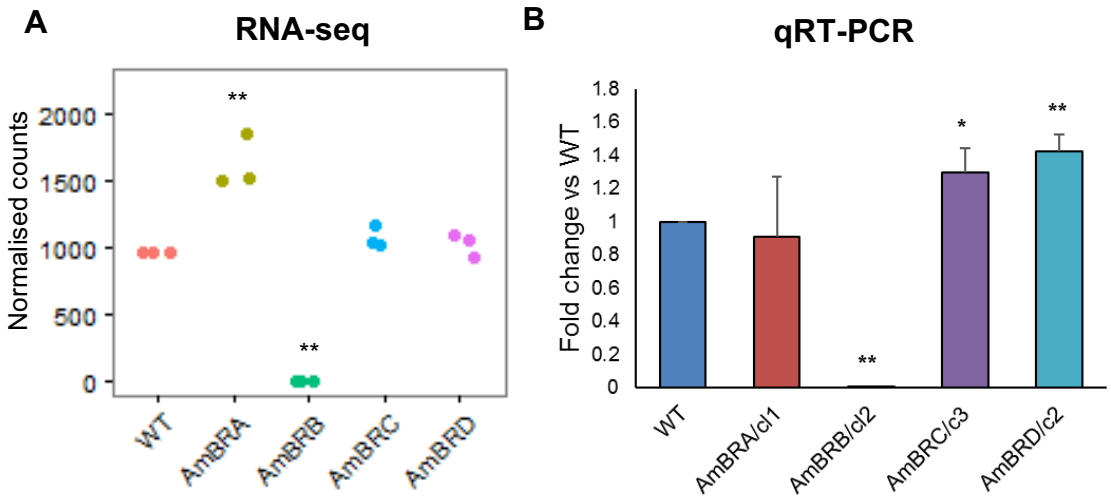


Figure 4-17: Expression changes in the miltefosine transporter, *LmxM.13.1530*. A) Expression changes as determined by RNA-seq. Normalised counts are plotted, with each dot representing an individual clone from each resistant line (and independent biological replicates for wild-type cells). Asterisks represent statistically significant differences in expression from wild-type (*P < 0.05, **P < 0.005). B) Fold change in expression as determined by qRT-PCR. Fold changes were calculated as $2^{(-\delta\delta Ct)}$, with mean values across three independent biological replicates (separate RNA extractions) shown. Error bars represent standard deviation. Asterisks indicate statistically significant differences in δCt from wild-type (P < 0.05).

4.2.2.5 Expression changes in polyamine and oxidative stress metabolism, and arginine and proline metabolism

AmB resistance has previously been linked to altered expression of genes involved in oxidative stress metabolism (Moreira *et al*. 2011, Purkait *et al*. 2012, Kumar *et al*. 2014). Therefore, I assessed whether changes in RNA abundance could be detected in 33 genes involved in polyamine, trypanothione, and glutathione metabolism, as well as ROS detoxification (Figure 1-5). Of these, 13 revealed significant changes with a $\log_2$(FC) of at least ± 0.5. In polyamine

biosynthesis, arginase (*LmxM.34.1480*) exhibited decreased expression in AmBRA and AmBRB (0.64-fold and 0.73-fold wild-type expression, respectively), whereas ornithine decarboxylase (*LmxM.12.0280*) showed 0.68-fold and 0.69-fold wild-type expression in AmBRB and AmBRD respectively. Spermidine synthase (*LmxM.04.0580*) exhibited decreased expression in all four lines, with fold changes of 0.62, 0.54, 0.64 and 0.73 in AmBRA, AmBRB, AmBRC and AmBRD, respectively. Hence there is evidence of decreased expression of genes in polyamine metabolism. On the other hand, moderate increases in expression of trypanothione reductase (*LmxM.05.0350*) were observed in AmBRB and AmBRC (1.54-fold and 1.55-fold).

Trypanothione is a hydride donor used in ROS detoxification by tryparedoxin peroxidases. Of five genes annotated as tryparedoxin peroxidases (*LmxM.15.1040*, *LmxM.15.1160*, *LmxM.23.0040*, *LmxM.26.0800* and *LmxM.26.0810*), only one (*LmxM.15.1160*) showed a significant difference with a $\log_2$(FC) of at least ± 0.5, a 0.69-fold change in AmBRA. Interestingly, AmBRB exhibited increased expression of this gene, with expression 1.35-fold that of wild-type parasites. By contrast, of six genes annotated as encoding tryparedoxins (*LmxM.08_29.1150*, *LmxM.08_29.1160*, *LmxM.30.1960* and *LmxM.30.1970*) or tryparedoxin-like proteins (*LmxM.08_29.1130* and *LmxM.08_29.1140*), more widespread changes were observed, with increases in expression of tryparedoxin-like protein *LmxM.08_29.1130* in AmBRA, AmBRB and AmBRD (of 1.27-fold, 1.70-fold and 1.25-fold, respectively) as well as in tryparedoxin 1 (*LmxM.08_29.1160*) in AmBRB (1.43-fold). On the other hand, expression of tryparedoxin 4 (*LmxM.30.1970*) was decreased in AmBRA, AmBRB and AmBRD (0.67-fold, 0.81-fold and 0.69-fold, respectively). However, it should be noted that in the case of multicopy gene families it is unclear how efficiently read mapping can distinguish between different copies, due to sequence similarity.

Besides the trypanothione/tryparedoxin system, other changes were noted particularly in superoxide dismutase genes. Iron superoxide dismutase *LmxM.08.0290* revealed decreased expression in AmBRA (0.69-fold), with smaller but still statistically significant decreases and increases in AmBRC and AmBRB (0.84-fold and 1.17-fold). Both iron superoxide dismutases *LmxM.31.1820* and

LmxM.31.1830 were universally decreased, with changes being more pronounced in *LmxM.31.1830* (0.55-fold, 0.51-fold, 0.59-fold and 0.62-fold in AmBRA, AmBRB, AmBRC and AmBRD, respectively) than in *LmxM.31.1820* (0.80-fold, 0.67-fold, 0.65-fold and 0.78-fold in AmBRA, AmBRB, AmBRC and AmBRD, respectively). Another superoxide dismutase (*LmxM.31.2630*) also showed decreases in AmBRB, AmBRC and AmBRD, with fold changes of 0.53, 0.64 and 0.66. Thus, overall there was a strong tendency towards loss of superoxide dismutase expression. Ascorbate peroxidase (*LmxM.33.0070*) did not show increased expression in any line, with decreased expression (0.66-fold) in AmBRB, in contrast to increased expression previously observed in AmB-resistant *L. donovani* (Kumar *et al*. 2014).

In the previous chapter, I noted changes in metabolites related to arginine and proline metabolism, which is in turn associated with polyamine metabolism. Therefore, I examined expression changes in genes annotated with functions related to this (GO:0006525, GO:0006526, GO:0006560, GO:0006561). Of these, four genes demonstrated significant changes with $\log_2$(FC) of at least ± 0.5. Two of these were annotated as involved in arginine metabolism. Acetylornithine deacetylase (*LmxM.07.0270*) showed decreased expression in AmBRA, AmBRB and AmBRC (0.67, 0.71 and 0.78, respectively). Glutamamyl carboxypeptidase (*LmxM.08_29.1570*) also showed decreased expression in AmBRA, AmBRC and AmBRD (0.52-fold, 0.74-fold and 0.59-fold). The reason for the annotation of this gene as related to arginine metabolism is unclear, but it is also annotated as having acetylornithine deacetylase activity, which may explain this association. Within proline biosynthesis, decreases were observed in delta-1-pyrroline-5-carboxylate dehydrogenase (*LmxM.03.0200*) in AmBRA (0.58-fold), AmBRB (0.71-fold) and AmBRD (0.71-fold). On the other hand, increases were observed in pyrroline-5-carboxylate reductase (*LmxM.13.1680*) in AmBRA (1.37-fold) and AmBRB (1.43-fold).

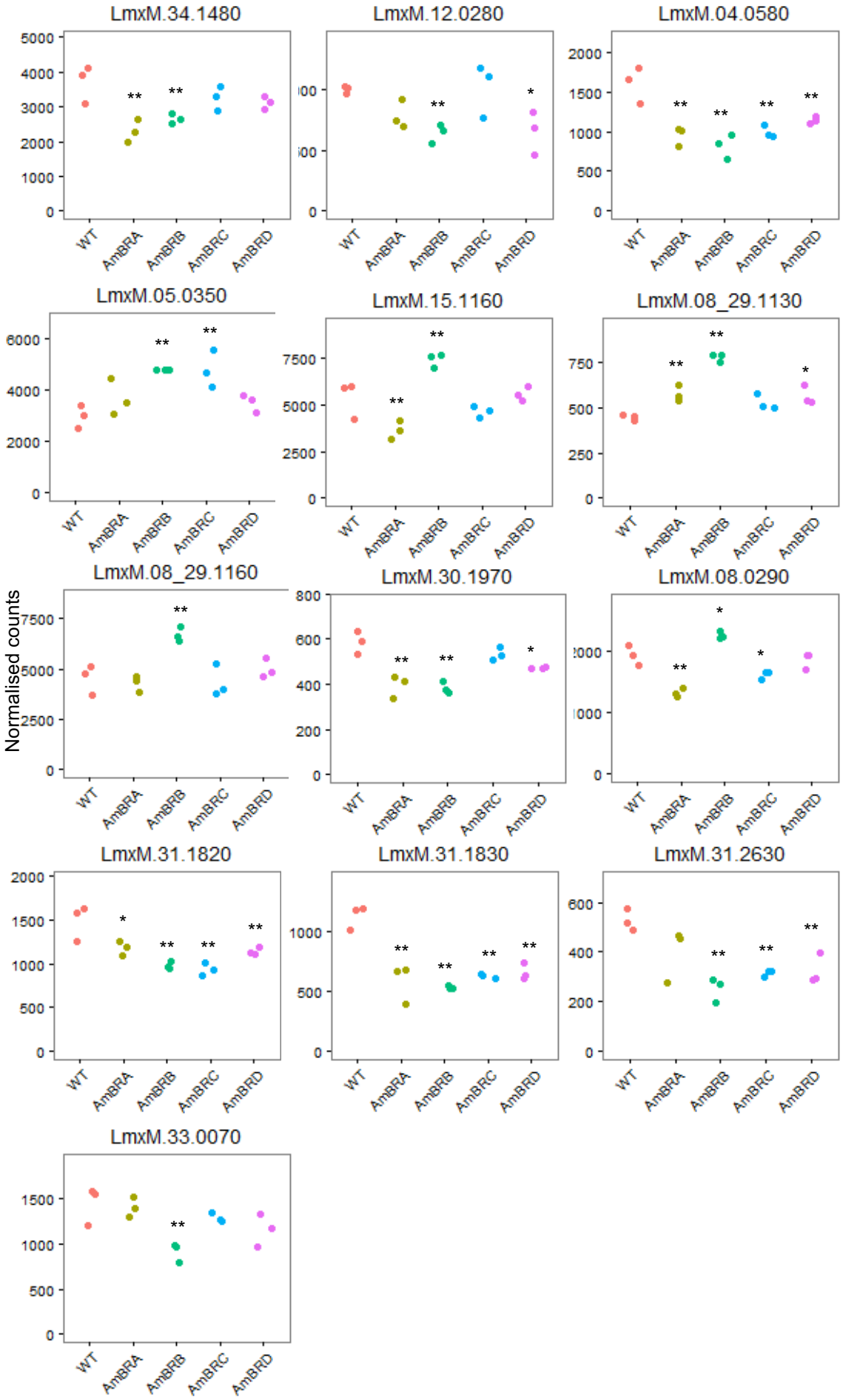

LmxM.34.1480
LmxM.12.0280
LmxM.04.0580
LmxM.05.0350
LmxM.15.1160
LmxM.08_29.1130
LmxM.08_29.1160
LmxM.30.1970
LmxM.08.0290
LmxM.31.1820
LmxM.31.1830
LmxM.31.2630
LmxM.33.0070
Normalised counts
WT
AmBRA
AmBRB
AmBRC
AmBRD

Figure 4-18: Expression changes in polyamine and oxidative stress metabolism. Normalised counts are plotted, with each dot representing an individual clone from each resistant line (and independent biological replicates for wild-type cells). Genes are plotted in the order of mention within the text. Asterisks represent statistically significant differences in expression from wild-type (*P < 0.05, **P < 0.005). Gene names are as follows: *LmxM.34.1480*: arginase; *LmxM.12.0280*: ornithine decarboxylase, putative; *LmxM.04.0580*: spermidine synthase, putative; *LmxM.05.0350*: trypanothione reductase; *LmxM.15.1160*: tryparedoxin peroxidase; *LmxM.08_29.1130*: tryparedoxin-like protein; *LmxM.08_29.1160*: tryparedoxin 1, putative; *LmxM.30.1970*: tryparedoxin 4, putative; *LmxM.08.0290*: iron superoxide dismutase A, mitochondrial; *LmxM.31.1820*: iron superoxide dismutase, putative; *LmxM.31.1830*: iron superoxide dismutase, putative; *LmxM.31.2630*: superoxide dismutase, putative; *LmxM.33.0070*: ascorbate peroxidase, putative.

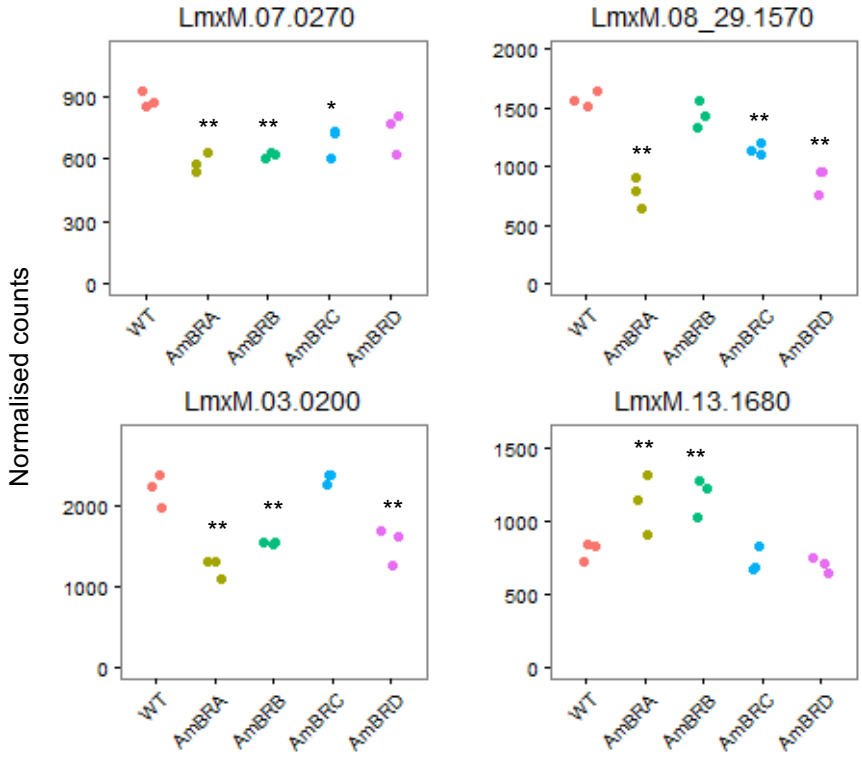


Figure 4-19: Expression changes in arginine and proline metabolism. Normalised counts are plotted, with each dot representing an individual clone from each resistant line (and independent biological replicates for wild-type cells). Genes are plotted in the order of mention within the text. Asterisks represent statistically significant differences in expression from wild-type (*P < 0.05, **P < 0.005). Gene names are as follows: *LmxM.07.0270*: acetylornithine deacetylase-like protein; *LmxM.08_29.1570*: metallo-peptidase, Clan MH, Family M18 (a.k.a. glutamamyl carboxypeptidase); *LmxM.03.0200*: delta-1-pyrroline-5-carboxylate dehydrogenase, putative; *LmxM.13.1680*: pyrroline-5-carboxylate reductase.

4.2.2.6 Expression changes in genes involved in mannogen biosynthesis

Given the pronounced decrease observed in the LC-MS metabolomics of poly-hexoses and GDP mannose, I investigated whether expression changes in this area of metabolism might underlie this phenotype. Many of the genes involved in mannogen biosynthesis do not appear to be annotated, but the steps to GDP-mannose synthesis are known, namely generation of mannose 6-phosphate from fructose 6-phosphate by phosphomannose isomerase (*LmxM.31.1580*),

isomerisation of this to mannose 1-phosphate by phosphomannomutase (*LmxM.36.1960*) and finally generation of GDP-mannose by GDP-mannose pyrophosphorylase (*LmxM.23.0110*). Moderate but statistically significant expression changes were noted in all three genes. In *LmxM.31.1580*, decreases were reported in AmBRB (0.69-fold) and AmBRD (0.74-fold). In *LmxM.36.1960*, decreases were also observed in AmBRA and AmBRB (0.69-fold and 0.80-fold, respectively). Finally, in *LmxM.23.0110*, expression in AmBRD was 0.74-fold that of wild-type. Whilst these changes are very modest, it is interesting that no significant changes were observed in AmBRC, which did not exhibit the reduction in mannogen-related metabolites seen in other lines. I examined whether any changes were evident at the genetic level in these genes. The only changes from wild-type were in the gene *LmxM.31.1580*, in which two heterozygous non-synonymous variants (C221R and T198A) present in wild-type cells became homozygous alternative in AmBRB/cl2. This presumably results from the loss of one copy of chromosome 31 (Figure 4-1) in this line.

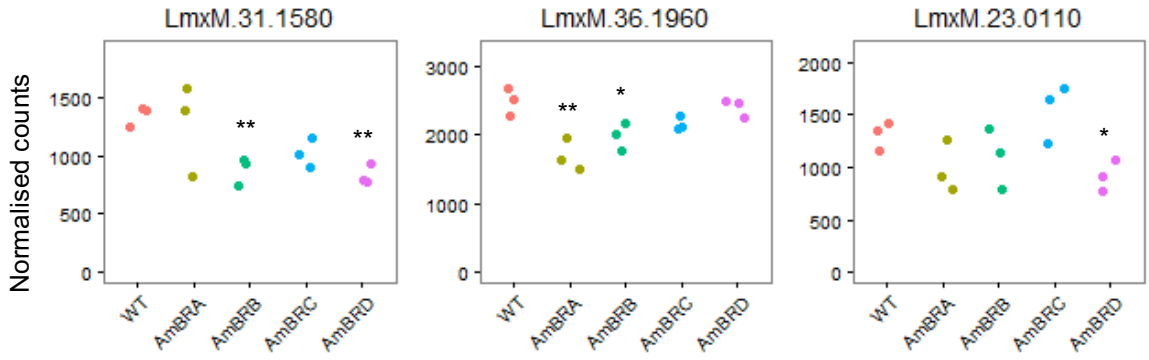


Figure 4-20: Expression changes in GDP-mannose biosynthesis genes. Normalised counts are plotted, with each dot representing an individual clone from each resistant line (and independent biological replicates for wild-type cells). Genes are plotted in the order of mention within the text. Asterisks represent statistically significant differences in expression from wild-type (*P < 0.05, **P < 0.005). Gene names are as follows: *LmxM.31.1580*: phosphomannose isomerase, putative; *LmxM.36.1960*: phosphomannomutase, putative; *LmxM.23.0110*: mannose-1-phosphate guanyltransferase (a.k.a. GDP-mannose pyrophosphorylase).

4.2.2.7 Expression changes in translation associated genes

On examining changes detectable in each line, it became evident that a large number of genes annotated with translation-related functions (GO term GO:0006412) exhibited differential expression, particularly in AmBRB. Of 231 genes retrieved with a search on TriTrypDB with this GO term, expression was detected in 224; of these, 148 showed differential expression (P < 0.05) in

AmBRB, 118 of which showed a $\log_2$(FC) of at least ± 0.5. AmBRA also exhibited a large number of differentially expressed genes (123), although far fewer (26) passed this fold-change threshold. In AmBRC and AmBRD, by contrast, differential expression was only observed in 30 and 25 genes, respectively, of which only 11 and 12 genes passed the fold-change threshold. Plotting these genes as a heatmap (Figure 4-21) clearly reveals this trend, with a large group of genes exhibiting reduced expression in all three AmBRB clones (in blue).

Therefore, in AmBRB, there was a clear skew towards negative fold-changes (Figure 4-22), which accounted for 127 out of 148 significantly differentially expressed genes. Amongst genes exhibiting reduced expression, the median $\log_2$(FC) was -0.71 (0.61-fold wild-type expression), with an interquartile range of -0.62 down to -0.80 (0.65-fold to 0.57). Therefore, few genes showed expression reduced to less than half that in wild-type cells; nevertheless, given the centrality of many of these proteins in an essential biological process (including structural constituents of the ribosome), even moderate fold changes may have important physiological effects.

I decided to test whether decreased expression of translation-associated genes in AmBRB influenced global rates of translation. In order to do this, I measured uptake of [^{35}S]-methionine by parasites over time, comparing AmBRB/cl2 with wild-type cells. In order to verify that this uptake was the result of translation-mediated incorporation of the radiolabel, I used cycloheximide (CHX), a known inhibitor of eukaryotic translational elongation, as a control. This revealed that while CHX itself dramatically reduced radiolabel uptake rates, there was no evidence of a reduced translation rate in AmBRB/cl2. I employed a linear model to examine the statistical significance of time and strain as factors in determining uptake, and while time was significant in all contexts, this confirmed that strain was not a significant factor in the absence of CHX ($P = 0.241$). Interestingly, however, there was a significant effect of strain in the presence of CHX ($P = 0.00888$). This might suggest a slightly higher susceptibility to CHX in AmBRB/cl2. Nevertheless, this rate difference in the presence of CHX was not very large (112 cpm/min in wild-type, 87.5 cpm/min in AmBRB/cl2).

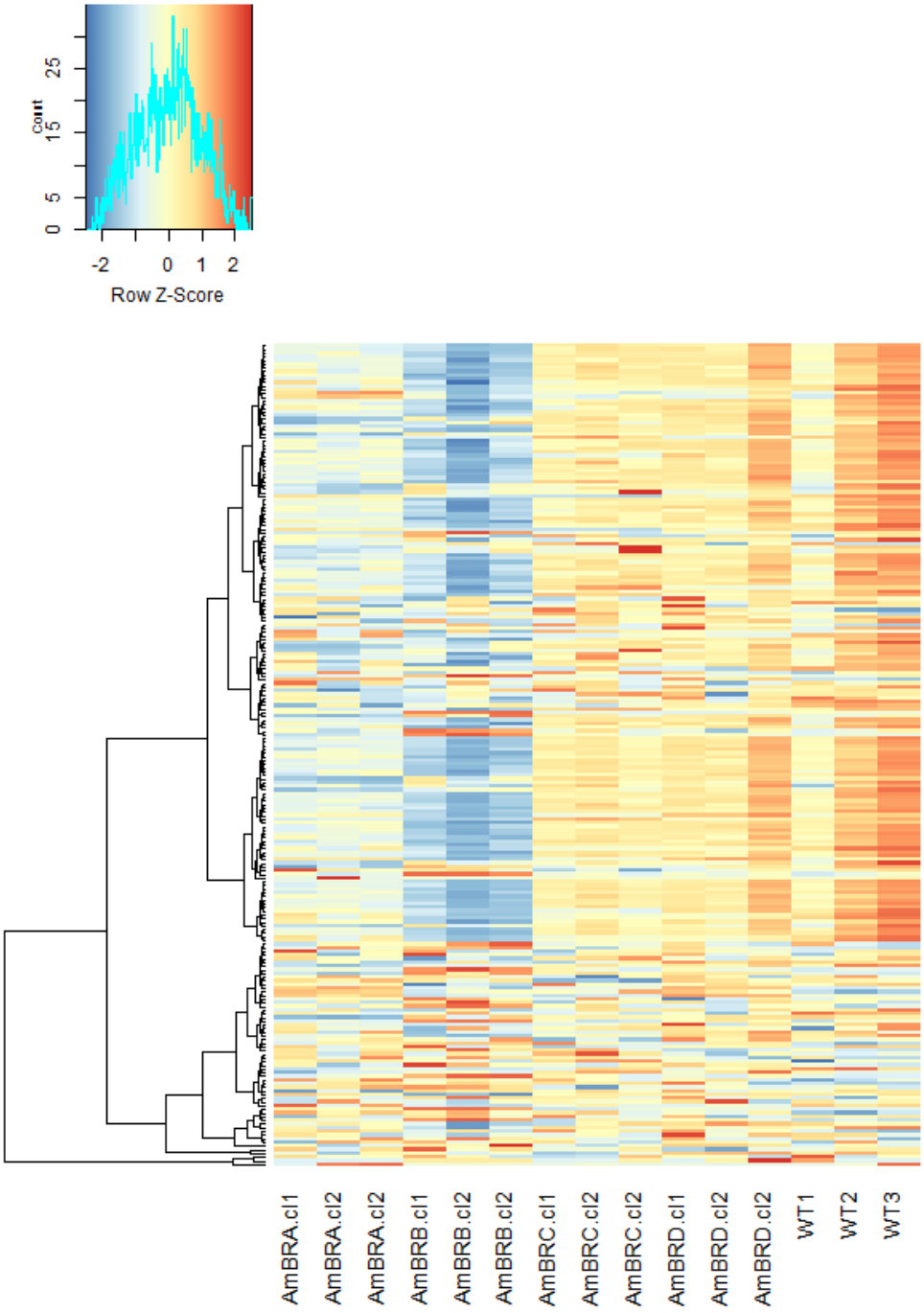


Figure 4-21: Heatmap of expression of translation-associated genes. Heatmap displaying expression of the 224 detectable genes that were retrieved with a search by GO term GO:0006412 (translation) in TriTrypDB. Each row represents one gene, and rows are subject to clustering. Values were subject to log transformation and colours are normalised by row, with the colour scale shown relating to these per-gene normalised log-transformed counts values. The heatmap was produced using the heatmap.2 function in the R package gplots.

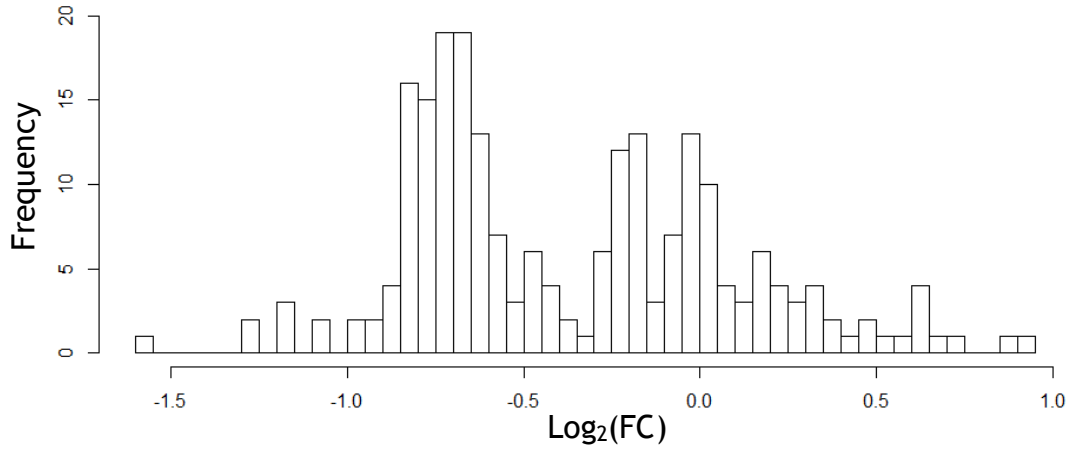


Figure 4-22: Distribution of fold changes in expression in translation-associated genes in AmBRB. The distribution of changes (subject to $\log_2$ transformation) is plotted as a histogram.

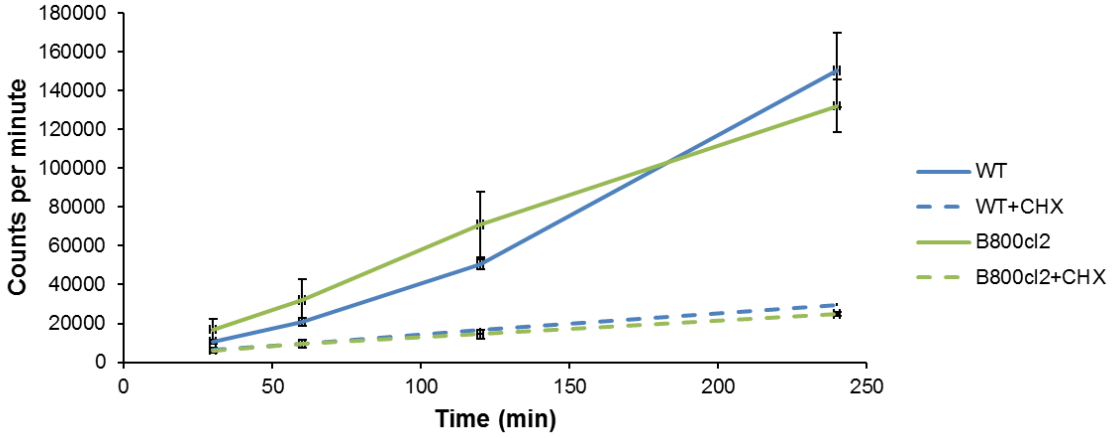


Figure 4-23: Global translation rates in wild-type and AmBRB/cl2. Uptake of [^{35}S]-radiolabelled methionine was measured by scintillation counting after harvesting of cells at specific times after radiolabel addition. Cycloheximide (CHX) was used as a translation-inhibited control.

4.2.3 Chronological order of mutations in AmBRB and AmBRC

For reasons that are elaborated upon in the Discussion section below, I made the assumption that the genotypic variations at positions 391 and 961 of the *SMT* locus in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 arose from the same mutational events which involved altered gene copy number and reductions in RNA abundance in these lines. However, this would appear to be separate from the T215G mutation observed only in AmBRC/cl3. Hence, two independent mutational events appear to arise at the *SMT* locus in AmBRC. By contrast, only one mutational event in sterol biosynthesis is observed for the other three lines. However, in AmBRB/cl2 I also observed deletion of the miltefosine transporter as a separate event of particular interest. Thus, two independent mutational

events are suspected to relate to resistance in AmBRB and AmBRC. I sought to determine the chronology of these events, and their influence on AmB sensitivity, by analysing cryogenically preserved cell lines collected at multiple points during resistance selection. These were uncloned, and therefore may represent mixed populations in which only a subset contain a given mutation.

In order to determine when changes arose at the *SMT* locus, I amplified *SMT* from extracted genomic DNA (note that both copies are amplified) and subjected amplicons to Sanger sequencing. With regard to the miltefosine transporter, I determined the time of deletion by PCR amplification and visual inspection of an amplified band by gel electrophoresis. In AmBRB (Figure 4-24A), this showed that both "heterozygous" genotypes in *SMT* became homozygous between 117 and 144 days (note that as amplification was not specific to either *SMT* copy, these genotypes refer to the combined alleles at these positions in both gene copies). On the other hand, the *MT* gene *LmxM.13.1530* could no longer be amplified after 75 days. Notably, the intensity of the band amplified from the 75 day sample was far lower than that from the 49 day sample and wild-type DNA, suggesting that it had already begun to disappear from the population at this stage (this difference was evident over three independent biological replicates). Statistical analysis indicated significant increases in IC_{50} between 49 and 75 days ($P = 7.26 \times 10^{-4}$), 75 and 93 days ($P = 0.00155$), 93 and 117 days ($P = 7.32 \times 10^{-4}$), 117 and 144 days ($P = 0.0126$), and 168 and 213 days ($P = 0.0485$) (note that within an uncloned population, different subsets have different IC_{50} values and that the proportions of these subsets may in turn be influenced by drug during the assay). Hence, resistance appeared to accumulate over multiple steps. This may suggest either that mutations take a while to become completely fixed within the population, or that these two changes alone do not completely account for total resistance in this line.

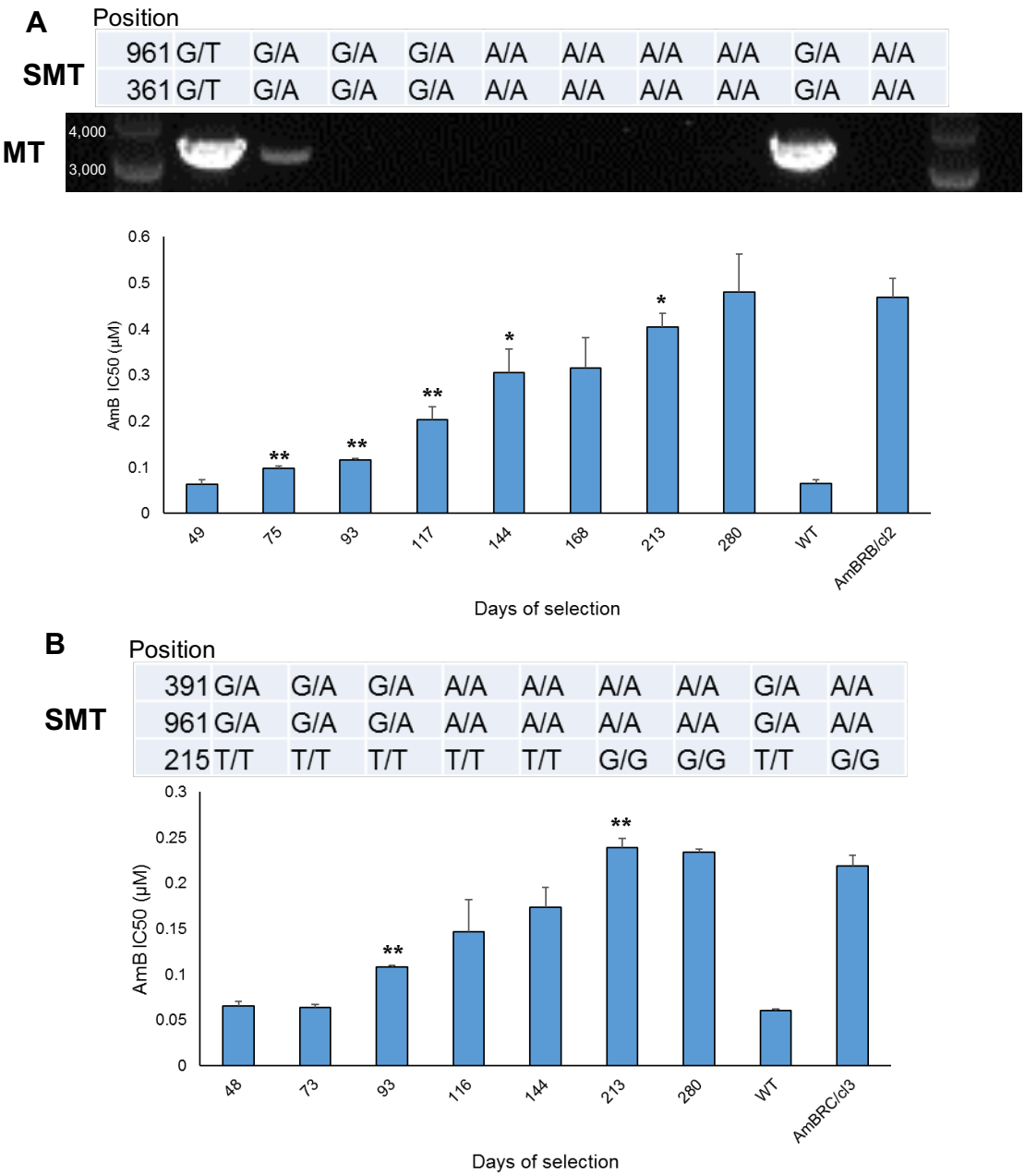


Figure 4-24: Mutation order during selection of AmBRB and AmBRC. Cryogenically stored (uncloned) subpassages during selection of resistance were tested for genotype and drug sensitivity. *SMT* genotypes were determined by PCR amplification and Sanger sequencing, with visual inspection of sequencing traces to determine genotype. AmB IC_{50} values represent the mean of four independent experiments, with error bars representing standard deviation. Asterisks represent statistically significant changes (*P < 0.05, **P < 0.05) from the previous subpassage (or compared to wild-type for the first subpassage). A) Changes in AmBRB, including genotyping of *SMT* positions 391 and 961, and the miltefosine transporter (MT). Deletion of *MT* was determined by PCR amplification, followed by 1% agarose gel electrophoresis (representative of three independent biological replicates). Bands to the far left and right represent DNA standards, with the size shown in base pairs in white (MT gene size = 3315 bp). B) Changes in AmBRC, including genotyping of *SMT* positions 391, 961 and 215.

In AmBRC (Figure 4-24B), the same transition from G/A to A/A happened at positions 391 and 961 between days 93 and 116 of selection. In contrast to this, the T215G substitution happened between days 144 and 213 of selection,

substantially after the initial changes to *SMT*. Resistance arose in fewer steps than for AmBRB, with significant increases in IC_{50} only between 93 and 116 days ($P = 2.47 \times 10^{-6}$), and between 144 and 213 days ($P = 0.00314$). The latter correlates with the emergence of the T215G substitution. Whilst the former does not correlate with a genetic change, it may nevertheless indicate that by 93 days a significant subpopulation of parasites have undergone the heterozygous to homozygous *SMT* change, allowing them to influence the IC_{50} recorded in drug assays, although with heterozygosity still present amongst the population as a whole.

4.3 Discussion

4.3.1 Changes in sterol biosynthesis genes in AmB resistance

One of the most distinctive and recurrent phenotypes associated with AmB resistance in both *Leishmania* and fungi is modification of sterol composition. In the previous chapter, I showed a dramatic reduction of the major sterol constituent, ergosta-5,7,24(28)-trienol, in all four AmB-resistant lines. Stability of the AmB resistance phenotype suggests that this was likely to have a genetic basis. In employing two NGS methods in parallel, I identified candidate changes in ergosterol biosynthesis genes that fit in terms of their predicted effects with the patterns observed in GC-MS data. AmBRA/cl1 demonstrated a clear accumulation of dienol sterols and loss of ring double bond conjugation, and this was accompanied by mutation in sterol C5-desaturase (*LmxM.23.1300*), the enzyme which introduces a 5(6) double bond required for the 5(6)-7(8) double bond conjugation observed in ergostane-type sterols. On the other hand, AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 all showed loss of C24-methylated sterols, with accumulation of cholestane-type sterols which retained 5(6)-7(8) double bond conjugation; this was accompanied by evidence of disruption at the locus for sterol C24-methyltransferase (*LmxM.36.2380*/*LmxM.36.2390*). Hence, in all cases, loss of one of the two primary features that appear to promote interaction of ergostane-type sterols with AmB was accompanied by genetic changes in the very enzyme that introduces those features.

4.3.1.1 G139R substitution in *SC5D*

AmBRA/cl1 possessed a substitution in *SC5D* gene *LmxM.23.1300* that was not observed in any other line. G139, while not entirely conserved in all species, is located within a histidine-rich motif that is one of multiple within this protein (Figure 4-3) that are believed to have a role in catalysis (Taton *et al*. 2000). The change from glycine to arginine is not conservative, with arginine not only introducing bulk, but also introducing an extra positive charge, which, given the evident importance of histidines in this enzyme mechanism (which can themselves be positively charged), may well interfere with normal function. The substitution is homozygous, being associated with a large LOH event on chromosome 23. The lack of 5(6)-7(8)-desaturation in this line provides strong circumstantial evidence for a functional impact of this SNP. Mutation in this enzyme has not previously been reported in *Leishmania*; however it has been noted in antifungal-resistant yeast. Chau and colleagues described two AmB and azole cross-resistant *C. albicans* isolates that exhibited nonsense mutations in the yeast homologue, *erg3* (Chau *et al*. 2005). On the other hand, two AmB-resistant *C. lusitaniae* strains exhibited decreased expression of this enzyme, where expression of other genes increased (Young *et al*. 2003). However, in this case, UV absorption was still partially retained, suggesting that at least some *erg3* function remained, and in any case, a causal link between AmB resistance and loss of *erg3* function was not demonstrated. Other previously identified mutations in AmB-resistant *Candida* species include a S258F substitution in *C. tropicalis* (Eddouzi *et al*. 2013) and an L193R substitution in *C. albicans* (Morio *et al*. 2012), as well as a truncation, also in *C. albicans* (Morio *et al*. 2012). In a single *C. albicans* line, substitutions K97E, L193P, V237A, A351V and A353T were all identified (Martel *et al*. 2010). The equivalent residue to *L. mexicana* G139 in *C. albicans erg3* is W228, thus it is in a similar region of the protein to the sites of several other previously described mutations (L193, V237, S258), suggesting that this region may be of particular importance for catalytic activity.

4.3.1.2 Complex changes at the *SMT* gene locus

In contrast to the mutation in *SC5D*, which is only observed in one line, I reported variation at the *SMT* locus in three separate lines, each appearing to be different. There are a number of threads of evidence relating to this: firstly,

SNPs in the coding sequence; secondly, evidence of copy number changes; thirdly, a large reduction in expression at the RNA level. A number of difficulties arise when interpreting these data. In particular, it is important to bear in mind that the similarity of the two gene copies, *LmxM.36.2380* and *LmxM.2390* (which differ in the reference genome only by the G391A substitution), means that distinction between the two copies on the basis of coding sequence alone is at best unreliable. Moreover, there is evidence of discontinuity of the alignment of even wild-type sequences (Figure 4-5) within the intergenic region between the two gene copies. This may well be due to incorrect assembly of the reference genome, possibly because of repetition in this region. Hence, there are limitations on how precise one can be at this stage with regard to the nature of changes, and how far one can discuss changes to individual *SMT* gene copies separately.

The first change noted was regarding SNPs within the coding sequence, particularly G391A and G961A, although T215G was also reported, uniquely in AmBRC/cl3. The T215G mutation, which causes a F72C amino acid substitution, appears to be within the putative sterol binding site and thus may interrupt enzymatic activity (Figure 4-4). However, I have clearly shown that this change is chronologically distinct from those at positions 391 and 961 (Figure 4-24). On the other hand, in both AmBRB and AmBRC, changes at these latter two positions happen together, suggesting that they are associated with the same overall mutational event. Changes at these sites relate to pre-existing variants in wild-type parasites. In *LmxM.36.2380,* G is found at both positions 391 and 961, whereas in *LmxM.36.2390,* A is found in both positions (although in the reference genome itself, *LmxM.36.2390* has a G at position 961). One should be cautious in attributing changes to one copy or another, for reasons explained above, but what is certain is that taking the two copies together, in wild-type cells both positions are G/A “heterozygous” (both alleles are present). On the other hand, in all three resistant lines exhibiting altered *SMT* expression, these positions are homozygous: AmBRB/cl2 and AmBRC/cl3 have an A391/A961 genotype, whereas AmBRD/cl2 has G391/A961. The resulting substitutions themselves (V131I and V321I) are themselves highly conservative, and so it is not necessarily the case that these changes themselves are functionally important. However, they are a useful indicator of changes occurring at this locus. It would seem to be an

extremely unlikely scenario whereby these changes occur due localised single nucleotide mutation, as it implies that identical mutations arise independently in one gene copy that are already present in the other. Far more likely, it is the result of broader structural variations at this locus.

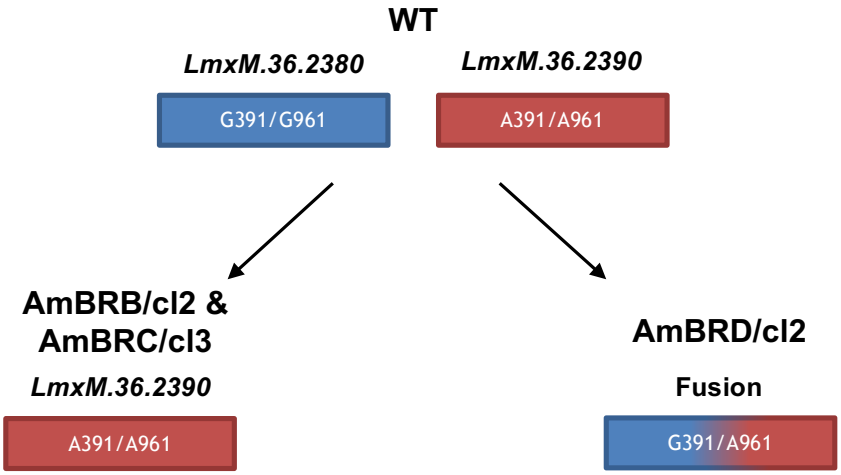


Figure 4-25: Initial model for changes in *SMT* genes. The wild-type *L. mexicana* genome possesses two tandemly arrayed copies of *SMT*, *LmxM.36.2380* and *LmxM.36.2390*. Based on evidence of structural variation, and genotypes at nucleotide positions 391 and 961, loss of *LmxM.36.2380* appears to occur in AmBRB/cl2 and AmBRC/cl3, whereas fusion of the two copies is a possible explanation for the genotype observed in AmBRD/cl2.

This is supported by evidence from WGS per-gene coverage data (Figure 4-8). The number of reads mapped to each gene was quantified and this revealed a complex picture in which coverage of *LmxM.36.2380* was decreased in AmBRB/cl2 and AmBRC/cl3, whereas reduced coverage of *LmxM.36.2390* was observed in AmBRC/cl3 and AmBRD/cl2, but actually increased in AmBRB/cl2. An important caveat here is that the tool used to quantify, htseq-count, excludes by default those reads with poor mapping quality (< 10), which includes reads that map equally well to multiple positions. Therefore, while changes are evident, these data are probably not reliable in quantitative terms. Nevertheless, copy number variation may explain the changes observed in SNPs. For example, if *LmxM.36.2380* possesses the genotype G391/G961 in wild-type parasites, with *LmxM.36.2390* having an A391/A961 genotype (as predicted by the SNP calls), the loss of G residues at these positions in AmBRB/cl2 and AmBRC/cl3 could arise from loss of *LmxM.36.2380* but retention of *LmxM.36.2390*. The AmBRD/cl2 genotype (G391/A961) is harder to explain, as in this model neither wild-type gene copy possesses this genotype. However, it may arise due to fusion of the two copies, facilitated by high sequence similarity

between the two. AmBRB/cl2 also shows evidence of a duplication in chromosome 20 (Figure 4-9), which, since its boundary appears to lie within the *SMT* region, likely arises as part of to the same structural variation event as changes to *SMT* in this line. Thus, it appears that the structural variation events within the *SMT* locus found in each line are unique to that line.

In each line showing evidence of structural changes at this locus, there is a large decrease in RNA expression. Based on qRT-PCR data, overall *SMT* expression in these lines ranges from 0.11-fold to 0.22-fold that of wild-type. Crucially, this is a larger decrease in magnitude than that which could be expected from loss of copy number alone (which one might expect to result in a 50% decrease in expression). There are two possible explanations for this: firstly, the structural variant itself leads to disruption of expression, for example due to deletion of splice sites or sites involved in promotion of RNA stability. Alternatively, there may be intrinsic expression differences between the gene copies, even in wild-type parasites, as a result of non-coding elements present in transcript UTRs. The 3'-UTRs of kinetoplastid transcripts are known to be of particular importance in gene regulation (Clayton 2014). Hence, it may be that the 3'-UTR of *LmxM.36.2390* is associated with poorer steady state transcript abundance (for example, through conferring decreased stability) than that of *LmxM.36.2380*. This would not only explain why expression is reduced in AmBRB and AmBRC, which only possess the coding sequence for *LmxM.36.2390*, but also in AmBRD, since while this has a *LmxM.36.2380*-derived G391 genotype, the 3'-end of the gene has a *LmxM.36.2390*-like A961 genotype, suggesting that it may also inherit the 3'-UTR from *LmxM.36.2390*. Clearly, this model is very speculative at this stage. For this reason, changes to this locus are investigated further in the next chapter.

A final point is that whilst loss of *SMT* expression in these lines is considerable (78-89%), the loss of C24-methylated sterols in these lines is even greater, from 94.3% of total sterol in wild-type parasites to 9.5%, 1.8% in AmBRB/cl2, and AmBRD/cl2, and <1% in AmBRC/cl3. This suggests that expression changes alone do not account for the change in sterol. Whilst it is possible that this relates to other mutations that have not yet been identified, another possibility is that the transcripts in these lines may themselves have reduced functionality, for

example altered translation efficiency. However, as I have not been able to quantitatively measure altered protein expression, this question cannot be answered with certainty.

4.3.1.3 Other changes in sterol biosynthesis gene expression

Whilst the changes described above are the most substantial with regard to sterol metabolism, a number of other genes were differentially expressed. Most of these represented relatively small fold changes and were associated with altered chromosome copy number (*LmxM.08_29.1830*, *LmxM.13.1620*, *LmxM.15.1460*, *LmxM.18.0020*, *LmxM.31.2320*, *LmxM.31.3310*). It should be noted, moreover, that while expression of both *LmxM.31.2320* and *LmxM.3310* was significantly reduced in AmBRB, AmBRC and AmBRD relative to wild-type, this reflects the fact that chromosome 31 is diploid in these lines, but triploid in AmBRA and wild-type. As the karyotype of even the parental line is expected to be unstable, therefore, it is cannot be concluded with certainty whether at the start of selection, chromosome 31 was diploid or triploid, and hence whether these changes reflect loss of expression in AmBRB, AmBRC and AmBRD or gain of expression in AmBRA and wild-type lines. Changes in three genes (*LmxM.21.1430*, *LmxM.34.1230* and *LmxM.36.2540*) could not be accounted for by chromosomal CNV. While these were for the most part also moderate, the largest change was in *LmxM.36.2540* (sterol C4-methyloxidase), and this was associated with the 50 kb duplication event in AmBRB. Interestingly, the increase in expression (3.9-fold) is greater than would be expected based on a doubling of gene copy alone, suggesting that the structural rearrangement involved may in some way promote increased RNA abundance through another mechanism, such as placing the gene in proximity to elements promoting greater RNA stability. Overall, there appears to be little pattern to the changes in expression in these lines, with both increases and decreases observed at different points within the pathway. In *C. parapsilosis* exhibiting mutations in *SC5D* gene *erg3*, expression of several ergosterol biosynthesis genes has been noted (Rybak *et al*. 2017), likely as a regulatory response to loss of pathway function. In *Leishmania*, by contrast, there is no evidence that such a regulated response occurs at the RNA level. Whilst these data do not rule out the possibility of regulation at the translational level, a previous proteomics analysis

of AmB-resistant *L. infantum* did not report any such changes (Brotherton *et al*. 2014).

4.3.2 Changes to other areas of lipid metabolism

4.3.2.1 Expression changes in lipid metabolism-associated genes

The LC-MS metabolomics experiment described in the previous chapter revealed multiple changes in lipid metabolism, particularly in AmBRB/cl2. Here I determined whether lipid metabolism might also be altered at the expression level. As with sterol metabolism, genes involved in other areas of metabolism showed a number of moderate changes in expression, both increases and decreases, most of which (*LmxM.03.0080*, *LmxM.03.0821*, LmxM.*08_29.1260*, *LmxM.13.0020*, *LmxM.13.0210*, *LmxM.18.0160*, *LmxM.17.1100*, LmxM.*18.0810*) where associated with changes in chromosome ploidy. Of those that could not be explained by chromosomal CNV (*LmxM.09.1040*, *LmxM.10.1320*, *LmxM.14.1360*, *LmxM.24.1810*, *LmxM.26.1670*, *LmxM.28.2100*, *LmxM.33.1070*), three genes showed a trend of decreased expression in all four lines. These were *LmxM.10.1320* (encoding a fatty acid desaturase), *LmxM.14.1360* (encoding myo-inositol-1-phosphate synthase) and *LmxM.33.1070* (class 3 lipase). As discussed with regard to the LC-MS metabolomics experiment, it is important to be cautious regarding changes that occur in all four resistant lines because no highly passaged control was used in this experiment, meaning that the possibility that these changes represent adaptation to culture cannot be excluded. Furthermore, myo-inositol-1-phosphate synthase actually showed 1.9-fold increased expression in a proteomics study of AmB-resistant *L. infantum* (Brotherton *et al*. 2014). Nevertheless, it is interesting in particular that these genes include a desaturase, given previous evidence of increased fatty acid saturation in AmB-resistant *L. donovani* (Mbongo *et al*. 1998), although such changes were not observed in an AmB-resistant *L. infantum* line (Fernandez-Prada *et al*. 2016). Furthermore, in the last chapter, I reported decreases in a number of lipids containing unsaturated fatty acid moieties. Whilst some of these (including [PC (14:1)] 1-(1E-tetradecenyl)-sn-glycero-3-phosphocholine and [PC (15:1)] 1-(1Z-pentadecenyl)-sn-glycero-3-phosphocholine) were specific to individual lines (particularly AmBRB/cl2), [FA (16:2)] N-hexadecyl-ethanolamine (which contains two fatty acyl double bonds) was decreased in all four lines. It is

feasible, therefore, that reduced expression of the fatty acid desaturase *LmxM.10.1320* contributes to these metabolic changes. Of final note is the gene *LmxM.09.1040*, encoding a phospholipid:diacylglycerol acyltransferase. While its increase in AmBRB is associated with chromosomal CNV, in AmBRD it is associated with a more local amplification event (Figure 4-9), further emphasising the impact such amplifications can have on the transcriptome.

4.3.2.2 Deletion of the miltefosine transporter, *LmxM.13.1530*

The most striking change in lipid metabolism is the 8 kb deletion observed in AmBRB/cl2, which results in total loss of expression of the miltefosine transporter, *LmxM.13.1530*, and is homozygous due to an associated LOH event. Mutations in this phospholipid flippase have previously been associated with MILT resistance, and indeed this line exhibits cross-resistance to MILT. Previously, an *L. infantum* line selected for AmB resistance exhibited a mutation (G433S) in this gene, and this was shown to be partially responsible for resistance to AmB, as well as to MILT (Fernandez-Prada *et al*. 2016). Interestingly, the mutation had only a minimal effect on MILT transport, in contrast to mutations observed in a line selected for resistance to MILT itself. Furthermore, this line was only 3.8-fold resistant to MILT, in contrast to the MILT-resistant line, which was 13.2-fold MILT-resistant (note that the MILT resistance-selected line showed 2.7-fold cross-resistance to AmB, which was also restored by re-expression of *MT*). Whilst I have not directly assessed MILT uptake in AmBRB/cl2, the complete absence of this gene would make it highly likely MILT transport is greatly affected. Deletion of *MT* in *L. donovani* abolished MILT transport, leading to 13.7-fold resistance to this drug (Pérez-Victoria *et al*. 2006). Given that AmBRB/cl2 is also an *MT* deletion mutant, it is surprising that resistance to MILT is so low (2.3-fold). This could reflect species-specific differences, although it may relate to secondary effects from other mutations present in this line. Research into acquired MILT resistance has focused strongly on the closely related species, *L. donovani* and *L. infantum*. Therefore, it is unclear whether conclusions drawn from studies using these organisms are relevant for other *Leishmania* species. In species of the subgenus *Viannia*, particularly *L. brasiliensis*, examples have been reported of natural resistance to MILT. This has been associated with low expression of MT and its associated binding partner Ros3, rather than sequence polymorphisms (Sánchez-Cañete *et al*. 2009, Obonaga *et al*. 2014). The role of MT within the

cell and in the context of MILT-resistance may therefore vary by species, something that may be worth investigating further.

Direct deletion of *MT* in *L. donovani* abolishes the inward translocation of labelled phosphatidylcholines and phosphatidylethanolamines from the outer leaflet of the plasma membrane to the inner leaflet (Weingärtner *et al*. 2010), so that in deletion mutants, these species are more exposed on the extracellular face of the membrane. However, in *MT*-knockout parasites, overall cellular phospholipid composition was not substantially affected. Moreover, sensitivity to AmB was unaffected by deletion of either MT or Ros3. In contrast to this, MILT- and AmB-resistant *L. infantum* exhibited extensive changes to lipid content (Fernandez-Prada *et al*. 2016). Importantly, however, there are similarities between these AmB- and MILT-resistant *L. infantum* lines in terms of the changes to lipid metabolism, notably an increase in inositol-phosphoceramides and some phosphatidylethanolamines, a reduction in some phosphatidylinositols, and no substantial changes to overall fatty acid composition. In AmBRB/cl2, the most substantial changes include reduced abundance of various phosphatidylcholines, phosphatidylethanolamines and sphingolipids. As suggested in the previous chapter, it is very difficult to directly compare these datasets due to the very different methodologies used. Therefore, a more comparable, phospholipid-focused analytical method would have to be employed to properly identify lipid changes associated with *MT* deletion in *L. mexicana*. However, the fact that *MT* deletion arises before *SMT* changes in AmBRB/cl2 suggests that it can mediate resistance even in the absence of sterol changes.

4.3.3 Broader metabolism, particularly polyamine and oxidative stress metabolism

Due to the high numbers of differentially expressed genes reported in each line, and the suspected role of apparently stochastic chromosomal CNV in many of these changes, I decided to focus on specific areas of metabolism either identified as altered in the LC-MS data, or that had previously been implicated in AmB resistance. In particular, given previously reported changes to expression of oxidative stress-associated genes (Purkait *et al*. 2012, Kumar *et al*. 2014), I focused on changes in genes either involved in polyamine and trypanothione metabolism, or directly associated with detoxification of ROS. As for lipid

metabolism, many of these changes were associated with chromosomal CNV. For example, spermidine synthase (*LmxM.04.0580*) showed decreased expression relative to wild-type in all four lines, but this was associated with triploidy of chromosome four in wild-type parasites, in comparison to diploidy in all resistant lines. As discussed above regarding chromosome 31, this may actually reflect a gain of copy in wild-type, given the expected karyotypic instability, rather than a loss of copy in the other lines. Decreases were observed in arginase (*LmxM.34.1480*) in AmBRA and AmBRB, as well as ornithine decarboxylase (*LmxM.12.0280*) in AmBRB and AmBRD. On comparison to metabolite levels, it is noteworthy that both L-arginine and L-ornithine are decreased in AmBRA/cl1, AmBRB/cl2 and AmBRD/cl2, but not AmBRC/cl3. It could be that the lack of changes in expression of either of these genes in AmBRC/cl3 is related to the lack of decreases in these metabolites in the same line; however, as arginase consumes L-arginine, it would not be expected that decreased expression of this enzyme would itself be a direct cause of lower L-arginine levels. Increased abundance of trypanothione reductase (*LmxM.05.0350*) in AmBRB and AmBRC was associated with increased copy number of chromosome five. However, these changes appear to have no relation to trypanothione levels, which were increased in AmBRA/cl1, AmBRC/cl3 and AmBRD/cl2, but not AmBRB/cl2.

With respect to genes directly involved in ROS detoxification, changes were moderate but heterogeneous. In the case of superoxide dismutases, multiple decreases were reported, as was the case for ascorbate peroxidase, an increase of which was previously reported in an AmB-resistant clinical isolate (Kumar *et al*. 2014). It is clear, therefore, that the signature of increased expression of genes relating to trypanothione metabolism and defence against oxidative stress is not observed in any of the AmB-resistant lines characterised here, in contrast to changes reported previously (Purkait *et al*. 2012). The previous study reported large increases, determined by qRT-PCR, from around twofold in ornithine decarboxylase and trypanothione synthase, to tenfold in cytoplasmic tryparedoxin peroxidase. In the context of the fold changes observed in the RNA-seq data obtained here, tenfold is a very large change. It is important to note, however, that whilst in this current study, wild-type and resistant lines are isogenic, the study by Purkait and colleagues compared only one resistant and one wild-type isolate, with no indication of the relatedness of the two lines.

Hence, it is possible that changes between the lines relate more to differences in genetic background rather than effects of AmB resistance selection. In a proteomics study that did use isogenic *L. infantum* lines (Brotherton *et al*. 2014), increased expression of trypanothione reductase (2.5-fold), a tryparedoxin peroxidase (2.7-fold) and an iron superoxide dismutase (2.4-fold) was reported, suggesting that translational regulation may play an important role.

Overall, numerous other areas of metabolism showed moderate changes in expression of multiple genes, but these changes were difficult to interpret in terms of potential effects. Despite striking changes observed in polysaccharide abundance in the LC-MS metabolomics data, no dramatic alterations to mannogen metabolism were apparent at the genomic or transcriptomic levels. I examined changes to arginine and proline metabolism; broadly speaking, however, these do not appear to relate to changes in metabolite levels, although the abundance of isomers that cannot be distinguished increases the complexity of interpretation. It may be that these changes are simply not reflected at the level of protein expression. Alternatively, it is possible that the sheer number of moderate changes to gene expression makes individual gene-metabolite relationships far more difficult to dissect.

4.3.4 Reduced expression of translation-associated genes

Where other expression changes described here bear little evidence of concerted regulation, there was a clear signature of downregulation of genes associated with translation in AmBRB. This appeared to be specific to cytoplasmic translation since of seven genes annotated as mitochondrial (*LmxM.04.0270*, *LmxM.08_29.0720*, *LmxM.18.0740*, *LmxM.24.0330*, *LmxM.25.0890*, *LmxM.36.0570*, *LmxM.36.2310*), only two showed differential expression in this line (*LmxM.08_29.0740* and *LmxM.18.0740*, encoding mitochondrial elongation factors Ts and Tu, respectively), and these were both increases. Of 148 genes with differential expression, 127 were decreased and of these, 106 were annotated as structural constituents of the ribosome (40 in the 40S subunit, 48 in the 60S subunit, and the remaining 18 lacking annotation for either). On the other hand, of the 21 genes increased, these included a few structural ribosomal proteins as well as various elongation factors and initiation factors. The median

relative expression of translation-associated genes exhibiting decreased expression was 0.61-fold that of wild-type. Whilst this is not a huge change in relative terms, the abundance of ribosomes and the centrality of translation as a cellular process, as well as the apparently regulated nature of the changes, made this a process of interest.

However, whilst the signature itself is clear, the causes and effects of these changes are not. Global translation rates did not differ between wild-type and AmBRB/cl2 parasites, although a mild difference was observed in the presence of cycloheximide. A reduced growth rate, which might be expected in cells exhibiting lower translation rates, was also not evident. Furthermore, because factors regulating ribosome biogenesis are unknown, I could not manipulate parasites either to restore wild-type expression of these genes in AmBRB/cl2 or to reduce wild-type expression, to determine if this could itself influence AmB resistance. Ribosome biogenesis is regulated by the TOR pathway in other organisms. *Leishmania* possesses several TOR-family enzymes (Madeira da Silva *et al*. 2010), creating the possibility of inhibition of wild-type ribosome biogenesis by this route, but rapamycin, the classical inhibitor of this pathway in other organisms, is a poor inhibitor of *Leishmania* TOR (Madeira da Silva *et al*. 2009). More potent inhibitors of *Leishmania* TOR have been developed (Diaz-Gonzalez *et al*. 2011), which may offer an opportunity in the future to try to recapitulate this phenotype in wild-type cells.

Regarding the role of such a phenomenon, disruption of the TOR pathway in yeast through both genetic methods and rapamycin-mediated inhibition has been shown to increase persistence in response to AmB, but not resistance (see 1.3.5.2 for a discussion of the differences). I did not explicitly test AmB persistence in our *Leishmania* lines. There is no standard assay for this, though likely it would have to involve exposure to high drug concentrations followed by cloning by limiting dilution to determine the number of viable parasites that can proliferate after drug is removed. Hence, it is possible that reduced ribosomal protein expression increases persistence in the face of high AmB concentrations.

Another possibility, however, is that rather than being an adaptation to AmB exposure, reduced expression of translation-associated genes is a response to increased gene load due to chromosomal CNV. By multiplying the number of

genes on each chromosome by the ploidy of that chromosome in a given line, one can calculate an estimate of the total number of copies of all genes (not accounting for sub-chromosomal amplifications and deletions). Whilst the wild-type genome possessed 17,871 gene copies, which was at least as much as AmBRA/cl1 (17,462), AmBRC/cl3 (17,830) and AmBRD/cl2 (17,380), the genome of AmBRB/cl2 had the highest number (18,546). Therefore, one would expect a higher load in terms of transcripts in this line (although the more moderate changes observed in AmBRA/cl1 cannot be explained by this). One hypothesis is that there is a feedback mechanism to prevent overproduction of protein by reducing ribosome biogenesis, thus lowering global per-transcript translation rates without necessarily reducing global per-cell translation rates (as measured in the [^{35}S]-methionine uptake assay). Such a feature could be important in coping with the high levels of copy number increases observed in *Leishmania* genomes. Notably, such a mechanism would also result in genes without CNV increases being less efficiently translated. However, given the absence of evidence of this, and the lack of an obvious effect of the expression changes, I was unable to investigate this phenomenon further. Nevertheless, it would be interesting to determine whether cellular ribosome content was substantially lower in AmBRB/cl2 than wild-type.

4.3.5 The relationship between genome and transcriptome

Since the advent of NGS technology, WGS has been applied to the study of *Leishmania*, both in investigation of its unusual genome biology and in the acquisition of drug resistance. RNA-seq, by contrast, has been applied only infrequently. Most of these studies have investigated cell cycle differentiation (Dillon *et al*. 2015, Dillon *et al*. 2015b, Fiebig *et al*. 2015, Fernandes *et al*. 2016). Only one study prior to this one has investigated drug resistance, focusing on MILT (Vacchina *et al*. 2016). Interestingly, the results obtained were in many ways similar to this one, with very large numbers of significantly differentially expressed genes, albeit with modest fold changes (less than twofold).

By pairing genome and transcriptome data, I have presented evidence that the extensive changes to expression levels are strongly influenced by alterations in karyotype. There is a clear influence of chromosomal ploidy changes on RNA abundance in all lines (Figure 4-13). To take an individual example of AmBRB,

this line has 3,411 differentially expressed genes relative to wild-type, 1,819 of which are increases and 1,592 of which are decreases. The WGS data predict that in AmBRB/cl2, there are nine chromosomes with increased copy number relative to wild-type, and three with decreases. In the chromosomes with increased copy, there are 1,663 genes, of which 1,362 show increased expression and 13 show decreased expression. In those chromosomes with decreased copy, out of 901 genes, 662 show decreased expression in comparison to three with increased expression. Thus, overall, 2,024 out of 3,411 differentially expressed genes are linked with a chromosomal copy number change. In fact, since these ploidy changes are derived after rounding values to the nearest integer, this does not take into account subtler effects of shifts in mosaic aneuploidy, which can result in non-integral ploidy of a population (Sterkers *et al*. 2012). Hence, the effect of chromosome CNV could be an underestimate. For example, in chromosome 16 a moderate increase in ploidy ratio relative to wild-type is observed in AmBRB/cl2 (Figure 4-1A) which does not result in a change in ploidy when expressed as an integer (Figure 4-1B). However, of 183 genes found on chromosome 16, 140 show increased expression in AmBRB, with no decreases.

One limitation with respect to experimental design is the choice to perform WGS in one clone per independent line, but to perform RNA-seq on three clones per line, using these as biological replicates. No significant changes in AmB IC_{50} were observed between clones derived from the same resistant line, suggesting that the mechanism of resistance was probably shared between clones. However, it was hoped that by sequencing separate clones this could account for biological variation that was not relevant to the resistance phenotype. Figure 4-11 shows close clustering of clones from the same line on the basis of transcriptome, and indeed changes of particular interest, such as loss of *SMT* expression and deletion of *MT*, were conserved across clones of the same line, providing support for this conjecture. However, from the perspective of using the data to relate the genome and the transcriptome, the instability of chromosome ploidy means that it is not valid to assume that ploidy changes will be conserved between individual clones. For this reason, when correlating RNA abundance and ploidy in Figure 4-13 I have used RNA data from only the clone subject to WGS. However, as there is only one replicate of each clone, it is statistically problematic to identify differentially expressed genes in one clone alone. Therefore,

investigation of the relationship between chromosomal CNV, calculated from individual clones, and the number of differentially expressed genes, determined based on use of multiple clones as biological replicates, may be hampered by the presence of inter-clonal ploidy differences. Moreover, as the wild-type line itself, while being clonal in origin, was not cloned immediately prior to sequencing, it is likely that the rate of mosaic aneuploidy is higher amongst this population. Hence, while one can begin to investigate the relationship between ploidy and expression in this dataset, the experimental design is not optimal for this purpose.

Recently, however, two studies have directly investigated the relationship between gene and chromosomal CNV and gene expression (Dumetz *et al*. 2017, Iantorno *et al*. 2017). The first of these tracked chromosomal CNV in *L. donovani* during infection of either sandflies or hamsters (Dumetz *et al*. 2017). Unlike local variants in nucleotide sequence, or indeed local CNV, chromosome copy number rapidly changed in a new environment. Unexpectedly, however, CNV did not appear to be random. For example, infection of the Nepalese *L. donovani* strain BPK282/0 in sandflies was associated with decreased ploidy of chromosome 33 and chromosome 35 in all three sandflies tested. The precise changes, and degree of ploidy variation, appeared to be strain-specific. The effect of chromosomal CNV on median gene expression in RNA-seq data was examined and for the most part was found to correlate; however, clusters of genes regulated in a CNV-independent manner were also found. Thus, while regulation of transcript abundance in response to differentiation can happen in a more gene-local fashion, there is some evidence that chromosomal CNV can represent deterministic adaptation to new environments. Separate work has since demonstrated other deterministic changes on *L. donovani* adaptation to *in vitro* culture, particularly in chromosomes five, nine, 23 and 26, with specific selection of certain chromosomal haplotypes also evident (Prieto Barja *et al*. 2017).

Another study focused on variation on a population level between different *L. tropica* isolates (Iantorno *et al*. 2017), revealing high variability in both chromosomal and gene copy number. On correlating this to changes in the RNA-seq-derived transcriptome, these copy numbers were found to explain a high

proportion (85%) of total variation in gene expression, suggesting that CNV is a dominant factor. Furthermore, some of the genes that exhibited highest variability in copy number were identified as transporters, something the authors highlighted in the context of the importance of this group of genes for drug sensitivity. Overall, therefore, a picture is emerging of CNV in *Leishmania* as highly important for population-based variation in gene expression, with aneuploidy also potentially having an adaptive role in cell cycle differentiation. In this context, the data presented in this chapter concur with a dominant role for chromosomal CNV in influencing expression differences between cell lines. However, there is little to suggest here that it plays a strong role in adaptation to drug exposure, perhaps because the resulting expression differences are too moderate to constitute a meaningful adaptive response to such a strong selection pressure. On the contrary, local gene CNV (both losses of copy number such as in *SMT* and *MT*, amplifications such as the duplication in chromosome 20 in AmBRB) appear to be more important in strongly influencing gene function and hence drug sensitivity.

4.4 Conclusions

In the experiments described in this chapter, I sought to determine a genetic basis for the phenotypic and metabolic alterations in four AmB-resistant *Leishmania* lines, as described in the previous chapter. In doing so, I identified mutations in two genes correlating with changes to sterol biosynthesis. One of these changes is a simple homozygous SNP, in *SC5D*, which is found only in one line. The other changes appear to be structural variants at the locus of the two *SMT* gene copies, arising independently in three separate lines. The nature of these changes is far less clear, reflecting the well-known challenges faced in characterising such structural changes from short read Illumina data.

In other areas of metabolism, by contrast, there are still many unknowns. It may be that deletion of *MT* in AmBRB/cl2 explains the extensive changes to lipid metabolism in this line (along with the cross-resistance to MILT), but this is unproven. Beyond this, extensive changes to both the metabolome and metabolism-associated genes are evident, but generally speaking I have not been able to relate the two. This may be because the interactions are complex, given the numbers of changes involved. Alternatively, however, it may be that the

most important changes to metabolism are regulated at the translational or post-translational level. Proteomics offers further insight into this, although data presented from a previous study (Brotherton *et al*. 2014) suggest that even this may only present a partial picture, given incomplete coverage of the proteome in such assays. Ribosome profiling is another technique that can characterise translational regulation in greater depth, and is now being applied to kinetoplastids (Jensen *et al*. 2014, Vasquez *et al*. 2014, Smircich *et al*. 2015). Hence, this may go further in the future to explain metabolite changes such as increased trypanothione biosynthesis that have not been accounted for here.

Based on these observations, one can begin to make an assessment of the utility of RNA-seq in understanding drug resistance in *Leishmania*. As previously found in a MILT-resistant line (Vacchina *et al*. 2016), a large proportion of the transcriptome may differ in abundance between wild-type and resistant parasites. However, these changes are largely moderate in magnitude, and generally speaking, with the notable exception of translation-associated genes, do not appear to be regulated in a concerted manner, arising more due to stochastic chromosomal CNV. Nevertheless, in certain specific cases it may provide important insights into the effects of changes at the genomic level, as is particularly evident in the case of *SMT*. Moreover, RNA-seq can offer many different types of information depending on the analysis tools used, and in the next chapter there is an example of how RNA-seq can provide insight where problems with the reference genome arise. Therefore, whilst the lack of conventional transcriptional regulation and the high degree of aneuploidy would appear to limit the utility of RNA-seq in *Leishmania*, in certain contexts it may provide crucial supportive information in appreciating the functional consequences of otherwise difficult to interpret lesions at the genomic level.

Overall, amongst a large number of changes at both the genome and transcriptome level, the most promising candidates for explanation of AmB resistance are the changes to *SC5D*, *SMT* and *MT*. Whilst alterations in *SMT* expression have been noted in AmB-resistant *L. donovani* (Pourshafie *et al*. 2004, Purkait *et al*. 2012), these have never previously been linked to genetic changes. Whilst a SNP in *MT* has also been reported in AmB-resistant *L. infantum* (Fernandez-Prada *et al*. 2016), the total deletion of this transporter is also a

novel genetic phenomenon in AmB-resistance. By contrast, this is the first report of mutated *SC5D* in AmB-resistant *Leishmania*. However, based on the data so far presented, these are only associations. Proof of their role in AmB resistance and associated phenotypes remains to be demonstrated.

5 Validation of the role of mutations in AmB resistance and further exploration of the sterol C24-methyltransferase locus

5.1 Introduction

In the previous chapter, in order to understand the genetic basis of extensive phenotypic and metabolic changes observed in four *L. mexicana* lines independently selected for AmB resistance, I subjected these lines to whole genome sequencing and RNA-seq transcriptomics. In doing so, I identified large numbers of variants both at the genetic level and in terms of expression. Many of these were difficult to correlate to AmB resistance or to other changes identified in chapter four. However, three changes were of specific interest: a SNP in sterol C5-desaturase, structural variations at the sterol C24-methyltransferase locus, and the deletion of the miltefosine transporter. In the former two cases, which relate to ergosterol biosynthesis genes, this is due to the established role of sterol changes in AmB resistance described here and elsewhere. The MT, on the other hand, has previously been implicated in AmB-MILT cross-resistance.

Whilst these changes stood out as of particular importance, the results presented so far do not explicitly demonstrate a functional role. The approach I present in this chapter, therefore, is to reintroduce genes into the resistant lines in which they are mutated, in order to determine whether this can restore wild-type phenotypes. First amongst these phenotypes is sensitivity to AmB itself; however, since PENT hypersensitivity appears to be a universally conserved phenotype amongst drug resistant lines, I also wished to determine which changes were responsible for this phenotype. Finally, as loss of ergostane-like sterols is a key feature of AmB resistance, I examined the effects of reintroducing these genes on sterol composition.

After describing these results, the remaining portion of this chapter then focuses in greater detail on the *SMT* locus. There are a number of reasons for this. Because C24-methyltransferase activity is not involved in mammalian sterol biosynthesis, SMT has been considered as a drug target, thus its biology is of interest (Haughan *et al*. 1995). Furthermore, of all the changes described, the

structural variations evident at the locus encoding the *SMT* gene copies are both the most widespread (being found in three separate lines) and the most difficult to understand. Whilst I presented a model in the previous chapter, this is lacking a full evidential basis and questions remain. In contrast, the mechanisms of disruption of the genes encoding *SC5D* (a simple homozygous mutation, G139R, in a site of putative importance for catalysis) and *MT* (total deletion of the gene) appear to be far clearer in terms of their functional impact. This difference in clarity arises because short read sequencing struggles to characterise large structural changes, particularly where there is repetition, as seen at this locus. The reads generated here in the WGS data are only 75 bp long, therefore the vast majority that map to the coding sequences of *SMT* (1062 bp) will not intersect with the only site that distinguishes *SMT* genes *LmxM.36.2380* and *LmxM.36.2390,* position 391. Paired-end sequencing helps by linking each read with another separated by a given range of insert sizes, thus aiding mapping of a read if its partner can be aligned to a unique position with greater certainty. Use of read pairs also helps with characterisation of structural variants, since each read in a pair may map to either side of a newly created join or deletion in the DNA. However, this again relies on unique mapping of individual pairs. Even with the use of tools specifically designed to interpret structural variants from these data, complex structural changes may not be resolved by Illumina short read sequencing alone. Therefore, in this chapter, I have combined inferences made from NGS data with various additional experimental techniques, including Sanger sequencing and qPCR, in order to understand these changes.

Precise characterisation of these structural changes is of medical importance for a number of reasons. Firstly, altered sterol C24-methyltransferase activity has been detected previously in *L. donovani* both at the level of sterol composition (Mbongo *et al*. 1998, Purkait *et al*. 2012) and transcript expression (Pourshafie *et al*. 2004, Purkait *et al*. 2012), as well as in three out of four *L. mexicana* lines selected here. Thus, loss of this activity appears to be particularly common as a mechanism of AmB resistance, including in a clinical isolate (Purkait *et al*. 2012). Secondly, the mechanism of resistance appears to be distinctive. Previous reports have shown that changes to SMT in *L. donovani* involve two distinct transcripts (as in *L. mexicana*) (Pourshafie *et al*. 2004, Purkait *et al*. 2012). Through use of PCR primers and Northern blot probes specific to the 3'-UTRs of

each transcript, as identified by 3'-rapid identification of cDNA ends (RACE), Pourshafie and colleagues demonstrated that in wild-type cells, one transcript, which they called SCMT1A, is more highly expressed than the other, SCMT1B. By contrast, in an AmB-resistant line, SCMT1A is absent, but expression of SCMT1B is greatly increased. Furthermore, by using PCR primers specific to the 5' splice leader RNA and the *SMT* coding sequence, they could detect evidence of proper *trans*-splicing in wild-type but not AmB-resistant cDNA. Similar changes were identified by Purkait and colleagues. In neither case was the genetic basis of this investigated. However, in the previous chapter I showed evidence that loss of total *SMT* RNA expression was associated with structural variants and possible evidence of CNV. Therefore, an explanation for the changes reported by Pourshafie and colleagues is that through a similar mechanism, the gene encoding SCMT1A is lost in AmB-resistant *L. donovani* genomes. If this is the case, the apparently conserved nature of changes to the *SMT* locus is striking. Whereas the changes to other sterol biosynthesis genes, *SC5D* (reported here) and *CYP51* (Mwenechanya *et al*. 2017), are SNPs, the *SMT* locus appears to be particularly susceptible to disruption by structural variation, putatively involving CNV.

CNV has previously been implicated in drug resistance in *Leishmania* (discussed in the introduction to Chapter 4). CNV events occur stochastically but their precise location is not random. The presence of extrachromosomal circular and linear amplicons within a population of *Leishmania* promastigotes has been assessed previously using sensitive PCR assays (Ubeda *et al*. 2014). It was found that many amplicons could be detected that were predicted from repeat elements interspersed throughout the *Leishmania* genome. The repeat elements detected through computational analysis of *L. major* and *L. infantum* genomes revealed a high degree of conservation between species, and high prevalence of sequences from two families of repeats, the small interspersed degenerate retroposons (SIDERs) 1 and 2. These families have been characterised previously across the genomes of *L. braziliensis*, *L. infantum* and *L. major*, with SIDER1 elements characterised as more heterogeneous than SIDER2 elements, and with the former shown to be derived from degenerated *ingi*/L1Tc-related elements (DIREs) (Smith *et al*. 2009). SIDERs appear to function both as sites for recombination and structural variation, and, when present in UTRs (particularly

3'-UTRs), as regulatory elements. In particular, SIDER2 elements within 3'-UTRs are associated with mRNA destabilisation and lower basal transcript abundance (Bringaud *et al*. 2007). SIDER1s, on the other hand, have been linked to translational control. A conserved 450 nucleotide element, later identified as a SIDER1 (Smith *et al*. 2009) was found in the 3'-UTRs of several genes, including members of the amastigote surface protein gene family, the amastins (Boucher *et al*. 2002). This element was required for increased translational efficiency on differentiation from promastigotes to amastigotes. Use of a luciferase reporter assay showed that this element was also required for increased translation in response to heat shock, but was not involved in changes in mRNA abundance in response to signals such as pH (McNicoll *et al*. 2005). Assessment of the regulatory role of SIDER1 elements has not been performed on a genome-wide scale, however, so it is unclear whether they have the same function at other loci.

The question of how specific structural variations arise is not merely an academic one. If certain loci contain sequence features that predispose them to structural instability, this means that the emergence of resistance due to disruption at such loci may be more likely to arise due to structural variation, rather than single nucleotide mutations. This is consequential from the perspective of screening for mutants. Screening for mutations in *SC5D*, such as G139R, can be accomplished simply by amplifying and sequencing the coding sequence. However, with the exception of the F72C mutation in AmBRC/cl3, the variants present at the *SMT* locus are not immediately indicative of functional change, being already present in wild-type sequences. Therefore, sequencing of the coding region alone may not be enough to detect novel variants in this gene related to drug resistance. Reduced *SMT* transcript abundance may be a conserved feature of these changes, but the instability of RNA makes expression analysis an undesirable approach. Understanding the nature of structural changes at this locus is therefore crucial for the design of tests that can detect resistance-associated genomic variants.

Chapter 5

5.2 Results

5.2.1 Generation and characterisation of episomal expression lines

In order to determine the phenotypic roles of the mutations described above, I transfected lines with plasmids expressing the genes of interest. The lines generated were as follows: for AmBRA/cl1, I transfected both the wild-type copy of *SC5D* gene *LmxM.23.1300*, as well as the mutated gene carrying the G415C nucleotide substitution, as amplified from the AmBRA/cl1 genome. Similarly, *SMT* was transfected into AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, with both wild-type *LmxM.36.2380* (T215/G391/G961), and *SMT* amplified from the individual lines (T215/A391/A961 in AmBRB/cl2, G215/A391/A961 in AmBRC/cl3, and T215/G391/A961 in AmBRD/cl2). Finally, the miltefosine transporter gene *LmxM.13.1530* was transfected into AmBRB/cl2. In addition, the same plasmid, pGL1132, was transfected into wild-type parasites containing the gene for green fluorescent protein (GFP) as a control for the effects of plasmid transfection and growth in the antibiotic selection marker for this plasmid, G418. I could observe green fluorescence in this line, confirming expression at the protein level.

In transfected lines, I assessed the degree of expression by qRT-PCR. For all *LmxM.13.1530* and *LmxM.23.1300* constructs, there was a mean of greater than tenfold increased expression in comparison to wild-type parasites across three biological replicates, although the cycle threshold (Ct) values were outside the linear range of the primers used (Supplementary file 2-2). Higher variability between lines was observed for the *SMT* constructs (Figure 5-1). Expression was significantly higher than wild-type in all cases ($P < 0.05$), ranging from 4.9-fold to 16.3-fold. However, there were differences both between strains and between copies within the same strain. For example, expression of wild-type *LmxM.36.2380* was higher than the copy derived from the resistant line itself in all cases, significantly in AmBRB/cl2 ($P = 2.68 \times 10^{-4}$) and AmBC/cl3 ($P = 0.0292$) but not in AmBRD/cl2 ($P = 0.0776$). Between strains, AmBRC/cl3 showed the highest degree of expression, followed by AmBRB/cl2 and then AmBRD/cl2. Therefore, there may be strain-specific differences in the ability to support plasmid-mediated expression. On the other hand, it is less clear why there might be differences between wild-type *LmxM.36.2380* (*SMT* WT) and resistant line-

derived (*SMT* R) *SMT* genes. The overexpression vectors are not integrated, hence in all cases there is likely to be heterogeneity in terms of copy number within a population as a whole. Whilst the sequence differences are very small (1-3 nucleotides out of 1062) this may have an influence. Alternatively, the AmBRB/cl2 and AmBRD/cl2 lines expressing *SMT* R were constructed at a later date, which may also affect results. It is possible, however, that this pattern is coincidental.

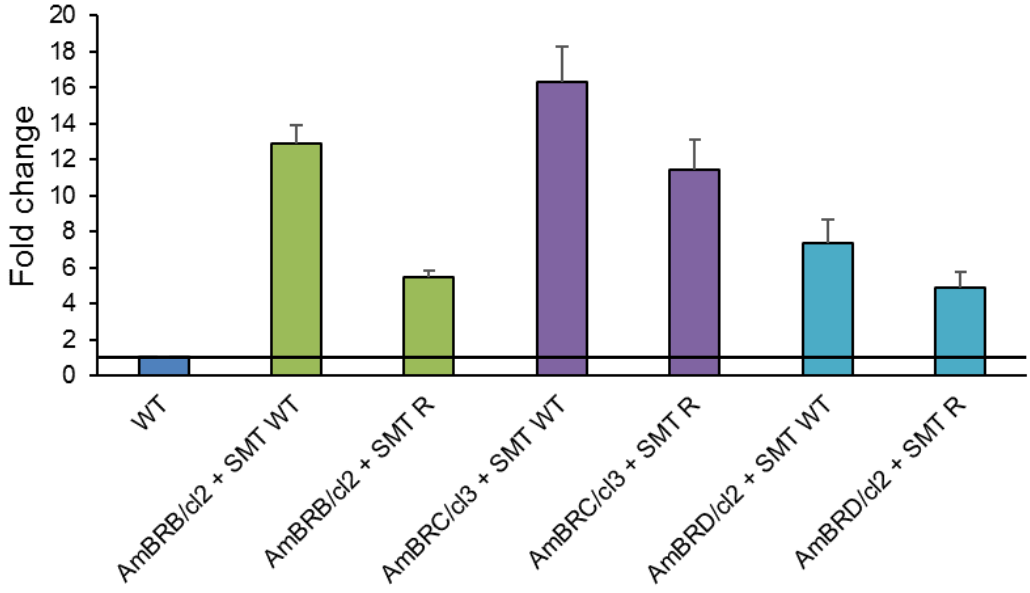


Figure 5-1: *SMT* expression in overexpression lines. Expression of *SMT* was compared between wild-type and transfected lines by qRT-PCR. Each resistant line was transfected with a plasmid encoding either the wild-type *SMT* gene *LmxM.36.2380* or the *SMT* sequence amplified from the resistant line itself (genotypes are described within the text). P values for overexpression lines for a Student *t*-test performed on δCt values are, from left to right, 5.50×10^{-7}, 2.28×10^{-5}, 1.15×10^{-5}, 1.92×10^{-6}, 8.71×10^{-5} and 2.72×10^{-4}. The vertical line marks a fold change of one (no change from wild-type).

5.2.1.1 Determination of sensitivity to AmB in overexpression lines

In order to determine a functional role of candidate genes in AmB resistance, I measured the AmB IC_{50} value in overexpression lines (Figure 5-2A, Table 5-1). In all lines transfected with wild-type gene copies, there was a significant difference in IC_{50} compared to the parental resistant line ($P < 0.05$). Introduction of *GFP* into wild-type cells had no effect on IC_{50}. AmBRA/cl1 expressing wild-type *SC5D* showed no difference in IC_{50} from wild-type parasites, suggesting that introduction of this gene completely restored wild-type AmB sensitivity. On the other hand, lines transfected with *MT* and wild-type *SMT* retained significantly higher IC_{50} values than wild-type parasites.

In order to determine the effect of mutations on gene function, I also measured the AmB IC_{50} of lines transfected with resistant line-derived gene copies. Transfection of *SC5D* carrying the G415C nucleotide substitution (encoding G139R) had no effect on IC_{50} compared to the parental AmBRA/cl1 line and the IC_{50} of this line was significantly different from AmBRA/cl1 expressing wild-type *SC5D* ($P = 1.79 \times 10^{-4}$). This suggests that enzyme function was totally lost in the mutated copy of this gene. In all cases where resistant line-derived *SMT* copies were transfected, there was significant restoration of drug sensitivity in comparison to the parental resistant line, although IC_{50} values remained significantly higher than wild-type parasites. Comparison between wild-type and mutant copies of the gene revealed no difference for AmBRB/cl2 ($P = 0.471$) and AmBRD/cl2 ($P = 0.263$). This suggests that differences at the G391A and G961A positions encoding V131I and V321I substitutions have little or no effect on enzyme function. On the other hand, there was a significant difference in IC_{50} ($P = 0.0161$) between AmBRC/cl3 expressing wild-type and AmBRC/cl3-derived copies of *SMT*. Whilst there are three differences in amino acid sequence between these two copies (V131I, V321I and F72C), the fact that the former two do not influence IC_{50} in AmBRB/cl2 provides evidence that the F72C mutation is responsible for this difference. This difference is fairly minor; however, it is worth noting that in this system both wild-type and resistant copies of the gene have higher expression than wild-type cells (16.3-fold and 11.4-fold), hence it may be that the effect of these variants on function is underestimated since loss of function may be compensated for by higher expression.

5.2.1.2 Determination of sensitivity to PENT in overexpression lines

As all resistant lines were reported to be hypersensitive to PENT, I determined whether reintroduction of genes would restore wild-type sensitivity (Figure 5-2B, Table 5-1). In this case, I only examined lines transfected with wild-type gene copies. In all cases, transfection of candidate resistance genes partially or fully restored PENT sensitivity to wild-type levels, with the effect being statistically significant ($P < 0.05$). The degree was variable, however. Introduction of *SMT* caused full restoration of PENT sensitivity in AmBRC/cl3 and AmBRD/cl2, whereas whilst *SMT* transfection in AmBRB/cl2 caused a 3.4-fold increase in IC_{50}, it remained hypersensitive in comparison to wild-type. On the other hand, transfection of *MT* into AmBRB/cl2 caused complete restoration of PENT

sensitivity, with a 6.9-fold increase in IC_{50}. In AmBRA/cl1, however, whilst introduction of SC5D significantly increased PENT IC_{50}, the effect was very marginal, with the IC_{50} of the transfected line being only 1.18-fold that of non-transfected AmBRA/cl1. Hence reintroduction of *SC5D* had far less of an influence on PENT sensitivity than *SMT*. Finally introduction of the vector containing the coding sequence for GFP actually increased sensitivity to PENT in wild-type cells. GFP is unlikely to have a specific effect on PENT sensitivity in *Leishmania* parasites, therefore it is more likely that this change is due either to the requirement on cells to maintain the plasmid, to the introduction of the neomycin resistance gene, or even non-lethal effects of the selection marker itself. The neomycin resistance gene is an aminoglycoside-specific kinase, but has previously been shown to influence gene expression and metabolism in mammalian cells, including evidence of reduced glycolytic activity (Valera *et al*. 1994).

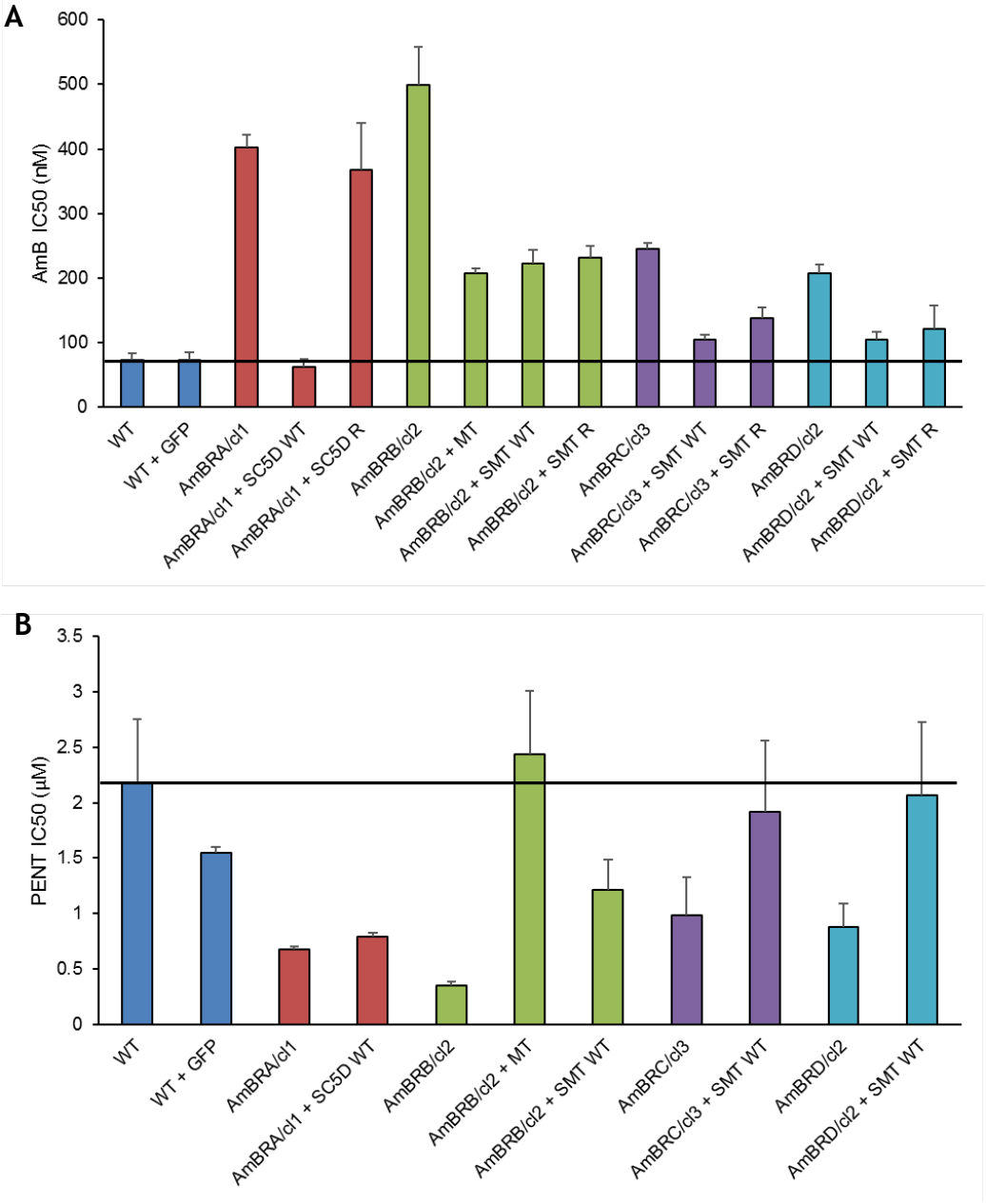


Figure 5-2: Drug sensitivity in AmB-resistant and episomal overexpression lines. Sensitivities to AmB (A) and PENT (B) were measured by Alamar blue assay. Values are plotted as the mean, with error bars representing standard deviation. The horizontal lines represent IC_{50} in wild-type parasites. See Table 5-1 for full description of the assay and P values.

Table 5-1: Drug sensitivity in AmB-resistant and episomal overexpression lines.

Sensitivities to AmB and PENT were measured by Alamar blue assay. Wild-type and AmB-resistant lines were compared to transfected cell lines containing either WT or resistant line-derived (R) gene copies (the latter only for AmB). P_1 values are derived by Student *t*-test comparison to wild-type parasites, P_2 by comparison between overexpression lines and the resistant line from which they are derived. IC_{50} values are left blank where assays were not performed, P values left blank are not applicable. For AmB assays, four biological replicates were performed with the exceptions of wild-type, AmBRB/cl2 and AmBRD/cl2 (n = 6), and AmBRB/cl2 + *MT*, AmBRB/cl2 + *SMT* WT and AmBRD/cl2 + *SMT* WT (n = 8). For PENT assays, due to high variability, five biological replicates were performed in all cases.

	AmB (nM)			PENT (μM)		
Strain	**IC_{50} ± SD**	**P_1**	**P_2**	**IC_{50} ± SD**	**P_1**	**P_2**
Wild-type	73.1 ± 9.94			2.18 ± 0.569		
Wild-type + *GFP*	72.6 ± 13.0	0.944		1.54 ± 0.0611	0.0369	
AmBRA/cl1	402 ± 19.4	4.06×10^{-10}		0.673 ± 0.0314	3.51×10^{-4}	
AmBRA/cl1 + *SC5D* WT	63.2 ± 10.9	0.175	8.26×10^{-8}	0.791 ± 0.0361	6.01×10^{-4}	5.59×10^{-4}
AmBRA/cl1 + *SC5D* R	367 ± 73.5	8.67×10^{-6}	0.389			
AmBRB/cl2	499 ± 58.4	7.42×10^{-9}		0.354 ± 0.0335	9.44×10^{-5}	
AmBRB/cl2 + *MT*	207 ± 7.99	2.64×10^{-12}	7.53×10^{-9}	2.44 ± 0.569	0.501	3.74×10^{-5}
AmBRB/cl2 + *SMT* WT	223 ± 21.4	2.14×10^{-9}	3.21×10^{-8}	1.21 ± 0.276	0.00876	1.28×10^{-4}
AmBRB/cl2 + *SMT* R	232 ± 18.0	8.56×10^{-8}	2.36×10^{-5}			
AmBRC/cl3	246 ± 9.26	3.24×10^{-9}		0.985 ± 0.342	0.00376	
AmBRC/cl3 + *SMT* WT	105 ± 8.30	7.81×10^{-4}	4.86×10^{-7}	1.91 ± 0.644	0.502	0.0216
AmBRC/cl3 + *SMT* R	138 ± 17.9	7.64×10^{-5}	3.89×10^{-5}			
AmBRD/cl2	207 ± 13.8	3.04×10^{-9}		0.877 ± 0.213	0.00134	
AmBRD/cl2 + *SMT* WT	105 ± 11.4	1.35×10^{-4}	3.47×10^{-9}	2.07 ± 0.659	0.781	0.00482
AmBRD/cl2 + *SMT* R	121 ± 36.0	0.0131	6.41×10^{-4}			

5.2.1.3 Effects of reintroducing resistance-associated genes on sterol composition

In order to determine whether reintroduction of candidate genes had an impact on sterol composition, I repeated the characterisation of sterol content by GC-MS as described in chapter 3. The overall pattern was retained in wild-type and resistant lines. In the discussion below, note that all P values discussed have been subjected to Benjamini-Hochberg multiple comparisons correction, with all values given in Supplementary file 3-1.

Comparison of wild-type and resistant parasite sterol extracts

Similar to the previous analysis of sterol composition, wild-type cells possessed an isomer of ergosterol, identified as ergosta-5,7,24(28)-trienol as the major sterol, with another, ergosta-5,22-dienol forming a substantial part (Table 5-3, Figure 5-3). As previously, sterol composition in AmBRA/cl1 was almost entirely made up of ergosta-5,22-dienol), although interestingly a substantial amount of cholesta-5,7,24-trienol was detected, something which would suggest that small amounts of C5-desaturase activity remain in this line. In AmBRB/cl2, AmBRC/3 and AmBRD/cl2, there was again loss of C24-methylation, with sterol composition dominated by cholestatrienols (83.7% in AmBRB/cl2, 89.4% in AmBRC/cl3 and 84.6% in AmBRD/cl2). In contrast to the previous experiment, substantial amounts of ergosterol were found in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, despite not being detectable above the 1% threshold in any case for the previous experiment in these lines.

Table 5-2: Sterol identification in GC-MS data.

Data are reported for sterols identified within *Leishmania* sterol extracts. Identification is either on the basis of matching to a standard or through comparison to NIST libraries, with the exception of ergosta-5,7,24(28)-trienol which was identified on the basis of comparison to literature (de Souza & Fernandes Rodrigues 2009, Andrade-Neto et al. 2016).

Retention time (min)	Major fragment mass *m/z*	Molecular ion *m/z*	Without TMS *m/z*	Putative identification	Basis of identification
10.87	368	459	386	Cholesterol	Match to standard
11.37	363	468	396	Ergosterol	Match to standard
11.36	349	455	382	Cholesta-5,7,22-trienol	NIST score 678
11.47	343	456	383	Cholesta-5,7,24-trienol	Match to desmosterol NIST score 699, wrong retention time
11.60	366	470	397	Ergosta-8,24(28)-dienol	NIST score 552
11.70	363	469	396	Ergosta-5,7,24(28)-trienol	Ergosterol NIST match score 804, wrong retention time, based on literature
11.82	343	470	397	Ergosta-5,22-dienol	NIST score 701
12.01	394	499	426	Lanosterol	Match to standard
12.38	377	483	410	4,4-dimethylcholesta-5,7,9(11)-trienol	NIST score 608

Cell line	Cholesterol	Ergosterol	Cholesta-5,7,22-trienol	Cholesta-5,7,24-trienol	Ergosta-8,24(28)-dienol	Ergosta-5,7,24(28)-trienol	Ergosta-5,22-dienol	Lanosterol	4,4-dimethyl cholesta-5,7,9(11)-trienol
Wild-type	2.74 ± 0.339		1.66 ± 0.333	2.99 ± 0.168		77.9 ± 1.21	13.5 ± 1.75		
AmBRA/cl1	2.23 ± 0.410			3.14 ± 0.959	1.46 ± 0.0313		92.2 ± 0.834		
AmBRA/cl1 + *SC5D* WT	2.20 ± 0.161		2.54 ± 0.988			83.4 ± 2.45	9.02 ± 0.786		1.09 ± 0.942
AmBRA/cl1 + *SC5D* R	1.50 ± 0.145			2.59 ± 0.648	1.39 ± 0.115		93.7 ± 0.721		
AmBRB/cl2	4.08 ± 1.20	1.48 ± 2.56	64.8 ± 4.20	18.9 ± 2.23		9.52 ± 0.283			
AmBRB/cl2 + *MT*	2.65 ± 0.476	3.73 ± 3.40	73.0 ± 1.24	12.3 ± 2.15		7.29 ± 0.470			
AmBRB/cl2 + *SMT* WT	4.26 ± 0.721		5.48 ± 1.91	2.78 ± 0.555		57.6 ± 2.31	12.9 ± 0.570		15.6 ± 1.58
AmBRB/cl2 + *SMT* R	3.93 ± 0.487		10.8 ± 1.99	5.08 ± 1.06		50.3 ± 2.57	11.4 ± 0.467		16.9 ± 3.26
AmBRC/cl3	1.88 ± 0.108	7.33 ± 0.733	70.3 ± 2.79	19.1 ± 3.47				1.01 ± 0.243	
AmBRC/cl3 + *SMT* WT	3.01 ± 0.188		5.70 ± 1.99	2.28 ± 0.538		66.7 ± 2.37	14.2 ± 0.956		7.07 ± 1.36
AmBRC/cl3 + *SMT* R	2.68 ± 0.585		28.7 ± 1.61	8.90 ± 1.08		50.9 ± 0.861	7.04 ± 0.552		
AmBRD/cl2	2.18 ± 0.153	9.51 ± 0.353	71.4 ± 3.29	13.2 ± 1.87		2.49 ± 2.16			
AmBRD/cl2 + *SMT* WT	2.70 ± 0.318		8.79 ± 2.71	2.09 ± 1.99		65.4 ± 4.71	13.7 ± 1.78		6.06 ± 0.497
AmBRD/cl2 + *SMT* R	2.07 ± 0.143		21.8 ± 5.46	7.22 ± 1.83		52.6 ± 4.90	7.76 ± 1.52		7.28 ± 1.93

Table 5-3: Sterol content of AmB resistant and episomal overexpression lines by GC-MS. Sterol content was determined by GC-MS, and values are represented as mean percentage of total sterol content (based on GC peak area) ± standard deviation. Values are excluded where the mean value for that sterol and line is less than 1%, on the basis that these values may be near the detection limit and unreliable.

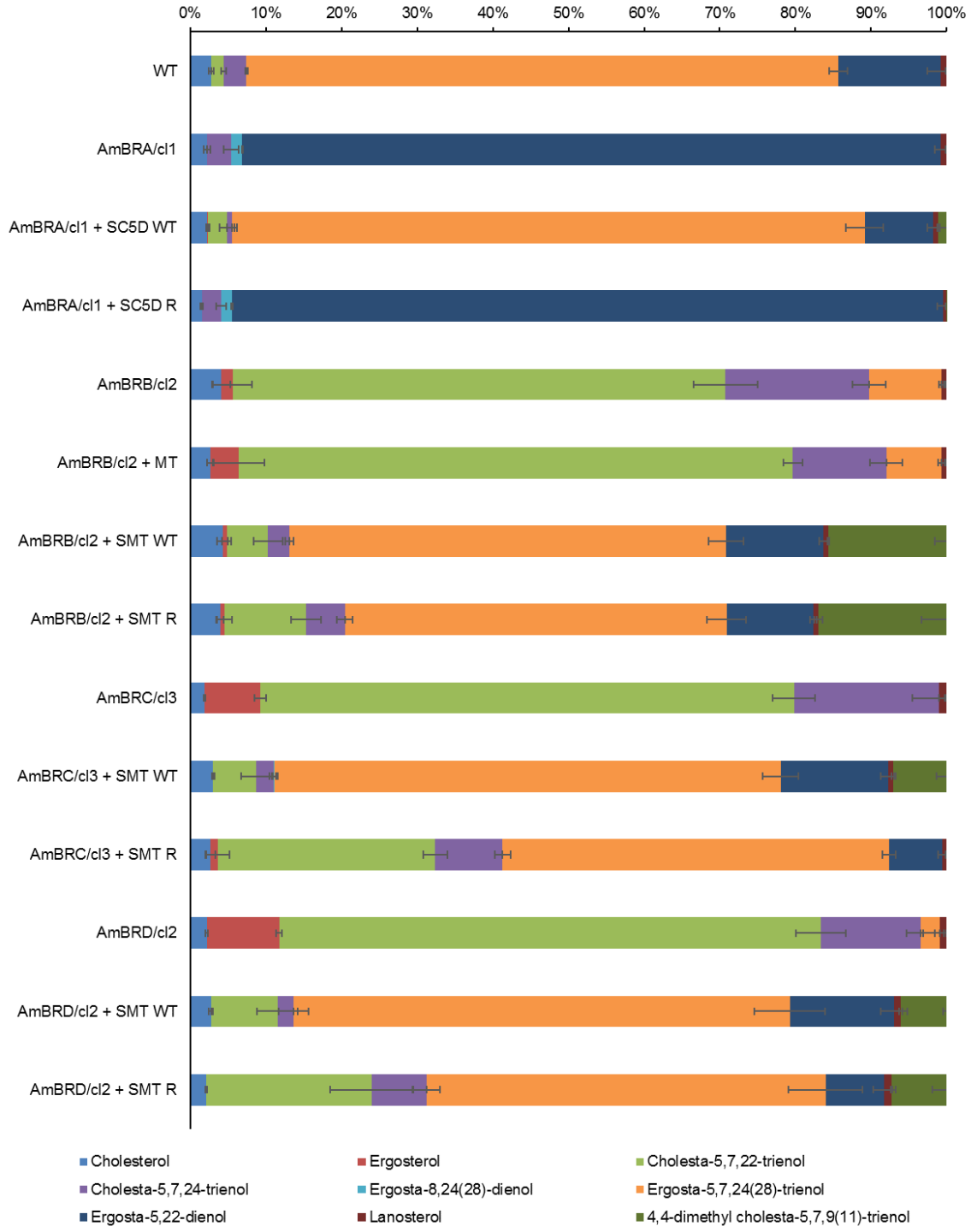


Figure 5-3: Sterol composition of AmB resistant and episomal overexpression lines by GC MS. Wild-type and AmB-resistant lines were compared to transfected cell lines containing either wild-type (WT) or resistant line-derived (R) gene copies. Stacked blocks represent the mean percentage of each sterol as a proportion of total sterol detected (based on GC peak area). Error bars represent the standard deviation across three biological replicates.

The effect of *SC5D* reintroduction on AmBRA/cl1

Reintroduction of wild-type *SC5D* into AmBRA/cl1 parasites dramatically altered the sterol composition, restoring it effectively to that of wild-type parasites, in keeping with the total restoration of wild-type AmB sensitivity in this line. One small difference was the presence of a sterol that was undetected in wild-type and all resistant lines, 4,4-dimethyl cholesta-5,7,9(11)-trienol, although this was only present at low detection levels (1.09%), and detected in only two out of three replicates. On the other hand, reintroduction of SC5D carrying the G139R amino acid substitution had no effect on sterol content, again fitting with the lack of an influence on AmB sensitivity. Given that this gene was greatly overexpressed (> 10-fold) in comparison to wild-type levels, it strongly suggests that enzyme activity is totally lost in this mutant version. One piece of evidence against this, however, is the small amount of cholesta-5,7,24-trienol detected in AmBRA/cl1 (3.14%). As SC5D activity is required for formation of the 5(6)-7(8)-double bond system, this would appear to suggest a small amount of activity of this enzyme remains.

The effect of *MT* reintroduction on AmBRB/cl2

Overall, there was little substantial effect of reintroduction of *MT* into AmBRB/cl2. Interestingly, however, there were small but significant changes, with cholesta-5,7,24-trienol showing a significant decrease, in relative abundance ($P = 0.0416$) from 18.9% to 12.3%, along with another significant decrease in abundance of the wild-type major sterol ergosta-5,7,24(28)-trienol ($P = 0.00537$) from 9.52% to 7.29%. Hence, *MT* overexpression does not restore wild-type sterol composition, if anything furthering its change from WT.

The effect of *SMT* reintroduction on AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2

As with reintroduction of wild-type *SC5D*, lines transfected with wild-type *SMT* showed large increases in content of C24-methylated sterols, with significant increases in all three transfected lines of ergosta-5,7,24(28)-trienol (to 57.6%, 66.7% and 65.4%, in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, respectively) and ergosta-5,22-dienol (undetectable in these three resistant lines, increasing to 12.9%, 14.2% and 13.7%, respectively). Notably, however, whilst ergosta-5,22-

dienol was restored to wild-type levels, ergosta-5,7,24(28)-trienol remained significantly lower than the 77.9% observed in wild-type parasites, in keeping with the lack of complete restoration of wild-type AmB sensitivity in these lines. This difference can partly be explained by the fact that whilst the major sterol in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2, cholesta-5,7,22-trienol, was greatly decreased after reintroduction of wild-type *SMT*, it remained significantly higher than in wild-type parasites. However, an extra sterol observed neither in wild-type nor resistant lines, 4,4-dimethyl cholesta-5,7,9(11)-trienol, was detected unexpectedly. This was also detected in AmBRA/cl1 transfected with wild-type *SC5D*, but here it was much more abundant, making up 15.6%, 7.07% and 6.06% of total sterol in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. Notably, this sterol does not possess C24-methylation.

Introduction of *SMT* copies derived from the resistant lines revealed similar restoration of sterol profile to wild-type *SMT*, confirming that in all cases, some enzyme function was retained despite variations in nucleotide sequence. There were, however, some differences. With regard to AmBRB/cl2, changes were of borderline significance, with lower levels of C24-methylated sterols: ergosta-5,7,24(28)-trienol ($P = 0.0416$) and ergosta-5,22-dienol ($P = 0.0495$) in AmBRB/cl2 expressing resistant line-derived *SMT* than wild-type *SMT*, and higher but not significantly different levels of cholesta-5,7,22-trienol ($P = 0.0528$) and cholesta-5,7,24-trienol ($P = 0.0527$). A similar pattern was observed with AmBRD/cl2, with reduced ergosta-5,7,24(28)-trienol ($P = 0.0539$) and ergosta-5,22-dienol ($P = 0.0245$) in the line expressing resistant line-derived *SMT* and rises in cholesta-5,7,22-trienol ($P = 0.0411$) and cholesta-5,7,24-trienol ($P = 0.0534$). Overall, however, differences in wild-type and resistant line-derived *SMT* are moderate in these lines but fit with a pattern where recovery of wild-type sterol composition is less for lines expressing *SMT* with the G391/G961 genotype than G391/A961 or A391/A961 genotypes. It should be considered, however, that whilst all episomal overexpression lines have substantially higher expression than wild-type parasites, in both cases the wild-type copies had higher expression than the resistant line derived ones (Figure 5-1).

In AmBRC/cl3, the differences between wild-type *SMT* overexpression and *SMT* carrying the T215G nucleotide substitution (in addition to having the A391/A961

genotype) were more pronounced. Ergosta-5,7,24(28)-trienol content was lower in resistant line-derived *SMT*-expressing parasites (P = 0.00138), as was ergosta-5,22-dienol (P = 0.00122), although the magnitude of these changes was similar to that observed in *SMT*-expressing AmBRB/cl2 and AmBRD/cl2. The higher level of cholesta-5,7,24-trienol was significant (P = 0.00227), but again this change was similar in magnitude to the other two lines. A much more substantial difference was observed in cholesta-5,7,22-trienol (P = 4.20×10^{-4}). Whereas the relative abundance of this sterol in resistant line-derived *SMT* compared to wild-type *SMT* was 2.0-fold and 2.5-fold higher, respectively, being of borderline significance in each case, there was a 5.0-fold difference in AmBRC/cl3 (5.7% with wild-type *SMT*, 28.7% in resistant line-derived SMT). Finally, there was a qualitative difference with regard to 4,4-dimethyl cholesta-5,7,9(11)-trienol, the sterol that was only detected in overexpression lines. Although this sterol was present at 7.07% in AmBRC/cl3 expressing wild-type *SMT*, it was not detected at all in AmBRC/cl3 expressing its own *SMT* gene copy. Therefore, the differences between lines overexpressing SMT with the wild-type *SMT* genotype of *LmxM.36.2380* and *SMT* derived from the resistant lines are both quantitatively and qualitatively more pronounced for AmBRC/cl3. These differences do not seem to be explicable by expression changes, and therefore it is very plausible that the T215G mutation, encoding the F72C mutation, is responsible for this difference.

5.2.2 Characterisation of specific *SMT* transcripts

5.2.2.1 Identification of transcript-specific sequences and design of qPCR primers

Having demonstrated that loss of *SMT* expression is responsible for AmB resistance, PENT hypersensitivity and altered sterol composition (although *MT* deletion also plays a role in one line in the former two cases), I wished to define further the exact nature of changes at this locus. One of the problems encountered previously is an evident error in assembly in the intergenic region between *SMT* gene copies, just downstream of *LmxM.36.2380* in the reference genome (see section 4.2.1.2). In order to overcome this issue, I used the RNA-seq reads to generate a *de novo* transcriptome, using the RNA-seq assembler, Trinity (v2.0.5, Grabherr *et al*. 2011). In order to generate maximum coverage,

reads were pooled from biological replicates for each strain (individual clones for resistant lines) in order to generate five libraries containing 20,514, 20,562, 20,471, 20,645 and 20,986 contigs for AmBRA, AmBRB, AmBRC, AmBRD and wild-type parasites respectively. A BLAST search was then performed in order to identify contigs matching to the coding sequence for *LmxM.36.2380*. This identified three transcripts in wild-type parasites, two each in AmBRA and AmBRD, and one each in AmBRB and AmBRC (Supplementary file 5-1). Suffixes within predicted transcript names ('_i1', '_i2' etc.) were given by the Trinity assembler and identify different predicted transcripts derived from the same line. I wished to determine whether there were differences in the UTRs between transcripts (Figure 5-4, Supplementary files 5-2 and 5-3). Alignment of 5'-UTRs revealed an almost identical sequence across all transcripts in all lines, with the exception of one transcript, AmBRD_i2, which was truncated relative to the others. All sequences except for this one began with part of the SL RNA sequence (5'-AGTATCAGTTTCTGTACTTTATTG-3') but not the full length (5'-AACTAACGCTATATAAGTATCAGTTTCTGTACTTTATTG-3'). Removing the partial SL RNA sequence showed that the remaining 234 nucleotides in the identical 5'-UTRs were identical to the 234 nucleotides upstream of *LmxM.26.2380* in the reference genome. Greater variation was displayed in 3'-UTR sequences, however, with variation in length from 263 nucleotides (WT_i2) up to 689 nt (AmBRC_i1). Of particular importance, there was a distinct cluster (AmBRA_i2, WT_i2 and WT_i3) that contained only transcripts from wild-type and AmBRA parasites, those lines which did not show disruption at the *SMT* locus.

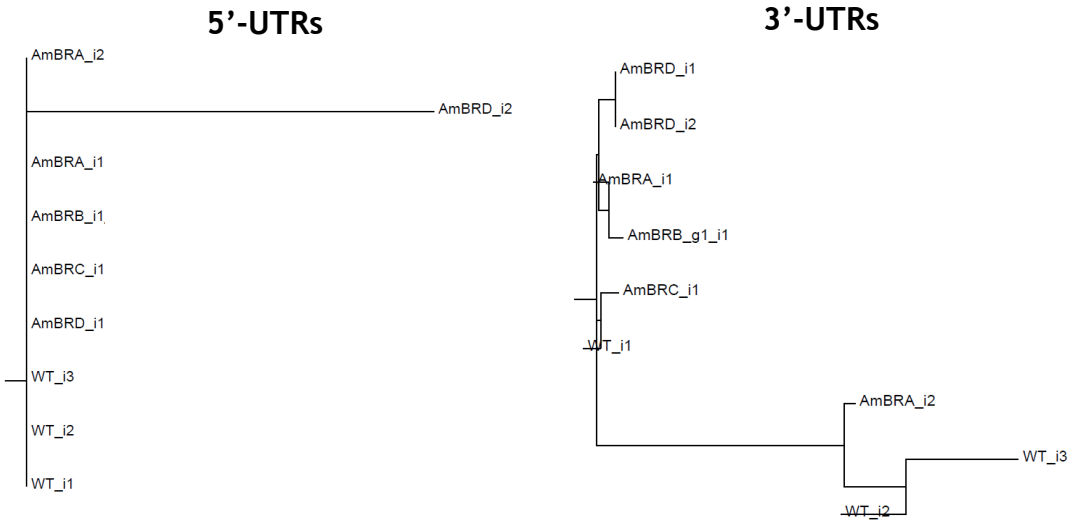


Figure 5-4: Phylograms of Trinity-derived *SMT* UTR sequences. Alignment of UTR sequences from Trinity-derived *SMT*-containing contigs was performed using Clustal Ω (Sievers *et al*. 2011). Phylogenetic tree data were generated using Clustal Ω with default settings, then visualised using the online tool Phylodendron (http://iubio.bio.indiana.edu/soft/molbio/java/apps/trees/).

Closer inspection of the aligned transcripts revealed that all transcripts were identical for the first 192 nucleotides (Figure 5-5), before diverging in a pattern that was in keeping with clustering in the phylogram. After this, there was a short region of sequence (48 nucleotides) identical between AmBRA_i2, WT_i2 and WT_i3, and a longer region of sequence (196 nucleotides) identical between all other transcripts. Downstream of these regions, there was greater variability in length and sequence content. In these unconserved regions, it is less clear how reliable the Trinity assembly is. Nevertheless, these regions of sequence identity are reproducible when assembled from biologically distinct samples. Hence, two types of transcript are evident from the Trinity assembly, of which one type is present in all lines, whereas the other is present only in wild-type and AmBRA transcriptomes.

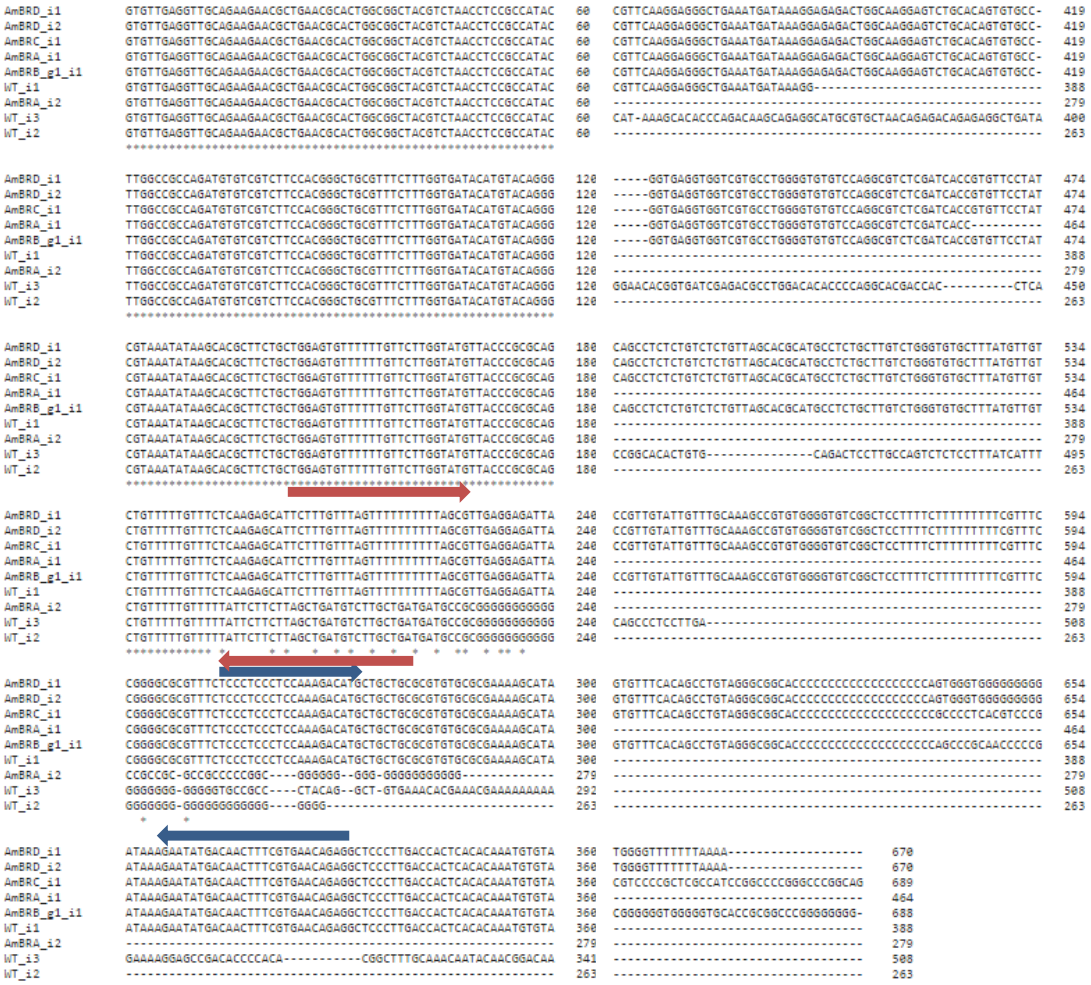


Figure 5-5: Alignment of 3'-UTRs from Trinity-derived *SMT* transcript sequences. Alignments were generated using Clustal Ω (Sievers *et al*. 2011). Red arrows below the alignment represent forward and reverse qPCR primer positions for the transcript type pertaining to AmBRA_i2, WT_i2 and WT_i3, blue arrows above the alignment represent primer positions for the transcript type present in all other lines.

In order to determine how these regions of homology related to the reference genome, I used a BLAST search. The first 192 nucleotides, which were shared between all transcripts, matched perfectly with chromosome 20 positions 955,254-955,445, which is immediately downstream of *LmxM.36.2390*. The first 117 nucleotides of this tract were also a perfect match to chromosome 20 positions 951,492-951,608, immediately downstream of *LmxM.36.2380*. The 196 nucleotide homology region common to all transcripts except AmBRA_i2, WT_i2 and WT_i3 matched perfectly to chromosome 20 positions 955,446-955,641; therefore, the first 388 nucleotides common to all transcripts except for these three are identical to the region immediately downstream of *LmxM.36.2390*. Downstream of this 196 nucleotide homology region, there were variations in sequence and length; however, these still revealed generally high homology to

the genomic region downstream of *LmxM.36.2390*. For example, 3'-UTRs found in AmBRB, AmBRC and AmBRD all demonstrated 100% sequence identity as far as position 955,872, 619 nucleotides downstream of *LmxM.36.2390*.

On the other hand, the 48 nucleotide region AmBRA_i2, WT_i2 and WT_i3 did not match to any region on chromosome 20 (the best hit in the *L. mexicana* genome was a partial alignment to chromosome 17 matching 30 out of 36 nucleotides). One possibility, therefore, is that this cluster of transcripts corresponds to *LmxM.36.2380*, which has a region of unassembled genome immediately downstream. Indeed, chromosome 20 positions 951,609-951,708 are marked by a string of 100 N nucleotides. This explains why 3'-UTRs match downstream of *LmxM.36.2380* for the first 117 nucleotides up to this point, but no further.

In order to target specific transcripts, I used the Trinity-derived predicted 3'-UTR sequences to design qPCR primers. To get the most reliable regions of sequence, I designed the primers to bind to regions that were identical between sequences within the same type of UTR. Hence for the 3'-UTR type present in all but AmBRA_i2, WT_i2 and WT_i3, I designed primers binding to the 196 nucleotide homology region specific to this transcript type (Figure 5-5). On the other hand, as the 48 nucleotide homology region specific to these three transcripts was too small to encompass both forward and reverse qPCR primers, I designed a forward primer binding to the 192 nucleotide homology region common to all transcripts, and a reverse primer within this 48 nucleotide region.

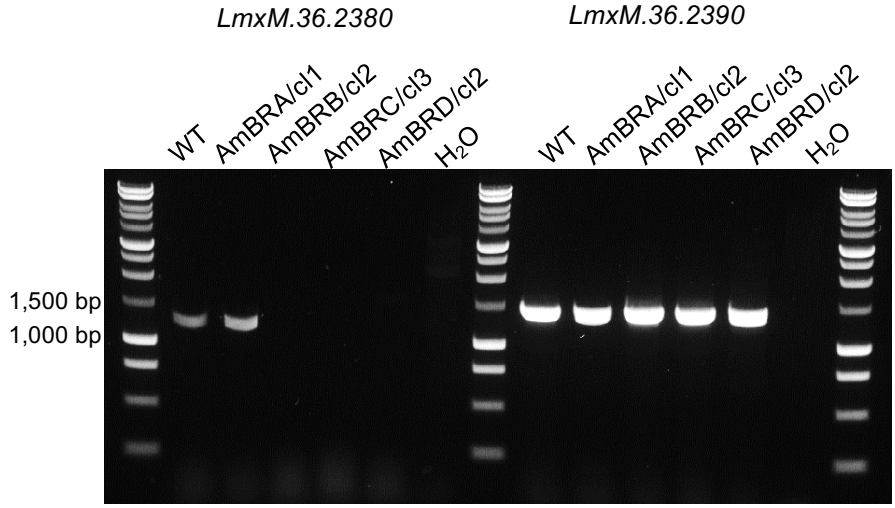


Figure 5-6: Amplification of coding sequences using transcript-specific reverse primers. Amplicons were derived by PCR using the forward primer used for gene amplification and cloning into the pGL1132 vector, and reverse primers binding to 3'-UTRs in a transcript-specific fashion, as described in the text. Products were visualised by electrophoresis using a 1% agarose gel. DNA standards (Promega 1 kb ladder) were run in parallel, with the positions of 1,500 bp and 1,000 bp standards indicated above. The left side of the gel represents amplification using a reverse primer specific to the 3'-UTR of the transcript type identified as *LmxM.36.2380*, the right using a primer specific to the *LmxM.36.2390*-like 3'-UTR.

In order to further validate the binding sites of the reverse primers, I amplified genomic DNA using these in combination with a forward primer at the start of the *SMT* CDS (Figure 5-6). In both cases, amplicons were 1-1.5kb in length. As predicted, whilst the one using a reverse primer binding to the Trinity transcript type found in all lines (containing the 196 bp homology region) could be amplified from genomic DNA in all cases, the other, using a reverse primer specific for the transcript type found only in wild-type and AmBRA transcriptomes (the 48 bp homology region), could only be amplified from these two lines. Sequencing of both amplicons derived from wild-type found that the amplicon derived using the wild-type/AmBRA-specific reverse primer possessed G residues at positions 391 and 961, whereas the other possessed A residues at these loci. Therefore, this explicitly relates the 48 bp homology region to *LmxM.36.2380*, and the 196 bp homology region to *LmxM.36.2390*. Furthermore, as predicted by the alignment of WGS data, it also demonstrates that *LmxM.36.2380* contains a G at position 961, *LmxM.36.2390* an A.

5.2.2.2 Testing of transcript-specific qPCR primers

I then used these primers to compare expression of *LmxM.36.2380* with *LmxM.36.2390* in wild-type cells. Total *SMT* was included for comparison. In

order to control for efficiency of amplification of different qPCR products, I normalised the data using a gDNA standard curve (remembering to correct for the fact that there are two expected *SMT* copies). Samples were then controlled for loading by division by a similar corrected value for GAPDH (Figure 5-7). This revealed that *SMT* was highly expressed (six-fold higher than cytosolic GAPDH). As predicted, there was a large discrepancy between transcript types, with expression of *LmxM.36.2380* being 28-fold higher than *LmxM.36.2390*. A surprising result was that the signal for mean normalised *LmxM.36.2380* expression was actually slightly higher than overall *SMT* expression. This was close to statistical significance, but did not pass this threshold (P = 0.0881). Possible explanations for this include non-specific amplification with *LmxM.36.2380* 3'-UTR-specific primers, but there was no evidence of this when dissociation curves were obtained (Supplementary file 2-2). Alternatively, since the difference was marginal, it may simply reflect the limits of precision of this assay. It is clear, however, that *LmxM.36.2380* dominates *SMT* expression in wild-type parasites.

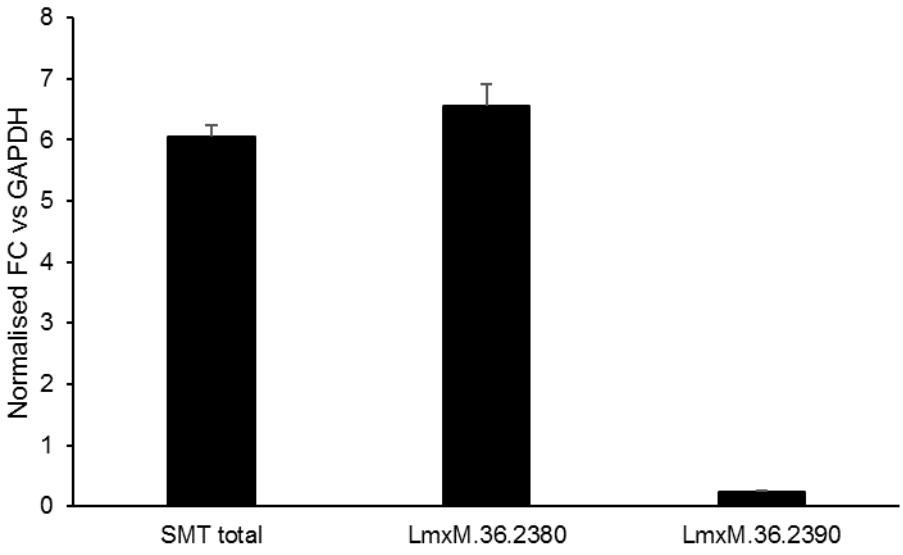


Figure 5-7: Expression of *SMT* genes in wild-type parasites. Expression was determined by qPCR. Ct values were converted into an equivalent input of genomic DNA by use of a standard curve of genomic DNA from 10 ng down to 0.078 ng. Samples were then controlled for loading by division of the ng genomic DNA equivalent by that of GAPDH. Mean ratios are plotted with error bars representing standard deviation.

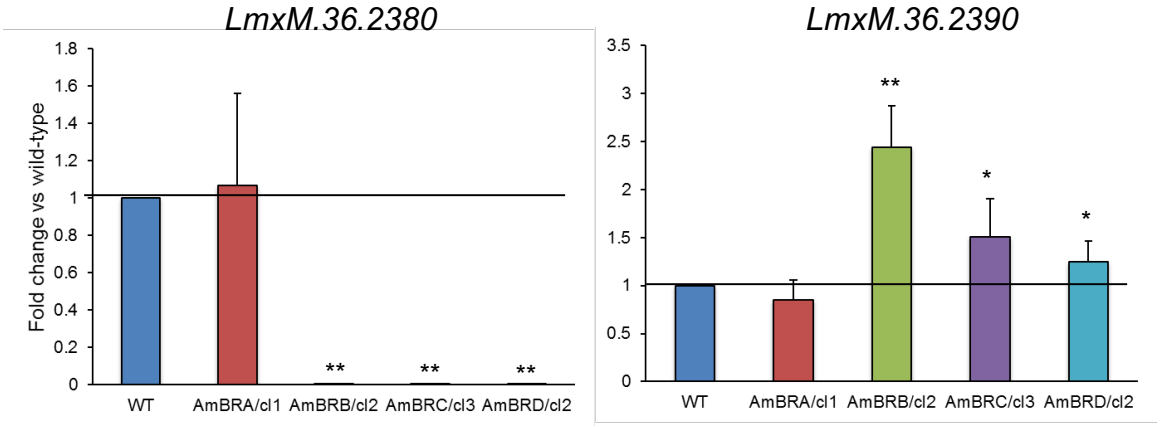


Figure 5-8: Transcript specific expression analysis in wild-type and AmB-resistant parasites. Expression was determined by qRT-PCR using primers specific to the 3'-UTRs of *LmxM.36.2380* (left) and *LmxM.36.2390* (right). Mean fold change values across three biological replicates are plotted, with error bars representing standard deviation. Asterisks mark statistically significant changes in δCt from wild-type (*P < 0.05, **P < 0.005). Horizontal lines represent no change compared to wild-type.

I employed the transcript-specific primers in order to determine expression of each transcript in wild-type and resistant parasites by qRT-PCR on parasite RNA extracts (Figure 5-8). Primers targeting the 3'-UTR of *LmxM.36.2380* revealed no change in expression of this transcript in AmBRA/cl1 (P = 0.999), whereas this transcript was effectively lost (Ct > 35 or not detected) in AmBRB/cl2 (P = 4.05 x 10^{-5}), AmBRC/cl3 (P = 2.21 x 10^{-4}) and AmBRD/cl2 (P = 0.00318). In contrast to this, the 3'-UTR of *LmxM.36.2390* was detectable in all lines, and in fact increases were observed. AmBRB/cl2 showed the biggest increase in expression of this transcript (2.44-fold, P = 4.71 x 10^{-4}) whereas AmBRC/cl3 and AmBRD/cl2 showed more modest but still significant increases of 1.51-fold (P = 0.0228) and 1.25-fold (P = 0.0430), respectively.

5.2.3 Characterisation of genome-level changes at the *SMT* locus

5.2.3.1 Validation of copy number changes by qPCR

In the previous chapter, I noted that the WGS suggested there were copy number changes in *SMT* genes but were unable to specify precisely what these changes were, due to limitations in the ability to distinguish between gene copies. I therefore employed a qPCR approach in order to quantify changes both in total *SMT* copy number and in transcript-specific 3'-UTR regions (Figure 5-9). As with qRT-PCR expression analysis, Ct values were normalised using cytosolic GAPDH

(*LmxM.36.2350*) as a loading control. As this is on the same chromosome (20) as the *SMT* genes, copy number changes are due to local variation rather than chromosomal aneuploidy. Comparison between lines revealed no significant changes in total *SMT* copy number in AmBRA/cl1 (P = 0.721), as well as in AmBRB/cl2 (P = 0.693). In contrast, AmBRC/cl3 and AmBRD/cl2 showed significant decreases of 0.53-fold (P = 0.00302) and 0.52-fold (P = 0.00256), respectively. Use of qPCR primers specific to the 3'-UTR of *LmxM.36.2380* again showed no change in AmBRA/cl1 (P = 0.312); however, in AmBRB/cl2 (P = 1.52 x 10^{-5}), AmBRC/cl3 (P = 5.78 x 10^{-6}) and AmBRD/cl2 (P = 4.39 x 10^{-4}), the target sequence was essentially undetectable (Ct > 35).

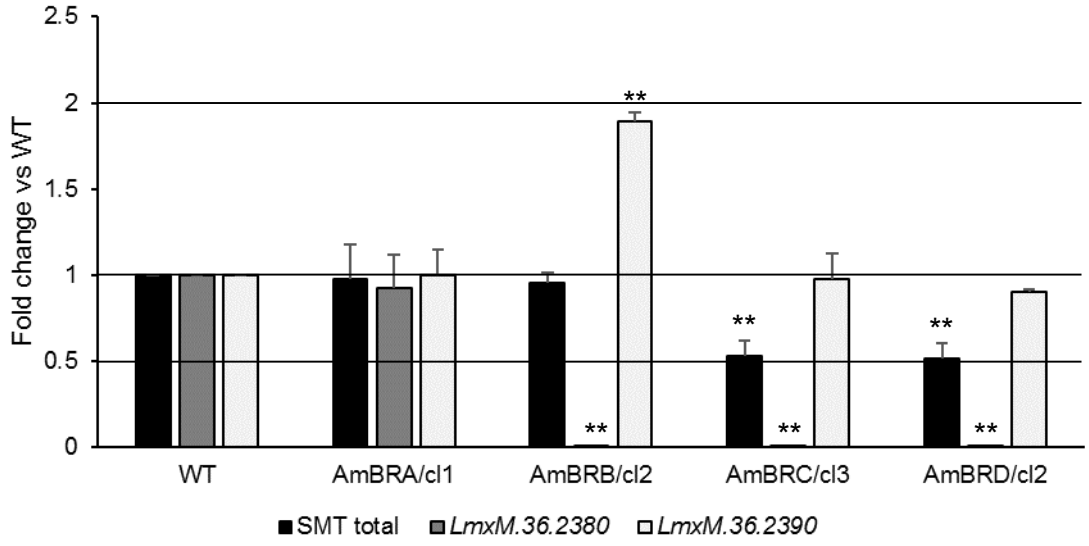


Figure 5-9: Copy number changes at the *SMT* genomic locus. Changes in genomic DNA copy number were determined by qPCR. Mean values of three biological replicates are plotted, with error bars representing standard deviation. Asterisks mark statistically significant changes in δCt from wild-type (**P < 0.005). Horizontal lines represent 50%, 100% and 200% of wild-type expression. Note that values relate to *fold change* relative to wild-type rather than absolute copy number.

In contrast to this, the 3'-UTR region of *LmxM.36.2390* was detectable in all lines, with no significant changes in AmBRA/cl1 (P = 0.928), AmBRC/cl3 (P = 0.621) and AmBRD/cl2 (P = 0.255). However, a 1.89-fold increase (P = 0.00105) was seen in AmBRB/cl2. Therefore, based on these data, it appears that at the genomic level, 50% loss of total *SMT* copies in AmBRC/cl3 and AmBRD/cl2 is explained by retention of *SMT* with an *LmxM.36.2390*-type 3'-UTR, but loss of *SMT* with an *LmxM.36.2380*-type 3'-UTR. Similarly, in AmBRB/cl2, the *LmxM.36.2380*-type 3'-UTR is absent; however, the *LmxM.36.2390*-type 3'-UTR

doubles in copy in this line, suggesting that the lack of overall change in *SMT* copy in AmBRB/cl2 results from loss of *LmxM.36.2380* associated with duplication of *LmxM.36.2390*.

This duplication of *LmxM.36.2390* in AmBRB/cl2 appears to be associated with a broader duplication event extending from this locus over a region of around 50 kb to *LmxM.36.2540* (sterol C4-methyloxidase). I wished to verify that such a duplication could be detected by qPCR, therefore I designed primers specific to the gene adjacent to the *SMT* locus, *LmxM.36.2400* (hypothetical protein), and the genes before and after the end of this duplication, *LmxM.36.2540* and *LmxM.2550* (hypothetical protein). In Figure 5-10, values are shown directly as 2^{-Ct}, representing relative expression without normalisation. This is in order to show the consistency of Ct values across different genes in wild-type cells, providing evidence that, with the notable exception of *SMT* (which has an expected haploid copy number of two) the basal haploid copy number of all other genes investigated here is the same. These data clearly show that whilst copy number is unchanged either side of the duplication event (in *LmxM.36.2350* and *LmxM.36.2550*, P = 0.603 and P = 0.485, respectively), the two genes within the duplication event, *LmxM.36.2400* and *LmxM.36.2540*, show 2.0-fold (P = 8.81 x 10^{-4}) and 1.9-fold (P = 0.00188) increases, respectively. This therefore confirms the observation made from per-gene coverage in the WGS data, as described in the previous chapter, that there is a large multi-gene duplication event on chromosome 20 in AmBRB/cl2. Duplication of the *LmxM.36.2390*-like 3'-UTR in this line, which is found in the intergenic region between *LmxM.36.2390* and *LmxM.36.2400*, provides further evidence that this duplication extends right into the *SMT* locus itself.

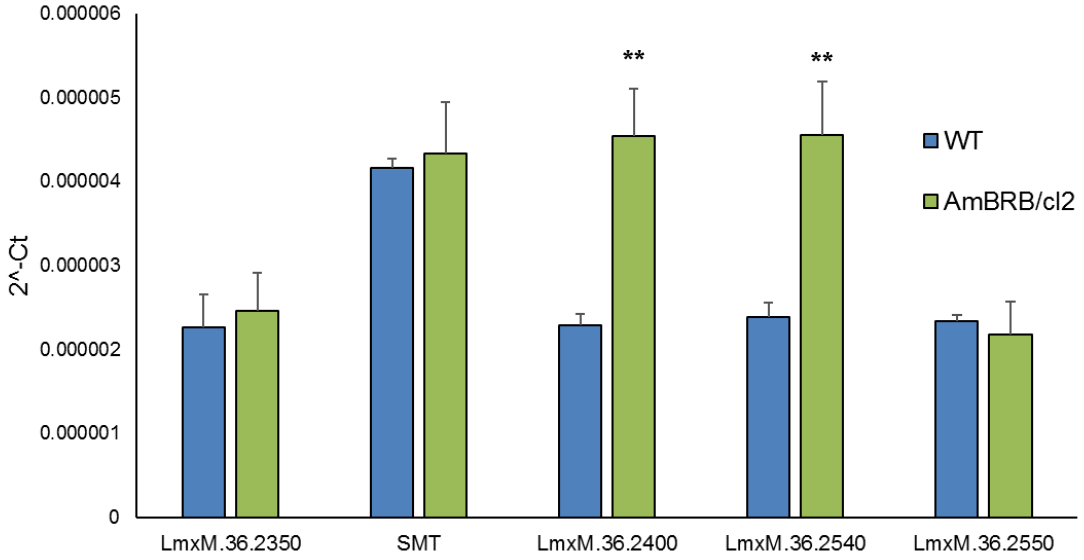


Figure 5-10: Confirmation of the regional duplication in chromosome 20 in AmBRB/cl2. Copy number changes were determined by qPCR of genomic DNA. In this graph, raw 2^{-Ct} values are plotted without normalisation to cytosolic GAPDH (loading is controlled based on measurement of input genomic DNA concentration alone). Mean values of three biological replicates are plotted, with error bars representing standard deviation. Asterisks mark statistically significant changes in δCt from wild-type (**P < 0.005). Values for *LmxM.36.2350* (cytosolic GAPDH) and *SMT* are derived from the same data as in Figure 5-9.

5.2.3.2 Sequencing of the intergenic region between *SMT* gene copies

It was apparent that the reference genome was incompletely assembled (both because of discontinuous coverage of WGS data and the string of 100 N nucleotides mentioned previously). Therefore, I decided to sequence this region using Sanger sequencing. I designed primers in order to amplify this region by locating the forward primer towards the 3'-end of the *SMT* coding sequence and the reverse primer at the 5'-end. This yielded a 3-4 kb band that could be amplified from WT and AmBRA/cl1 but not AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 genomic DNA. I then sequenced this region (Supplementary file 5-4 for sequence) from wild-type parasites and found it to be 2,888 bp, as opposed to 2,700 bp in the reference genome (excluding the regions of flanking coding sequence in the amplicon). Data from four different clones of the PCR amplicon were used in order to ensure accuracy and identify polymerase errors. I then realigned the WGS data to a corrected version of the genome with this region replacing the intergenic region in wild-type parasites in order to test whether this gave a more continuous alignment.

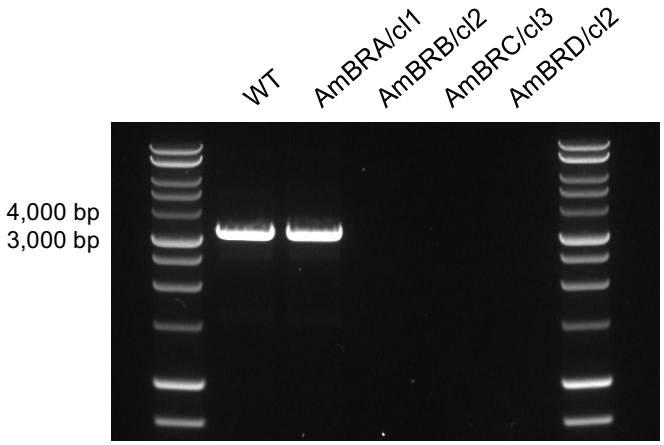


Figure 5-11: Gel image of the *SMT* intergenic region amplicon. Primers designed within the *SMT* coding sequences to amplify across the intergenic region between *SMT* copies were used to amplify by PCR this region from genomic DNA extracted from wild-type and AmB-resistant parasite lines. Samples were then visualised by electrophoresis with a 1% agarose gel and DNA standards (Promega 1 kb ladder). The position of the 3,000 bp and 4,000 bp standards are marked.

Visualisation of this alignment using IGV (Figure 5-12) showed that in wild-type and AmBRA/cl1 WGS data, continuous coverage was achieved across this whole region, in contrast to the original reference genome (Figure 4-5). There was one point just downstream of *LmxM.36.2380*, at which coverage dropped off markedly (shown with a black arrow). However, even the position of lowest coverage (951,722, 231 bp downstream of *LmxM.36.2380*) retained six-fold coverage in wild-type and nine-fold coverage in AmBRA/cl1, in contrast to the original reference in which a substantial region of no coverage was seen. This is associated with a poly(C) tract, of which during sequencing of multiple clones I found evidence of variable length, potentially explaining lower coverage at this site. AmBRB/cl2, however, did show a small gap in alignment, with no coverage from positions 951,681-951,709, and only one- or twofold coverage to 951,738. This covers the base positions 190-247 bp downstream of *LmxM.36.2380*. As the first 192 bases downstream of all *SMT* transcripts detected in the Trinity assembly are identical, the region which I have associated specifically with the 3'-UTR of *LmxM.36.2380*-like transcripts encompasses the tract 191-238 bp downstream of this gene copy. Hence, AmBRB/cl2 specifically lacks coverage at this region. Note the genomic coordinates described here appear to be identical across multiple reference genome releases on TriTrypDB (as far as version 34) but may not be in the future.

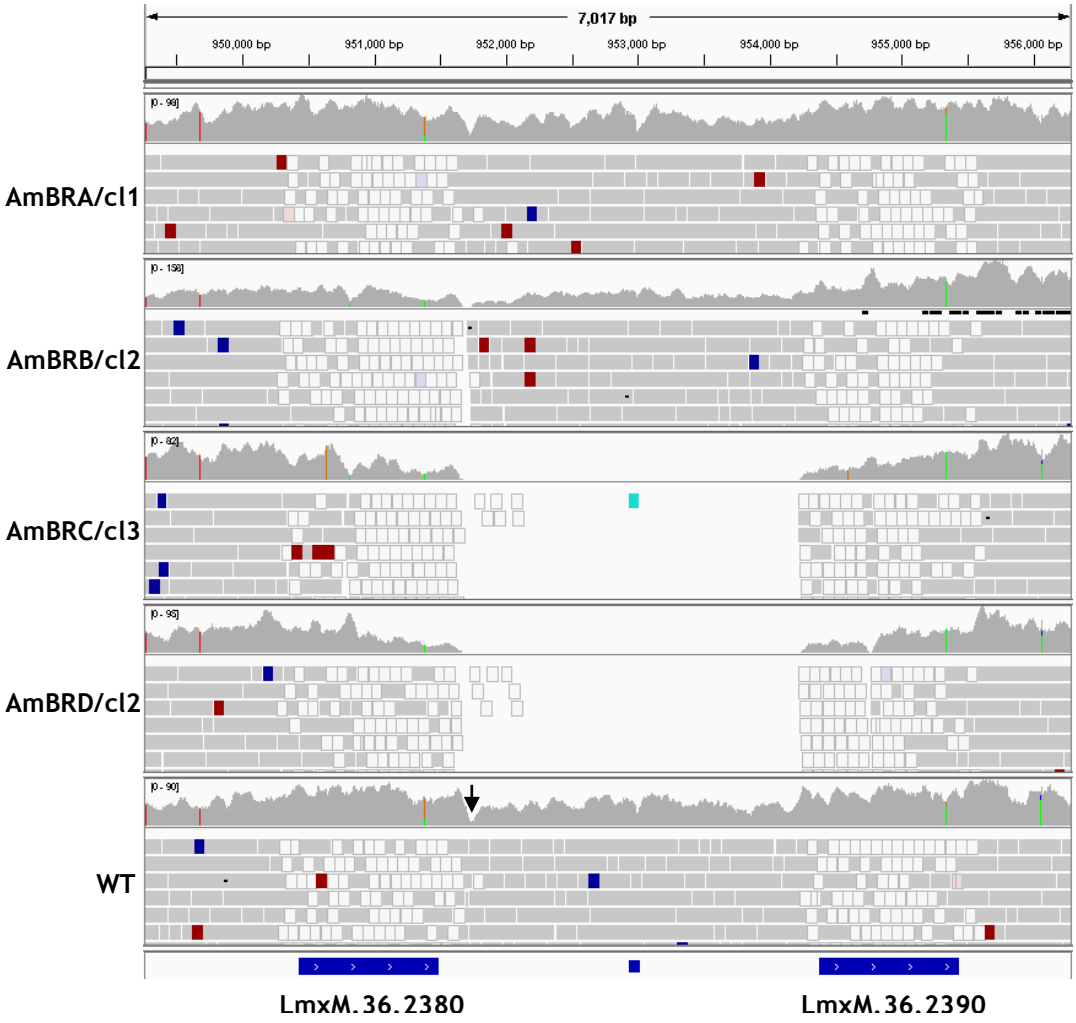


Figure 5-12: Alignment of WGS data to the *L. mexicana* genome with a corrected *SMT* intergenic region. The genomic region of *SMT* genes *LmxM.36.2380* and *LmxM.36.2390*, as visualised using the Integrative Genomics Viewer (Robinson *et al.* 2011), with the intergenic region between *SMT* gene copies replaced by the one sequenced using Sanger sequencing. Grey blocks represent aligned reads with mapping quality greater than 0, white blocks with mapping quality of 0. Blocks in other colours typically represent non-concordantly mapped reads. A black arrow marks the area where wild-type coverage drops off, as described in the text. See Figure 4-2 for a full description of the visualisation.

On the other hand, much larger sequence gaps were evident in AmBRC/cl3 and AmBRD/cl2, extending from a similar point as for AmBRB/cl2 for approximately 3 kb to just upstream of *LmxM.36.2390*. However, the presence of reads that map to a number of positions makes it difficult to determine the correct location of break points. Figure 5-13 therefore shows the same alignment with all reads with a mapping quality of 0 excluded in order to remove reads which map non-uniquely to this region. In this case, gaps can be seen in AmBRA/cl1 and wild-type DNA in regions of the coding sequences, presumably where uniquely assigning read positions is not possible. Interestingly, the gap observed in AmBRB/cl2 is also substantially wider, suggesting that the very narrow gap

observed in Figure 5-12 may be artificially narrowed by reads which can map to either *LmxM.36.2380* or *LmxM.36.2390* (or indeed the 192 bp downstream which are shared between both). Interestingly, viewed in this way, the alignments of AmBRC/cl3 and AmBRD/cl2 now look different, specifically with better coverage of *LmxM.36.2390* in AmBRC/cl3 and of *LmxM.36.2380* in AmBRD/cl2. This fits with the *LmxM.36.2380*-like genotype, G391, in AmBRD/cl2, compared to the *LmxM.36.2390*-like genotype, A391, in AmBRC/cl3.

Sequencing of the intergenic region also allowed us to test the accuracy of assembly of the 3'-UTRs from *LmxM.36.2380*-like transcripts. Despite variable 3'-UTR lengths (279 nucleotides for AmBRA_i2, 263 for WT_i2 and 508 for WT_i3), in all cases there was an identical match to the first 230 nucleotides of sequence. This is actually less than the 240 nucleotide region shared across these transcripts (the 192 nucleotide non-transcript-specific region plus the 48 nucleotide transcript-specific region). In the Trinity assembly, the last ten nucleotides of the 48 nucleotide homology region are a string of G residues, whereas in the sequenced intergenic region, a string of C residues are present. It is therefore possible that Trinity encounters problems with homopolymer tracts, although I did not find any specific examples of this.

With regard to the 5'-UTR sequences, there is evidence that in this case the Trinity assembly may be incorrect. Whilst the 5'-UTR sequence conserved across all transcripts was identical to the 234 nucleotides upstream of *LmxM.36.2380* in the reference genome (after SL RNA had been removed), alignment of this to the *LmxM.36.2390* upstream 234 nucleotides (part of the intergenic sequence) showed identity only for the first 159 nucleotides immediately upstream of this coding sequence, upstream of which there was little similarity. The AmBRD transcript with the truncated 5'-UTR did not match any better to that of *LmxM.36.2390* than the others. Therefore, it may be that because the relative abundance of *LmxM.36.2380* is so much higher and the 5'-UTR sequences are relatively short, Trinity omitted the 5'-UTR of *LmxM.36.2390* from the assembly, and because of the large, almost identical region of the CDS in between UTRs, it fixed *LmxM.36.2380*-like 5'-UTRs onto all transcripts.

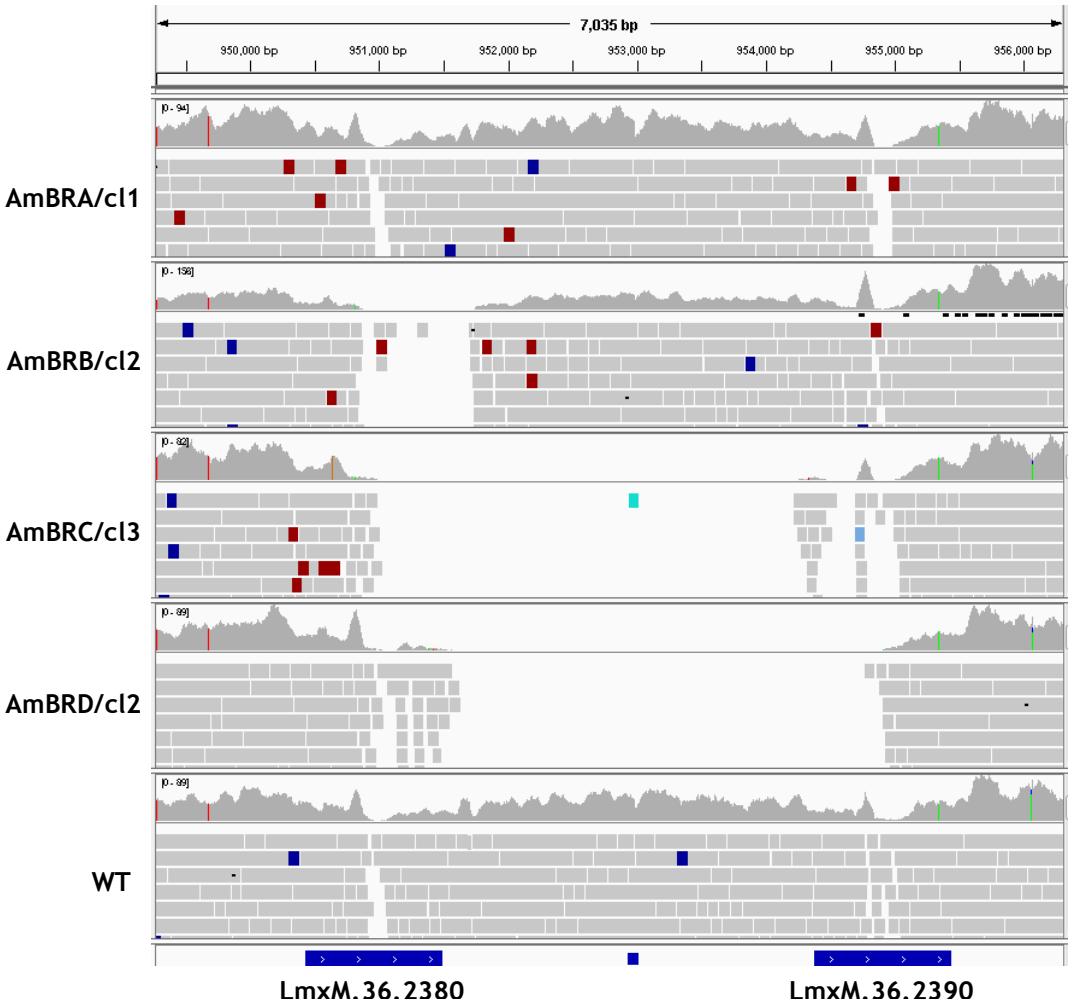


Figure 5-13: Alignment to the *L. mexicana* corrected *SMT* region applying a mapping quality threshold. This visualisation was produced in the same way as Figure 5-12 with the exception that reads with a mapping quality of 0 were excluded. See Figure 4-2 for a full description of the visualisation.

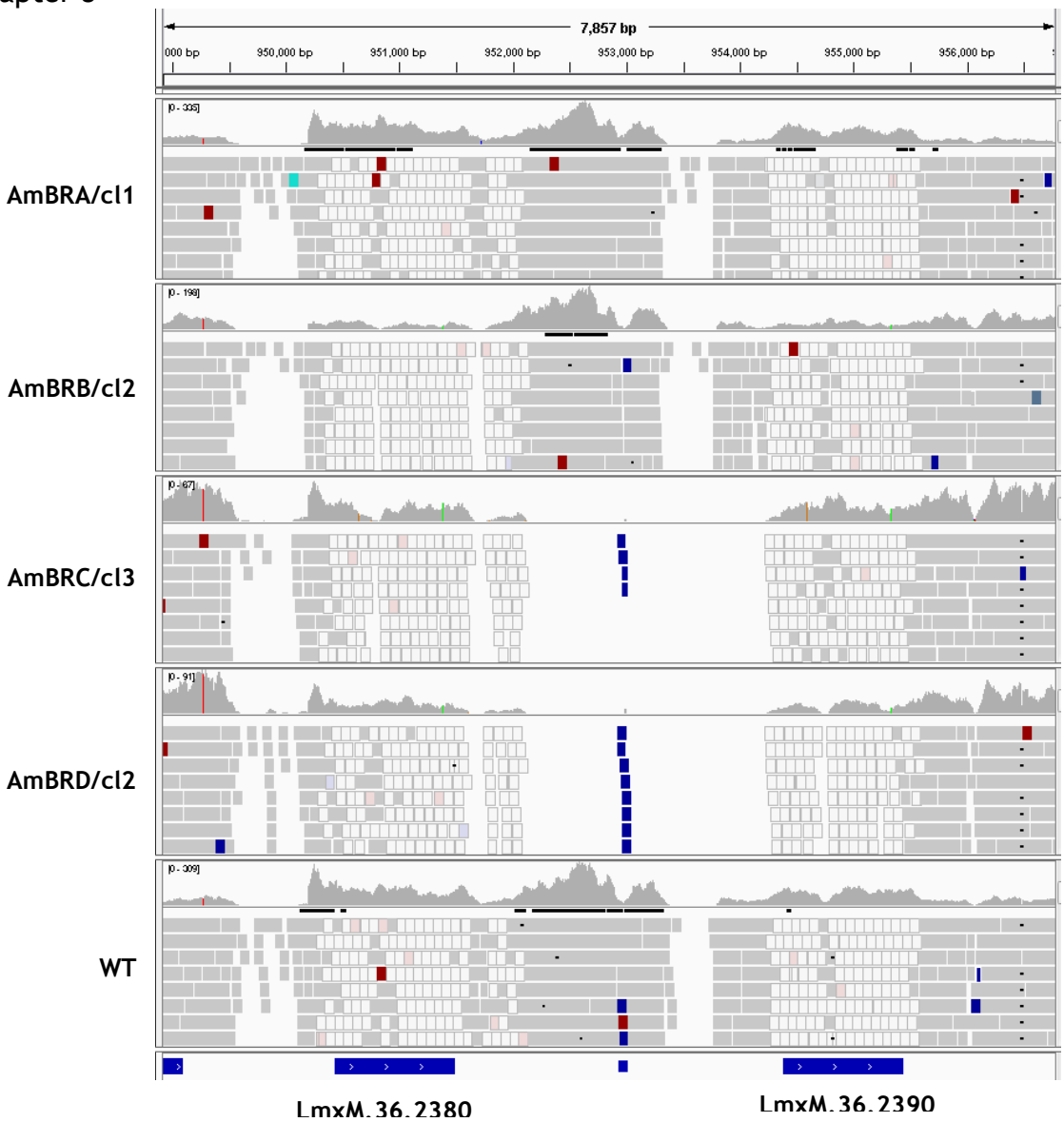


Figure 5-14: Alignment of RNA-seq data to the *L. mexicana* genome with a corrected *SMT* intergenic region. The genomic region of *SMT* genes *LmxM.36.2380* and *LmxM.36.2390*, as visualised using the Integrative Genomics Viewer (Robinson *et al*. 2011), with the intergenic region between *SMT* gene copies replaced by the one sequenced using Sanger sequencing. Grey blocks represent aligned reads with mapping quality greater than 0, white blocks with mapping quality of 0. Blocks in other colours typically represent non-concordantly mapped reads. See Figure 4-2 for a full description of the visualisation.

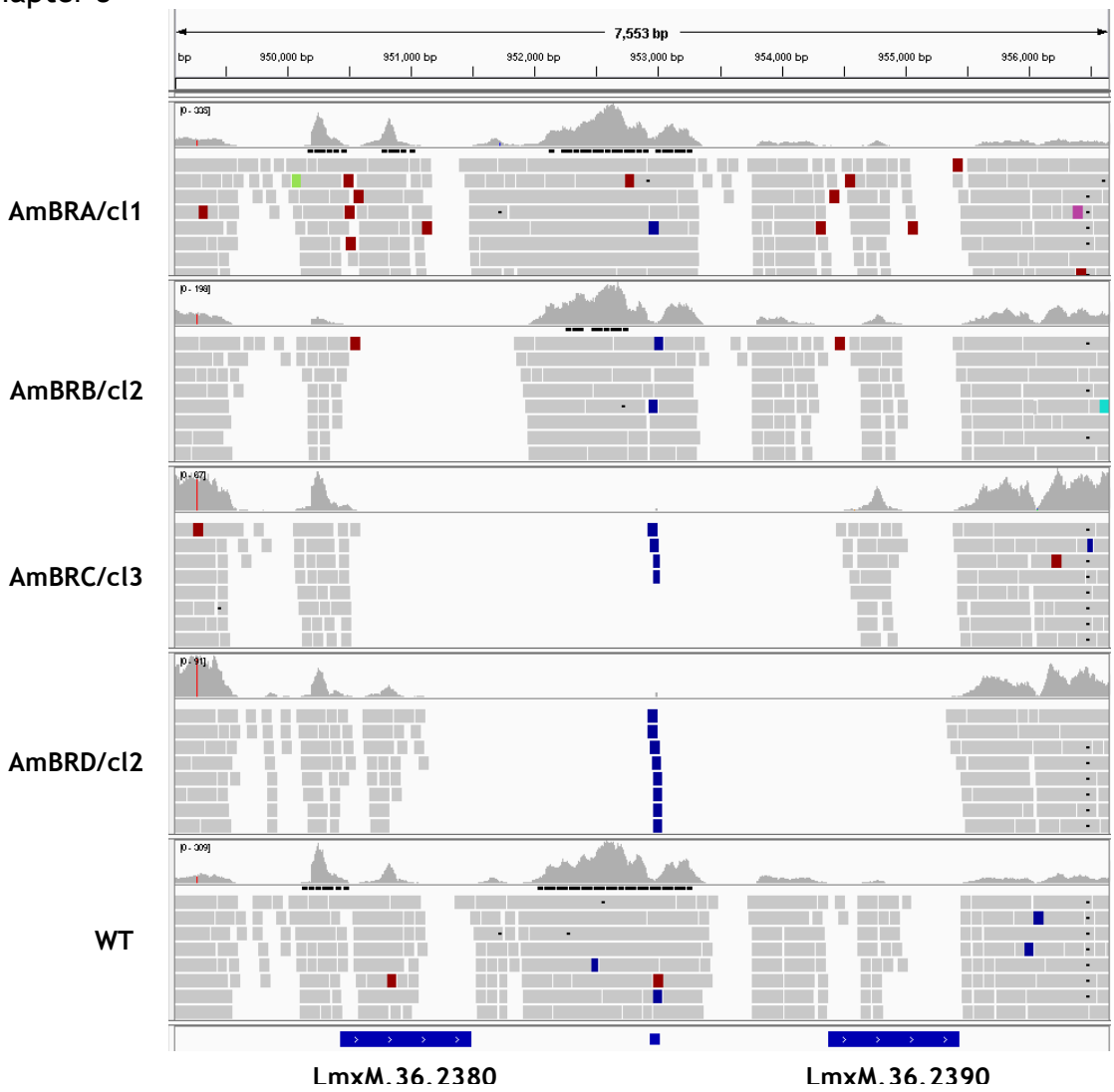


Figure 5-15: Alignment of RNA-seq data to the *L. mexicana* corrected *SMT* region applying a mapping quality threshold. This visualisation was produced in the same way as Figure 5-14 with the exception that reads with a mapping quality of 0 were excluded. See Figure 4-2 for a full description of the visualisation.

Finally, I also realigned RNA-seq data to the corrected genome (Figure 5-14). Similar to the WGS data, there is a large loss of coverage in the intergenic region in AmBRC/cl3 and AmBRD/cl2, with a smaller gap downstream of *LmxM.36.2380* in AmBRD/cl2. Thus, these data show loss of the 3'-UTR of *LmxM.36.2380* in AmBRC/cl3 and AmBRD/cl2, but shows that whilst there is evidence of disruption in AmBRB/cl2, a large part of it is still retained. As with the WGS data I also sought to visualise the alignment without ambiguously mapped reads, by using a mapping quality threshold of 0 (Figure 5-15). This shows loss of coverage particularly in the CDS where sequences may be identical. The "island" of coverage in the middle of each CDS likely arise due to the G391A difference, allowing unique assignment of reads. Therefore, whilst this "island" of coverage

is found in both coding sequences in AmBRA/cl1 and wild-type parasites, it is only present in *LmxM.36.2390* in AmBRB/cl2 and AmBRC/cl3 (which have the A391 genotype) and *LmxM.36.2380* in AmBRA/cl1 (which has the G391 genotype). Interestingly, however, it shows that coverage is retained in the 5'-UTR of *LmxM.36.2380* in these lines whereas coverage of the 5'-UTR of *LmxM.36.2390* is lost in AmBRC/cl3 and AmBRD/cl2. AmBRB/cl2, on the other hand, shows retention of coverage of both 5'-UTRs, as well as large parts of the intergenic region, but, crucially, loss of coverage immediately downstream of *LmxM.36.2380*, which is where the target sequences of the *LmxM.36.2380*-specific primers are located.

5.2.3.3 Identification of SIDER1-like repetitive sequences in the vicinity of *SMT*

Initial efforts to solve the problem with the reference genome in the intergenic region involved using a BLAST search of this region against the Trinity assemblies. Surprisingly, this returned hits in the 3'-UTR of *LmxM.36.2540*, the gene encoding sterol C4-methyloxidase and the gene at the other extreme from *SMT* within the duplicated region of chromosome 20 in AmBRB/cl2. An explanation for this was unclear for some time; however, during construction of the corrected intergenic region, BLAST against the reference genome revealed, alongside a partial match to the *SMT* intergenic region of the reference sequence itself, two further hits, both on chromosome 20 with an E score of 0.0. One of these matched in antisense orientation from positions 992,402-991,947 (462 bp match, with six gaps and two nucleotide differences), the other in sense orientation from positions 1,011,188-1,011,632 (450 bp match, with five gaps and three nucleotide differences). The former of these occurs between *LmxM.36.2470* and *LmxM.36.2480*. The latter, however, occurs between *LmxM.36.2540* and *LmxM.36.2550*. This appears to explain why the intergenic region in the reference genome had a BLAST match to *LmxM.36.2540* Trinity transcripts: the presence of this 450 nucleotide tract at both loci. The position of the 462 nucleotide sequence relative to the reference genome (as seen in Figure 5-12) is from chromosome 20 951,677-952,138.

Table 5-4: Top BLAST results for a putative SIDER1 element in *SMT* intergenic sequence *L. infantum* and *L. major.*

The 462 bp segment described in the text was used as a query in a BLAST search of *L. infantum* JPCM5 and *L. major* Friedlin genomes. Hits were then compared to a list of annotated SIDER1 positions in these two species (Smith *et al*. 2009). "Length" refers to the total length of the aligned BLAST hit, "Matches" to the number of matching nucleotides within that hit.

Species	Position (strand)	Upstream gene	Length	Matches	Score	E	Associated SIDER1 position
L. infantum	Chromosome 36 954,555-954,982 (+)	*LinJ.36.2510*	431	373	509	2e-143	954,529-954,991
L. infantum	Chromosome 36 997,034-996,605 (-)	*LinJ.36.2610*	431	371	504	2e-143	997,059-996,609
L. infantum	Chromosome 36 1,015,791–1,016,054 (+)	*LinJ.36.2670*	265	232	325	5e-88	1,015,774-1,016,049
L. major	Chromosome 36 1,006,079-1,005,648 (-)	*LmjF.36.2480*	433	369	486	2e-136	1,006,107-1,005,652
L. major	Chromosome 36 1,024,686-1,025,106 (+)	*LmjF.36.2540*	422	360	475	4e-133	1,024,672-1,025,122
L. major	Chromosome 36 962,553-962,972 (+)	*LmjF.36.2380*	432	364	464	7e-130	962,536-962,990

In order to characterise further this repeated sequence, I used a BLAST search against the genomes of *L. infantum* and *L. major* (Table 5-4). As a query sequence, I chose the 462 nucleotides of the intergenic sequence that matched to *L. mexicana* chromosome 20 positions 992,402-991,947. In both cases, the top three matches were to regions downstream of the orthologues of *LmxM.36.2380*, *LmxM.36.2470* and *LmxM.36.2540* (*LinJ.36.2510*, *LinJ.36.2610* and *LinJ.36.2670* in *L. infantum*, *LmjF.36.2380*, *LmjF.36.2480* and *LmjF.36.2540* in *L. major*, respectively). Both of these genomes have previously been searched for SIDER elements (Smith *et al*. 2009), thus I was able to compare these hits to the position of SIDER elements in this genome. Whilst no hits corresponded to SIDER2 elements, all overlapped with annotated SIDER1 elements. Hence, there is strong evidence that the three ~450 bp elements found in *L. mexicana* are SIDER1 elements conserved across multiple species

5.2.4 Evidence of stage-specific gene regulation of *SMT*

In order to understand the role the changes to *SMT* described could play in AmB resistance in a clinical context, it is important to determine whether such structural changes might have the same impact in the mammalian amastigote stage as in promastigotes. Previously published RNA-seq data (Fiebig *et al*. 2015) comparing *L. mexicana* promastigotes and intracellular amastigotes show that there is a 0.46-fold change in expression of *LmxM.36.2380* ($P = 9.49 \times 10^{-8}$) on differentiation from promastigotes to intracellular amastigotes, whereas there is no change in expression of *LmxM.36.2390* (expression in amastigotes is 1.26-fold that of promastigotes, P = 0.180). This could suggest that the relative functional importance of these two transcripts is quite different in amastigotes from promastigotes, with *LmxM.36.2390* playing a greater role in the amastigote.

Due to the previously explained difficulties in differentiating between transcripts, and the problems with the reference genome in the 3'-UTR of *LmxM.36.2380*, I wished to confirm whether these changes were reproducible by qRT-PCR. Therefore, I extracted RNA from intracellular amastigotes within primary murine macrophages (Figure 5-16). I initially examined overall *SMT* expression using non-gene copy-specific primers. This revealed a decrease in expression in amastigotes to 0.49-fold that of promastigotes ($P = 3.00 \times 10^{-4}$). An even greater decrease was reported using primers specific to the *LmxM.36.2380*

3'-UTR, with amastigote expression 0.33-fold that of promastigotes (P = 1.42 x 10^{-5}). In contrast to this, a slight increase in expression was observed in *LmxM.36.2390*, with amastigote expression 1.14-fold that of wild-type, but this was not statistically significant (P = 0.373). Therefore, the changes reported in the RNA-seq data published by Fiebig and colleagues are reproducible by qRT-PCR in our hands.

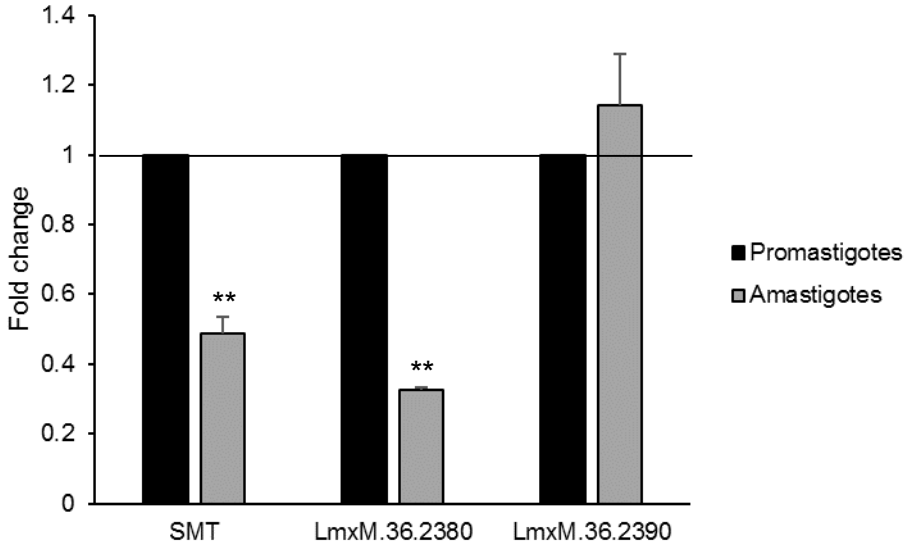


Figure 5-16: Stage-specific expression changes in *SMT* gene copies. Expression of *SMT* overall (using non-gene copy-specific primers) and individual gene copies (using primers targeted to the 3'-UTRs) was measured between promastigotes and intracellular amastigotes by qRT-PCR. Values are the mean fold-changes across three biological replicates, with error bars representing standard deviation. Asterisks represent significant differences in expression between promastigotes and amastigotes (**P < 0.005) and the horizontal line represents promastigote expression.

I wished to determine directly whether the differences in expression between wild-type and AmB-resistant promastigotes were also evident in amastigotes (Figure 5-17). As in promastigotes, overall *SMT* expression was unchanged in AmBRA/cl1 (P = 0.716) but decreased significantly in AmBRB/cl2 (P = 0.0270), AmBRC/cl3 (P = 0.00106) and AmBRD/cl2 (P = 0.00164). It was evident, however, that the magnitude of these changes was less than in promastigotes. In the previous chapter, I demonstrated by qRT-PCR that the respective fold-changes in overall *SMT* expression in promastigotes in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 were 0.22-fold, 0.12-fold and 0.11-fold. In amastigotes, however, the equivalent changes were 0.46-fold, 0.30-fold and 0.18-fold. Thus the degree to which *SMT* expression is reduced in amastigotes appears to be less than that in promastigotes.

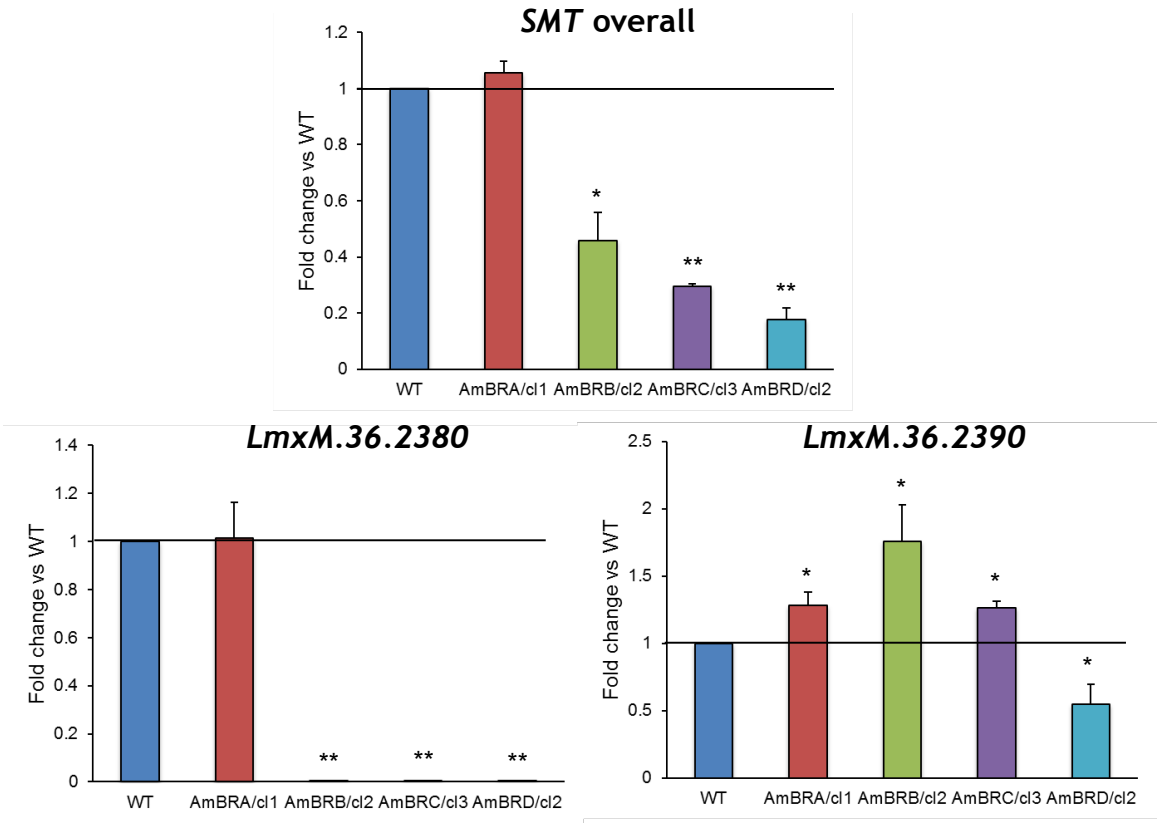


Figure 5-17: Differential expression of *SMT* genes in wild-type and AmB-resistant lines as intracellular amastigotes. Overall and transcript-specific *SMT* expression was measured by qRT-PCR. Values are the mean fold-changes across three biological replicates, with error bars representing standard deviation. Asterisks represent significant differences in expression from wild-type parasites (*P < 0.05, **P < 0.005) and the horizontal line represents wild-type expression.

On a transcript-specific basis, expression of transcripts containing the *LmxM.36.2380* 3'-UTR was unchanged in AmBRA/cl1 (P = 0.960) whereas in AmBRB/cl2 (P = 0.00164), AmBRC/cl3 (P = 6.31 x 10^{-6}) and AmBRD/cl2 (P = 1.95 x 10^{-4}) transcripts were essentially undetectable (Ct > 35), similar to promastigotes. Expression of transcripts containing the *LmxM.36.2390* 3'-UTR, however, showed significantly different expression from wild-type in all four resistant lines. AmBRA/cl1, AmBRB/cl2 and AmBRC/cl3 showed increases of 1.28-fold (P = 0.0105), 1.76-fold (P = 0.0143) and 1.26-fold (P = 0.0396). Surprisingly, AmBRD/cl2 actually showed a decrease in expression to 0.55-fold that of wild-type (P = 0.0121). In this assay, I also included uninfected macrophages as a control, to verify that amplification from macrophage RNA was not a significant factor. Average Ct values for GAPDH, overall *SMT*, *LmxM.36.2380*-specific and *LmxM.36.2390*-specific primers were 37.1, 34.1, 36.6 and 37.8, although amplification was not detected in all samples. As these are similar to values for water-only negative controls (Ct > 35 or undetectable), all

primers were deemed to be specific for *Leishmania* RNA targets. In order to verify that RNA was still efficiently extracted from uninfected controls, I tested all samples from one replicate using primers specific to the macrophage 18S rRNA. Signal was high (Ct < 14) in all cases including uninfected macrophages, thus demonstrating that the low or undetectable signal using *Leishmania*-specific primers in uninfected macrophages was not due to low RNA yield in these samples.

5.3 Discussion

5.3.1 Restoration of phenotypes in episomal overexpression lines

Having identified three candidate genes with a putative role in resistance in the previous chapters, here I assessed their role in three phenotypes, AmB resistance, PENT hypersensitivity and changes to sterol composition. In all cases, ectopically expressed genes were expressed well above wild-type levels as expected using this episomal expression system. It was evident, however, that levels of expression varied, and indeed transfection of *SMT* genes into an AmBRD/cl2 background resulted in substantially lower degrees of overexpression than the same vectors transfected into AmBRB/cl2. This may reflect differences in the genetic background of these lines and tolerance of the consequences of maintaining the plasmid, such as the expression of the neomycin resistance gene (as referred to in section 5.2.1). I did not try to measure plasmid copy number, although relative quantification could be achieved by qPCR on DNA extracts, and could confirm or reject copy number as a possible basis for these differences.

5.3.1.1 The role of candidate genes in AmB resistance

In all cases, transfection of wild-type gene copies caused at least partial restoration of AmB sensitivity, confirming a role for mutations in these genes in the resistance phenotype. In the case of AmBRA/cl1, reintroduction of *SC5D* caused total restoration of sensitivity, whereas transfection of the copy carrying the G139R mutation found in this line had no influence. Therefore, it is reasonable to argue that this mutation is solely responsible for the 7.3-fold AmB resistance phenotype in this line. Since *SC5D* with the G139R mutation did not influence AmB sensitivity even in the context of more than tenfold overexpression compared to wild-type cells, it strongly suggests that gene

function is almost or totally lost. In this context, it is interesting that a small amount of cholesta-5,7,24-trienol was detectable in AmBRA/cl1 in the GC-MS assay described in this chapter, despite a requirement for SC5D activity in the genesis of the 5(6)-7(8) double bond system. This sterol was not, however, detectable above a threshold of 1% abundance in the GC-MS assay in chapter 3, thus its detection may not be robust or fully reliable. Furthermore, given the difficulty of accurate sterol identification without standards, a misidentification cannot be ruled out (although the fact that this sterol increases in the other three resistant lines is consistent with the lack of C24-methylation in cholesta-5,7,24-trienol). Thus whilst the G139R mutation clearly severely reduces SC5D activity, there is still a question about whether this loss is absolute.

In AmBRB/cl2, it appears that changes to both *SMT* and *MT* contribute to the resistance phenotype. I did not attempt to assess whether transfection of both would completely restore AmB sensitivity, however, so it is not impossible that other factors contribute. I discussed deletion of *MT* in the context of other reported changes to this gene in the previous chapter; the data presented in this chapter confirm that indeed *MT* deletion plays a role, possibly through an influence on phospholipid content as suggested previously (Fernandez-Prada *et al*. 2016).

Finally, I have confirmed a role for changes at the *SMT* locus in resistance in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. Altered expression and structural variation associated with these changes are discussed in greater depth below. What is evident, however, is that in no instance did SMT transfection result in complete restoration of AmB sensitivity to wild-type levels. There are two possible explanations for this. The first is that other unrelated factors contributed to the resistance phenotype. In the case of AmBRB/cl2, which possesses the *MT* deletion, this is certainly at least a partial explanation. In AmBRC/cl3 and AmBRD/cl2, however, I have not identified other changes that can be said for sure to contribute to resistance. The other possibility is that episomal overexpression of *SMT* genes does not fully compensate in itself for loss of wild-type SMT function. In particular, as only the coding sequence is introduced into the vector, expression is driven by other UTRs, with the 5'-UTR derived from phosphoglycerate kinase B and the 3'-UTR from phosphoglycerate

kinase A. I did not directly determine protein levels in these lines so it is possible that translation of *SMT* genes was inefficient from this vector. However, this was not an issue with *SC5D*, despite use of the same vector, and green fluorescence could clearly be seen in wild-type parasites expressing GFP in this system. Alternatively, it is possible that the UTRs contribute to gene function in other ways, a possibility to which I will return later in this discussion.

In all cases where resistant line-derived *SMT* copies were transfected, some restoration of AmB sensitivity was observed, thus suggesting that all copies retained activity. Transfection of AmBRB/cl2 and AmBRD/cl2-derived copies, with respective genotypes of A391/A961 and G391/A391, revealed no significant difference from wild-type *SMT* gene *LmxM.36.2380* (G391/G961) in the ability to restore AmB sensitivity, suggesting that these SNPs themselves do not directly contribute to the resistance phenotype in these lines. By contrast, transfection of AmBRC/cl3 with the *SMT* gene found in that line (possessing the T215G mutation encoding the F72C substitution in addition to having an AmBRB/cl2-like A391/A961 genotype) had significantly less of a restorative effect on AmB sensitivity than wild-type *LmxM.36.2380*. This suggests that the F72C mutation does influence enzyme function. However, one cannot say from these data how strong an effect it has. Whilst the difference is modest (wild-type SMT reduces the AmB IC_{50} from 246 nM to 105 nM, whereas AmBRC/cl3-derived SMT only reduces it to 138 nM), one must consider that in both cases *SMT* expression is more than 10-fold that of wild-type parasites. Thus, the effect of F72C on SMT function may be much more pronounced than these data suggest, but the high degree of overexpression could compensate for this difference. Hence, this assay is not optimal for determining the precise impact of SNPs on enzyme function. In order to quantitatively examine this, it would be necessary either to reintroduce endogenous wild-type expression (which could be a difficult task given the structural changes at this locus) or to express and purify the SMT protein with various substitutions, determining activity in an *in vitro* assay system.

5.3.1.2 The genetic basis of PENT hypersensitivity

Because hypersensitivity to PENT was found to be a conserved phenotype across all resistant lines (both in this study and elsewhere), I tested whether episomal overexpression lines showed restoration of wild-type PENT sensitivity. Having

already assessed whether resistant-line derived copies could still restore AmB sensitivity (and hence whether gene function was retained), I focused only on the wild-type copies. In all cases, there was a significant increase in PENT IC_{50}. The extent of this increase varied, however. Overexpression of *SMT* caused complete restoration of wild-type sensitivity in AmBRC/cl3 and AmBRD/cl2, whereas in AmBRB/cl2 there was partial 3.4-fold restoration. In contrast to this, whilst overexpression of *SC5D* in AmBRA/cl1 caused a significant increase in IC_{50}, this increase was very marginal (1.18-fold). This suggests, surprisingly, that whilst altered sterol biosynthesis was responsible for PENT hypersensitivity in the lines with loss of SMT function, mutation in *SC5D* was not the direct cause in AmBRA/cl1. Overexpression of *MT* caused complete restoration of wild-type sensitivity in AmBRB/cl2, in contrast to the only partial restoration seen in AmB sensitivity. Again, this is surprising, given that loss of SMT function also contributed to PENT hypersensitivity. This may be a result of overexpression, with dose-dependent effects of *MT* expression even above wild-type levels – indeed, it would be interesting to test whether *MT* overexpression in wild-type lines leads to PENT resistance.

These data suggest that two separate changes associated with AmB resistance, altered sterol biosynthesis and loss of MT function, result in PENT hypersensitivity, further underscoring what appears to be a conserved negative correlation in sensitivity between the two drugs. The common thread in this would appear to be altered membrane composition, something that has indeed been reported in PENT-resistant lines (Basselin *et al*. 1997). Specifically, membrane fluidity increased in these lines, with membrane polar lipids and intracytoplasmic neutral lipids decreased in PENT resistance, but overall phospholipid content increased. As noted before, increased membrane fluidity has been observed both in AmB-resistant (Mbongo *et al*. 1998) and PENT-resistant *Leishmania*, an unexpected result given the negative correlation between sensitivity to the two. This suggests that the determinants of sensitivity may be more complex than a single membrane property. However, with regard to MT, the changes to lipid composition reported by Basselin and colleagues are interesting because they suggest that lipid distribution within the cell may be important; as a lipid flippase, it is feasible that the influence of MT on PENT

sensitivity operates through a effects on distribution either between cellular compartments or within the lipid bilayer.

Regarding the link between PENT sensitivity and sterol composition, the difference in effect of *SMT* and *SC5D* overexpression is striking. It suggests that the presence of sterol C24-methylation is protective against PENT, whereas 5(6)-7(8) desaturation has little influence. In the AmB-resistant *L. mexicana* line previously shown to have a mutation in *CYP51*, PENT hypersensitivity was also found to be a result of altered sterol biosynthesis (Mwenechanya *et al*. 2017). This raises the question of how PENT hypersensitivity arises in AmBRA/cl1. Given that AmB resistance and changes to sterol composition appear to be mediated entirely by the G139R mutation in SC5D, it suggests that the direct cause of sterol changes and of PENT hypersensitivity are separate. Given the degree of conservation of PENT hypersensitivity amongst AmB-resistant lines, however, it would be extremely surprising if it was just a chance occurrence in AmBRA/cl1 that had nothing to do with selection for AmB resistance. Therefore, it is possible that some secondary change, unidentified here, arises to compensate for loss of SC5D activity. For example, there may be changes to lipid composition or distribution to accommodate the large increase in ergosta-5,22-dienol, which is expected to have a less planar ring structure than the major wild-type sterol, ergosta-5,7,24(8)-trienol (as well as ergosterol itself). I have not found any such change, but it could be interesting in the future to focus more specifically on lipid metabolism, in particular those changes in AmBRA/cl1 that are not reversed on reintroduction of *SC5D*. One point to note, however, is that introduction of the plasmid itself appears to cause a slight but significant decrease in IC_{50}. Hence, the increased IC_{50} observed on *SC5D* reintroduction may be understated due to counteracting effects of the plasmid itself. Even if this is influential, however, the magnitude of influence of *SC5D* is still far less than that of *MT* or *SMT*.

5.3.1.3 The effect of candidate gene overexpression on sterol composition

As expected, reintroduction of wild-type copies of sterol biosynthesis genes caused at least partial restoration of sterol composition. In keeping with the AmB sensitivity data, overexpression of wild-type *SC5D* results in essentially total recovery of wild-type sterol composition, whereas transfection with the

G139R mutant appears to have no influence. On the other hand, transfection of *MT* into AmBRB/cl2 only has a very marginal effect on sterols, with no evidence of a shift towards wild-type composition. This is unsurprising as MT is not involved in sterol biosynthesis, but confirms that *MT* deletion contributes to AmB resistance through a sterol-independent mechanism.

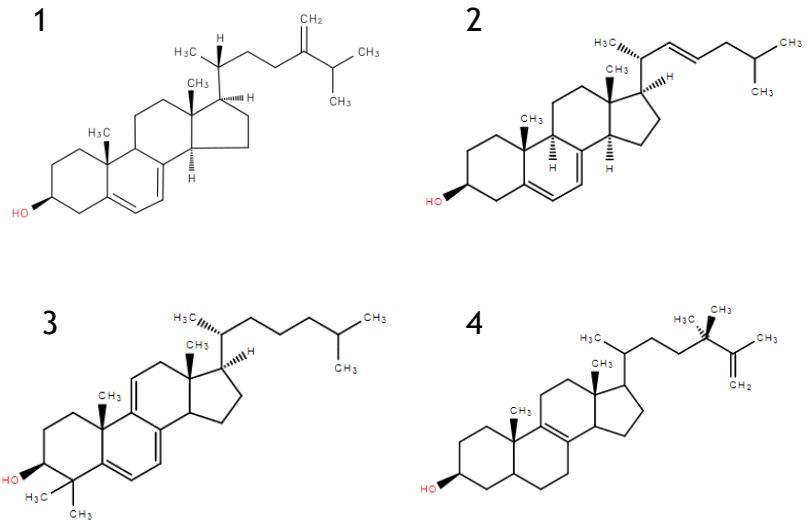


Figure 5-18: Sterol structures described in the text. 1) Ergosta-5,7,24(28)-trienol (the major sterol in wild-type parasites). 2) Cholesta-5,7,22-trienol (the major sterol in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2). 3) 4,4-dimethylcholesta-5,7,9(11)-trienol (the sterol accumulating in *SMT*-overexpression lines). 4) 24,24-dimethylergosta-8,25(27)-dienol (side-product of *T. brucei* SMT activity *in vitro*, Zhou *et al*. 2006).

In contrast to *SC5D*, however, *SMT* overexpression in all cases only resulted in partial restoration of AmB sensitivity, with higher levels retained of cholesta-5,7,22-trienol and cholesta-5,7,24-trienol, which lack C24-methylation. This may underlie the residual resistance to AmB observed in these lines. The possible reasons for incomplete restoration were discussed in section 5.3.1.1 but are unclear. However, there was one important qualitative difference between SMT overexpressors, and their parental resistant lines, namely the presence of another sterol, 4,4-dimethylcholesta-5,7,9(11)-trienol (Figure 5-18, sterol 3). This was marginally detectable in AmBRA/cl1 overexpressing *SC5D*, but was far more abundant in resistant lines expressing *SMT*. The origin of this sterol is unclear. One possibility is that it arises due to *SMT* expression well above that of normal wild-type levels. As this sterol lacks C24-methylation, however, it is unclear how it could arise through increased activity of SMT. 4,4-

dimethylcholesta-5,7,9(11)-trienol is an isomer of 4,4-dimethylcholesta-8,14,24-trienol, the substrate of C14-sterol reductase (Figure 1-7). The RNA-seq data showed a moderate decrease in expression of this enzyme (*LmxM.31.2320*) associated with reduced chromosomal copy in AmBRB, AmBRC and AmBRD. Theoretically, this could result in an increase in 4,4-dimethylcholesta-8,14,24-trienol, which could then be converted into 4,4-dimethylcholesta-5,7,9(11)-trienol through the nonspecific action of other enzymes. However, as C14-sterol reductase acts upstream of SMT, this should be an expected phenotype in the resistant lines themselves, as well as the overexpression lines. Overexpression of *MT* from the same vector system did not have this result, therefore it is not an effect of the plasmid itself. More likely it is a direct effect of *SMT* overexpression, which may influence the regulation pathway in complex ways. In this context, it is further evidence that factors other than the coding sequence itself (such as the UTRs or broader genomic context) are required for normal functioning of SMT.

With regard to the different genotypes of *SMT* that were transfected, whilst there were marginal differences between AmBRB/cl2 and AmBRD/cl2 transfected with wild-type *LmxM.36.2380* (G391/G961) in comparison to resistant line-derived *SMT* (genotypes A391/A961 and G391/A961, respectively), these could possibly relate to expression differences, and in any case the impact is not large. This provides further evidence that the coding sequence differences between *LmxM.36.2380* and *LmxM.36.2390* do not have a major influence on enzyme function. These differences were more pronounced for AmBRC/cl3, however, which has the additional T215G (F72C) substitution. Not only was this variant less effective at restoring wild-type sterol composition than wild-type SMT, it also did not result in accumulation of 4,4-dimethylcholesta-5,7,9(11)-trienol. If the F72C substitution reduces catalytic efficiency of the enzyme, the effective increase in SMT activity relative to wild-type would be less in parasites overexpressing the F72C SMT copy than other copies. In a scenario whereby 4,4-dimethylcholesta-5,7,9(11)-trienol accumulates due to a large increase in SMT activity, this would explain the absence of this sterol in AmBRC/cl3 expressing the mutated copy. Alternatively, since there is evidence that the F72C SNP may be found in the sterol binding site, this could result from altered enzyme specificity. Characterisation *in vitro* of SMT from *T. brucei* suggests that several

products can be generated by this enzyme (Zhou *et al*. 2006). All of these products are, as expected, C24-methylated, but these do include the dimethylated product, 24,24-dimethylergosta-8,25(27)-dienol (Figure 5-18, sterol 4). This is not an isomer of 4,4-dimethylcholesta-5,7,9(11)-trienol, but it is feasible that *in vivo* other desaturase activities could convert it into one. Furthermore, whilst the dimethylation is at the wrong end of the sterol molecule, it is not inconceivable that this sterol could be misidentified in our data, without the availability of a standard. These data highlight the challenging nature of interpreting changes to sterol composition, where identifications may be uncertain and enzymes may have multiple substrates and products.

5.3.2 A model for structural changes at the *SMT* locus

In the previous chapter, I described a putative model for the types of changes happening at the locus of the two *SMT* genes, *LmxM.36.2380* and *LmxM.36.2390*. However, given the evident problem in the reference genome downstream of *LmxM.36.2380* (which included a string of 100 N residues, thus demonstrating incomplete assembly), and given the high sequence homology between the two coding sequences, I was limited in the ability to provide precise description of the changes. Having characterised the transcripts in greater detail, however, I was able to design qPCR probes to allow characterisation of changes on a gene copy-specific basis. This has allowed more precise definition of a model for the types of changes happening at this locus, as well as genomic features that may contribute to the occurrence of such events.

5.3.2.1 Trinity assembly of *SMT* transcripts

The initial transcript characterisation came with the generation of Trinity *de novo* transcriptome assemblies from the RNA-seq data. Because at first it was difficult to judge the accuracy of these assemblies, I decided to focus on regions that were conserved across multiple transcripts. This revealed a high degree of reproducibility between datasets, with differences within the coding sequences only observed at sites of identified SNPs (T215G, G391A, G961A). The 5'-UTRs were also identical, with the exception of one out of two identified AmBRD transcripts. In all cases except for this AmBRD transcript, part of the SL RNA was present at the start of the 5'-UTR, suggesting that proper splicing occurs in all

transcripts. Interestingly, the full length SL RNA was not present in any of the contigs; conceivably this could result from difficulties arising from modification of the 5'-cap interfering with sequencing of the extreme 5' end of transcripts. It is difficult to establish whether the truncated 5'-UTR in the AmBRD assembly represents a real variant transcript or simply an artefact of Trinity assembly, however, as it only appears in one sample. Comparison to the sequenced intergenic region, however, suggested that the Trinity assembly may have mistakenly classified all 5'-UTRs as identical, when in fact there were large differences between the upstream regions of *LmxM.36.2380* and *LmxM.36.2390*. As discussed, this may arise due to the far higher abundance of *LmxM.36.2380*.

Alignment of the 3'-UTRs revealed high reproducibility, with 100% identity for the first 192 nucleotides. Beyond this, the *LmxM.36.2390*-like transcripts showed extensive homology to downstream sequences of *LmxM.36.2390* in the reference genome, whereas the first 38 out of 48 nucleotides in the *LmxM.36.2380*-like transcript homology region were identical to those in the intergenic region obtained by Sanger sequencing. This suggests that Trinity was reproducible and accurate in assembly for the breadth of transcripts, but was more likely to make errors at the edges of transcripts. I did not examine the reason for this in detail, although feasibly the action of exonucleases, either *in vivo* or due to RNase contamination, could lower or increase the variability of coverage at the edges of transcripts. A key outcome, however, is that focusing on the reproducible regions of the assembled 3'-UTRs allowed not only classification of transcripts, but successful design of transcript-specific primers. *De novo* assembly, therefore, can be an important use of RNA-seq data in kinetoplastids, beyond differential expression analysis.

5.3.2.2 The relationship between genomic changes and loss of *SMT* expression

Through sequencing of amplicons using transcript-specific primers, transcript types identified in Trinity were explicitly associated with the genotypes of *SMT* coding sequences. In the reference genome, *LmxM.36.2380* and *LmxM.36.2390* differ by the G391A substitution. Hence sequencing revealed that in wild-type parasites, transcripts with the 48 nucleotide homology region (note that only the first 38 nucleotides of this appear to be correct) in the Trinity assembly relate

explicitly to *LmxM.36.2380*, whereas those with the 196 nucleotide homology region relate to *LmxM.36.2390* (which also possesses a G961A substitution relative to *LmxM.36.2390* in the reference genome). On this basis, it is evident that in wild-type promastigotes, *LmxM.36.2380* is far more highly expressed than *LmxM.36.2390*. Therefore, in order to drastically deplete SMT activity, it appears that the favoured mechanism is to lose transcripts containing the *LmxM.36.2380*-like 3'-UTR.

The parallels with previously observed changes in *L. donovani* are immediately evident. Pourshafie and colleagues previously described two *SMT* transcripts in *L. donovani*, one of which, *SCMT 1A*, was far more abundant in wild-type cells than the other, *SCMT 1B* (Pourshafie *et al*. 2004). Because this study was performed before the publication of an *L. donovani* reference genome, the authors could not specifically relate these transcripts to individual gene copies, in the way that has been done here. Interestingly, in the currently available reference genome for *L. donovani* (*L. donovani* BPK282A1) only one *SMT* gene copy, LdBPK_362510.1, is apparent. However, the closest related species, *L. infantum*, has two copies, as does a more recent unpublished assembly of *L. donovani* based on Pacific Biosciences long read sequencing (Hideo Imamura, Institute of Tropical Medicine Antwerp, Belgium, personal communication). Thus, the single gene copy may well be an assembly error, and the two *SMT* transcripts with distinct UTR sequences described by Pourshafie and colleagues more likely relate to two separate gene copies. In this context, therefore, the loss of *SCMT 1A* in AmB-resistant *L. donovani*, as shown by Northern blot, is similar to the loss of *LmxM.36.2380* in our AmB-resistant *L. mexicana* lines; in both cases, the dominant transcript in wild-type parasites becomes undetectable. Also in both cases, there is increased expression of the other transcript (*SCMT 1B* or *LmxM.36.2390*), although whilst this was not quantified in *L. donovani*, it appears to be more pronounced (based on Northern blot) than in our case. Interestingly, the two *SMT* 3'-UTR sequences described in *L. donovani* in this paper also shared 182 nucleotides of homology downstream of the stop codon, remarkably similar to the 192 nucleotides described here. The authors did not publish the UTR sequences they found for *L. donovani SMT*, therefore a direct comparison of similarities is not possible in this case. Overall, however, changes at the transcript level appear to be remarkably similar.

Indeed, changes fitting this pattern were identified by semi-quantitative RT-PCR in an AmB-resistant *L. donovani* clinical isolate (Purkait *et al*. 2012). One transcript, called SCMT A, was reported as present in wild-type cells but absent in resistant lines, although the authors did not show data for this claim. The other, SCMT B, showed a 2.5-fold increase in resistant parasites. Loss of ergosterol and ergosta-5,7,24(28)-trienol, with accumulation of cholesta-5,7,24-trienol in this line is also consistent with loss of SMT activity.

Where our observations do differ is with respect to the 5'-UTRs, which Pourshafie and colleagues described as differing between transcripts in *L. donovani*, with evidence of lack of splicing in AmB-resistant parasites. The authors found that a forward primer specific to the SL RNA and a reverse primer within the *SMT* coding sequence produced an amplicon from wild-type, but not AmB-resistant, cDNA. Notably, however, these data did not include a positive control that would yield a band in resistant cells (such as using a CDS-specific forward primer). In our case, all 5'-UTRs in the Trinity showed evidence of SL RNA sequence, including in all AmB-resistant lines. However, as the 5'-UTR of *LmxM.36.2390* appears to be omitted from both wild-type and AmBRA assemblies, one cannot say from the data available whether the 5'-UTR of *LmxM.36.2390* contains the SL RNA sequence. In any case, however, the transcript found in AmBRB, AmBRC and AmBRD appears to be of the *LmxM.36.2380* type, hence lack of splicing does not appear to be a consideration here.

Therefore, loss of the more abundant *SMT* transcript (or more specifically in all cases described, loss of the more abundantly expressed 3'-UTR) has been repeatedly observed as the mechanism of loss of SMT activity. This study, however, is the first one to identify the genomic basis of such changes, and reveals multiple routes. In all cases, there is loss of the *LmxM.36.2380*-like 3'-UTR at the level of the genome as well as the transcriptome. In AmBRC/cl3 and AmBRD/cl2, this is clearly associated with loss of *SMT* copy number. However, the genotypes of *SMT* found in these two lines are different: both have an A at position 961, but AmBRC/cl3 has an A at position 391, whereas AmBRD/cl2 has a G. These data are consistent in each case with a model whereby homology between the two sequences causes the gene copies to be fused in such a way as

to link the 5'-UTR of *LmxM.36.2380* with the 3'-UTR of *LmxM.36.2390*. One possible mechanism by which these changes arise is microhomology-mediated end joining (MMEJ), a form of double strand break repair that has previously been shown to occur in *Leishmania* in a *MRE11*-dependent, *RAD50*-independent manner (Laffitte *et al*. 2016b), as a possible cause of chromosomal translocations. In this mechanism, resection of DNA by exonucleases to generate single-stranded DNA ends at a double-stranded break generates regions of homology that allow binding and repair across the break. In this scenario, the different products of AmBRD/cl2 and AmBRC/cl3 may be generated either because of the position of the initial breakpoint (before or after position 391) or because after MMEJ the resulting G391/A391 mismatch is repaired in contrasting ways by the cellular DNA repair machinery.

This model fits with the data regarding qPCR at both the genomic (CNV) and RNA (expression) levels, as well as the NGS data in terms of the coverage across this region. There are a number of weaknesses, however. For example, one cannot rule out the possibility that other types of transcript exist. Whilst this did not appear to be the case based on the Trinity assembly, the erroneous assembly of 5'-UTRs in the wild-type assembly (specifically, the lack of a *LmxM.36.2390*-like UTR in any transcript) indicates that Trinity may struggle, particularly where transcripts differ strongly in abundance and where there are extensive regions of homology, both of which are the case here. Indeed, transcript types which cannot be detected (or differentiated between) using our transcript-specific primers may have been missed. A counter-argument to this is that despite very similar primer efficiencies between the GAPDH and total *SMT* primers (Supplementary file 2-2), the 2^{-Ct} values for *SMT* in wild-type DNA (Figure 5-9) are approximately double those for GAPDH, suggesting a haploid copy number of two. Furthermore, all changes in copy number in *SMT* (both total *SMT* and using 3'-UTR-specific primers) were by a factor of two (increases and decreases), strongly suggesting an even number of copies in wild-type DNA. However, I have also not fully considered the possibility of heterozygous structural changes at this locus and the influence these may have. Given the potential for instability in this genomic region, it is feasible that even in wild-type DNA the number or nature of *SMT* copies may differ substantially between copies of chromosome 20.

If this were the case, it would make interpreting these lesions extremely complicated. Hence, this remains only a putative model.

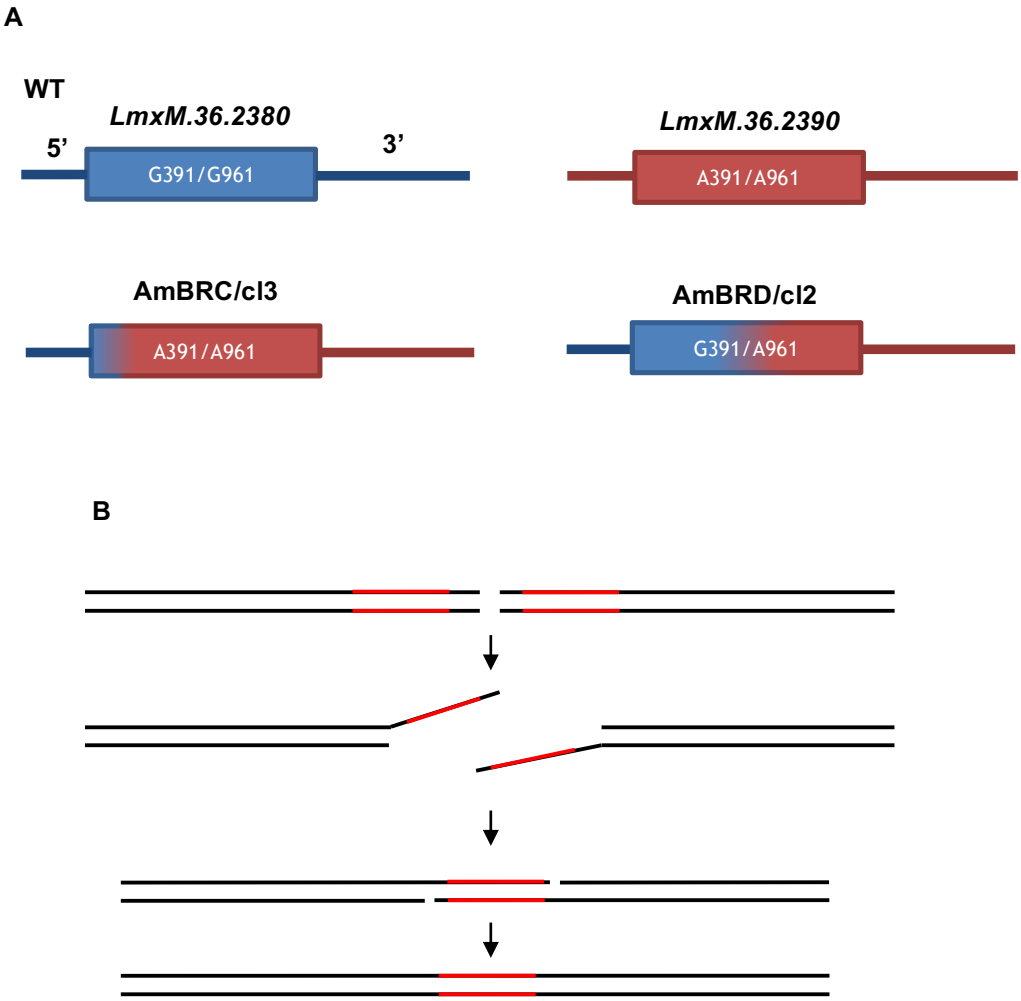


Figure 5-19: A model for structural changes in AmBRC/cl3 and AmBRD/cl2. A) A model for two types of transcripts found in wild-type (top row), and the ones arising in AmBRC/cl3 and AmBRD/cl2 (bottom row). The boxes represent the coding sequences, with genotypes of positions 391 and 961 marked. The lines either side of the box represent the UTRs, with blue representing *LmxM.36.2380*-type UTRs and red representing *LmxM.36.2390*-like UTRs. B) The mechanism of microhomology-mediated end joining (MMEJ). Double-stranded breaks are resected by endonucleases, exposing single-stranded complementary regions (red). These bind, allowing DNA repair machinery to repair the resulting single-stranded nicks in the DNA backbone, to form a new, shorter, region of genome.

In AmBRB/cl2, the lesion appears to be even more complex. Whilst the 3'-UTRs present in both the Trinity assembly and the qPCR data appear to be solely of the *LmxM.36.2390* type, there appears to be coverage of both types of 5'-UTRs in the NGS data (Figure 5-12, Figure 5-13, Figure 5-15). As the copy number of transcripts containing the *LmxM.36.2390*-type 3'-UTR appears to double, it stands to reason, therefore, that each one has a different type of 5'-UTR associated with it, whilst both CDSs have the A391/A961 genotype. To complicate matters further, AmBRB/cl2 retains a greater degree of coverage in the intergenic region than AmBRC/cl3 and AmBRD/cl2 (with the missing coverage being largely restricted to the target of the *LmxM.36.2380*-like 3'-UTR-specific primers) and it also possesses a multigenic duplicated region. As described before, the intergenic region and the region just downstream of *LmxM.36.2540* (the last gene in this duplication) both contain putative SIDER1 elements. These elements appear to be retained in AmBRB/cl2, and it is likely that they are involved in mediating the duplication event, given their location at the start and end of the duplicated region. However, I have not demonstrated the exact nature of the duplication, although its apparent stability after multiple passages without drug might argue against its being extrachromosomal. The fact that an intergenic region could not be amplified from this line suggests that although copy number overall is unchanged, gene copies are no longer found in tandem (indeed even if gene copies were adjacent but convergently or divergently arranged, an amplicon would still be expected due to either forward or reverse primers binding to two, convergent positions). Thus, were the duplication not to be extrachromosomal, Figure 5-20 describes one putative model for AmBRB/cl2 that is consistent with the data. It remains, however, highly speculative.

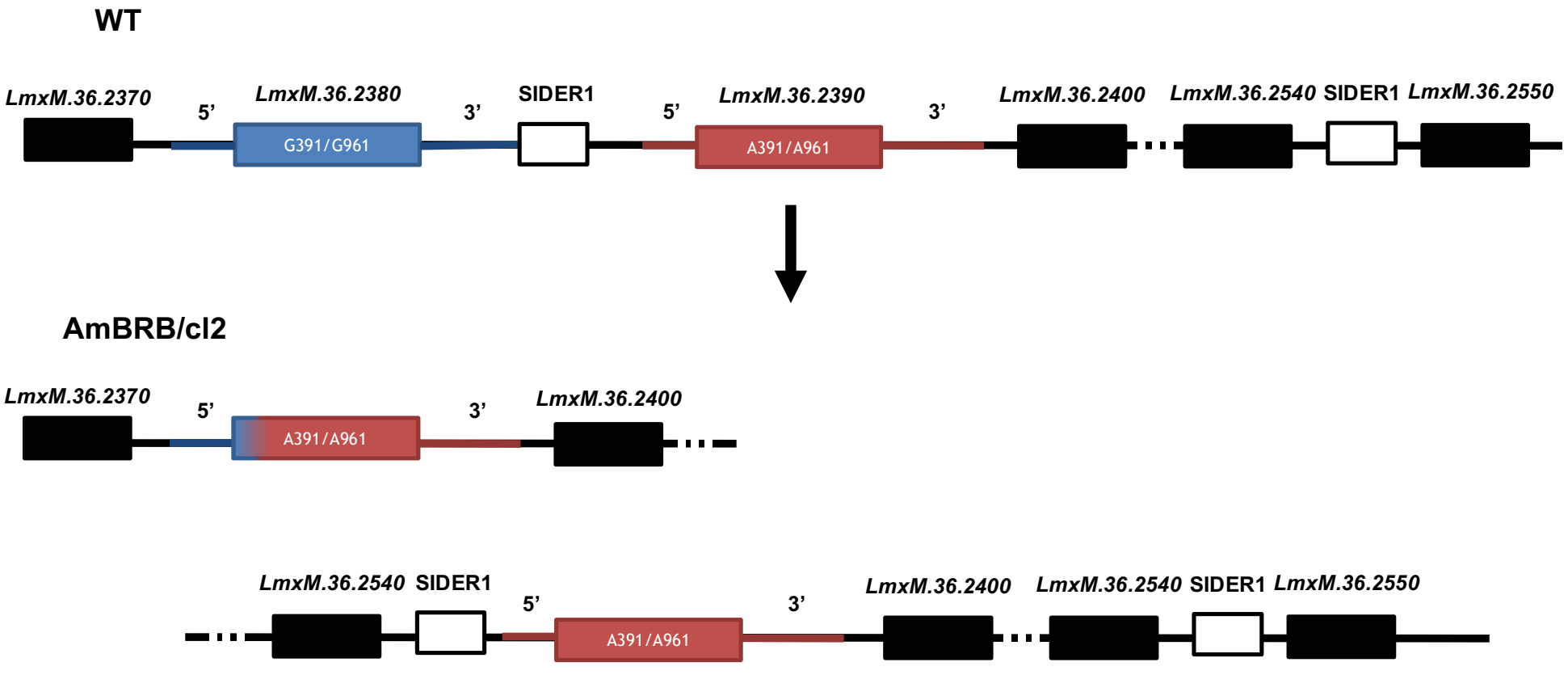


Figure 5-20: Putative model for the duplication event in AmBRB/cl2. The top part represents the putative *SMT* locus, as well as SIDER1 elements in the intergenic region and ~ 50 kb downstream. The gapped line between *LmxM.36.2400* and *LmxM.36.2540* represents multiple genes not depicted here. Note that in this model, the region within AmBRB/cl2 is contiguous, but is shown in two lines here for the sake of visualisation.

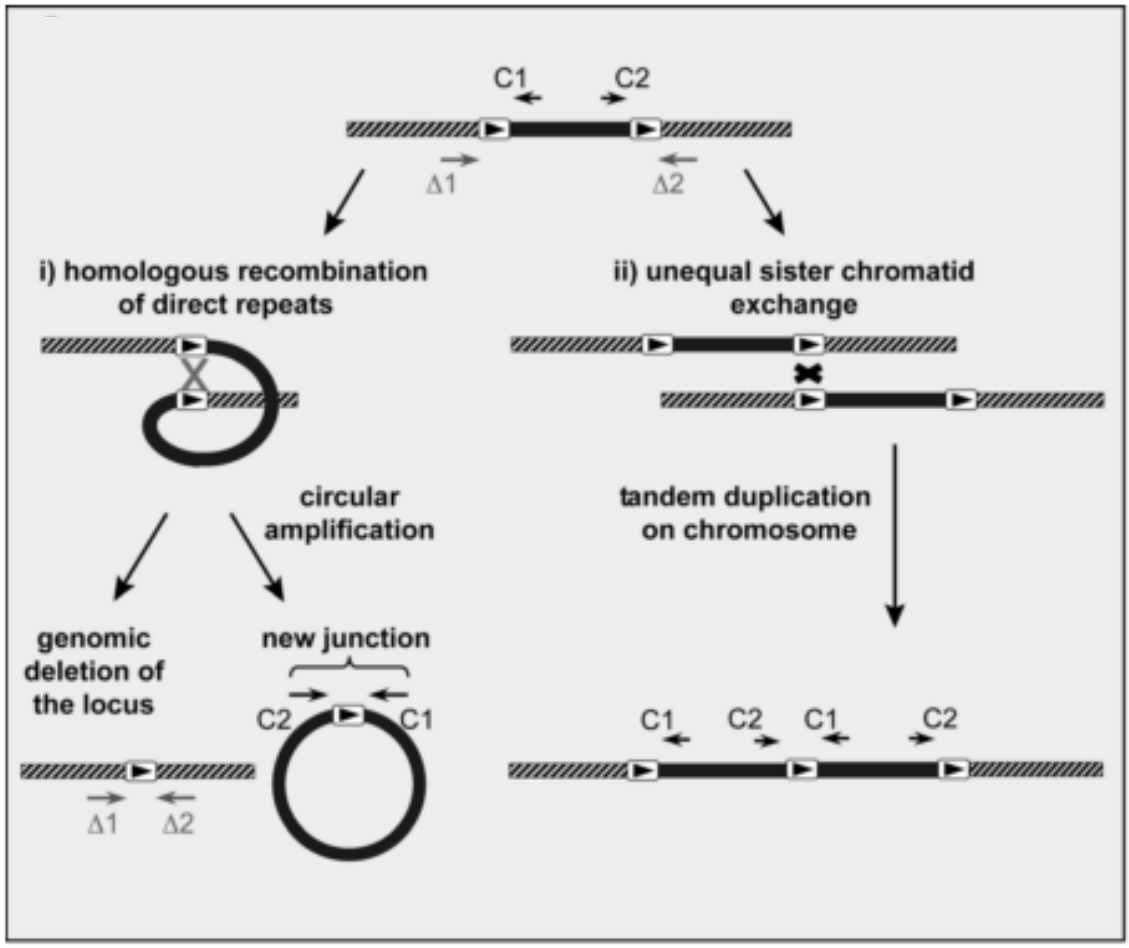


Figure 5-21: Models of tandem repeat-mediated amplification. Amplification can arise through either generation of extrachromosomal circular amplicons (i) or tandem duplication on the chromosome (ii). Amplifications can be detected using primers C1 and C2 due to the formation of new junctions. Similarly, non-conservative generation of circular chromosome, leading to deletion of the amplified region on the chromosome itself, can be detected through amplification from primers Δ1 and Δ2. Source: Ubeda *et al*. 2014.

In this model, the duplication is mediated by SIDER1 homology. However, as part of this duplication, much of the intergenic region (upstream of the SIDER1 element) is lost, possibly through a mechanism similar to that for AmBRC/cl3, with the same result of an *SMT* sequence with an A391/A961 genotype, an *LmxM.36.2380*-like 5'-UTR and an *LmxM.36.2390*-like 3'-UTR. On the other end of the duplicated region, however, repair using SIDER1 homology results in joining of the region upstream of the second SIDER1 element (including *LmxM.36.2540*) to the region downstream of the first SIDER1 element (*LmxM.36.2390*), in this case preserving both the genotype and both the UTRs of the original *LmxM.36.2390*. This model explains a number of features, including preservation of overall copy number, duplication of the *LmxM.36.2390*-like 3'-UTR and loss of the *LmxM.36.2380*-like 3'-UTR, coverage of both types of 5'-UTR, duplication of the region of *LmxM.36.2400* to *LmxM.36.2540*, the inability

to amplify an intergenic region, and the lack of evidence of gaps in coverage between *LmxM.36.2370* and *LmxM.36.2380*, *LmxM.36.2390* and *LmxM.36.2400*, and *LmxM.36.2540* and *LmxM.36.2550*. However, as I did not successfully amplified a newly formed *LmxM.36.2540-LmxM.36.2390* intergenic region (as predicted by this model), full empirical evidence is lacking.

Work by Ubeda and collegues has demonstrated a role for direct repeats, including SIDERs, in the generation of both extrachromosomal circles and linear amplicons, with *RAD51*-mediated homologous recombination required for the former, but not the latter (Ubeda *et al*. 2014). Such events occur spontaneously in the absence of selection pressure. The models they present describe either non-conservative amplification in the context of circular amplicon formation (resulting in the deletion of the intra-repeat sequence, as described in the previous paragraph) or tandem duplication on the chromosome, as a result of unequal exchange of sister chromatids. These can be differentiated in theory through detection (in the former case, but not the latter) of the newly formed junction between regions upstream and downstream of the amplified sequence (here, the loci of *LmxM.36.2370* and *LmxM.36.2550*). Whilst this is not something I have yet attempted, there are a number of arguments against extrachromosomal circular amplicon formation. Firstly, coverage of 5'-UTRs from both *LmxM.36.2380* and *LmxmM.36.2390* suggests that *SMT* copies are not identical in AmBRB/cl2, as might be expected if amplification were to arise due to circular amplicon formation. Secondly, maintenance of an extrachromosomal amplicon within a population of cells would be expected to require continuous selection pressure. This pressure may come from the requirement to retain those genes within the region lost on the chromosome itself. However, since loss of copy number is observed in AmBRC/cl3 and AmBRD/cl2, and RNA expression is decreased not only in these lines but also in AmBRB/cl2, it is difficult to imagine that selection pressure would exist for anything more than a single copy of this region.

In contrast, a linear intrachromosomal amplicon might be expected to be more stable, and the model proposed above is consistent with the presence of both 5'-UTR types. Such an amplification event, mediated by SIDER1 elements, does not on its own explain deletion of the *LmxM.36.2380*-derived 3'-UTR. If a linear

amplification event were to arise spontaneously, it would be expected to duplicate *LmxM.36.2390* (lying within the intra-repeat region) but as a conservative event, should also retain the wild-type *SMT* locus. However, if this were followed by the same process described for AmBRC/cl3 and AmBRD/cl2 (Figure 5-19), the result would be the arrangement described for AmBRB/cl2 in Figure 5-20. Moreover, such an event would result in destruction of the inter-*SMT* SIDER element, helping to stabilise the amplification. Interestingly, the work by Ubeda and colleagues in predicting amplicons based on repeats identified SIDER1-flanked sequences in both *L. major* and *L. infantum* within which *LmjF.36.2390*/*LinJ.36.2520* (*LmxM.36.2390* orthologues) and *LmjF.36.2540*/*LinJ.36.2670* (*LmxM.36.2540* homologues) were the first and last genes, respectively, further demonstrating conservation across species (Ubeda *et al*. 2014).

5.3.3 Regulation of *SMT* and the need for two copies

Whether the models described for the precise changes occurring to *SMT* are correct, I have clearly demonstrated that expression is dominated in wild-type promastigotes by transcripts containing the *LmxM.36.2380* 3'-UTR and that loss of this 3'-UTR is key to resistance. The 5'-UTR is of less importance because AmBRB has a large loss of expression, despite there being evidence of both types of 5'-UTR in this line. Consistent with evidence from previous RNA-seq data (Fiebig *et al*. 2015), there is decreased expression of *SMT* on differentiation from promastigotes to amastigotes, driven by decreased *LmxM.36.2380*, where *LmxM.36.2390* expression is unchanged. This is consistent with reduced expression of a wide range of other sterol biosynthesis genes (see section 1.2.4.3). Interestingly, there was evidence of increased expression of the *LmxM.36.2390*-like 3'-UTR relative to wild-type promastigotes in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. These changes were moderate, and it is unclear if they relate to regulated changes or simply arise due to genetic or epigenetic alterations in the regional environment of *SMT* (although in the case of AmBRB/cl2 this is clearly associated with copy number increase). In amastigotes, increased expression of the *LmxM.36.2390*-like 3'-UTR relative to wild-type was observed in AmBRA/cl1 (despite no evidence of genomic changes), AmBRB/cl2 and AmBRC/cl3, whereas in AmBRD/cl2 there was actually a decrease. Again, the cause of these changes or their significance (they are fairly modest in

magnitude) is not fully clear. It should be noted that because of the genomic changes, these cannot be said precisely to be expression changes in *LmxM.36.2390 per se*, since the apparent gene fusion events mean that whilst the *SMT* genes found in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 contain the *LmxM.36.2390*-like 3'-UTR, they do not relate to the described wild-type gene copies in their entirety, including both genotypes and UTRs.

RNA abundance in kinetoplastids is influenced particularly by elements within the 3'-UTR, consistent with these data. Therefore, it may reasonably be supposed that differences in 3'-UTRs influence both basal transcript abundance in promastigotes and differential expression on differentiation into amastigotes. Whilst SIDER2 elements have been linked strongly to transcript stability (Bringaud *et al*. 2007), what little is known about the role of SIDER1 elements appears to involve promotion of translation on differentiation from promastigotes to amastigotes, as seen in the context of the amastin genes (Boucher *et al*. 2002, Smith *et al*. 2009). By contrast, one clearly sees in the case of *LmxM.36.2380*, which has a downstream SIDER1 element, that transcript abundance decreases on differentiation. Whilst this element was not included in the Trinity assembly, it was evident that the last 10 nucleotides of the 48 nucleotide homology region shared between *LmxM.36.2380*-like transcripts was incorrect relative to the sequenced intergenic region (poly(G) instead of poly(C)) suggesting difficulties in assembly. Moreover, continuous coverage in the NGS data extended across the SIDER1 sequence, so it is reasonable to suggest that this element is included within the 3'-UTR of these transcripts. I have not evaluated translation in this study. Nevertheless, it appears that the SIDER1 element is associated with very different functional regulation in this context than in that of the amastins. Therefore, the influence of SIDER1 elements may depend strongly on the local sequence context. Notably, *LxmM.36.2540* did not show decreased expression on differentiation to amastigotes (Fiebig *et al*. 2015), despite the downstream presence of a highly homologous SIDER1 element. It would therefore be valuable to determine further the regulatory role of this SIDER1 element associated with *SMT*, both at the transcriptional and translational levels. Use of reporter assays, similar to those for the amastin SIDER1 (Smith *et al*. 2009) are a feasible option, although endogenous protein tagging, using transcript-specific regions to target to specific gene copies, may

represent an alternative option that does not take the element out of its local genomic context.

A final question relates to the gene copy number of *SMT*: why have two copies, when you can have one? The occurrence of two copies is conserved in the reference genomes of *L. mexicana*, *L. major*, *L. infantum*, *L. aethiopica*, *L. arabica* and *L. gerbilli*, albeit not in *L. braziliensis*, although given the problems already described for *L. mexicana*, these numbers are not certain to be correct in all cases (as argued for *L. donovani*). The formation of multiple gene copies can arise because it allows for divergence of gene function; however, in this case, the coding sequences are almost identical, and our evidence does not suggest an influence on gene function in an episomal overexpression context (although subtle differences in catalytic parameters cannot be discounted). In the absence of gene-specific promoters, increased copy number may also be a mechanism to ensure higher expression; yet expression here is clearly dominated by one copy. Perhaps one clue lies in the phenotype of resistant lines overexpressing *SMT*: despite considerable overexpression in these lines, neither wild-type AmB sensitivity nor sterol composition were fully recovered. It should be noted, however, that whilst loss of the *LmxM.36.2380*-like 3'-UTR was *the* change observed in AmB-resistant lines with reduced *SMT* expression, I did not reintroduce *SMT* associated with this 3'-UTR. It is perfectly possible, therefore, that there is a specific role played by this UTR that cannot be compensated for by high SMT activity alone. One such function could be mRNA localisation. A similar mechanism has been argued for *Leishmania* ß-tubulin genes, whose multiple transcript types appear to undergo independent differential regulation (Coulson *et al*. 1996). The authors noted the presence of a motif (5'-GGATC-N_{120}-AATGC-3') which they argued was similar to an 'RNA zipcode' localising ß-cytoplasmic actin mRNA in fibroblasts (5'-GGACT-N_{40}-AATGC-3'). They presented no empirical evidence of this, however. More recently, subcellular localisation of mRNAs has been directly confirmed in *T. cruzi* (Alves *et al*. 2013), suggesting that kinetoplastids may be capable of such a mechanism in principle. The authors of this study used fluorescence *in situ* hybridisation to track mRNAs within the parasites, an approach that should in principle be applicable here. Different transcript localisation could in turn lead to different compartmentalisation of the protein product, in turn providing access to

variable substrates, or interacting with other proteins, such as enzymes downstream in the same pathway or regulatory factors.

5.4 Conclusions

In this chapter, I have directly demonstrated a role for changes to our three candidate genes, *SC5D*, *SMT* and *MT*, in AmB resistance. As expected in the case of *SC5D* and *SMT*, both involved in sterol biosynthesis, resistance appears to derive from changes in sterol composition arising from these lesions. This is not the case for *MT*, hence in accord with previous analysis (Fernandez-Prada *et al*. 2016), AmB resistance associated with *MT* lesions arises through a sterol-independent mechanism.

There are still some questions relating to the effect of SNPs on enzyme function. Whilst the total restoration of wild-type AmB sensitivity and sterol composition on reintroduction of *SC5D* into AmBRA/cl1, compared to the lack of change on reintroducing the G139R-substituted copy (despite overexpression) strongly argues in favour of complete loss of function in this mutant, the identification of low levels of 5(6)-7(8)-desaturated sterols in AmBRA/cl1 questions this conclusion. It may require use of standards to rule out the possibility that this sterol is misidentified. In *SMT*, no SNPs appear to engender total loss of enzymatic function. However, the extent of functional effects cannot be determined in a system in which genes are overexpressed to different degrees, albeit above wild-type levels in all cases. *In vitro* expression and characterisation (as has previously been achieved for the enzyme in *T. brucei*, Zhou *et al*. 2006) will be an interesting approach to determine effects of mutation on both activity and specificity, particularly for the F72C substitution found in AmBRC/cl3.

Further insight has also been gained in this chapter into the basis of PENT hypersensitivity associated with AmB resistance. Due to the conservation of both sterol changes and PENT hypersensitivity in AmB resistance, a link between the two is unsurprising. Thus, reintroduction of *SMT* partially or totally restored wild-type PENT sensitivity in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2. A similar role for *CYP51* was confirmed previously (Mwenechanya *et al*. 2017). The minimal impact of *SC5D* reintroduction into AmBRA/cl1, however, was

unexpected. This suggests that in this line, PENT hypersensitivity arises from separate mutations to AmB resistance, although such changes may well arise in compensation for sterol changes. The biggest effect of gene reintroduction on PENT hypersensitivity, however, was with *MT*. This is of particular biological significance, as it raises the possibility that there may be a similar inverse correlation between sensitivity to PENT and MILT. A previous study reported no change in PENT sensitivity in an *L. infantum* line selected for MILT resistance (carrying a homozygous nonsense mutation in *MT*) (Vacchina *et al*. 2016). Therefore, it may be that the influence of MT on PENT sensitivity may be dependent on genetic background (including species-species variation, as well as other resistance-associated mutations).

Finally, I attempted extensive characterisation of the structural changes observed at the *SMT* locus in three resistant lines and their influence on gene expression. The challenge of defining events at this locus is considerable. In addition to the weaknesses of NGS in this regard, one encounters large tracts of non-unique sequence, regarding both the coding sequences themselves and SIDER1-like elements. I have tackled this question through a combination of methods but uncertainties remain; in particular, whilst I corrected an obvious error in the intergenic region between *SMT* copies, one cannot rule out broader problems with the ordering of genes within this locus that might sharply undermine our conclusions. Nevertheless, I have defined with some certainty that structural rearrangements occur in all cases, and that a defining feature of these changes is disruption of the 3'-UTR sequence I have here associated with *LmxM.36.2380*. I have also identified a putative role for sequence homology, arising not only due to tandem gene duplication but also SIDER1 elements, in local genomic instability, something which may well be conserved in other *Leishmania* species (as the SIDER1 elements are themselves conserved). Last of all, I have found evidence of transcript-specific regulation on differentiation from promastigotes to amastigotes. This is important because it influences our expectations of how resistance may arise in the mammalian host. However, it also opens up a range of possibilities to study the unusual regulatory biology of *Leishmania* in the context of a core metabolic enzyme, resistance-associated gene and putative drug target.

6 General discussion

6.1 Understanding ergosterol biosynthesis in *Leishmania*

The two sterol biosynthesis genes I have identified in this study join *CYP51* (Mwenechanya *et al*. 2017), meaning that there are now three separate genes in this pathway associated with AmB resistance in *Leishmania*. One cannot say whether this is an exhaustive list, or whether mutations in other genes can give rise to resistance – mutations in C8-sterol isomerase (*erg2*), for example, have previously been identified in AmB-resistant *C. albicans* (Vincent *et al*. 2013). This is likely to depend on the essentiality of individual genes for survival, both *in vitro* and in the mammalian host. Indeed, this essentiality is likely to influence the types of mutations that can arise. Most evidence points to complete loss of functionality in G139R-mutated SC5D, although there is some contradictory evidence in the GC-MS sterol analysis (see section 5.2.1.3). On the other hand, regarding the *CYP51*-mutated AmB-resistant line (Mwenechanya *et al*. 2017), substantial amounts of both ergosterol and 4,4-dimethylcholesta-8,14-24-trienol (the product of reaction) were observed, strongly suggesting that enzymatic activity is retained in spite of the N176I mutation. Similarly all lines in this study carrying *SMT* lesions demonstrate retention of detectable C24-methylated sterol. In the former case, studies in *L. major* and *L. donovani* have revealed *CYP51* essentiality to be species-specific, or possibly related to specific growth conditions that were not readily discernible in the relevant studies (Xu *et al*. 2014, McCall *et al*. 2015). The essentiality of *CYP51* in *L. mexicana* has not been determined; however, the fact that some activity was retained in the N176I-mutated line (Mwenechanya *et al*. 2017) could suggest total loss of function is not possible. This cannot be determined on the basis of one mutated line, however, as mutation is a stochastic process and the N176I substitution may simply have arisen first and become fixed in the population. Furthermore, the presence of a related enzyme, CYP5122A1, which is essential in *L. donovani*, further complicates this picture as its specific function has not been determined. In the case of SMT, there is evidence of a "growth-sparking" role of C24-methyltransferase activity in *T. brucei* (Haubrich *et al*. 2015), although there is some evidence against this in *Leishmania* promastigotes (Haughan *et al*. 1995). Therefore, it would be interesting to determine which genes are truly essential in sterol biosynthesis in *L. mexicana*. As CRISPR-Cas9 is now available as a tool

for genome editing in *L. mexicana* (Beneke *et al*. 2017), there is an opportunity to do this on a broader scale. CRISPR-Cas9 systems are particularly suited to disruption of multicopy genes, something which is an issue with *SMT*, but also more broadly given the potential for polyploidy in *Leishmania*. Such an approach could yield further insight into the effects of disrupting individual genes on drug sensitivity and virulence.

Targeted disruption of individual genes would also provide a far cleaner genetic background in order to assess the effects of these changes. Nevertheless, assessment of the effects of gene disruptions in these resistant lines can provide insight into sterol metabolism in *Leishmania*. For example, in the conventional pathway for ergosterol biosynthesis as depicted in Figure 1-7, sterol C5-desaturase activity acts before sterol C24-methyltransferase activity. Therefore, unsurprisingly, the predominant sterol in AmBRA/cl1, the *SC5D* mutant, is C24-methylated (ergosta-5,22-dienol). On the other hand, if this pathway were to be strictly correct, loss of SMT activity should also result in loss of 5(6)-7(8)-desaturation, as SMT acts before SC5D in the pathway. This is not observed, however, as cholesta-5,7,22-trienol and cholesta-5,7,24-trienol are the dominant sterols in these lines. Therefore, SC5D must be able to act on sterols lacking C24-methylation. This has been proposed previously for other trypanosomatids (Roberts *et al*. 2003). Furthermore, in both cases, the major sterol is desaturated at the C22 position. This is similar to ergosterol, but differs from the major wild-type sterol identified in *L. mexicana*, ergosta-5,7,24(28)-trienol. As the substrates of both SMT and SC5D in the conventional pathway (zymosterol and episterol) are C24-desaturated, this may also suggest that the last two stages of this pathway, which shift the double bond from the 24(28) position (ergosta-5,7,24(28)-trienol) to the 22(23) position (ergosterol) happen irrespective of upstream disruptions. Standards would be required to ascertain whether these double bond positions are correctly identified, however.

One unanswered question is in regard to the genesis of another sterol, found only in lines overexpressing sterol biosynthesis genes: 4,4-dimethylcholesta-5,7,9(11)-trienol. Removal of C4-methyl groups from the sterol ring structure occurs upstream of both SMT and SC5D. If identified correctly, therefore, it is unclear why restoration of these enzymes, particularly SMT, would result in

accumulation of a sterol that has not undergone either C4-demethylation and C24-methylation. I have postulated that this could be due to higher activity than wild-type parasites, which in the case of SMT could lead to either some sort of feedback to earlier in the pathway or accumulation of an additional product of the reaction (such as a C24-dimethylated sterol, misidentified as a C4-dimethylated sterol). However, the fact that wild-type sterol composition is only partially restored (with a lower ratio of C24-methylated to unmethylated sterols than wild-type parasites) argues against straightforward 'overactivity' as a cause. Overexpression in a non-native context, therefore, may not fully compensate for wild-type expression driven by endogenous UTRs. Alternatively, reintroduction of sterol biosynthesis genes onto the background of AmB resistance may lead to unexpected effects due to the presence of compensatory mutations, adapted to resistance-associated sterol changes. This provides further incentive for targeted disruption of genes, combined with more extensive use of sterol standards. Such a study would elucidate what appears to be a complex pathway, with evidence of multiple substrates and possibly multiple products for individual enzymes. Given the potential for development of ergosterol biosynthesis as a drug target (for example, the CYP51 inhibitor posaconazole has previously been used to treat successfully *L. infantum* CL, Paniz Mondolfi *et al*. 2011), a better understanding of this pathway is essential in predicting the efficacy of targeting different stages in this pathway.

6.2 The clinical relevance of this study

In this study, I have focused primarily on the mechanisms of resistance *in vitro*, primarily in the promastigote (insect) stage of the *L. mexicana* life cycle. Therefore, it is important to consider the relevance for the potential emergence of resistance in clinical settings. At present, evidence of resistance amongst clinical *Leishmania* populations is restricted to a few isolated reports (Srivastava *et al*. 2011, Purkait *et al*. 2012). One source of acquired resistance that seems particularly feasible is in the context of VL-HIV coinfection, in which poor parasite clearance after treatment may lead to repeated cycles of drug exposure and relapse. Decreased AmB susceptibility in *L. infantum* has been observed in this context (Di Giorgio *et al*. 1999), although another study found no evidence of altered AmB susceptibility in the same clinical context even after multiple treatment cycles (Lachaud *et al*. 2009). Therefore, the feasibility of significant

emergence of AmB resistance remains unproven, although it cannot be ruled out. Moreover, particularly in the Indian context, emergence of antimonial resistance appears to have resulted in parasites that are more resistant to oxidative stress (an important feature in AmB-mediated killing) and more immunomodulatory than their antimony-sensitive counterparts, factors that may facilitate further acquisition of drug resistance in these parasites. Moreover, I have confirmed a link previously demonstrated (Fernandez-Prada *et al*. 2016) between mutations in the miltefosine transporter with MILT-AmB cross-resistance. Recent identification of miltefosine transporter polymorphisms in MILT-resistant clinical isolates (Srivastava *et al*. 2017) further highlights this threat, and here I observed direct evidence of loss of miltefosine transporter function prior to sterol changes in AmBRB. Decades of experience of AmB use in pathogenic fungi suggests that acquired AmB resistance is rare, at least when compared to other antifungal drugs, although it is known and can be selected in the laboratory. However, many fungal infections are opportunistic, and AmB is reserved for only the most severe cases; thus, the selection pressures faced by *Leishmania*, which differs in both these regards, may be very different, questioning lessons drawn from use of AmB as an antifungal.

With regard to this study, therefore, one can make a number of observations. In chapter 3, I found limited evidence of hypersensitivity to oxidative stress, in contrast to previous studies in *L. mexicana* and *C. albicans* (Mwenechanya *et al*. 2017, Vincent *et al*. 2013). Thus, greater susceptibility to oxidative stress does not appear to be an essential accompaniment to selection of AmB resistance, although the basis of hypersensitivity to methylene blue requires further investigation. I also saw little evidence of cross-resistance to other antileishmanials as an inevitable feature of AmB resistance. Furthermore, at least one line, AmBRC/cl3, retained comparable levels of infectivity in primary macrophages to wild-type parasites, as well as the ability to survive *in vivo*. This is in spite of disruption of sterol biosynthesis through both large-scale genomic structural variation and F72C substitution in SMT. Whilst the other three lines showed significant loss of infectivity, one cannot distinguish the effects of AmB adaptation on virulence from the effects of serial passaging. Thus, the evidence from this study of intrinsically prohibitive fitness costs of AmB resistance is limited. One potentially useful weakness that may well be an inevitable trade-

off, however, is PENT hypersensitivity, which I found in all lines. PENT is already available for clinical use. Hence, should AmB-resistant *Leishmania* emerge, greater potency of PENT may make this drug a useful treatment option. The role of disruption of the miltefosine transporter in this phenotype is of considerable interest, given that loss of function in this gene may already be in the process of emergence due to widespread MILT administration.

Even if changes associated with AmB resistance *in vitro* are compatible with survival *in vivo*, it is not certain that the phenotype will be sufficient to result in clinical problems such as relapse. Whilst relapse due to infection by parasites selected for AmB resistance has never been shown in an experimental model, one such *L. mexicana* line did show insensitivity to AmB treatment *in vivo*, where the same treatment produced a temporary delay in lesion growth in wild-type-infected mice (Al-Mohammed *et al*. 2005). However, this line exhibited 16-fold AmB resistance relative to wild-type parasites, whereas the ones selected here range in relative sensitivity between 2.6-fold and 8.4-fold (although the latter is similar to the level found by Purkait and colleagues in a clinical isolate, Purkait *et al*. 2012). Indeed, the only one of these lines retaining measurable resistance as intracellular amastigotes, AmBRC/cl3, demonstrated only a twofold higher IC_{50} than wild-type in this stage (although I have argued in section 3.3.3.2 that in the other lines, loss of infectivity and intracellular replication may well be a confounding factor).

With respect to mutations in the sterol biosynthesis pathway, it is evident that not all have equal effect. Reintroduction of *SC5D* into AmBRA/cl1 caused a 6.4-fold decrease in sensitivity to AmB. Reintroduction of *CYP51* into previously selected AmB-resistant *L. mexicana* (Mwenechanya *et al*. 2017) caused an even bigger change, from approximately 2.5 μM down to wild-type levels of 50-100 nM. On the other hand, reintroduction of *SMT* only caused a 2-2.5-fold decrease in AmB IC_{50}. This could be attributable to a number of factors, namely evidence of residual enzyme activity in resistant parasites and incomplete restoration of wild-type sterol composition in these overexpression lines. It is also likely that different modifications to the predominant sterols in these lines result in different effects on interactions with AmB. Indeed, only the sterols in *SMT*-mutated lines retain the 5(6)-7(8) conjugated double bond system, essential for

planarity in the ergosterol ring structure. Therefore, it is evident that different modifications to sterol composition have different degrees of influence on AmB sensitivity.

Another level of consideration is the potential for biological variation in the effect of these changes between life cycle stages. For example, increased reliance on uptake of environment-derived cholesterol in comparison to endogenous synthesis of ergosterol and related sterols has been documented on differentiation of log-phase promastigotes into metacyclic promastigotes (Yao & Wilson 2016) and amastigotes (Bouazizi-Ben Messaoud *et al*. 2017, Jara *et al*. 2017). Moreover, decreased RNA abundance of a number of ergosterol biosynthesis genes has been documented in *L. mexicana* amastigotes in comparison to promastigotes (Fiebig *et al*. 2015), notably *LmxM.11.1100* (*CYP51*) but not *LmxM.23.1300* (*SC5D*). I have also validated changes reported in *SMT* genes: *LmxM.36.2380* the dominant transcript in promastigotes, decreases in amastigotes whilst *LmxM.36.2390* remains unchanged. This offers, therefore, an alternative explanation of resistance patterns observed in intracellular amastigotes. Due to increased relative importance of transcripts containing the *LmxM.36.2390*-like 3'-UTR in amastigotes, the loss of transcripts containing the *LmxM.36.2380*-like 3'-UTR, as observed in AmBRB/cl2, AmBRC/cl3 and AmBRD/cl2 may not in itself be enough to effect resistance. Rather, it is the secondary mutation in *LmxM.36.2390*, F72C, which reduces SMT activity in amastigotes sufficiently to influence AmB sensitivity. A conclusion of this is that, although sequence repetition may well predispose the SMT locus to genomic instability and the sorts of structural changes I have described here, secondary mutations may also be required in order to influence AmB sensitivity *in vivo*.

The mutations in *SMT* found in AmBRC/cl3 appear not to compromise parasite fitness, and are associated with resistance within primary macrophages. However, this line is only moderately (~four-fold) resistant in the promastigote stage, and twofold in the amastigote stage. The question, therefore, is whether this represents a clinically meaningful change. Whilst one cannot answer this with certainty, the following might be offered. First, all analyses of drug resistant *L. donovani* have pointed towards loss of sterol C24-methyltransferase activity, including in a clinical isolate (Mbongo *et al*. 1998, Pourshafie *et al*.

2004, Purkait *et al*. 2012). Secondly, there is increasing evidence that resistance to AmB may arise through multiple mutations, including sterol changes, but also lesions in the miltefosine transporter, increased resistance to oxidative stress, and effects in other areas of metabolism (see 1.3.5.2 for a full discussion). Crucially, evidence has been presented here and elsewhere (Purkait *et al*. 2012, Fernandez-Prada *et al*. 2016) that multiple such adaptations can occur cumulatively in the same line, fitting with the pattern seen previously for resistance to both antimony and MILT. Therefore, it is a reasonable prediction that, were resistance to emerge in a clinical context, a similar pattern would appear: multiple changes may be involved, including alterations to sterol composition such as loss of C24-methylation. Whilst in themselves, these changes may not be enough to cause treatment failure in all cases, they may nonetheless influence relapse rates, in turn providing the opportunity for further accumulation of mutations. AmB administration has only recently become more widespread; however, this process may already have begun with selection of relevant changes through exposure to other drugs, notably Sb^V and MILT.

6.3 Genomes, transcriptomes and metabolomes

In Chapter 4, I described application of NGS to characterise numerous genomic and transcriptomic changes observed in the four AmB-resistant lines. In doing so, I highlighted the role of CNV, particularly at the chromosomal level but also in the context of sub-chromosomal amplifications, in shaping the transcriptome. Heterogeneity in gene copy number was recently shown to dominate variability in expression levels in *L. tropica* populations (Iantorno *et al*. 2017). Similarly, in our data, large numbers of moderate but statistically significant changes were detected across individual resistant lines and wild-type parasites, most of which could be correlated with changes in chromosomal CNV.

Whilst the link between karyotypic flexibility and gene expression is increasingly recognised as a key factor in intraspecific diversity, its functional implications are yet to be explored. Evidently, extrapolating this to variation at the level of protein abundance is challenging, given the important role played by translational regulation. Proteomics offers the most direct route to this, although the emergence of ribosome profiling as a technique to study translational regulation may provide a broader understanding at a whole-genome

level, due to its potential for higher coverage. Use of the latter technique in *T. brucei* (Jensen *et al.* 2014) has demonstrated that in the context of life cycle differentiation, changes in transcript abundance and translational efficiency frequently act cooperatively to amplify differences in protein production relative to changes in RNA expression alone. Whilst one might expect that an increase in transcript abundance might lead to a proportionate increase in protein production, without the per-transcript translational efficiency being affected, feedback mechanisms due to altered abundance of either individual or overall transcripts may serve to complicate further the influence of karyotypic changes. I argued for the presence of such a mechanism in section 4.3.4 with regard to the expression of translation-associated genes.

However, one can also directly investigate effects of transcriptome remodelling further downstream than protein production, by investigating their influence on cellular metabolism itself. I have attempted to do this in a pathway-by-pathway fashion, correlating the WGS and RNA-seq datasets discussed in chapter 4 with the untargeted LC-MS metabolomics assay in chapter 3. Whilst I did find changes at all three levels in the same pathway, I was unable to interpret these for the most part in terms of functional effects of changes at the genomic or transcriptomic levels on the abundance of individual metabolites. This may be because the relatively modest changes to transcript abundance did not affect the activity levels of the enzyme products of these genes; this is something a more comprehensive understanding of proteomic changes would provide (including, perhaps, posttranslational modifications). Incomplete coverage of the metabolome also confounds interpretation. However, it may be that by introducing so many expression changes, each relatively small in its effect, the metabolome is remodelled in such a complex way as to make correlation on the level of individual genes and metabolites largely unfeasible.

What evidence is available, therefore, that metabolism is substantially affected by transcriptomic variability? If the relationship between metabolome and transcriptome (and by extension, genome) is highly complex, it is likely to require analysis at the level of the whole system. During the course of this thesis, I have not attempted such analysis, in part because the sophisticated modelling required is complex, and the data were collected without the

intention of performing a systems analysis. For example, whilst WGS and metabolomics profiling were performed on individual clones, RNA-seq used separate clones for each biological replicate. As discussed in chapter 4, given the expected instability of chromosomal copy number, one cannot assume that there is no ploidy variation between individual clones of the same AmB-resistant line. Therefore, whilst one can directly compare the clone sequenced in WGS with RNA-seq data for that same clone (as I have done), these efforts are hampered by the lack of biological replicates from the same clone. Secondly, and relatedly, this instability also means that one would have to obtain genomic, transcriptomic and metabolomics data from identical populations of cells. In practice this means obtaining all three from the same flask of parasites. Again, whilst I did this for RNA and DNA extractions from the same clone, metabolite extractions were performed as part of an entirely distinct experiment.

In spite of these caveats, it is interesting to compare the two PCA plots derived from the transcriptomics and metabolomics data (reproduced in Figure 6-1). There is a striking similarity between the overall arrangement of parasite lines within these plots. Specifically, AmBRB(/cl2) is separated from the others in both cases by principal component (PC) 1 (*x*-axis), which accounts for 34.3% and 32% of variance in the metabolome and transcriptome, respectively. The other clones cluster more closely but are separated along PC2 (*y*-axis), which accounts for 16.9% and 25% of variance in the metabolome and transcriptome, respectively. One difference is that AmBRA/cl1 and AmBRD/cl2 cluster closely but are separated primarily by PC1 in the metabolomics data, whereas in the transcriptomics data, both PCs separate these lines. Otherwise, the arrangement of lines is similar. It is important to note that whilst in both cases the data are subject to log-transformation, they are unscaled. This can lead to dominance of loadings by a small number of functionally related changes (e.g. loss of expression of one gene leading to a large change in another, highly abundant metabolite), rather than overall variance. However, subjecting the metabolomics data to additional scaling (both Pareto and range scaling, Figure 6-2) failed to change the overall distribution, suggesting that these patterns are largely robust. Evidently, the exact PCA plots are highly dependent on the data included, and whilst all detected genes were included for the transcriptome PCA plot (most of which are unrelated to metabolism), only those metabolites

passing initial filtering were used to construct the metabolome PCA plot. Clearly, these observations do not tell us anything specific about the relationship between transcriptome and metabolome, nor do they imply causality, but they do support the possibility of a connection between the two.

It is therefore possible to imagine an experiment to generate the data for an in-depth analysis of this relationship. Ideally, there would be maximal variability at the karyotypic level, with minimal variability in nucleotide sequence. It could be enough to isolate individual clones from a wild-type population of parasites, expand these and then extract DNA, RNA and metabolites. This should yield low interclonal sequence variability, but one would still expect variations in karyotype. Indeed, by deriving populations from individual clones, one would expect to observe substantially lower karyotypic mosaicism than in the parental population, thereby allowing easier classification of chromosomal ploidy (performing such combined analysis on single cells would be ideal, but is well beyond current technological capacities). Integrative analysis may allow identification of the complex relationships between gene copy number, transcript expression and metabolite abundance.

If these relationships prove to be substantive, it would suggest that karyotypic diversity is a major factor underlying metabolic heterogeneity within populations. As metabolism is expected to influence sensitivity to a wide range of environmental factors, from the host environment and nutrient availability through to drug exposure, such heterogeneity would allow for a high degree of stochastic inter- and intra-host metabolic diversity, improving the resilience of *Leishmania* populations to a range of challenges. To go a step further, if models based on such data were robust enough, they may allow prediction of the metabolic effects of specific ploidy changes, and even resulting phenotypic effects. Because the host environment may place specific pressures on parasite populations, resulting in more deterministic ploidy shifts (Dumetz *et al*. 2017), it would be important to aim as close to the *in vivo* context as possible. The challenge of producing enough parasite material (particularly for metabolomics) is, however, a significant limitation. Axenic amastigotes could potentially offer a compromise in this regard, although their strength as a model in this respect is still subject to some debate.

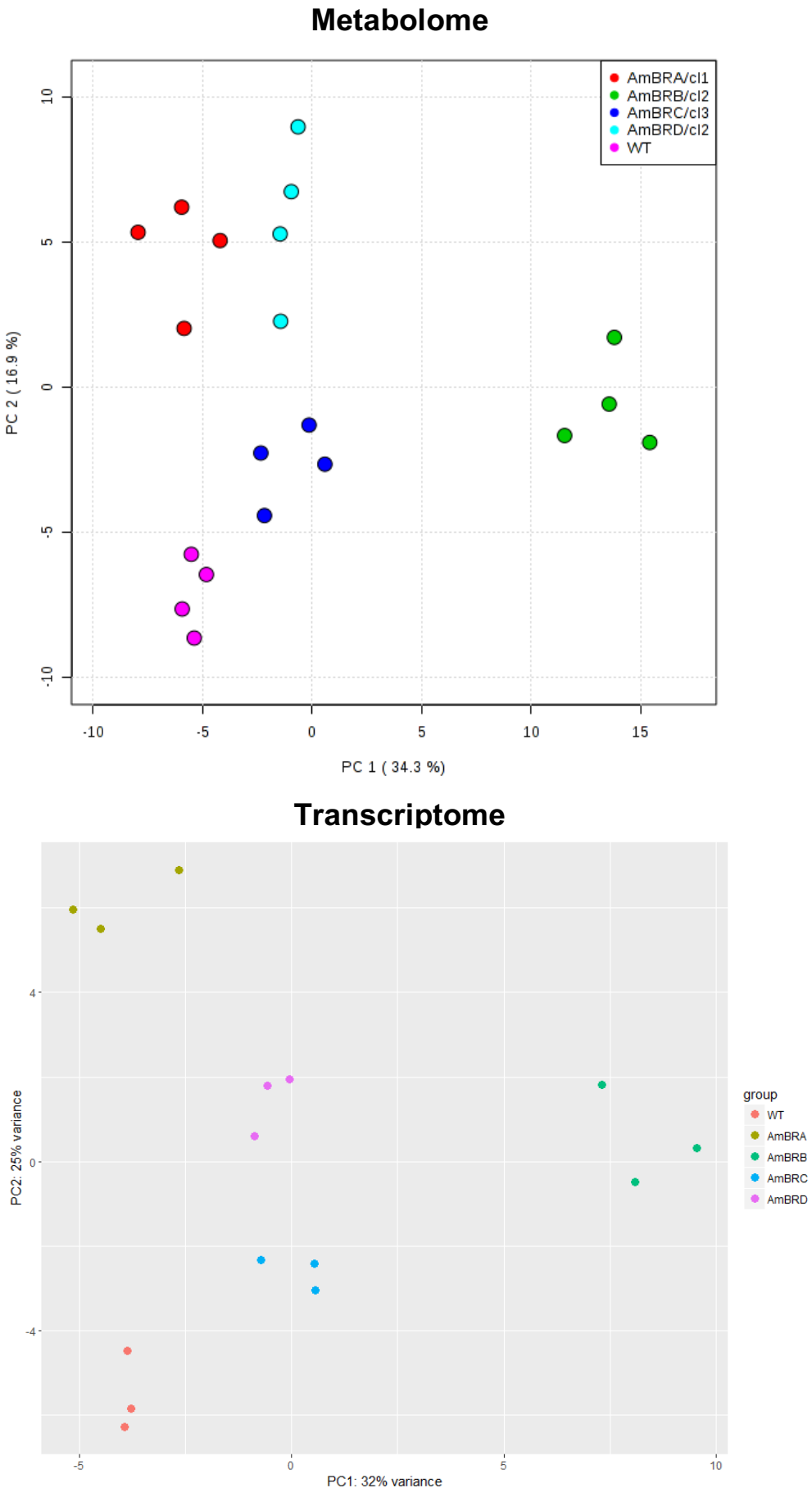


Figure 6-1: Reproduction of PCA plots for genomic and transcriptomic data. Plots are identical to those in Figure 3-9 and Figure 4-11. Note the different colour scale in each case.

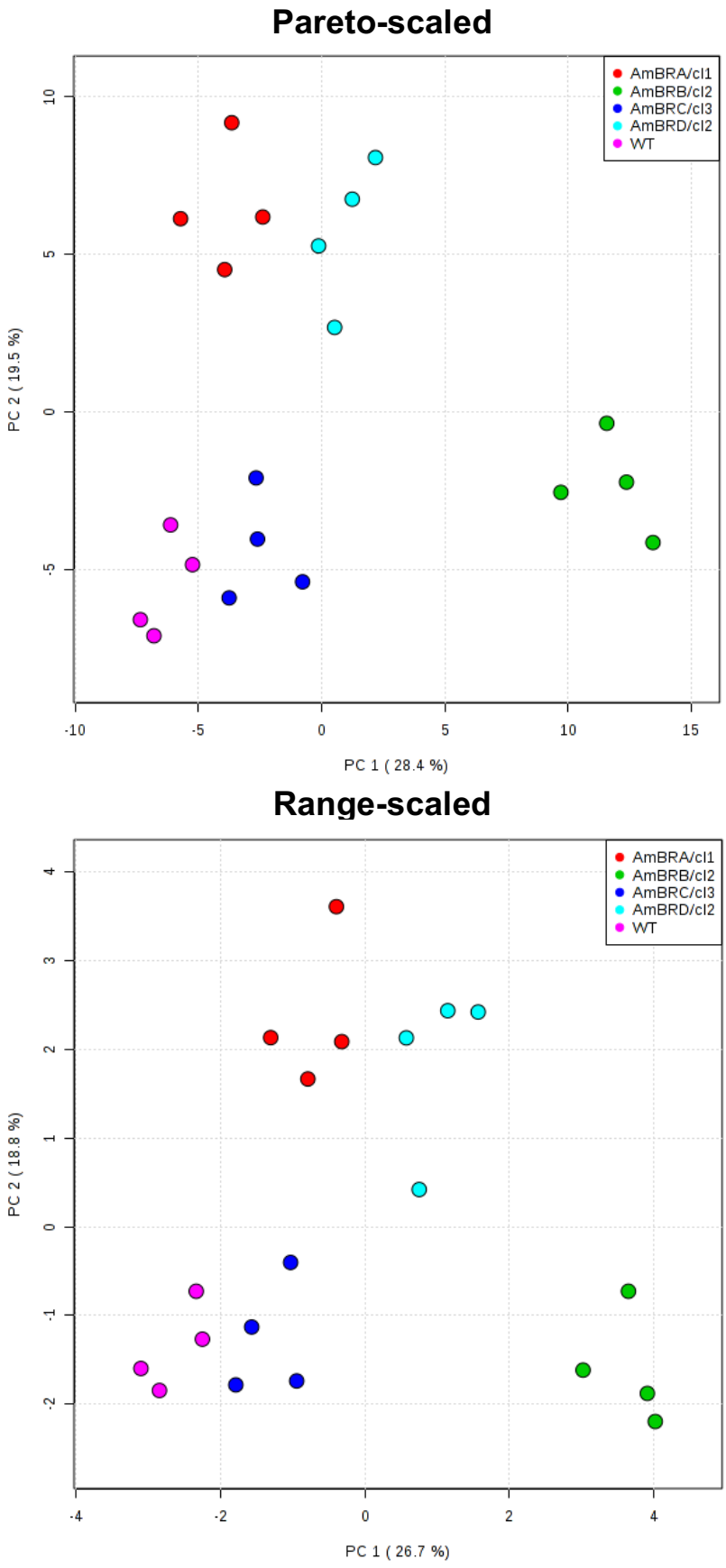


Figure 6-2: Reproduction of metabolomics PCA-plots with different scaling methods. PCA plots were generated using Metaboanalyst 3.0 software (Xia et al. 2015) as for Figure 3-9, introducing different scaling methods as labelled. Range scaling proceeds by adjusting differences in signal intensities according to the total biological range for a given metabolite. The Pareto method scales relative to the standard deviation of that metabolite.

Chapter 6

6.4 General conclusions

During the course of the work described in this thesis, I successfully selected resistance in four independently derived lines, and characterised phenotypic, metabolic and genetic changes. This is the first such study to compare multiple lines selected for AmB resistance under the same conditions in parallel. Through this approach, I was able to dissect which changes are essential accompaniments of AmB resistance, and which may be more related to stochastic changes emerging during the selection process. Indeed, this revealed that most of the changes I identified, including loss of infectivity, altered sensitivity to other antileishmanial drugs, and large numbers of changes to metabolism, chromosomal ploidy and gene expression, are not reproduced between lines. The list of conserved changes is much shorter. As I argued with regard to both the LC-MS metabolomics and RNA-seq transcriptomics, one needs to take care in interpreting even these changes, as I did not control in these cases for deterministic effects of long-term passaging. Moreover, some of these conserved changes may simply arise by coincidence, given the large number of parameters I have measured. This makes it harder to predict what phenotypes any emergent AmB-resistant *Leishmania* may display, but weakens the notion that such parasites would inevitably have severe costs to fitness or vulnerability to dangers such as oxidative stress.

Similarly, I have identified a dominant role for chromosomal and gene CNV in reshaping the transcriptome in a manner which appears to be largely random. Whilst the RNA-seq data proved to be critical to our understanding of the effects of genomic changes to *SMT* on gene function, largely this poses a considerable limitation on identification of functionally relevant changes in transcript abundance. Nevertheless, I have argued in this chapter that there may well be an important connection between the transcriptome and the metabolome in *Leishmania* parasites, which may contribute to population diversity and robustness. Further generation and analysis of empirical data are required to interrogate this relationship.

Amongst this high level of noise, I have focused on changes to three genes: sterol C5-desaturase, sterol C24-methyltransferase, and the miltefosine transporter. I have validated a role for all three in AmB resistance, as well as

demonstrated their variable contribution to PENT hypersensitivity, a phenotype universally reported in all AmB-resistant *Leishmania* lines tested. SC5D and SMT both influence AmB sensitivity through their role in sterol biosynthesis, demonstrating that AmB resistance can emerge through mutation in at least three sterol biosynthesis genes (including *CYP51*). This is the first such study to report a mutation in *SC5D* as responsible for AmB resistance in *Leishmania*; by contrast, whilst altered *SMT* expression has been described in this context before, this study is the first to show the genomic basis of such changes. Similarly, single nucleotide mutation in *MT* has been described in AmB-resistant *L. infantum*, but here complete deletion is described. Moreover, by connecting this deletion to PENT hypersensitivity, PENT is added to AmB and MILT as antileishmanial drugs whose efficacy is influenced by this transporter. By demonstrating that PENT hypersensitivity is associated with both sterol-dependent and sterol-independent mechanisms of AmB resistance, it further strengthens the prediction that any future emergence of AmB-resistant *Leishmania* would be hypersensitive to this drug.

Finally, this work has contributed to our understanding of the genomic locus of *SMT*, through correction of problems with the sequence itself, investigation of the roles and regulation of the individual gene copies, and demonstration of how this locus may change during selection of drug resistance. I have shown that the relative expression of the two transcript types differs substantially between cell cycle stages, suggesting that additional disruption beyond loss of *LmxM.36.2380* may be required for resistance in the mammalian host. The F72C substitution in AmBRC/cl3 provides one such further mutation. I have also argued for a potential role of repetitive sequence elements in promoting genomic instability. Given that no fitness costs could be associated with loss of SMT function, the potential for such instability within clinical populations should be of some concern. From a biological perspective, there are many unanswered questions regarding regulation of SMT abundance (at both RNA and protein levels), and the fact that episomal overexpression cannot in itself fully replicate normal gene function.

The lessons from decades of experience in control of parasites and other pathogens using chemotherapeutic agents contend that no drug is free from the

risk of emergent resistance. The current administration of AmB as a single drug, and frequently as a single injection, in efforts to control a parasite with exceptional genomic plasticity and adaptability, only reinforces this risk. Our work presented here offers predictions about how AmB resistance may arise, and may even provide a head start in identifying and containing such a threat.

List of References

Adak, S. & Datta, A.K., 2005. *Leishmania major* encodes an unusual peroxidase that is a close homologue of plant ascorbate peroxidase: a novel role of the transmembrane domain. *The Biochemical Journal*, 390, pp.465–474.

Akarid, K. et al., 2004. *Leishmania major*-mediated prevention of programmed cell death induction in infected macrophages is associated with the repression of mitochondrial release of cytochrome c. *Journal of Leukocyte Biology*, 76(July), pp.95–103.

Akhoundi, M. et al., 2016. A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Neglected Tropical Diseases*, 10(3), pp.1–40.

Akopyants, N.S. et al., 2009. Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science*, 324(5924), pp.265-268.

Al-Mohammed, H., 2005. Production and characterization of stable amphotericin-resistant amastigotes and promastigotes of *Leishmania mexicana*. *Antimicrobial Agents and Chemotherapy*, 49(8), pp.3274-3280.

Al-Salem, W.S. et al., 2016. Cutaneous leishmaniasis and conflict in Syria. *Emerging Infectious Diseases*, 22(5), pp.931–933.

Alvar, J. et al., 2012. Leishmaniasis worldwide and global estimates of its incidence. *PloS ONE*, 7(5), e35671.

Alves, L.R. et al., 2013. MRNA localization mechanisms in *Trypanosoma cruzi*. *PLoS ONE*, 8(12), e81375.

Amirrajab, N. et al., 2016. *In vitro* activities of six antifungal drugs against *Candida glabrata* isolates : an emerging pathogen. *Jundishapur Journal of Microbiology*, 9(5), e36638.

Anders, S., Pyl, P.T. & Huber, W., 2015. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), pp.166–169.

Andrade-Neto, V.V. et al., 2016. Imipramine alters the sterol profile in *Leishmania amazonensis* and increases its sensitivity to miconazole. *Parasites & Vectors*, 9(1), 183.

Andrade-Neto, V.V. et al., 2011. The pharmacological inhibition of sterol biosynthesis in *Leishmania* is counteracted by enhancement of LDL endocytosis. *Acta tropica*, 119(2-3), pp.194–198.

Andreoli, T.E., 1974. The structure and function of amphotericin B-cholesterol pores in lipid bilayer membranes. *Annals of the New York Academy of Sciences*, 235, pp.448–468.

Ariyanayagam, M.R. & Fairlamb, A.H., 2001. Ovothiol and trypanothione as antioxidants in trypanosomatids. *Molecular and Biochemical Parasitology*, 115, pp.189–198.

Arthington-Skaggs, B.A. et al., 1999. Quantitation of ergosterol content: novel method for determination of fluconazole susceptibility of *Candida albicans*. *Journal of Clinical Microbiology*, 37(10), pp.3332–3337.

Azizi, H., Müller-McNicoll, M. & Papadopoulou, B., 2017. SIDER2 retroposon-mediated mRNA decay in *Leishmania* is coupled to translation. *International Journal for Parasitology*, 47(6), pp.305-310.

Bacchi, C.J. et al., 1994. Combination chemotherapy of drug-resistant *Trypanosoma brucei rhodesiense* infections in mice using DL-α-difluoromethylornithine and standard trypanocides. *Antimicrobial Agents and Chemotherapy*, 38(3), pp.563–569.

Baiocco, P. et al., 2009. Molecular basis of antimony treatment in leishmaniasis. *Journal of Medicinal Chemistry*, 52(8), pp.2603–2612.

Bari, V.K. et al., 2015. Plasma membrane proteolipid 3 protein modulates amphotericin B resistance through sphingolipid biosynthetic pathway. *Scientific Reports*, 5, p.9685.

Basselin, M. et al., 2002. Resistance to pentamidine in *Leishmania mexicana* involves exclusion of the drug from the mitochondrion. *Antimicrobial Agents and Chemotherapy*, 46(12), pp.3731–3738.

Basselin, M., Badet-Denisot, M.A. & Robert-Gero, M., 1998. Modification of kinetoplast DNA minicircle composition in pentamidine-resistant *Leishmania*. *Acta Tropica*, 70(1), pp.43–61.

Basselin, M., Coombs, G.H. & Barrett, M.P., 2000. Putrescine and spermidine transport in *Leishmania*. *Molecular and Biochemical Parasitology*, 109(1), pp.37–46.

Basselin, M. & Robert-Gero, M., 1998. Alterations in membrane fluidity, lipid metabolism, mitochondrial activity, and lipophosphoglycan expression in pentamidine-resistant *Leishmania*. *Parasitology Research*, 84(1), pp.78–83.

Basu, J. et al., 2006. Gluconate induces generation of reactive oxygen species and nitric oxide via phosphoinositide 3-kinase and mitogen-activated protein kinase activation in *Leishmania*. *Antimicrobial agents and chemotherapy*, 50(5), pp.1788–1797.

Bates, P.A., 2007. Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. *International Journal for Parasitology*, 37(10), pp.1097–1106.

Belenky, P., Camacho, D. & Collins, J.J., 2013. Fungicidal drugs induce a common oxidative-damage cellular death pathway. *Cell Reports*, 3(2), pp.350–358.

Bellocchio, S. et al., 2005. Liposomal amphotericin B activates antifungal resistance with reduced toxicity by diverting Toll-like receptor signalling from TLR-2 to TLR-4. *Journal of Antimicrobial Chemotherapy*, 55(2), pp.214–222.

Beneke, T. et al., 2017. A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *Royal Society Open Science*, 4(5), p.170095.

Berg, M. et al., 2015. Experimental resistance to drug combinations in *Leishmania donovani*: Metabolic and phenotypic adaptations. *Antimicrobial Agents and Chemotherapy*, 59(4), pp.2242–2255.

Berg, M. et al., 2013. Metabolic adaptations of *Leishmania donovani* in relation to differentiation, drug resistance, and drug pressure. *Molecular Microbiology*, 90(2), pp.428–442.

Bhandari, V. et al., 2014. Elucidation of cellular mechanisms involved in experimental paromomycin resistance in *Leishmania donovani*. *Antimicrobial Agents and Chemotherapy*, 58(5), pp.2580–2585.

Bojsen, R. et al., 2016. A common mechanism involving the TORC1 pathway can lead to amphotericin B-persistence in biofilm and planktonic *Saccharomyces cerevisiae* populations. *Scientific Reports*, 6, p.21874.

Bolard, J., Seigneuret, M. & Boudet, G., 1980. Interaction between phospholipid bilayer membranes and the polyene antibiotic amphotericin B. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 599(1), pp.280–293.

Bouazizi, H. et al., 2017. Changes in lipid and fatty acid composition during intramacrophagic transformation of *Leishmania donovani* complex promastigotes into amastigotes. *Lipids*, 52(5), pp.433-441.

Boucher, N. et al., 2002. A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3′-untranslated region element. *Journal of Biological Chemistry*, 277(22), pp.19511–19520.

Bringaud, F. et al., 2007. Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. *PLoS Pathogens*, 3(9), pp.1291–1307.

Britto, C. et al., 1998. Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene*, 222, pp.107–117.

Brotherton, M.-C. et al., 2013. Proteomic and genomic analyses of antimony resistant *Leishmania infantum* mutant. *PloS ONE*, 8(11), p.e81899.

Brotherton, M.-C. et al., 2014. Quantitative proteomic analysis of amphotericin B resistance in *Leishmania infantum*. *International journal for parasitology. Drugs and Drug Resistance*, 4(2), pp.126–132.

Canuto, G.A.B. et al., 2014. Multi-analytical platform metabolomic approach to study miltefosine mechanism of action and resistance in *Leishmania*. *Analytical and Bioanalytical Chemistry*, 406(14), pp.3459–3476.

Carrero-Lérida, J. et al., 2009. Intracellular location of the early steps of the isoprenoid biosynthetic pathway in the trypanosomatids *Leishmania major* and *Trypanosoma brucei*. *International Journal for Parasitology*, 39(3), pp.307–314.

Carter, K.C. et al., 2005. Sodium stibogluconate resistance in *Leishmania donovani* correlates with greater tolerance to macrophage antileishmanial responses and trivalent antimony therapy. *Parasitology*, 131(6), pp.747–757.

Carter, N.S., Berger, B.J. & Fairlamb, A.H., 1995. Uptake of diamidine drugs by the P2 nucleoside transporter in melarsen-sensitive and -resistant *Trypanosoma brucei brucei*. *Journal of Biological Chemistry*, 270(47), pp.28153–28157.

Castro, H. & Tomás, A.M., 2013. Thiol peroxidases of trypanosomatids. *Antioxidants & Redox Signaling*, 10(9), pp. 1593–1606.

Chau, A.S. et al., 2005. Inactivation of sterol $\Delta^{5,6}$-desaturase attenuates virulence in *Candida albicans*. *Antimicrobial Agents and Chemotherapy*, 49(9), pp.3646–3651.

Cingolani, P. et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), pp.80–92.

Clayton, C. & Shapira, M., 2007. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Molecular and Biochemical Parasitology*, 156(2), pp.93–101.

Coelho, A.C. et al., 2012. Multiple mutations in heterogeneous miltefosine-resistant *Leishmania major* population as determined by whole genome sequencing. *PLoS Neglected Tropical Diseases*, 6(2), p.e1512.

Coler, R.N. et al., 2015. From mouse to man: safety, immunogenicity and efficacy of a candidate leishmaniasis vaccine LEISH-F3+GLA-SE. *Clinical & Translational Immunology*, 4(4), p.e35.

Colotti, G. & Ilari, A., 2011. Polyamine metabolism in *Leishmania*: From arginine to trypanothione. *Amino Acids*, 40(2), pp.269–285.

Cotton, J.A., 2017. The expanding world of human leishmaniasis. *Trends in Parasitology*, 33(5), pp.341–344.

Coulson, R.M.R. et al., 1996. Differential expression of *Leishmania major* ß-tubulin genes during the acquisition of promastigote infectivity. *Molecular and Biochemical Parasitology*, 82(2), pp.227–236.

Creek, D.J. et al., 2012. Metabolomic analysis of trypanosomatid protozoa. *Molecular and Biochemical Parasitology*, 181(2), pp.73–84.

Creek D.J. et al. 2012b. IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. *Bioinformatics*, 28(7), 1049-1049.

Cubbon, 2009. Metabolomic applications of HILIC-LC-MS. *Mass Spectrometry Reviews*, 47(12), pp.987–992.

Das, M. et al., 2013. Miltefosine-unresponsive *Leishmania donovani* has a greater ability than miltefosine-responsive *L. donovani* to resist reactive oxygen species. *The FEBS Journal*, 280(19), pp.4807–4815.

Das, S. et al., 2014. Asymptomatic *Leishmania* infections in northern India: A threat for the elimination programme? *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 108(11), pp.679–684.

Deep, D.K. et al., 2017. Increased miltefosine tolerance in clinical isolates of *Leishmania donovani* is associated with reduced drug accumulation, increased infectivity and resistance to oxidative stress. *PLoS Neglected Tropical Diseases*, 11(6), e0005641.

Delarze, E. & Sanglard, D., 2015. Defining the frontiers between antifungal resistance, tolerance and the concept of persistence. *Drug Resistance Updates*, 23, pp.12–19.

Diaz-Gonzalez, R. et al., 2011. The susceptibility of trypanosomatid pathogens to PI3/mTOR kinase inhibitors affords a new opportunity for drug repurposing. *PLoS Neglected Tropical Diseases*, 5(8), e1297.

Van Dijk, E.L., Jaszczyszyn, Y. & Thermes, C., 2014. Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research*, 322(1), pp.12–20.

Dillon, L.A.L. et al., 2015. Simultaneous transcriptional profiling of *Leishmania major* and its murine macrophage host cell reveals insights into host-pathogen interactions. *BMC Genomics*, 16(1), 1108.

Dillon, L.A.L. et al., 2015b. Transcriptomic profiling of gene expression and RNA processing during *Leishmania major* differentiation. *Nucleic Acids Research*, 43(14), pp.6799-6813.

Dinesh, N. et al., 2014. Exploring *Leishmania donovani* 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) as a potential drug target by biochemical, biophysical and inhibition studies. *Microbial pathogenesis*, 66, pp.14–23.

Downing, T. et al., 2011. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome research*, 21(12), pp.2143–56.

Doyle, M.A. et al., 2009. LeishCyc: a biochemical pathways database for *Leishmania major*. *BMC systems biology*, 3, 57.

Dufernez, F. et al., 2006. The presence of four iron-containing superoxide dismutase isozymes in Trypanosomatidae: Characterization, subcellular localization, and

phylogenetic origin in *Trypanosoma brucei*. *Free Radical Biology and Medicine*, 40(2), pp.210–225.

Dupont, S. et al., 2012. Ergosterol biosynthesis: a fungal pathway for life on land? *Evolution*, 66(9), pp.2961–2968.

Eddouzi, J. et al., 2013. Molecular mechanisms of drug resistance in clinical *Candida* species Isolated from Tunisian hospitals. *Antimicrobial Agents and Chemotherapy*, 57(7), pp.3182–3193.

El-Sayed, N.M. et al., 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science*, 309(5733), pp.404–409.

Emami, S., Tavangar, P. & Keighobadi, M., 2017. An overview of azoles targeting sterol 14α-demethylase for antileishmanial therapy. *European Journal of Medicinal Chemistry*, 135, pp.241–259.

Fadda, A. et al., 2014. Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Molecular Microbiology*, 94(2), pp.307–326.

Fernandes, M.C. et al., 2016. Dual transcriptome profiling of *Leishmania*-infected human macrophages reveals distinct reprogramming signatures. *mBio*, 7(3), e00027-16.

Fernández, M.M., Malchiodi, E.L. & Algranati, I.D., 2011. Differential effects of paromomycin on ribosomes of *Leishmania mexicana* and mammalian cells. *Antimicrobial Agents and Chemotherapy*, 55(1), pp.86–93.

Fernandez-Prada, C. et al., 2016. Different mutations in a P-type ATPase transporter in *Leishmania* parasites are associated with cross-resistance to two leading drugs by distinct mechanisms. *PLoS Neglected Tropical Diseases*, 10(12), e0005171.

Fiebig, M., Kelly, S. & Gluenz, E., 2015. Comparative life cycle transcriptomics revises *Leishmania mexicana* genome annotation and links a chromosome duplication with parasitism of vertebrates. *PLOS Pathogens*, 11(10), e1005186.

Forastiero, A. et al., 2013. *Candida tropicalis* antifungal cross-resistance is related to different azole target (Erg11p) modifications. *Antimicrobial Agents and Chemotherapy*, 57(10), pp.4769–4781.

García-Hernández, R. et al., 2012. *Leishmania donovani* develops resistance to drug combinations. *PLoS Neglected Tropical Diseases*, 6(12), e1974.

Garrison, E. & Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Genest, P.A. et al., 2015. Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Research*, 43(4), pp.2102–2115.

Ghosh, A. K. et al., 2015. Metabolic reconfiguration of the central glucose metabolism: A crucial strategy of *Leishmania donovani* for its survival during oxidative stress. *The FASEB Journal*, 29(5), 2081–2098.

Ghosh, M., Roy, K. & Roy, S., 2013. Immunomodulatory effects of antileishmanial drugs. *Journal of Antimicrobial Chemotherapy*, 68(12), pp.2834–2838.

Gilead Sciences, 2016. Gilead Sciences and the World Health Organization Announce Five-Year Visceral Leishmaniasis Collaboration. Press release, September 27, 2016. Available at http://www.gilead.com/news/press-releases/2016/9/gilead-sciences-and-the-world-health-organization-announce-fiveyear-visceral-leishmaniasis-collaboration, accessed 25/10/2017.

Di Giorgio, C. et al., 1999. Flow cytometric assessment of amphotericin B susceptibility in *Leishmania infantum* isolates from patients with visceral leishmaniasis. *Journal of Antimicrobial Chemotherapy*, 44(1), pp.71–76.

Gourbal, B. et al., 2004. Drug uptake and modulation of drug resistance in *Leishmania* by an aquaglyceroporin. *Journal of Biological Chemistry*, 279(30), pp.31010–31017.

Grabherr, M.G. et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), pp.644–652.

Gray, K.C. et al., 2012. Amphotericin primarily kills yeast by simply binding ergosterol. *Proceedings of the National Academy of Sciences of the United States of America*, 109(7), pp.2234–9.

Guha, R. et al., 2014. Antimony resistant *Leishmania donovani* but not sensitive ones drives greater frequency of potent T-regulatory cells upon interaction with human PBMCs: role of IL-10 and TGF-β in early immune response. *PLoS Neglected Tropical Diseases*, 8(7), e2995.

Hansen, C. et al., 2011. Reduction of Sb(V) in a human macrophage cell line measured by HPLC-ICP-MS. *Biological Trace Element Research*, 144(1-3), pp.234–243.

Haubrich, B.A. et al., 2015. Discovery of an ergosterol-signaling factor that regulates *Trypanosoma brucei* growth. *Journal of Lipid Research*, 56(2), pp.331–341.

Haughan, P.A., Chance, M.L. & Goad, L.J., 1995. Effects of an azasterol inhibitor of sterol 24-transmethylation on sterol biosynthesis and growth of *Leishmania donovani* promastigotes. *The Biochemical Journal*, 308, pp.31–38.

Hendrickx, S. et al., 2016. Evidence of a drug-specific impact of experimentally selected paromomycin and miltefosine resistance on parasite fitness in *Leishmania infantum*. *Journal of Antimicrobial Chemotherapy*, 71(7), pp.1914-1921.

Hendrickx, S. et al., 2015. In vivo selection of paromomycin and miltefosine resistance in *Leishmania donovani* and *L. infantum* in the Syrian hamster model. *Antimicrobial Agents and Chemotherapy*, 59(8), 4714-4718.

Hsuchen, C.C. & Feingold, D.S., 1973. Selective membrane toxicity of the polyene antibiotics: studies on lecithin membrane models (liposomes). *Antimicrobial Agents and Chemotherapy*, 4(3), pp.309–315.

Iantorno, S. et al., 2017. Gene Expression in *Leishmania* is Regulated Predominantly by Gene Dosage. *mBio*, 8(5), e01393-17.

Robinson, J.T. et al., 2011. Integrative Genomics Viewer. *Nature Biotechnology*, 29, 24-26.

Imai, K. et al., 2017. Non-invasive diagnosis of cutaneous leishmaniasis by the direct boil loop-mediated isothermal amplification method and MinION™ nanopore sequencing. *Parasitology International*, 67(1), 34-37.

Imamura, H. et al., 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *eLife*, 5, e12613.

Imbert, L. et al., 2012. Identification of phospholipid species affected by miltefosine action in *Leishmania donovani* cultures using LC-ELSD, LC-ESI/MS, and multivariate data analysis. *Analytical and Bioanalytical Chemistry*, 402(3), pp.1169–1182.

Jara, M. et al., 2017. Macromolecular biosynthetic parameters and metabolic profile in different life cycle stages of *Leishmania braziliensis*: amastigotes as a functionally less active stage. *PLoS ONE*, 12(7), e0180532.

Jaramillo, M. et al., 2011. *Leishmania* repression of host translation through mTOR cleavage is required for parasite survival and infection. *Cell Host and Microbe*, 9(4), pp.331–341.

Jeddi, F. et al., 2014. Heterogeneity of molecular resistance patterns in antimony-resistant field isolates of *Leishmania* species from the western Mediterranean area. *Antimicrobial Agents and Chemotherapy*, 58(8), pp.4866–4874.

Jensen, B.C. et al., 2014. Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics*, 15(1), 911.

Jhingran, A. et al., 2009. Paromomycin: Uptake and resistance in *Leishmania donovani*. *Molecular and Biochemical Parasitology*, 164(2), pp.111–117.

Jiménez-Jiménez, C. et al., 2008. Δ24 (25)-sterol methenyltransferase: Intracellular localization and azasterol sensitivity in *Leishmania major* promastigotes overexpressing the enzyme. *Molecular and Biochemical Parasitology*, 160(1), pp.52–59.

Kalb, V.F. et al., 1987. Primary structure of the P450 lanosterol demethylase gene from *Saccharomyces cerevisiae*. *DNA*, 6(6), pp.529–537.

Kamina, A.D. & Williams, N., 2017. Ribosome assembly in trypanosomatids: a novel therapeutic target. *Trends in Parasitology*, 33(4), pp.256–257.

Kaul, S., Sharma, S.S. & Mehta, I.K., 2008. Free radical scavenging potential of L-proline: evidence from in vitro assays. *Amino Acids*, 34(2), pp.315–320.

Kaye, P. & Scott, P., 2011. Leishmaniasis: complexity at the host-pathogen interface. *Nature Reviews. Microbiology*, 9(8), pp.604–615.

Kim, D. et al., 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, 11(9), pp.1650–1667.

Koek, M.M. et al., 2011. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics*, 7(3), pp.307–328.

De Koning, H.P., 2001. Uptake of pentamidine in *Trypanosoma brucei brucei* is mediated by three distinct transporters: implications for cross-resistance with arsenicals. *Molecular Pharmacology*, 59(3), pp.586–592.

Kovářová, J. & Barrett, M.P., 2016. The pentose phosphate pathway in parasitic trypanosomatids. *Trends in Parasitology*, 32(8), 622-634.

Kuhls, K. et al., 2011. Comparative microsatellite typing of new world *Leishmania infantum* reveals low heterogeneity among populations and its recent old world origin. *PLoS Neglected Tropical Diseases*, 5(6), e1155.

Kumar, A. et al., 2014. Ascorbate peroxidase, a key molecule regulating amphotericin B resistance in clinical isolates of *Leishmania donovani*. *Antimicrobial Agents and Chemotherapy*, 58(10), pp.6172–6184.

Kumar, D. et al., 2009. In vitro susceptibility of field isolates of *Leishmania donovani* to miltefosine and amphotericin B: correlation with sodium antimony gluconate susceptibility and implications for treatment in areas of endemicity. *Antimicrobial Agents and Chemotherapy*, 53(2), pp.835–838.

Lachaud, L. et al., 2009. Parasite susceptibility to amphotericin B in failures of treatment for visceral leishmaniasis in patients coinfected with HIV type 1 and *Leishmania infantum*. *Clinical Infectious Diseases*, 48(2), e16.

Laffitte, M.-C.N. et al., 2016. Chromosomal translocations in the parasite *Leishmania* by a MRE11/RAD50-independent microhomology-mediated end joining mechanism. *PLoS Genetics*, 12(6), e1006117.

Laffitte, M.-C.N. et al., 2016. Deep-sequencing revealing mutation dynamics in the miltefosine transporter gene in *Leishmania infantum* selected for miltefosine resistance. *Parasitology Research*. 115(10), 3699-3703.

Laskay, T., Van Zandbergen, G. & Solbach, W., 2003. Neutrophil granulocytes - Trojan horses for *Leishmania major* and other intracellular microbes? *Trends in Microbiology*, 11(5), pp.210–214.

Layer, R.M. et al., 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology*, 15(6), p.R84.

Leprohon, P. et al., 2009. Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant *Leishmania infantum*. *Nucleic Acids Research*, 37(5), pp.1387–99.

Li, H. et al., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), pp.1754–1760.

Lira, R. et al., 1999. Evidence that the high incidence of treatment failures in Indian kala-azar is due to the emergence of antimony-resistant strains of *Leishmania donovani*. *The Journal of Infectious Diseases*, 180(2), pp.564–567.

Lockhart, S.R. et al., 2017. Simultaneous emergence of multidrug-resistant *Candida auris* on 3 continents confirmed by whole genome sequencing and epidemiological analyses. *Clinical Infectious Diseases*, 64(2), 134-140.

Lorente, S.O. et al., 2004. Novel azasterols as potential agents for treatment of leishmaniasis and trypanosomiasis. *Antimicrobial Agents and Chemotherapy*, 48(8), pp.2937–2950.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), pp.1–34.

Lamy-Freund, M.T., Ferreira, V.F.N. & Schreier, S., 1984. Mechanism of Inactivation of the polyene antibiotic amphotericin B - evidence for radical formation in the process of autooxidation. *The Journal of Antibiotics*, 38(6), pp.753–757.

Madeira da Silva, L. et al., 2009. Regulated expression of the *Leishmania major* surface virulence factor lipophosphoglycan using conditionally destabilized fusion proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), pp.7583–7588.

Madeira da Silva, L. & Beverley, S.M., 2010. Expansion of the target of rapamycin (TOR) kinase family and function in *Leishmania* shows that TOR3 is required for acidocalcisome biogenesis and animal infectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26), pp.11965–11970.

Madoui M.-A. *et al.*, 2015. Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics*, 16, 327.

Marshall, J.A. & Nes, W.D., 1999. Isolation and characterization of an active-site peptide from a sterol methyl transferase with a mechanism-based inhibitor. *Bioorganic & Medicinal Chemistry Letters*, 9, 1533-1536.

Martel, C.M. et al., 2010. Identification and characterization of four azole-resistant erg3 mutants of *Candida albicans*. *Antimicrobial Agents and Chemotherapy*, 54(11), pp.4527–4533.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), pp.10-12.

Mathur, R. et al., 2015. Elevated ergosterol protects *Leishmania* parasites against antimony-generated stress. *The FASEB Journal*, 29(10), 4201-4213.

Matrangolo, F.S. V et al., 2013. Comparative proteomic analysis of antimony-resistant and-susceptible *Leishmania braziliensis* and *Leishmania infantum chagasi* lines. *Molecular and Biochemical Parasitology*, 190(2), pp.63–75.

Maugeri, D.A. et al., 2003. Pentose phosphate metabolism in *Leishmania mexicana*. *Molecular and Biochemical Parasitology*, 130(2), pp.117–125.

Mbongo, N. et al., 1998. Mechanism of amphotericin B resistance in *Leishmania donovani* promastigotes. *Antimicrobial Agents and Chemotherapy*, 42(2), 352-357.

McCall, L.-I. et al., 2015. Targeting ergosterol biosynthesis in *Leishmania donovani*: essentiality of sterol 14alpha-demethylase. *PLOS Neglected Tropical Diseases*, 9, p.e0003588.

Mccarthy, A., 2010. Third generation DNA sequencing: Pacific Biosciences' single molecule real time technology. *Chemistry and Biology*, 17(7), pp.675–676.

McConville, M.J. & Naderer, T., 2011. Metabolic pathways required for the intracellular survival of *Leishmania*. *Annual review of microbiology*, 65, pp.543–561.

Mesa-Arango, A.C. et al., 2014. The production of reactive oxygen species Is a universal action mechanism of amphotericin B against pathogenic yeasts and contributes to the fungicidal effect of this drug. *Antimicrobial Agents and Chemotherapy*, 58(11), pp.6627–38.

Mo, C. & Bard, M., 2005. A systematic study of yeast sterol biosynthetic protein-protein interactions using the split-ubiquitin system. *Biochimica et Biophysica Acta*, 1737, pp.152–160.

Molina, I. et al., 2014. Randomized trial of posaconazole and benznidazole for chronic Chagas' disease. *The New England Journal of Medicine*, 370, pp.1899–1908.

Moon, S. et al., 2011. A win-win solution?: A critical analysis of tiered pricing to improve access to medicines in developing countries. *Globalization and Health*, 7, p.39.

Moraes, C.B. et al., 2014. Nitroheterocyclic compounds are more efficacious than CYP51 inhibitors against *Trypanosoma cruzi*: implications for Chagas disease drug discovery and development. *Scientific Reports*, 4, p.4703.

Moreira, W., Leprohon, P. & Ouellette, M., 2011. Tolerance to drug-induced cell death favours the acquisition of multidrug resistance in *Leishmania*. *Cell Death & Disease*, 2(9), e201.

Morio, F. et al., 2012. Amino acid substitutions in the *Candida albicans* sterol δ5,6-desaturase (Erg3p) confer azole resistance: Characterization of two novel mutants with impaired virulence. *Journal of Antimicrobial Chemotherapy*, 67(9), pp.2131–2138.

Mukherjee, A. et al., 2009. The gamma-glutamylcysteine synthetase gene of *Leishmania* is essential and involved in response to oxidants. *Molecular Microbiology*, 74(4), pp.914–27.

Mukherjee, B. et al., 2013. Antimony-resistant but not antimony-sensitive *Leishmania donovani* up-regulates host IL-10 to overexpress multidrug-resistant protein 1. *Proceedings of the National Academy of Sciences of the United States of America*, 110(7), pp.E575–E582.

Mukherjee, M., Basu Ball, W. & Das, P.K., 2014. *Leishmania donovani* activates SREBP2 to modulate macrophage membrane cholesterol and mitochondrial oxidants for establishment of infection. *International Journal of Biochemistry and Cell Biology*, 55, pp.196–208.

Mukhopadhyay, R. et al., 2011. Characterisation of antimony-resistant *Leishmania donovani* isolates: biochemical and biophysical studies and interaction with host cells. *International Journal for Parasitology*, 41(13-14), pp.1311–1321.

Munday, J.C. et al., 2014. *Trypanosoma brucei* aquaglyceroporin 2 is a high-affinity transporter for pentamidine and melaminophenyl arsenic drugs and the main genetic determinant of resistance to these drugs. *Journal of Antimicrobial Chemotherapy*, 69(3), pp.651–663.

Mwenechanya, R. et al., 2017. Sterol 14α-demethylase mutation leads to amphotericin B resistance in *Leishmania mexicana*. *PLOS Neglected Tropical Diseases*, 11(6), p.e0005649.

Myler, P. J. et al., 1999. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proceedings of the National Academy of Sciences of the United States of America*, 96, pp.2902–2906.

Nakagawa, Y. et al., 2014. Effect of sterol side chain on ion channel formation by amphotericin B in lipid bilayers. *Biochemistry*. 53(19), pp.3088-3094

Neumann, A. et al., 2016. Membrane sterols modulate the binding mode of amphotericin B without affecting its affinity for a lipid bilayer. *Langmuir*, 32 (14), pp.3452–3461.

Obonaga, R. et al., 2014. Treatment failure and miltefosine susceptibility in dermal leishmaniasis caused by *Leishmania* subgenus *Viannia* species. *Antimicrobial Agents and Chemotherapy*, 58(1), pp.144–152.

Ogata, H. et al., 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1), pp.29–34.

Oliver, S. et al., 1998. Systematic functional analysis of the yeast genome. *Trends in Biotechnology*, 16(9), pp.373–378.

Oliveros, J.C., 2007-2015. Venny, An interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html, accessed 27/02/2018.

Ortiz-Gómez, A. et al., 2006. Farnesyl diphosphate synthase is a cytosolic enzyme in *Leishmania major* promastigotes and its overexpression confers resistance to risedronate. *Eukaryotic Cell*, 5(7), pp.1057–1064.

Paila, Y.D., Saha, B. & Chattopadhyay, A., 2010. Amphotericin B inhibits entry of *Leishmania donovani* into primary macrophages. *Biochemical and Biophysical Research Communications*, 399(3), pp.429–433.

Paniz Mondolfi, A.E. et al., 2011. Successful treatment of old world cutaneous leishmaniasis caused by *Leishmania infantum* with posaconazole. *Antimicrobial Agents and Chemotherapy*, 55(4), pp.1774–1776.

Parkhomchuk, D. et al., 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research*, 37(18), e123.

Pérez-Victoria, F.J. et al., 2006. Phospholipid translocation and miltefosine potency require both *L. donovani* miltefosine transporter and the new protein LdRos3 in *Leishmania* parasites. *Journal of Biological Chemistry*, 281(33), pp.23766–23775.

Perry, M.R. et al., 2013. Chronic exposure to arsenic in drinking water can lead to resistance to antimonial drugs in a mouse model of visceral leishmaniasis. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49), pp.19932–19937.

Pinto, W.J. & Nes, W.R., 1983. Stereochemical specificity for sterols in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 258(7), pp.4472–4476.

Popp, C. et al., 2017. Competitive fitness of fluconazole-resistant clinical *Candida albicans* strains. *Antimicrobial Agents and Chemotherapy*, 61(7), e00584-17.

Pourshafie, M. et al., 2004. Cloning of S-Adenosyl-l-Methionine : C-24-Δ-sterol-methyltransferase (ERG6) from *Leishmania donovani* and characterization of mRNAs in wild-type and amphotericin B-resistant promastigotes. *Antimicrobial Agents and Chemotherapy*, 48(7), 2409-2414.

Prajapati, V.K. et al., 2012. *In vitro* antileishmanial drug susceptibility of clinical isolates from patients with Indian visceral leishmaniasis - status of newly introduced drugs. *American Journal of Tropical Medicine and Hygiene*, 87(4), pp.655–657.

Prieto Barja. P. et al., 2017. Haplotype selection as an adaptive mechanism in the protozoan pathogen *Leishmania donovani*. *Nature Ecology & Evolution*, 8(3), e00599-17.

Purkait, B. et al., 2012. Mechanism of amphotericin B resistance in clinical isolates of *Leishmania donovani*. *Antimicrobial Agents and Chemotherapy*, 56(2), pp.1031–1041.

Purkait, B. et al., 2015. Up-regulation of silent information regulator 2 (Sir2) is associated with amphotericin B resistance in clinical isolates of *Leishmania donovani*. *Journal of Antimicrobial Chemotherapy*, 70(5), 1343-1356.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/, accessed 27/02/2018.

Ralton, J.E. et al., 2003. Evidence that intracellular ß1-2 mannan is a virulence factor in *Leishmania* parasites. *Journal of Biological Chemistry*, 278(42), pp.40757-40763.

Ramos, H. et al., 1996. Amphotericin B kills unicellular leishmanias by forming aqueous pores permeable to small cations and anions. *Journal of Membrane Biology*, 152(1), pp.65–75.

Räz, B. et al., 1997. The Alamar Blue assay to determine drug sensitivity of African trypanosomes (*T.b. rhodesiense* and *T.b. gambiense*) *in vitro*. *Acta Tropica*, 68(2), pp.139–47.

Razonable, R.R. et al., 2005. Secretion of proinflammatory cytokines and chemokines during amphotericin B exposure is mediated by coactivation of Toll-like receptors 1 and 2. *Antimicrobial Agents and Chemotherapy*, 49(4), pp.1617–1621.

Reinert, K. et al., 2015. Alignment of next-generation sequencing reads. *Annual Review of Genomics and Human Genetics*, 16, pp.133–151.

Rijal, S. et al., 2013. Increasing failure of miltefosine in the treatment of Kala-azar in Nepal and the potential role of parasite drug resistance, reinfection, or noncompliance. *Clinical Infectious Disease*, 56(11), pp.1530–1538.

Rimmer, A. et al., 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics*, 46(8), 912-918.

Ritz, C. & Streibig, J.C., 2005. Bioassay analysis using R. *Journal of Statistical Software*, 12(5).

Roberts, C.W. et al., 2003. Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. *Molecular and Biochemical Parasitology*, 126(2), pp.129–142.

Rogers, M.B. et al., 2011. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Research*, 21(12), pp.2129–2142.

Rybak, J.M. et al., 2017. Loss of C-5 sterol desaturase activity results in increased resistance to azole and echinocandin antifungals in a clinical isolate of *Candida parapsilosis*. *Antimicrobial Agents and Chemotherapy*, 61(9), e00651-17.

Saint-Pierre-Chazalet, M. et al., 2009. Membrane sterol depletion impairs miltefosine action in wild-type and miltefosine-resistant *Leishmania donovani* promastigotes. *Journal of Antimicrobial Chemotherapy*, 64(5), pp.993–1001.

Salek, R.M. et al., 2013. The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience*, 2, p.13.

Sánchez-Cañete, M.P. et al., 2009. Low plasma membrane expression of the miltefosine transport complex renders *Leishmania braziliensis* refractory to the drug. *Antimicrobial Agents and Chemotherapy*, 53(4), pp.1305–1313.

Dos Santos Ferreira, C. et al., 2003. Thiol-induced reduction of antimony(V) into antimony(III): A comparative study with trypanothione, cysteinyl-glycine, cysteine and glutathione. *BioMetals*, 16(3), pp.441–446.

Sau, K. et al., 2003. The antifungal drug amphotericin B promotes inflammatory cytokine release by a toll-like receptor- and CD14-dependent mechanism. *Journal of Biological Chemistry*, 278(39), pp.37561–37568.

Saunders, E.C. et al., 2011. Isotopomer profiling of *Leishmania mexicana* promastigotes reveals important roles for succinate fermentation and aspartate uptake in Tricarboxylic Acid Cycle (TCA) anaplerosis, glutamate synthesis, and growth. *Journal of Biological Chemistry*, 286(31), pp.27706–27717.

Scheltema, R.A. et al., 2011. PeakML/mzMatch: A file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Analytical Chemistry*, 83(7), pp.2786–2793.

Seifert, K. et al., 2007. Inactivation of the miltefosine transporter, LdMT, causes miltefosine resistance that is conferred to the amastigote stage of *Leishmania donovani* and persists *in vivo*. *International Journal of Antimicrobial Agents*, 30(3), pp.229–235.

Semini, G. et al., 2017. Changes to cholesterol trafficking in macrophages by *Leishmania* parasites infection. *MicrobiologyOpen*, 6, e00469.

Shaked-Mishant, P. et al., 2001. Novel intracellular Sb^{V} reducing activity correlates with antimony susceptibility in *Leishmania donovani*. *Journal of Biological Chemistry*, 276(6), pp.3971–3976.

Shaw, C.D. et al. 2016. *In vitro* selection of miltefosine resistance in promastigotes of *Leishmania donovani* from Nepal: genomic and metabolomic characterization. *Molecular Microbiology*, 99(6), 1134-48.

Shekhova, E., Kniemeyer, O. & Brakhage, A.A., 2017. Induction of mitochondrial ROS production by itraconazole, terbinafine and amphotericin B as a mode of action against *Aspergillus fumigatus*. *Antimicrobial Agents and Chemotherapy*, 61(11), e00978-17.

Sievers, F. et al., 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), pp.539–539.

Singh, a K., Papadopoulou, B. & Ouellette, M., 2001. Gene amplification in amphotericin B-resistant *Leishmania tarentolae*. *Experimental Parasitology*, 99(3), pp.141–147.

Singh, O.P. et al., 2016. Elimination of visceral leishmaniasis on the Indian subcontinent. *The Lancet Infectious Diseases*, 16(12), e304–e309.

Smircich, P. et al., 2015. Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. *BMC Genomics*, 16(1), 443.

Smith, M., Bringaud, F. & Papadopoulou, B., 2009. Organization and evolution of two SIDER retroposon subfamilies and their impact on the *Leishmania* genome. *BMC Genomics*, 10, 240.

Sokol-Anderson, M. & Sligh, J., 1988. Role of cell defense against oxidative damage in the resistance of *Candida albicans* to the killing effect of amphotericin B. *Antimicrobial Agents and Chemotherapy*, 32(5), pp.702-705.

Sokol-Anderson, M.L., Brajtburg, J. & Medoff, G., 1986. Amphotericin B-induced oxidative damage and killing of *Candida albicans*. *The Journal of Infectious Diseases*, 154(1), pp.76–83.

Song, J. et al., 2016. Pentamidine is not a permeant but a nanomolar inhibitor of the *Trypanosoma brucei* aquaglyceroporin-2. *PLoS Pathogens*, 12(2), e1005436.

de Souza, W. & Rodrigues, J.C.F., 2009. Sterol biosynthesis pathway as target for anti-trypanosomatid drugs. *Interdisciplinary Perspectives on Infectious Diseases*, 2009, 642502.

Srivastava, P. et al., 2011. Unusual case of resistance to amphotericin B in visceral leishmaniasis in a region in India where leishmaniasis is not endemic. *Journal of Clinical Microbiology*, 49(8), pp.3088–3091.

Srivastava, S. et al., 2017. Laboratory confirmed miltefosine resistant cases of visceral leishmaniasis from India. *Parasites & Vectors*, 10, 49.

Sterkers, Y. et al., 2012. Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. *Molecular Microbiology*, 86(1), pp.15–23.

Sundar, S. & Chakravarty, J., 2013. Leishmaniasis: an update of current pharmacotherapy. *Expert Opinion on Pharmacotherapy*, 14(1), pp.53–63.

Sundar, S. et al., 2000. Failure of pentavalent antimony in visceral leishmaniasis in India: report from the center of the Indian epidemic. *Clinical Infectious Diseases*, 31, pp. 1104-1107.

Sundar, Shyamgarwal, D., Rai, M. & Murray, H.W., 2010. Single-dose lipisomal amphotericin B for visceral leishmaniasis in India. *The New England Journal of Medicine*, 362(2), pp.504–512.

T'Kindt, R. et al., 2010. Towards an unbiased metabolic profiling of protozoan parasites: Optimisation of a *Leishmania* sampling protocol for HILIC-orbitrap analysis. *Analytical and Bioanalytical Chemistry*, 398(5), pp.2059–2069.

Tetaud, E. et al., 2002. A new expression vector for *Crithidia fasciculata* and *Leishmania*. *Molecular and Biochemical Parasitology*, 120(2), pp.195–204.

Thakur, C.P., Mitra, D.K. & Narayan, S., 2003. Skewing of cytokine profiles towards T helper cell type 2 response in visceral leishmaniasis patients unresponsive to sodium antimony gluconate. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 97(4), pp.409–412.

Thomas, S. et al., 2009. Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics*, 10(1), 152.

Todd, B.L. et al., 2006. Sterol regulatory element binding protein is a principal regulator of anaerobic gene expression in fission yeast. *Molecular and Cellular Biology*, 26(7), pp.2817–2831.

Trapnell, C. et al., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), pp.511–515.

Ubeda, J.-M. et al., 2014. Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite *Leishmania*. *PLoS Biology*, 12(5), e1001868.

Vacchina, P. et al., 2016. Multifactorial basis for *in vitro* acquisition of miltefosine resistance in *Leishmania donovani* : a genomic appraisal. *Antimicrobial Agents and Chemotherapy*, 60(7), pp.4089-4100.

Valera, A. et al., 1994. Expression of the neomycin-resistance (*neo*) gene induces alterations in gene expression and metabolism. *Human Gene Therapy*, 5, pp.449-456.

Vanaerschot, M. et al., 2011. Antimonial resistance in *Leishmania donovani* is associated with increased *in vivo* parasite burden. *PLoS ONE*, 6(8), e23120.

Vanaerschot, M. et al., 2013. Drug-resistant microorganisms with a higher fitness - can medicines boost pathogens? *Critical Reviews in Microbiology*, 39(4), pp.384-394.

Vandeputte, P. et al., 2007. Reduced susceptibility to polyenes associated with a missense mutation in the ERG6 gene in a clinical isolate of *Candida glabrata* with pseudohyphal growth. *Antimicrobial Agents and Chemotherapy*, 51(3), pp.982–990.

Vasquez, J.-J. et al., 2014. Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Research*, 42(6), pp.3623–3637.

Verma, S., Mehta, A. & Shaha, C., 2011. CYP5122A1, a novel cytochrome P450 is essential for survival of *Leishmania donovani*. *PloS ONE*, 6(9), e25273.

Vincent, B.M. et al., 2013. Fitness trade-offs restrict the evolution of resistance to amphotericin B. *PLoS Biology*, 11(10), e1001692.

Waterman, M.R. & Lepesheva, G.I., 2005. Sterol 14α-demethylase, an abundant and essential mixed-function oxidase. *Biochemical and Biophysical Research Communications*, 338(1), pp.418–422.

Wei, Y. et al., 2001. The hydroxyanilide fenhexamid, a new sterol biosynthesis inhibitor fungicide efficient against the plant pathogenic fungus *Botryotinia fuckeliana* (*Botrytis cinerea*). *Pest Management Science*, 57(11), pp.1060–1067.

Weingärtner, A. et al., 2010. Disruption of the lipid-transporting LdMT-LdRos3 complex in *Leishmania donovani* affects membrane lipid asymmetry but not host cell invasion. *PLoS ONE*, 5(8), e12443.

Weischenfeldt, J. & Porse, B., 2008. Bone marrow-derived macrophages (BMM): Isolation and applications. *Cold Spring Harbor Protocols*, 3(12).

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.

Wiseman, H., 1993. Vitamin D is a membrane antioxidant: ability to inhibit iron-dependent lipid peroxidation in liposomes compared to cholesterol, ergosterol and tamoxifen and relevance to anticancer action. *FEBS Letters*, 326, pp.285–288.

World Health Organization, 2015. Kala-Azar Elimination Programme: report of a WHO consultation of partners. Geneva, Switzerland, 10-11 February 2015. Available at http://apps.who.int/iris/handle/10665/185042, accessed 27/02/2018.

Wyllie, S., Cunningham, M.L. & Fairlamb, A.H., 2004. Dual action of antimonial drugs on thiol redox metabolism in the human pathogen *Leishmania donovani*. *Journal of Biological Chemistry*, 279(38), pp.39925–39932.

Xia, J. et al., 2015. MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Research*, 43(W1), pp.W251–W257.

Xu, W. et al., 2014. Sterol biosynthesis is required for heat resistance but not extracellular survival in *Leishmania*. *PLoS Pathogens*, 10(10), p.e1004427.

Xu, X. et al., 2001. Effect of the structure of natural sterols and sphingolipids on the formation of ordered sphingolipid/sterol domains (rafts). *Journal of Biological Chemistry*, 276(36), pp.33540–33546.

Yang, G., Choi, G. & No, J.H., 2016. Antileishmanial mechanism of diamidines involves targeting kinetoplasts. *Antimicrobial Agents and Chemotherapy*, 60(11), pp.6828–6836.

Yao, C. & Wilson, M.E., 2016. Dynamics of sterol synthesis during development of *Leishmania* spp. parasites to their virulent form. *Parasites & Vectors*, 9(1), 200.

Young, L.Y., Hull, C.M. & Heitman, J., 2003. Disruption of ergosterol biosynthesis confers resistance to amphotericin B in *Candida lusitaniae*. *Antimicrobial Agents and Chemotherapy*, 47(9), pp.2717–2724.

Younsi, M. et al., 2000. Amphotericin B resistance and membrane fluidity in *Kluyveromyces lactis* strains. *Antimicrobial Agents and Chemotherapy*, 44(7), pp.1911–1916.

Zhou, W. et al., 2006. Mechanistic analysis of a multiple product sterol methyltransferase implicated in ergosterol biosynthesis in *Trypanosoma brucei*. *Journal of Biological Chemistry*, 281(10), pp.6290–6296.

Zilberstein, D., Liveanu, V. & Gepstein, A., 1990. Tricyclic drugs reduce proton motive force in *Leishmania donovani* promastigotes. *Biochemical Pharmacology*, 39(5), pp.935–940.

Zilka, A. et al., 2001. Developmental regulation of heat shock protein 83 in *Leishmania*. *Journal of Biological Chemistry*, 276(51), pp.47922–47929.

Zulfiqar, B., Shelper, T.B. & Avery, V.M., 2017. Leishmaniasis drug discovery: recent progress and challenges in assay development. *Drug Discovery Today*, 22(10), pp.1516–1531.