CHAPTER 6

CONCLUSIONS

The aim of this thesis was to investigate genome plasticity and the effects of hybridization on *L. tropica* genome structure and function. I have demonstrated that like *L. infantum* and *L. major*, *L. tropica* is also capable of performing genetic exchange. In the Introduction, I have delineated three possible models of reproduction that this genetic exchange might follow: asexual, sexual, and parasexual. I hereby discuss each in light of my findings, and explore to what extent my objective of describing genetic exchange in this species has been achieved.

6.1. Population genetics in *L. tropica*

6.1.1. Heterozygosity and reproduction

As discussed in Chapter 2, a characteristic feature of the *L. tropica* species complex is the great genetic heterogeneity encountered. As I have presented in that chapter, both MLST and WGS data show that isolates covering the entire geographic distribution of the species often bear genome-wide differences in patterns of homozygosity and heterozygosity that can be attributed to differences in the frequency of intercrossing between individual "clones" or "clonal lineages". Previous studies that have analysed in detail microsatellite data from a much larger sample

set than that considered in this thesis (Schwenkenbecher, Wirth et al. 2006, Krayter, Bumb et al. 2014) have found three broad clusters that broadly speaking overlap geographically with those identified in the analyses performed in this thesis: one cluster for Israel/Palestine, one for Africa and the Galilee region, and one for Asia/India. The genetic distance between isolates within each of these clusters was often quite large in the studies mentioned above.

It is important to bear in mind that the concept of "clones" in *Leishmania* species comes with some caveats given the large variation in ploidy seen within individual culture-adapted isolates (Sterkers, Crobu et al. 2014). Changing genome structure appears to confer an added level of genetic variability that may serve an adaptive purpose in the downregulation or upregulation of specific genes. The functional effects of this plasticity in genome structure were presented in Chapter 3 and are summarized in Section 6.2.1.

From my analysis of observed allele frequencies with respect to Hardy-Weinberg equilibrium expectations, it became evident that individuals carrying higher heterozygosity than expected, and therefore with low inbreeding coefficients (F_{IT}), are widespread throughout Asia and the Indian subcontinent. On the other hand, the majority of individuals with lower heterozygosity than expected were typically found in Northern and Eastern Africa, and a restricted area of the Middle East region, mainly Israel and Palestine. Denser sampling from neighbouring countries could further elucidate patterns of genetic exchange in these populations, given the apparent overlap between different parasite clusters that could be associated with interbreeding.

Previous studies have suggested the possibility that during the evolutionary history of *L. tropica* a hybridization event associated with an Out-of-Africa origin for this species gave rise to the most common circulating strains found in Asia and India, which have therefore lower inbreeding coefficients. Based on the evidence I presented, I do not exclude this possibility, but warn against the practice of drawing conclusions from the results of inbreeding analyses that mainly simply establish that panmixia is not observed in this species, and can only hint at a directional skew towards inbreeding or outbreeding. The exact quantification of gene flow between population units will require the application of more advanced techniques such as coalescent-based approaches and a more detailed knowledge of the epidemiologically relevant demographic units in this pathogen species.

6.1.2. Wahlund effects and reproduction

As discussed in the Introduction, Wahlund effects have often been invoked to explain patterns of genetic variation in *Leishmania* species. Wahlund effects are defined as the apparent lack of heterozygosity seen in populations where there is population substructuring. The individual subpopulations themselves may be in Hardy-Weinberg equilibrium, and therefore follow panmictic assumptions, but sampling irrespective of these subpopulation boundaries will artificially inflate the expected heterozygosity due to different equilibrium allele frequencies being present in each subpopulation, thus increasing the difference between observed and expected heterozygosity.

Wahlund effects can therefore explain low heterozygosity even when there is significant interbreeding. Genotyping of the limited number of isolates considered in this study shows that homozygous markers are often quite diverse in terms of the number of different allelic sequence variants (Table 2.3), which would confirm the presence of significant substructuring. Given that the inbreeding coefficient calculation did not take into consideration allele sharing between individuals, a DAPC, which is informed by the number of shared alleles between individuals, was performed to gather information on genetic distances between isolates and the number of distinct clusters. Most of the isolates fell into three distinct clusters, which were validated by phylogenetic analysis of the concatenated sequence data.

One of the isolates, L810 from Northern Israel, proved problematic, with conflicting results between the phylogenetic analysis and DAPC. This isolate can be taken as a case in point to further present the unique epidemiological scenarios exemplified by a vector-borne disease such as *Leishmania*, and specifically a species such as *L. tropica* that has been shown to have both a urban/periurban anthroponotic cycle, and a rural zoonotic cycle, depending on the geographical region considered (see Section 1.1.3). This isolate was isolated from a *P. arabicus* sand fly in Northern Israel (Jacobson, Eisenberger et al. 2003, Schnur, Nasereddin et al. 2004), in a region were transmission of *L. tropica* was previously thought to be uniquely due to *P. sergenti*. These and following studies demonstrating its preference for a different vector species and the presence of two distinct transmission cycles in a restricted geographical area (Svobodova, Votypka et al. 2006) exemplify the complexity of the epidemiological setting in endemic areas, and

how parameters such as the presence of a sylvatic reservoir or the biting habits of different vector species, which may be more or less prone to bite again following a first infectious feed, may increase or reduce the chance of co-infection of different strains of *L. tropica* and thus the possibility of mating in the midgut of the insect vector.

A deeper understanding of gene flow between demographic units of *L. tropica* may be offered by studies that tackle population genetics within an epidemiologically relevant setting. Although the clonal theory has focused on the concept of evolutionarily stable "clones" which are transmitted from person to person within epidemic foci, this theory fails to provide useful information on the frequency of hybridization events in natural population of *Leishmania* species, which are now known to occur. Specifically, no distinction is made between lack of hybridization due to physiological boundaries in the capacity of different strains to perform genetic exchange, on one hand, and lack of hybridization due to epidemiological reasons on the other, such as presence of transmission cycles by different vector species, geographic barriers, or even transmission intensity, which for instance may reduce the chances of two different parasite strains meeting within the vector if transmission is low, such as in those areas associated with a sylvatic zoonotic cycle where CL due to *L. tropica* is known to be very sporadic.

6.1.3. Population genetics and models of reproduction

In summary, the population genetics evidence reviewed in this and preceding chapters, and the data I presented in Chapter 2 suggest that genetic exchange consistent with Mendelian inheritance may be present in *L. tropica*, given the pattern of allele sharing observed in isolates from different geographical clusters, and the presence of abundant heterozygosity in a subset of these samples. The presence of meiotic homologous recombination was suggested by previous studies (Krayter, Alam et al. 2014), and the observation of heterozygous markers in linkage with homozygous markers in the samples analyzed in this thesis would seem to confirm this suspicion. In a different species of *Leishmania*, recombination breakpoints were directly detected by WGS in field samples isolated from sand flies (Rogers, Downing et al. 2014). These parasite lines were likely produced by an outcrossing event between two divergent lines with subsequent inbreeding within and between hybrid lines, both of which would produce the genomic pattern of patchy heterozygosity that was observed. However, processes associated with meiotic homologous recombination other than crossing over, such as gene conversion, are suspected to be present in *Leishmania*, although the process lacks molecular characterization, and may give rise to stretches of homozygous regions where heterozygous regions are expected (Akopyants, Kimblin et al. 2009). In human genomes, for which considerable more evidence is available, the rate of gene conversion seems to be higher than the rate of crossovers (Jeffreys and May 2004, Gay, Myers et al. 2007).

The presence of meiotic homologous recombination, in the form of either gene conversion or chromosomal crossing over, would be in line with either the parasexual or sexual model of reproduction as put forward in the Introduction. Although our evidence for recombination is limited, the presence of both homozygous and heterozygous regions is apparent from the WGS, with predominance of either allele 1 or allele 2 in homozygous form in a few isolates and the same alleles present in short stretches in heterozygous form in other isolates (Figure 2.9). This genomic survey of genetic variability in these *L. tropica* isolates provided a useful conceptual framework to further prioritize different strains for experimental crosses, and revealed important differences in the resolving power of WGS technologies when compared to traditional MLST, which only captures variation in a small fraction of the genome (only approximately 15 kb were amplified by the PCR probes used in Chapter 2), and is prone to bias depending on which coding or non-coding regions are selected for genotyping.

6.2. Genome plasticity in *L. tropica*

6.2.1. Variation in somy and effects on transcription

As I presented in Chapter 3, extensive variation in chromosome number is seen in *Leishmania* species, and the *L. tropica* isolates considered in this study were no exception. Cloning of individual cells in a subset of isolates and their paired WGS and RNA-seq analysis revealed mosaicism in the number of chromosomes within a

single *in vitro* adapted parasite population. Whether this mosaicism is also seen to the same extent in natural parasite populations isolated from either individual hosts or individual sand flies is unknown, as all previous studies concern *in vitro* adapted parasites (Sterkers, Crobu et al. 2014). Mitotic divisions associated with parasite proliferation *in vitro* may be intrinsically less likely to have balanced segregation of the chromosomes: i*n vitro* promastigote forms are known to be pleomorphic and display a range of shapes and sizes, are developmentally arrested, and do not proceed to the infectious metacyclic stage (Sacks 1989, Schuster and Sullivan 2002).

Regardless of whether these findings are consistent in field isolates that have never been adapted to growth in culture medium, the results I discuss in Chapter 3 demonstrate that *L. tropica* is in principle capable of tolerating extensive aneuploidy, confirming previous reports of aneuploidy in this and other *Leishmania* species (Bastien, Blaineau et al. 1992, Sterkers, Lachaud et al. 2011). These reports have also been validated by WGS and other experimental approaches (Cruz, Titus et al. 1993, Rogers, Hilley et al. 2011). In previous fluorescence *in situ* hybridization (FISH) studies, asymmetric nuclear chromosome allocations were seen in mitotically dividing cells (Sterkers, Lachaud et al. 2011). The mechanism by which aneuploidy arises was thus postulated to be a defect in chromosomal replication, which explained the odd number of chromosomes seen in all asymmetrically dividing cells. Cloning of individual cell lines from the original population showed that any combination of somy was observed to occur, each chromosome being present in at least two of the following conditions: monosomy, disomy, and trisomy.

In the set of isolates considered in this study, I found most cloned and uncloned lines to be near-diploid, although variation at the level of individual chromosome number was common. One clonal line appeared to be near-triploid both by modelling of the expected haploid read depth and by inspection of allele frequencies on individual chromosomes. The fact that a near triploid clone was generated from a near diploid isolate suggests that considerable standing cellular variation in karyotype exists within culture-adapted isolates.

Aneuploidy is generally deleterious in most eukaryotic species. In cancer cells, for example, karyotype alterations are associated with oncogenesis(Giam and Rancati 2015), while if they occur in human germline cells, they are associated with developmental defects that are often lethal (Hassold and Hunt 2001). The hypothesis I aimed to test in Chapter 3 involved determining whether supernumerary chromosomes are effectively downregulated at the level of steady state transcript levels to compensate for the increased gene dosage. The transcriptional machinery in *Leishmania* lacks the ability to regulate transcriptional initiation, and all genes are constitutively expressed. However, several RNA-binding motifs in the 3' UTR of mRNA transcripts have been found to determine the half-life of specific transcripts via RNA-binding proteins (See Section 3.1). A possible mechanism for reduced steady state transcript levels of supernumerary chromosomes can be in theory attributed to cis-acting RNA-binding proteins, which are also more highly expressed by virtue of being on the same chromosome, and therefore may function in a "balancing act" to proportionally reduce steady state transcript levels to those seen in the disomic state. If the majority of RNA-binding

proteins however function in trans, steady state transcript levels of genes on supernumerary chromosomes would still be higher than expected in the disomic condition.

The results I present in Chapter 3 suggest that gene expression on supernumerary chromosomes behaves in a dose dependent manner, with higher somy associated with higher steady state mRNA levels. Gene dosage effects entirely explain the patterns observed at the chromosome level in two clonal lines generated from the same field isolate, where somy differed at a subset of chromosomes. This leaves open the question as to how *L. tropica* can cope with the increased expression of a large number of genes to maintain cellular homeostasis. Further regulation at the level of translation into protein may explain this ability, possibly via the activity of chaperones such as seen in the heat shock response (Spath, Drini et al. 2015) or via RNA-binding proteins that regulate translation activity via sequestration or modification of the mRNA transcript, while leaving the total level of mRNA transcript (which is what is captured by RNA-seq) present in the cell unchanged.

6.2.2. Variation in gene copy number and effects on transcription

In addition to describing changes in expression due to aneuploidy, one of the aims of Chapter 3 was also to identify which genes were most differentially expressed (DE) between the isolates considered. Remarkably, the most highly represented category in the DE gene set included transmembrane proteins with a transporter function. These are known to play an important role in shuttling

nutrients and other compounds in and out of the cell, thus affecting parasite fitness in unfavourable environmental conditions, such as when a drug is present in the extracellular environment. Many structural variants that we found to be differentially expressed span loci that have been previously implicated in drug resistance, for instance at the FT1 transporter locus in antifolate drug resistance (Ouameur, Girard et al. 2008). In order to further dissect gene dosage effects at the sub-chromosomal levels, two clonal lines of the same isolate were again compared to identify copy number variants (CNVs) that differed between the two lines. Many of the top 30 DE genes identified from the entire sample set of isolates fell into genomic regions that appeared to be large CNVs in these two clones, spanning dozens of genes.

Relative shifts in the read depth ratio between the two clonal lines showed that genes within these CNVs were consistently upregulated in the clone in which the CNV had higher relative read depth. Gene dosage effects thus again largely accounted for differences in expression of genes within CNVs; however, at a minority of genes the trend was opposite to that expected from differences in read depth. In order to address the presence of regulatory mechanisms determining transcript abundance independently of the copy number in the genome context, a screen identifying significant differences between DNA and RNA read depth was performed at a genome-wide SNP level, looking for allele-specific gene expression. The presence of only one allelic variant in RNA transcripts at heterozygous loci would by definition be a violation of gene dosage rules, which otherwise seem to reliably predict in most cases the amount of transcription seen for a given gene. A

large number of SNPs in coding regions gave highly significant p-values, and inspection of the DNA and RNA sequencing reads overlapping those positions indeed showed that only one allele was present in the RNA data, whereas the DNA data showed equal proportions of two different alleles.

Allele-specific gene expression is a previously uncharacterized phenomenon in *Leishmania* and represents an important exception to the predictive power of gene dosage as an explanation for RNA steady state levels. While gene dosage generally can reliably predict transcript levels at a given locus, there are additional layers of regulation that can, in select cases, dictate which allelic variant is actually expressed. Given the overlap between the most significant hits from the allele-specific expression analysis of the two clones, it is highly unlikely that the same gene conversion event has independently occurred in two different cultured parasite populations at the same exact genomic locus, and must instead reflect the presence of variation in DNA sequence affecting transcript abundance. Our data, however, cannot distinguish between sequence variation affecting RNA stability, translational efficiency, or variation in conserved binding motifs mediating interactions between specific gene transcripts and RNA-binding proteins.

In conclusion, I have provided an in depth analysis of patterns of gene expression in *L. tropica*, and confirmed that this species can tolerate significant plasticity in both chromosome number and intrachromosomal structural variants. CNVs appear to be an important mechanism for parasites to upregulate or downregulate, via either amplifications or deletions, the expression of specific genes.

6.2.3. Genome structure and models of reproduction

While the focus of Chapter 3 was on the effects of genome structure on transcription, some conclusions can be drawn on reproductive strategies in *L. tropica*. In particular, the observation that considerable variation in genome structure is observed within individual isolates after being grown *in vitro*, which can therefore be considered true "mosaics", complicates the concept of "clonal lineages" which is central to the clonal theory. Indeed, if the parasite population within a given host or vector is composed by an assemblage of individual parasites with different patterns of aneuploidy, and this variation in genome structure is indeed functional, as suggested by the fact that expression of the genes on these chromosomes increases in a dose-dependent manner, then this added layer of genetic variation must be carefully taken into consideration when determining relationships between different isolates.

Moreover, long runs of homozygosity (LROH) were seen in both cloned and uncloned isolates, resembling the patterns of patchy heterozygosity mentioned in Section 6.1.3 that have been associated with hybridization in *L. donovani* complex isolates from Turkey (Rogers, Downing et al. 2014). Such blocks of homozygosity alternating with heterozygous tracts can be explained by the presence of homologous recombination in natural populations of *L. tropica*. The two models, parasexual and sexual, are therefore supported by the variation in genome structure presented in Chapter 3, with aneuploidy expected to appear through mitotic divisions following cell fusion. The difference between these two models lie in the

type of cell that can perform cell fusion: if the cell is a haploid or near-haploid gamete, then the reproductive strategy would effectively be sexual and involve a meiotic step, during which homologous recombination occurs; if the cell is a somatic cell with a ploidy greater than *n* (where *n* is the set of haploid chromosomes, equal to 36 in *L. tropica*), then no meiosis has occurred and the reproductive strategy would be considered parasexual, although recombination may still be present. The evidence presented here is unfortunately insufficient to draw conclusions on whether a sexual or parasexual cycle is more likely.

6.3. Genetic exchange in *L. tropica*

6.3.1. Sand fly infections and hybridization

In Chapter 4 and Chapter 5 I describe the experimental crosses that were performed in *L. tropica* using combinations of different drug resistant parasite lines, and the results of WGS of 10 different hybrid lines. I hereby provide a short summary of the hybrid recovery rate seen in these crosses and the implications this has for hybridization in natural populations of *L. tropica*.

Out of 7 different cross combinations, only one gave hybrid lines that retained double drug resistance following passage into a new culture flask. These results suggest that only certain combinations of parasite lines have mating competency, or that certain pairings may be more fertile than others: an intriguing possibility that the data seems to suggest is that lines with abundant heterozygosity may have

decreased fertility compared to lines in which homozygosity is more abundant across the whole genome. Heterozygosity and homozygosity can in this case be considered as a proxy measure for outcrossing and inbreeding. None of the crosses in which at least one of the two parental lines was largely heterozygous (see Chapter 2) produced any viable hybrid lines. If isolates that fell into Cluster 3 can be considered "outbred" due to their elevated heterozygosity, as previous analyses seem to suggest, then they must have recently arisen trough an hybridization event between two different "inbred" strains. "Inbred" strains may have greater reproductive potential than "outbred" strains: such a situation is the opposite of that seen in heterosis, or hybrid vigour, where hybrids have greater fitness than either parental line. A condition in which hybrids have a reduced capability to undergo sexual or parasexual reproduction is consistent with the establishment of post-zygotic barriers to interbreeding between parental lines and the new hybrid lines, which may therefore facilitate speciation.

6.3.2. Genomic consequences of hybridization

The main aim of Chapter 5 was to describe the changes observed during hybridization at a genome level. One of the most important objectives was to establish patterns of inheritance of the genetic material as it was passed on from the parental lines to the offspring. Previous crosses in *L. major* and in *L. infantum* found biparental inheritance at a number of genomic markers. We expand these analyses to include inheritance of all biallelic loci, resolution of the gametic phase in all

hybrid lines, somy estimation for all chromosomes, and identification of *de novo* variants associated with hybridization.

All hybrid lines were near-diploid, as were the two parental lines. The only chromosomes showing evidence of aneuploidy were chromosomes 4, 23, and 31. While chromosome 4 was disomic in both parental lines, chromosomes 23 and 31 were trisomic and tetrasomic, respectively. Interestingly, all chromosomes were homozygous in the parents, while all chromosomes in the hybrid lines were heterozygous. Phasing of disomic chromosomes confirms biparental inheritance, and while multisomic chromosomes cannot be phased with traditional approaches, the same pattern of heterozygosity was seen in all hybrid lines, suggesting that meiotic processes may consistently reduce the number of multisomic chromosomes and in a sense "reset" aneuploidy.

As evidenced from allelic plots, the inheritance of markers on disomic chromosomes is largely Mendelian. A minority of SNP positions appeared to be violating Mendelian inheritance rules. A more accurate analysis of the distribution of these SNPs in the genome could not be perfomed due to the lack of annotation of the reference genome used. An automated annotation pipeline has been developed for *L. tropica* and will be deployed for future studies. Many *de novo* structural variants were also detected in the hybrids. Herein I have limited myself to providing a list of the different types of structural variants detected and have postponed a more detailed analysis of the distribution of these variants with respect to coding regions to subsequent studies that will be informed by an accurate genome annotation.

6.3.3. Hybridization and models of reproduction

Tracking the inheritance of whole chromosomes can shed some light on the processes governing *L. tropica* biparental inheritance following cell fusion. All trisomic and tetrasomic chromosomes in the hybrid lines were heterozygous (with allele frequency peaks at 0.33 and 0.67 if trisomic, and 0.5 if tetrasomic). This consistency across several independent mating events suggests that non-random segregation of the chromosomes occurs during hybridization. The presence of a parasexual cycle would involve cell fusion followed by random loss of certain chromosomes during mitotic division. This was not observed in the data presented in Chapter 5. Consistently, each hybrid line inherits at least one chromosome from each parent, suggesting a more complicated mechanism being at work than concerted loss of chromosomes during mitotic divisions following a cell fusion event between near-diploid cells, as predicted by the parasexual model. By the combinatorial formula, the probability that the 33 disomic chromosomes that were phased in the hybrids each inherited a single chromosome from each parent if they originated from a cell fusion between the two diploid parents and subsequently lost two of the four chromosomes at random is approximately $1x10^{-6}$. The probability decreases if we account for potential aneuploidy that arises in that step, providing convincing evidence that a meiotic haploid or near haploid stage is very likely.

These results are indeed in line with what is expected from a meiotic stage being present in *L. tropica*. The obligatory generation of haploid gametes would explain why all hybrid lines inherit at least one chromosome from each parent.

Chromosomes that are trisomic in the parental lines cannot be split evenly in the resulting gametes, thus resulting in two possible near haploid gametes with either a monosomic or a disomic version of these chromosomes. Both of these possibilities are in fact seen in the hybrid lines for chromosome 23, with some hybrids being disomic (1 + 1) and others being trisomic (2 + 1). Chromosomes that are tetrasomic in the parental lines can be split evenly in the resulting gametes, and seem to be preferentially passed on as disomic chromosomes in the gametes (2 + 2). This would explain the maintenance of tetrasomy at chromosome 31 in most *L. tropica* isolates analysed in this thesis, and the fact that all hybrids had a single peak at 0.5 in the allele frequencies for this chromosome.

The exception to this remarkable consistency is given by chromosome 4, which is trisomic in two of the hybrid lines. Such isolated cases of aneuploidy in the offspring when the parental lines have balanced ploidy could arise during the *in vitro* growth phase rather than as a result of chromosomal segregation during meiosis. Indeed, over prolonged *in vitro* culture, the consistency in chromosome number seen across independent mating events could be lost due to asymmetric cell divisions (Section 6.2.1).

6.4. Future directions

Additional crosses of *L. tropica*, especially backcrosses involving the F1 hybrids, will help generate enough recombination events in the progeny to break associations between even closely linked markers. As described in Section 5.1, near

isogenic lines produced by a backcross of the F1 hybrids to one of the parental lines provide some of the best tools for forward genetic mapping of traits of interest in several model organisms. *L. tropica* has very few *in vitro* or *in vivo* models of pathogenesis, although infection in hamsters has been described as a possible model of viscerotropic disease. The establishment of simple, quantitative, biologically relevant phenotype readouts will help screen a large number of hybrid clones for presence or absence of the trait of interest. A high throughput, accurate, and reliable phenotyping platform will be necessary to obtain informative results.

In the short term, more sensitive approaches to detect recombination events on disomic chromosomes in the presence of extensive homozygosity will be developed. These statistical methods need to take into consideration the unique genomic features of *Leishmania* and other kinetoplastid parasites, and will focus on detecting recombinant haplotype blocks between heterozygous SNPs in different F1 hybrids. The exact location of the crossover won't be known if it falls within a stretch of homozygous SNPs, but evidence that a recombination has occurred between the two heterozygous SNPs can be gathered.

The development of *ad hoc* phasing approaches for multisomic chromosomes will prove to be difficult given the large number of possible phasing solutions. Phasing and detection of recombination are both areas of active research in bioinformatics, and in humans these benefit from the presence of large projects such as the HapMap or the 1000 Genomes Project that provide a steady influx of relevant information for this type of analyses.

Crosses between parasite lines engineered with different fluorescent markers will be used to identify the meiosis competent stage in *L. tropica*. A cross between green and red fluorescent lines can point to the location within the midgut where hybrids, which will fluoresce yellow, make their first appearance. Morphological observations on the type of cells expressing red, green, or both fluorescent markers can narrow down identification of the promastigote stage undergoing meiosis, as has been done for *T. brucei* crosses.

The reference genome that was generated utilizing optically mapped data and assembled independently of synteny with other *Leishmania* spp. will be annotated using a combination of coding region prediction tools and tools that identify homology with other species such as *L. major*. Such an annotated genome will be used to further dissect structural changes and the location of new single nucleotide variants that occur during hybridization. The presence of abundant RNA-seq data for this species will allow discovery and annotation of transcription start and end sites.

In terms of genome plasticity and the effects of structural changes on transcription, spike-in RNA-seq of lines with polyploid genomes will allow absolute quantitation of RNA levels in diploid, triploid, and tetraploid lines. By comparing the relative gene expression observed within experimental samples to a known amount of a reference RNA transcript spiked into the preparation, these type of studies will be able to better inform our understanding of how *Leishmania* is able to cope with polyploidy and aneuploidy.

In plants, aneuploidy seems to be associated with dampening of transcription to avoid deleterious effects on cellular homeostasis. The effects of ploidy on transcription can be better understood in terms of absolute, rather than relative, mRNA levels. The exciting prospect of performing single-cell WGS and RNA-seq, which has now been achieved in several unicellular and multicellular eukaryotes, will allow resolution of the dynamics associated with mosaic aneuploidy. The possibility of performing single-cell WGS on parasites from a clinical or field isolate without prior culturing can reveal the extent of mosaic aneuploidy *in natura*.

Lastly, large-scale population genetics studies of *L. tropica* and other *Leishmania* species are currently underway. These will provide crucial information for arriving at a more accurate and effective systematics in *Leishmania*. Evolutionary relationships between different species or subspecies can be explored with these large datasets. However, to reach a more detailed understanding of the population dynamics in active foci of transmission, dense sampling from a localized area in conjunction with adequate epidemiological surveys will need to be carried out in order to assess the relative frequency of hybridization in different endemic settings.