EMBO reports

Manuscript EMBO-2017-44216

Identification of the centromeres of *Leishmania major*: revealing the hidden pieces

Maria-Rosa Garcia-Silva, Lauriane Sollelis, Cameron Ross MacPherson, Slavica Stanojcic, Nada Kuk, Lucien Crobu, Frédéric Bringaud, Patrick Bastien, Michel Pagès, Artur Scherf, and Yvon Sterkers

Corresponding author: Yvon Sterkers, UMR MIVEGEC CNRS 5290 - IRD 224 - Universitè Montpellier

Review timeline:

Submission date:	17 March 2017
Editorial Decision:	04 April 2017
Revision received:	03 July 2017
Editorial Decision:	21 July 2017
Revision received:	12 August 2017
Editorial Decision:	15 August 2017
Revision received:	15 August 2017
Accepted:	28 August 2017

Editors: Achim Breiling

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 04 April 2017

Thank you for the submission of your research manuscript to EMBO reports. We have now received reports from the three referees that were asked to evaluate your study, which can be found at the end of this email.

As you will see, all three referees highlight the potential interest of the findings. However, all three referees have raised a number of concerns and suggestions to improve the manuscript, or to strengthen the data and the conclusions drawn. As the reports are below, I will not detail them here, but we think it will be most critical to address point 1 of referee #1 (proper analysis of ChIP-data; this relates to point 2 of referee #3) and point 1 of referee #3 (expression of LmKKT1) during the revision.

Given the constructive referee comments, we would like to invite you to revise your manuscript with the understanding that all referee concerns must be addressed in the revised manuscript and in a point-by-point response. Acceptance of your manuscript will depend on a positive outcome of a second round of review. It is EMBO reports policy to allow a single round of revision only and acceptance or rejection of the manuscript will therefore depend on the completeness of your responses included in the next, final version of the manuscript.

REFEREE REPORTS

Referee #1:

In this short manuscript, the authors report the identification of centromere sequences in Leishmania major, a kinetoplastid parasite that has an unconventional set of kinetochore proteins (called KKT proteins). Although centromeres were previously mapped in two other kinetoplastids (Trypanosoma cruzi and Trypanosoma brucei), the position of centromeres remained enigmatic in Leishmania. To address this question, the authors first characterized the LmKKT1 protein and showed that it is indeed a kinetochore protein based on its localization pattern. They subsequently performed a ChIP-seq analysis and successfully mapped the position of centromeres in all 36 chromosomes. Identification of centromeres in Leishmania represents a significant step toward understanding the reported mosaic aneuploidy that may have relevance to its parasitic life style. This manuscript also reports the first characterization of unconventional kinetochore proteins outside T. brucei, showing the functional conservation of unconventional kinetochore proteins in different kinetoplastids. I therefore support the publication of this important manuscript in EMBO Reports once my following concerns are fully addressed in the revised manuscript.

Major comments:

1. The authors have to analyze their ChIP-seq data properly. In the manuscript, they "subtracted" the INPUT reads from IP reads, which is not a standard practice in ChIP analyses. They have to either report fold-enrichment (IP reads divided by INPUT reads) for proper normalization or report both IP reads and INPUT reads in separate graphs (e.g. see https://www.ncbi.nlm.nih.gov/pubmed/23230266, https://www.ncbi.nlm.nih.gov/pubmed/24582333, https://www.ncbi.nlm.nih.gov/pubmed/27384170).

2. Despite the authors' analyses, L. major genome clearly encodes putative orthogs for KKT5 (LmjF.06.0200) and KKT12 (LmjF.24.1400).

3. The authors used HHpred and reported that the LmKKT1 and TbKKT1 have similarity to a DEAD box-helicase (E-value 99) and a TOG domain (E-value 16). Although potentially interesting, the similarity score is not significant at all. In HHpred, "the E-value is the average expected number of non-homologous proteins with a score higher than the one obtained for the database match. An E-value much lower than 1 therefore indicates statistical significance" (taken from the HHpred website: https://toolkit.tuebingen.mpg.de/hhpred/help_ov). Without any supporting data, they should take out the analysis.

Minor comments:

1. "interphasic cells" should be "interphase cells".

2. "... this suggests that the repeat elements were not lost in L. major and T. cruzi but rather appeared in T. brucei". This is a premature conclusion. One would need to know the centromere structure in the common ancestor of the three trypanosomatids at the very least to make such a statement.

3. Given the high level of synteny among trypanosomatids, it would be interesting to analyze whether the position of the centromeres in L. major is conserved in T. brucei and T. cruzi.

Referee #2:

The nature of centromeres in the protozoan parasite Leishmania has yet to be resolved. Techniques which have led to the identification of centromeric regions in other organisms, including related kinetoplastids, have not yielded reliable data when applied to Leishmania. Here, the authors have used a protein (LmKKT1), inferred from the recently identified Trypanosoma brucei kinetochore complex, as a probe in ChIP-Seq experiments. In doing so, they identify single short regions (4 - 8kb) on each chromosome which they interpret as being the centromeric location.

This work represents an important step towards understanding chromosome segregation in Leishmania. The data are intriguing, if somewhat preliminary. The conclusions would be much more robust if another of the kinetochore proteins had been used in parallel and a similar data set produced. Having said this, the results presented will be of interest to those in the field, and act to stimulate further research. Below I have listed a number of mainly minor points which the authors should also address.

1. The introduction could be better formatted. It is slightly stilted and gives the impression of having been written in a hurry.

2. The frequently used phrase 'chromosome allotment' is one that I have not come across - is it specific to the field?

3. The data presented in Table 1 do not add much to the paper - move to supplementary information?

4. I found Fig 2A confusing at first, and Fig 2B difficult to visualise in detail. Perhaps get rid of Fig 2A, and increase the size and uniformity of Fig 2B.

5. The authors should comment more on the data relating to chromosome 29 and add more detail.

6. I was confused by the usage in the text of IUPAC representation for alternative nucleotide sequence. Could the authors make this clearer, as there is possible ambiguity at first glance with the amino acid one letter code (Bs not withstanding). Could the authors also comment more on the statistical validity of their interpretations.

6. Figure 4 needs more detail. This should include a more comprehensive legend and better markers on the figure (eg use of arrows). Do the authors have any views why they identify only 8 - 12 spots with LmKKT1-GFP?

7. The discussion relating to the co-localisation of LmKKT-1-hits with regions shown to be required for mitotic stability (page 6/7) has a rather anecdotal feel to it. Can the authors be more specific and detailed.

8. In the discussion, the authors make inferences about the possibility of regional centromeres in Leishmania. Without further data, they need to be careful not to over-interpret their data.

Referee #3:

The manuscript "Identification of the centromeres of Leishmania major: revealing the hidden pieces" by Garcia-Silvia and others used bioinformatics, ChIP-seq, and FISH-IF to identify a putative non-canonical kinetochore protein, similar to those earlier found in Trypanosma brucei. These exciting findings provide experimental confirmation that these kinetoplastid kinetochore proteins are indeed conserved between divergent Trypanosomatida species. Centromere and their associated kinetochore proteins are enigmatic in their fast pace of evolution, their presence or absence in specific clades, as well as chromosomal distribution pattern. All of this while facilitating one of the most important biological events: faithful chromosome segregation. This centromere paradox was further complicated by the discovery of KKTs in kinetoplastids. This highlight the needs to expand the search for centromere and kinetochore components in diverse species to better understand how and maybe even why these proteins evolve the way they do. Indeed, this manuscript provides exciting new insights in the evolution patterns of KKTs. Although the authors overall present a strong manuscript, a few concerns need to be addressed, as listed below.

Major concerns:

1. The authors use an expression vector with GFP-tagged LmKKT1, but the authors do not mention what kind of promoter was driving the expression of LmKKT1. If LmKKT1 was overexpressed, this could result in ectopic localization of LmKKT1, similar to what is observed for overexpressed CENP-A/cenH3 in human cell lines, as shown by the Almouzni, Dalal, and Cleveland labs. Also the

presence of the endogenous LmKKT1 might alter the localization of GFP-tagged LmKKT1. Comparing the expression levels of the transfected LmKKT1 versus genes within the same polycistronic arrays of genes where the endogenous LmKKT1 resides, would allow the authors to discuss the relative expression levels.

2. For their ChIP-seq experiment the authors use a complex protocol using both MNase to digest Leishmania chromatin, followed by additional sonication of mononucleosome-depleted chromatin. In addition, the authors first cross-linked their samples before fragmentation. If it is technically possible, native ChIP could provide more enrichment of LmKKT1 at its chromatin association sites with reduced background noise induced by crosslinking. Furthermore, MNase digestion should be sufficient for ChIP-seq. In case the ChIP'ed DNA is too long for sequencing by synthesis, it can be treated for a second time with MNase. Furthermore, how ChIP-seq data is binned will affect how peaks are found. The authors should elaborate on the rationale for their ChIP-seq protocol and analysis.

3. For the TRF analysis the authors use a very stringent parameters. Satellite tandem repeat sequences are known for their heterogeneity, both in sequence composition as well as in minor indels. To accommodate for these features in their cross-eukaryote study, Melters et al 2013 used very leanient parameters (match=1, mismatch=1, indels=2 with a maximum period size of 750). Also, the search for tandem repeats should not be limited to LmKKT1 binding sites, but should encompass the entire genome. The results from this search can be overlaid with the LmKKT1 ChIP-seq data.

4. The FISH-IF experiments performed by the authors provide great inside in how GFP-LmKKT1 and LmKKT1 binding sequences of chromosome 13 and 27 behave throughout Leshmania's cell cycle. Figure 4 could be improved by showing each channel separately (green, red, DAPI, merged, and DIC) with each image clearly marked what it shows. Furthermore, it is rather surprising that the LmKKT1 binding sequences do not co-localize with certain GFP-LmKKT1 foci, but rather appear to be adjacent to each other. The authors should address this observation.

5. In the discussion the authors mention that centromeric sequences are early replicating in eukaryotes in general. Next, they argue that data about replication timing in Leishmania is scarce and controversial. Nevertheless, this does not withhold the authors to use these data to argue that they indeed found the centromeres by LmKKT1 ChIP-seq. First, in humans alpha-satellite sequences tend to be late replicating, contradicting the statement by the authors. Second, if the authors state that the replication timing data for Leishmania is controversial, their reliance on using the data that does exist as supporting arguments should be equally cautious.

Minor concerns:

1. The authors express GFP-tagged LmKKT1 in Leishmania cells. One concerns that might arise is when introducing a gene, is that it creates a distinct phenotype. For instance, overexpression of the canonical kinetochore protein CENP-C in DT40 cells results in cytokinesis defects (Fukagawa 1999). The authors should discuss any GFP-tagged LmKKT1 induced phenotypes.

2. The authors identify LmKKT1 to be the homologue of TbKKT1 based on 36% identity and 53% similarity. This finding would be further highlighted if these two proteins were aligned with a graphical representation of their relative homologue, including emphasizing the two conserved domains (DEAD box-helicase domain and TOG-domain). In Akiyoshi and Gull 2014 KKT1s were found in several other kinetoplastids. This would allow the authors to determine the rate of evolution of KKT1. Problems with identifying canonical kinetochore proteins is well known, as highlighted by Meraldi et al 2007 and various papers from the Henikoff lab, where putative orthologs are restricted to short sequences, such as the CENP-C motif in CENP-C proteins. Maybe this problem also exist in KKTs if they were equally fast evolving.

3. In the discussion, the authors mention that transcription of centromere DNA is unexpected, but recently pervasive transcription happens at all centromeres, as has been shown by various labs. For instance, Athwal et al 2015 shows that ectopic CENP-A predominantly goes to transcription start sites. Molina et al 2016, Quenet & Dalal 2014, Koo et al 2016, Grenfell et al 2016, and Blower 2016

are just a few of the most recent papers describing centromeres being transcribed.

4. In the discussion the authors mention that Leishmania diverged early in trypanosomatid evolution. This gives the appearance that Trypanosma is the base of the Trypanosomatida tree and that Leishmania is the derived branch. But the phylogenetic trees of kinetoplastids show a bifurcation of the branches that harbor either Lieshmania or Trypanosma.

5. Centromere morphology is rather diverse. Budding yeast has a genetic centromere, as do its very close relatives, whereas most other eukaryotes have regional centromeres. These regional centromeres are commonly characterized by the presence of large arrays of tandem repeat sequences, but unique sequences are frequently found as well, such as in various fungi species, many chicken chromosomes, and incidental horse and orangutan chromosomes. Finally, there are the holocentric chromosomes, a feature that has evolved at least 15 times. Mosaic aneuploidy as found in Leishmania is a very interesting observation and might be the result of less than optimal functioning kinetochores. Disfunctioning canonical kinetochores do result in aneuploidy. Maybe the authors could discuss briefly the potential implications of their enrichment levels of LmKKT1 on Leishmania chromosomes as well as the presence of minor peaks in their ChIP-seq data. It is intriguing to contemplate the potential of a unstable Leishmania kinetochore permitting the rise of mosaic aneuploidy, maybe even with promiscuous kinetochore seeding on the chromosomes, as shown by the presence of the minor peaks in the ChIP-seq data.

1st Revision - Authors' response 03 July 2017

Referee #1:

Major comments:

1. The authors have to analyze their ChIP-seq data properly. In the manuscript, they "subtracted" the INPUT reads from IP reads, which is not a standard practice in ChIP analyses. They have to either report fold-enrichment (IP reads divided by INPUT reads) for proper normalization or report both IP reads and INPUT reads in separate graphs (e.g. see https://www.ncbi.nlm.nih.gov/pubmed/23230266,https://www.ncbi.nlm.nih.gov/pubmed/24582333, https://www.ncbi.nlm.nih.gov/pubmed/27384170).

We thank the reviewer for their suggestion but would like to point out that the task at hand was to separate noise from signal. To this end, a simple subtraction after normalization was sufficient. Many ChIP-Seq experiments might benefit from alternative mathematical treatment and we acknowledge the more recognizable metric of fold-change. As such, we have revised this version of the manuscript by adapting figure 2 and its legend to describe fold-change smoothed over the entire genome using a 200 nts window, updating the Materials and methods, and rephrasing several sentences in the text. We should stress that using one method or another gave very similar results. And therefore this FC analysis does not change the data obtained here.

2. Despite the authors' analyses, L. major genome clearly encodes putative orthogs for KKT5 (LmjF.06.0200) and KKT12 (LmjF.24.1400).

True; they are not annotated as such in TritrypDB, but they can be found using a BLAST search. Thank you for mentioning it, we have modified the text and updated Table 1, which has been transferred to the Supplementary data as suggested by referee#2.

3. The authors used HHpred and reported that the LmKKT1 and TbKKT1 have similarity to a DEAD box-helicase (E-value 99) and a TOG domain (E-value 16). Although potentially interesting, the similarity score is not significant at all. In HHpred, "the E-value is the average expected number of non-homologous proteins with a score higher than the one obtained for the database match. An E-value much lower than 1 therefore indicates statistical significance" (taken from the HHpred website: https://toolkit.tuebingen.mpg.de/hhpred/help_ov). Without any supporting data, they should take out the analysis.

We have presented this analysis because we carried out a database search for remote homologues that could be really divergent at primary sequence level due to the fact that they are not conserved among the eukaryotic clade. This search revealed the possible presence of an interesting divergent DEAD box domain. Regarding the validity of our hits, we followed the guidelines given in **https://toolkit.tuebingen.mpg.de/hhpred/help_faq#correct_match,** which stated that "the probability is a more sensitive measure than the E-value". That's why we presented this parameter and not the E-value. However, we agree that we do not have any experimental data to confirm these bioinformatics data, so we removed this from the manuscript and from Figure 1.

Minor comments:

1. "interphasic cells" should be "interphase cells".

Corrected throughout the manuscript.

2. "... this suggests that the repeat elements were not lost in L. major and T. cruzi but rather appeared in T. brucei". This is a premature conclusion. One would need to know the centromere structure in the common ancestor of the three trypanosomatids at the very least to make such a statement.

We agree that this may be considered as too speculative. We just wanted to say that, since *Leishmania* diverged early in trypanosomatid evolution, before the split between *T. brucei* and *T. cruzi*, the most parcimonious hypothesis is that only one event occurred rather than two events; *i.e.* that the repeat elements appeared in *T. brucei* rather than were lost both in *L. major* and *T. cruzi*. Yet, we removed the sentence in this version.

3. Given the high level of synteny among trypanosomatids, it would be interesting to analyze whether the position of the centromeres in L. major is conserved in T. brucei and T. cruzi.

We found that the centromeres of *Leishmania* and *T. brucei* are essentially not syntenics. We have looked if the genes which flank the centromeres in LmjF and in *T. brucei* have orthologues in the counterpart genome and if these genes flank a centromere. In most cases, we did found the orthologous genes but they are not located close to a centromere with one exception, on LmjF chr.12 and Tb Chr.27: the flanking genes for the chr. #12 centromere in Leishmania are LmjF12.0510 and 0520; their orthologues in T. brucei are Tb927.1.3560 and Tb927.1.3820, the first one is separated from the centromere of Tb Chr. #1 by one gene, whereas the second one is nearby but separated by an rDNA cluster, several putative genes and two SSRs. We have added this information in the manuscript as well as in a new supplemental figure S2).

[Data not included in review process file.]

Figure S2: The sole other example of a synteny concerns Tb Chr.4 and LmjF Chr.34. FAZ protein orthologues (Tb927.4.3740 and LmjF34.0680/0690) are neighboring one side of a centromere in each species. Yet, on the other side of the centromere, the orthologue of Tb927.4.3750 in L. major is not on Chr.34 but on Chr.17 (LmjF.17.0180); and LmjF.34.0670 and 0660 have no orthologues in *T. brucei*, and the orthologue of LmjF34.0650 is located on Tb Chr.10, very far from its centromere. However, we thought all this is by far too specific, so we did not include the latter analysis in the manuscript. Regarding *T. cruzi*, we did not perform a compared analysis since the genome data are much more fragmentary, and all data were obtained before the assembly of the genome sequence. Therefore, precise localizations of the centromeres are not available.

Referee #2:

Minor points:

1. The introduction could be better formatted. It is slightly stilted and gives the impression of having been written in a hurry.

We have edited the introduction.

2. The frequently used phrase 'chromosome allotment' is one that I have not come across - is it specific to the field?

Actually 'chromosome allotment' is specific to the field and was coined to explain the basis of how mosaic aneuploidy occurs in *Leishmania*

3. The data presented in Table 1 do not add much to the paper - move to supplementary information?

Done

4. I found Fig 2A confusing at first, and Fig 2B difficult to visualise in detail. Perhaps get rid of Fig 2A, and increase the size and uniformity of Fig 2B.

Accordingly to the reviewer's suggestions, we got rid of Fig2A and increased the size and uniformity of Fig2B (now Fig. 2)

5. The authors should comment more on the data relating to chromosome 29 and add more detail.

Chromosome 29 is the only chromosome for which a clear peak had not been obtained by ChIP; on the contrary, this is a very "noisy" dataset. This is even more visible now, since, as suggested by Reviewer #1, we have used the fold-change method instead of the subtraction method. The subtraction method allowed us to distinguish a major peak and a minor peak, both being separated by 10 kb and sharing sequence similarities. However, we should stress that the peaks identified then were smaller than those on the other chromosomes. One possible explanation might be that the centromeric region of chromosome 29 has not been properly assembled in the sequence available in TritrypDB. This same explanation had been alleged by Akiyoshi & Gull (Cell 2014) for Chr. 9, 10 and 11 of T. brucei. In order to analyze this, we have tried to PCR-amplify the intergenic regions which bear both peaks on Chr.29. Several of the expected PCR fragments in both regions could not be obtained, indicating errors in the sequence assembly. We have added these details in the Results section as well as in the Supplemental material.

6. I was confused by the usage in the text of IUPAC representation for alternative nucleotide sequence. Could the authors make this clearer, as there is possible ambiguity at first glance with the amino acid one letter code (Bs not withstanding). Could the authors also comment more on the statistical validity of their interpretations.

Since the referee found this confusing, we have specified the use of the IUPAC representation for nucleotides in the manuscript and in the legend of Figure 3.

Re: statistics, as claimed by Wenxiu Ma et al in Nat Protocols in 2014, MEME-ChIP is a web-based tool for analyzing motifs in large DNA or RNA data sets which can be used to analyze peak regions identified by ChIP-seq, providing a comprehensive picture of the DNA or RNA motifs that are enriched in the input sequences. Our aim was not to obtain statistically significant hits even if MEME-ChIP searches rank the hits according to E-values; it was rather to obtain sequences that were present in low copy number and in the genome and which colocalize with the centromeres we identified using ChIP-seq. The fact that we have been able to obtain two 'sentences' that detect the centromeres of 19/36 chromosomes is very original, important and opens new avenues to determine how the centromeres are defined in Leishmania.

7. Figure 4 needs more detail. This should include a more comprehensive legend and better markers on the figure (eg use of arrows). Do the authors have any views why they identify only 8 - 12 spots with LmKKT1-GFP?

We have improved Figure 4. In particular, we have followed suggestions of Referee#3 by showing each channel separately (green, red, DAPI, merged, and DIC) with each image clearly marked what it shows. “only 8 - 12 spots with LmKKT1-GFP” is correlated with the number of dense plaques observed by electron microscopy and regarded as kinetochores. This was mentioned in the manuscript.

8. The discussion relating to the co-localisation of LmKKT-1-hits with regions shown to be required for mitotic stability (page 6/7) has a rather anecdotal feel to it. Can the authors be more specific and detailed.

The discussion has been detailed with respect to the two sequences which had been found to confer mitotic stability in the 'pre-genomics era': "'pre-genomics' works had shown that the 'right' end of Chr. #1 conferred mitotic stability to an artificial chromosome made of construct repeats and of this 'right' end, comprising a subtelomeric ~20kb cluster of 272-bp repeats [22]: from the mapping data available, it clearly appears that the LmKKT1 binding site is located precisely at the 'left' end of the ~272-bp repeat cluster (see Results). Similarly, the LmKKT1 binding site on Chr. #19 correlates well with a 30-kb region involved in the mitotic stability of an extra chromosome originating from the mirror duplication of one subtelomeric end of this chromosome [23, 24]. By crossing the sequence and mapping data from 2000 to those in TriTrypDB, we could infer that the centromere identified here using ChIP-seq is precisely located in a ~10kb region devoid of CDSs but rich in poly(dA)n, poly(dT)n, poly(dC)n and poly(dG)n, which was suspected at the time to be responsible for this mitotic stability [24]."

9. In the discussion, the authors make inferences about the possibility of regional centromeres in Leishmania. Without further data, they need to be careful not to over-interpret their data.

We did not aim to over-interpret our data, but just to discuss them in the light of what is known about the centromeres in other organisms. However we agree that we do not have sufficient data in that sense, so we suppressed all inferences concerning the possibility of regional centromeres in Leishmania.

Referee #3:

Major concerns:

1. The authors use an expression vector with GFP-tagged LmKKT1, but the authors do not mention what kind of promoter was driving the expression of LmKKT1. If LmKKT1 was overexpressed, this could result in ectopic localization of LmKKT1, similar to what is observed for overexpressed CENP-A/cenH3 in human cell lines, as shown by the Almouzni, Dalal, and Cleveland labs. Also the presence of the endogenous LmKKT1 might alter the localization of GFP-tagged LmKKT1. Comparing the expression levels of the transfected LmKKT1 versus genes within the same polycistronic arrays of genes where the endogenous LmKKT1 resides, would allow the authors to discuss the relative expression levels.

We agree that the referee rose what could have been a serious concern. First, we have to recall here that in Trypanosomatids, there are no conventional RNA pol II promoters for activating transcription of protein coding genes: (i) transcription of mRNAs is polycistronic, and the primary transcripts are further processed by trans-splicing and polyadenylation to yield monocistronic mature transcripts; (ii) the UTRs, in particular the 3’UTR, are important in the regulation of the expression levels. Yet, most of the regulation is post-transcriptional and the amount of mRNA is weakly informative on the amount of proteins. We have chosen to express the KKT-GFP using an episomal expression vector because it is the most common way to express and to localize proteins in this organism. We agree that we have used an artificial system, using the UTRs of a tubulin gene, which allows obtaining a visible signal in most cases. Nevertheless, several evidences give us confidence in the reliability of our data.

(i) In *T. brucei*, the localizations of the KKTs were obtained by insertion of the tag into the genome, yet conserving one endogenous copy of the gene on the other homologue (Akiyoshi & Gull, Cell 2014). These authors have obtained reliable results since the KKT-binding sites localized precisely at the previously defined centromeres. We therefore do not consider that there might be an alteration of the localization of GFP-tagged LmKKT1 by the endogenous copy.

(ii) The mutants and wild-type cells show no phenotypic differences, whether in the growth rate, cell morphology, Nucleus/Kinetoplast division patterns and motility. Moreover, we have placed the GFP tag in N and C-terminal of the protein and obtained the same localizations for both (Figure 1). Finally, the localization obtained in *Leishmania* is the same as that in *T. brucei* (Akiyoshi et al, Cell 2014), including the pattern of relocalization toward the spindle pole at the end of mitosis, typical of a kinetochore protein (Figure 1). We therefore strongly believe that the localization obtained here for LmKKT1 is correct.

(iii) Similarly to what was observed in human cells where an excess of CENP-A accumulates at non-centromeric locations in the genome, our system could have given a non-interpretable pattern with multiple binding sites. However, on the contrary, we obtained a single very clear binding site per chromosome, similar to what was obtained in T. brucei and very different from artefactual ChIP-seq patterns obtained by Lacoste et al. (2014) for example.

(iv) Given that the protein levels are not well correlated with mRNA levels, only western blots would allow controlling the protein levels of LmKKT1-GFP and the endogenous polycistronic gene array; but we do not have antibodies available.
All in all, we consider that the ectopic expression of LmKKT1 does not introduce any bias in the results obtained here. We added one sentence about this concern in the Results.

2. For their ChIP-seq experiment the authors use a complex protocol using both MNase to digest Leishmania chromatin, followed by additional sonication of mononucleosome-depleted chromatin. In addition, the authors first cross-linked their samples before fragmentation. If it is technically possible, native ChIP could provide more enrichment of LmKKT1 at its chromatin association sites with reduced background noise induced by crosslinking. Furthermore, MNase digestion should be sufficient for ChIP-seq. In case the ChIP'ed DNA is too long for sequencing by synthesis, it can be treated for a second time with MNase. Furthermore, how ChIP-seq data is binned will affect how peaks are found. The authors should elaborate on the rationale for their ChIP-seq protocol and analysis.

Regarding our complex protocol using both MNase and sonication: we have first tried protocols using either MNase digestion or sonication, but we were never satisfied with the results. So we decided to combine both techniques to treat the chromatin of L. major and then obtained satisfactory sheared DNA fragments. Regarding the use of crosslinked ChIP vs. native ChIP: we are aware of the advantages and problems of both approaches. Briefly, the rationale for using cross-linked ChIP here is that native ChIP is particularly appropriate when histones are targeted, but native ChIP is generally not suitable when non-histone proteins (even those that do not themselves bind DNA) are studied; cross-linked ChIP is then preferred there (taken from: ChIP with Native Chromatin: Advantages and Problems Relative to Methods Using Cross-Linked Material Bryan Turner. Copyright © 2001, Institut national de la santé et de la recherche médicale (INSERM) Bookshelf ID: NBK7099PMID: 21413358). About the rationale of the ChIP-seq analysis, please see response to referee #1 (comment 1). Briefly, we used the fold-change method which gave a clear peak for all chromosomes but Chr. #29, for which we secondarily used the subtraction method with better results. This is discussed about the comment #5 of referee #2. We elaborated about this point (Chr. #29) and the rationale of our protocol in the Results section of the manuscript and in the supplemental material.

3. For the TRF analysis the authors use a very stringent parameters. Satellite tandem repeat sequences are known for their heterogeneity, both in sequence composition as well as in minor indels. To accommodate for these features in their cross-eukaryote

study, Melters et al 2013 used very leanient parameters (match=1, mismatch=1, indels=2 with a maximum period size of 750).
Also, the search for tandem repeats should not be limited to LmKKT1 binding sites, but should encompass the entire genome. The results from this search can be overlaid with the LmKKT1 ChIP-seq data.

Regarding the leniency of the parameters, we had started our analysis by using more lenient parameters (match=2, mismatch=3, indels=5 with a maximum period size of 1000), which are the most lenient parameters available on the TRF website. Of note, the less stringent parameters proposed by the referee cannot be set up in the menu proposed by TRF, even in the advanced search. Metlers et al. had first used the basic parameters proposed in TRF, like we did, to obtain the repeated sequences for each species of interest, and then used WU-BLASTN with parameters M = 1, N = -1, Q = 3, R = 3, W = 10 (with post-processing from various Perl scripts) to produce a set of 'global' and 'local' clusters of repeats in each species. Actually, (i) WU-BLAST is not available anymore on the Internet, and (ii) our search parameters allow to find dinucleotides repeated 10 times and 10 mers repeated <3 times, which in our view is lenient enough for the detection of centromeric repeats. Using the same parameters allowed us to state that the centromeres have a much more repeated nature in *T. brucei* than in *Leishmania major*. Selecting less stringent parameters only allowed retrieving more minor repetitions (28 instead of 4, as shown in the Table below for Chr.4). We therefore prefer to not present the data obtained using these less stringent parameters.

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
371--420	20	2.4	21	70	16	56	40	12	40	8	1.72
460--505	2	23.0	2	100	0	92	0	0	50	50	1.00
1109--1152	15	2.9	16	76	10	58	6	40	43	9	1.63
2643--2700	21	3.0	20	60	18	57	36	37	25	0	1.57
2646--2711	21	3.3	19	59	14	54	34	37	27	0	1.57
2643--2720	20	3.8	21	63	19	56	32	39	28	0	1.57
6999--7050	25	2.0	27	74	7	69	26	30	32	9	1.89
7569--7615	23	2.1	22	72	4	57	6	27	51	14	1.67
8476--8531	26	2.2	26	70	6	65	0	23	42	33	1.54
8535--8606	25	3.3	22	58	24	56	1	34	38	25	1.65
9071--9121	22	2.4	21	73	3	60	3	39	47	9	1.55
9103--9145	21	2.0	22	78	13	59	2	23	62	11	1.40
9352--9382	16	2.0	15	87	6	50	9	16	70	3	1.26
10514--10567	2	27.0	2	76	0	73	46	42	5	5	1.50
10991--11022	12	2.7	11	82	17	50	0	0	81	18	0.70
12722--12765	6	7.3	6	97	0	78	38	0	61	0	0.96
12718--12773	18	3.1	18	81	0	82	39	3	57	0	1.16
12718--12773	4	14.0	4	80	0	72	39	3	57	0	1.16
12899--12937	2	19.5	2	83	0	63	41	51	7	0	1.31
12986--13051	22	2.8	24	56	11	65	53	22	19	4	1.64
13365--13403	18	2.2	18	77	9	51	30	10	48	10	1.70
14065--14103	20	2.0	19	75	5	51	30	25	41	2	1.69
14173--14232	31	1.9	32	73	6	80	30	13	48	8	1.71

15588--15653	24	3.0	22	66	12	68	7	46	22	22	1.77
15684--15742	24	2.4	25	66	8	61	10	42	27	20	1.84
16135--16218	36	2.4	34	61	20	78	5	36	26	30	1.80
16492--16522	12	2.6	12	89	0	52	25	0	70	3	1.02
17121--17174	24	2.3	24	73	0	68	24	12	51	11	1.72

We used the TRF analysis both in *Leishmania major* and *Trypanosoma brucei* essentially with the aim of comparing the data obtained in both species and with what is published about the structure of the centromeres in *T. brucei*. Searches for tandem repeats encompassing whole chromosomes using the default parameters (2 7 7, 80, 10, 50, 500) are available through TriTrypDB for *Leishmania* and *Trypanosoma*. Examples for Chr. 1 and 4 are given below. Tandem repeats are relatively scarce in the Leishmania genome and are not found in the centromeres. Hence, we have shown only the non-coding regions flanking the KKT binding sites in both species in Table S2 and S4.

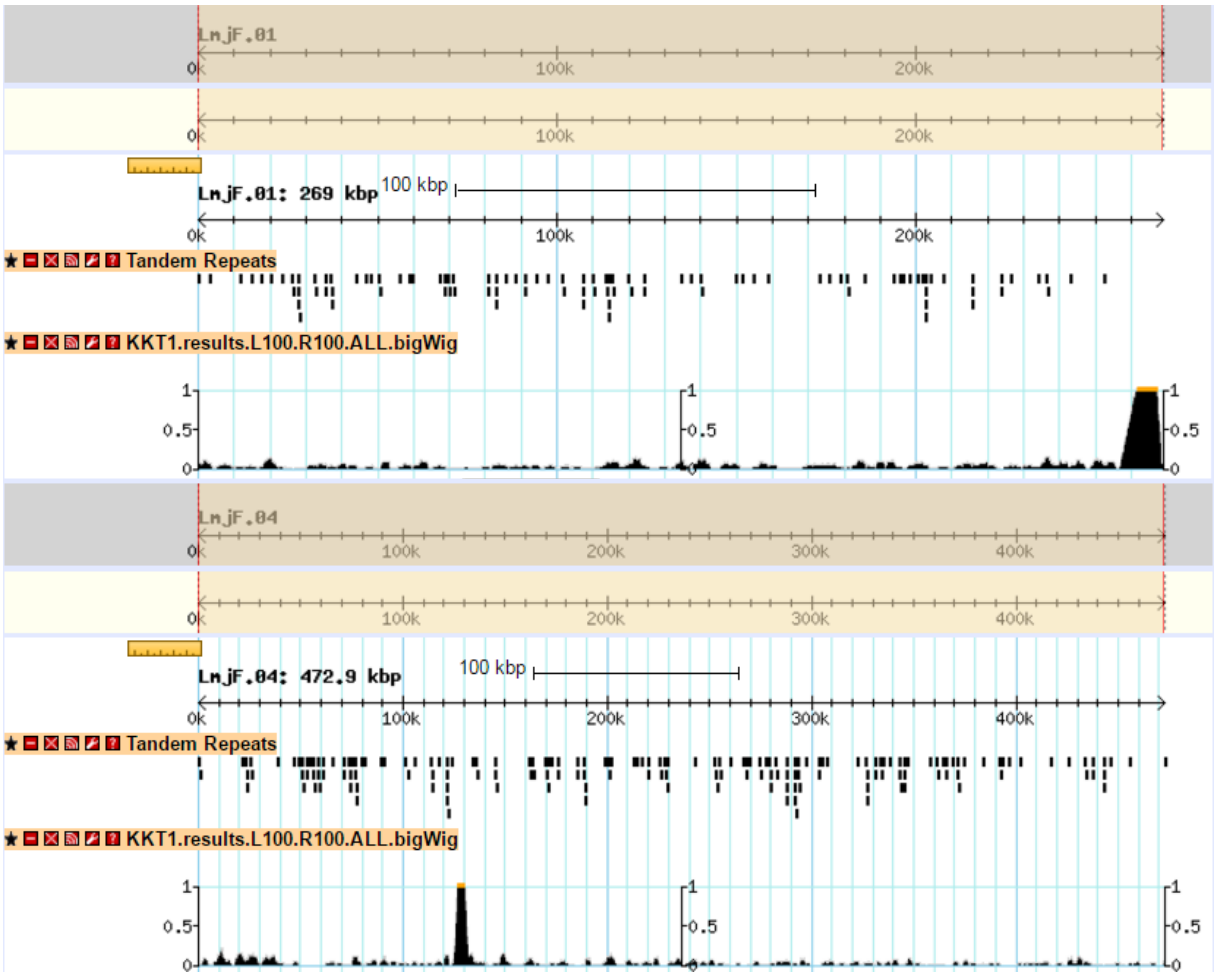


Figure legend: 101 and 174 repeats are found for LmjF Chr. #1 and 4 respectively (see figure below). For chr. #1, no tandem repeats were found close to the LmKKT1 binding site, whereas 4 were found for Chr. #4. This is what was presented in Table S2. Tandem repeats of LmjF.30 are now shown in supplemental figure S3.

3. The FISH-IF experiments performed by the authors provide great inside in how GFP-LmKKT1 and LmKKT1 binding sequences of chromosome 13 and 27 behave throughout Leshmania's cell cycle. Figure 4 could be improved by showing each channel separately (green, red, DAPI, merged, and DIC) with each image clearly marked what it shows. Furthermore, it is rather surprising that the LmKKT1 binding sequences do not co-localize with certain GFP-LmKKT1 foci, but rather appear to be adjacent to each other. The authors should address this observation.

We improved figure 4 as suggested by Referee#3. Regarding the absence of strict co-localization of the spots obtained by FISH with certain GFP-LmKKT1 foci, we partly disagree with the referee. The LmKKT1 binding sequences do not colocalize with most GFP-LmKKT1 foci in interphase cells. By contrast, they perfectly co-localize for Chr.27 (Fig. 4A) and they are adjacent to each other for Chr.13 (Fig. 4B). This is concordant with the probes used for FISH. Indeed, in order to obtain a good sensitivity (for a good proportion of the cells to be labeled, see Sterkers et al. Cell Microbiol. 2011), we have to use probes that span a large sequence of the targeted chromosome, hence available BACs, cosmids or PCR fragments targeting repeated sequences. And good BAC and cosmid libraries are not that common for this organism. Since the centromeres of Leishmania are not based on repeats, we had to design PCR probes from repeat sequences located further on the chromosome. Thus, for Chr.13, the identified KKT1 binding site (coordinates LmjF.13:143300..145099) and the DNA probe target (LmjF.13:91772..118155) are distant by >30 kb. By contrast, for Chr.27, we targeted the rRNA genes of which the coordinates are Lmj.27:989640..1060406, close to the LmKKT1-binding peak (LmjF.27:983200..987599). We have added a sentence in the legend's figure to comment on this point.

4. In the discussion the authors mention that centromeric sequences are early replicating in eukaryotes in general. Next, they argue that data about replication timing in Leishmania is scarce and controversial. Nevertheless, this does not withhold the authors to use these data to argue that they indeed found the centromeres by LmKKT1 ChIP-seq. First, in humans alpha-satellite sequences tend to be late replicating, contradicting the statement by the authors. Second, if the authors state that the replication timing data for Leishmania is controversial, their reliance on using the data that does exist as supporting arguments should be equally cautious.

The reviewer is right, our writing was ambiguous and our reasoning twisted. We have therefore erased this entire paragraph from the Discussion, as well as the reported data from these authors from Fig. 3B.

Minor concerns:

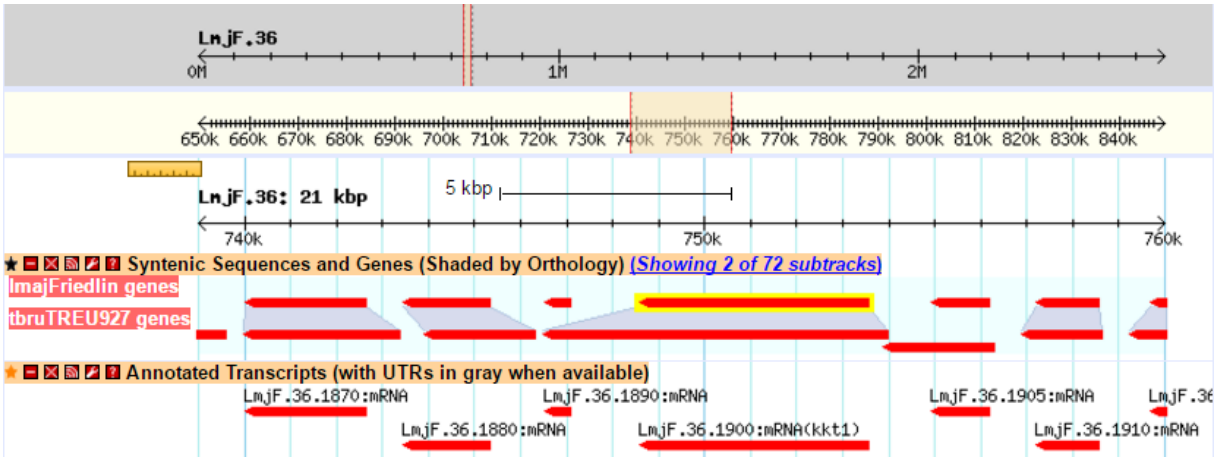
1. The authors express GFP-tagged LmKKT1 in Leishmania cells. One concerns that might arise is when introducing a gene, is that it creates a distinct phenotype. For instance, overexpression of the canonical kinetochore protein CENP-C in DT40 cells results in cytokinesis defects (Fukagawa 1999). The authors should discuss any GFP-tagged LmKKT1 induced phenotypes.

We have discussed this as a response to Major concern 1, here above.

2. The authors identify LmKKT1 to be the homologue of TbKKT1 based on 36% identity and 53% similarity. This finding would be further highlighted if these two proteins were aligned with a graphical representation of their relative homologue, including emphasizing the two conserved domains (DEAD box-helicase domain and TOG-domain). In Akiyoshi and Gull 2014 KKT1s were found in several other kinetoplastids. This would allow the authors to determine the rate of evolution of KKT1. Problems with identifying canonical kinetochore proteins is well known, as highlighted by Meraldi et al 2007 and various papers from the Henikoff lab, where putative orthologs are restricted to short sequences,

such as the CENP-C motif in CENP-C proteins. Maybe this problem also exist in KKTs if they were equally fast evolving.

We considered LmKKT1 (LmjF.36.1900) to be the homologue of TbKKT1 (Tb927.10.6330) because it is annotated as such in the database, and because they are syntenic (see screen shot from TriTrypDB below). We agree that the percentages of similarity and identity are rather low and by themselves do not allow to consider LmjF36.1900 to be the homolog of Tb927.10.6330. We have modified the manuscript accordingly.



Regarding the second part of the comment of referee #3, Referee #1 legitimately argued that the presence of the domains were findings from in silico search (HH pred), without any supporting data, and that we should therefore take out this analysis. We have followed his advice and remove the sequence domain features from the manuscript. As for the evolution rate of KKTs, although this is a highly interesting track to follow, we believe that it would be out of focus for the present paper.

3. In the discussion, the authors mention that transcription of centromere DNA is unexpected, but recently pervasive transcription happens at all centromeres, as has been shown by various labs. For instance, Athwal et al 2015 shows that ectopic CENP-A predominantly goes to transcription start sites. Molina et al 2016, Quenet & Dalal 2014, Koo et al 2016, Grenfell et al 2016, and Blower 2016 are just a few of the most recent papers describing centromeres being transcribed.

We agree with referee #3 and have modified the manuscript according to his comment.

4. In the discussion the authors mention that Leishmania diverged early in trypanosomatid evolution. This gives the appearance that Trypanosma is the base of the Trypanosomatida tree and that Leishmania is the derived branch. But the phylogenetic trees of kinetoplastids show a bifurcation of the branches that harbor either Lieshmania or Trypanosma.

In term of evolution of the trypanosomatids, the prevailing view is an early divergence of the Leishmania genus and the monophyly of the Trypanosoma genus. However, since our sentence was misunderstood by the referee, and since this concern was also raised by Referee #1 (minor comment 2), in view of the weakness of the data available, we preferred to erase this comment from the Discussion.

5. Centromere morphology is rather diverse. Budding yeast has a genetic centromere, as do its very close relatives, whereas most other eukaryotes have regional centromeres. These regional centromeres are commonly characterized by the presence of large arrays of tandem repeat sequences, but unique sequences are frequently found as well, such as in various fungi species, many chicken chromosomes, and incidental horse and orangutan chromosomes. Finally, there are the holocentric chromosomes, a feature that has evolved at least 15 times. Mosaic aneuploidy as found in Leishmania is a very interesting observation and might be the result of less than optimal functioning kinetochores. Disfunctioning canonical kinetochores do result in aneuploidy. Maybe the authors could discuss briefly the potential implications of their enrichment levels of LmKKT1 on Leishmania chromosomes as well as the presence of minor peaks in their ChIP-seq data. It is intriguing to contemplate the potential of a unstable Leishmania kinetochore permitting the rise of mosaic aneuploidy, maybe even with promiscuous kinetochore seeding on the chromosomes, as shown by the presence of the minor peaks in the ChIP-seq data

We agree that mosaic aneuploidy found in Leishmania is a very interesting observation and might be the result of what we called a "permissive segregation"; and we acknowledge the interest of the referee for seeding hypotheses from our data for this unique feature of Leishmania. However, our main finding is the presence of a single LmKKT binding site per chromosome, which does not allow us to speculate about less than optimal functioning kinetochores. Moreover, the best hypothesis re: mosaic aneuploidy in Leishmania is rather a defect in chromosomal replication allowing the gain or loss of chromosome copies at mitosis (Sterkers et al., Mol Microbiol 2012). As regards the secondary peaks, we do not consider there is sufficient evidence to consider them as potential "secondary" centromeres, therefore we fear that discussing about this would lead to over interpreting our data.

2nd Editorial Decision 21 July 2017

Thank you for the submission of your revised manuscript to our editorial offices. We have now received the reports from the referees that were asked to re-evaluate your study. As you will see, all three referees now support the publication of your manuscript in EMBO reports.

Before we can proceed with formal acceptance, I have the following editorial requests that need to be addressed in a final revised version:

Please have the manuscript revised by a native English speaker (as also indicated by referee #2).

For a Short Report, results and discussion must be combined into one section (Results and Discussion). Please do that for your manuscript text.

Please shorten the abstract to below 176 words and provide it written in present tense.

The supplementary material could be presented differently. As you have only 4 main figures, you could show some of these data as expanded view items. Expanded View items will be displayed in the main HTML of the paper in a collapsible format. You can submit up to 5 images or tables as Expanded View. Please follow the nomenclature Figure/Table EV1, Figure/Table EV2 etc. The figure legend for these should be included in the main manuscript document file in a section called Expanded View Figure Legends after the main Figure Legends section, and these items need to be uploaded as single files.

All additional Supplementary material should be supplied as a single pdf labeled Appendix. The Appendix includes a table of content on the first page, all figures and tables and their legends. Currently, you have uploaded 4 tables as Tables S2-S5 (but marked as data sets). If these are

supplementary tables, please move them to the Appendix. If these are data sets, please rename them, and put a legend on the first tab of the xls files. Or should some of these be shown as EV tables? The Appendix currently contains many items not labeled as Appendix Table or Appendix Figure (e.g. Supplemental material on LmjF and Supplemental material on retrotransposons). All these items need a nomenclature and a call out in the manuscript text. If these are figures or tables, please name them accordingly. If these are datasets (maybe all the sequence info?), please upload them as datasets as indicated above, and remove them from the Appendix.

Finally, after these changes, please update the callouts in the manuscript text, using our style, i.e. Figure/Table EVx, Appendix Figure/Table Sx, Dataset Sx, and delete the text about supplementary files on pages 21-22 of the manuscript.

It is not clear (to non-experts) what the grey scale inserts in Figures 1 and 4 are. Please describe these in the legend. In 1C, there seem to be two grey scale images overlapping. Please clearly separate them, and explain what they are.

Please also use the same format for all figure legends.

The text in Fig 2 and 3A is small and hard to read at 100%. Please use bigger fonts and arrange the figures in a more comprehensive way. Please refer to our guidelines for figure preparation.

I look forward to seeing the final revised version of your manuscript when it is ready. Please let me know if you have questions or comments regarding the revision.

REFEREE REPORTS

Referee #1:

In the revised manuscript, the authors fully addressed all of my previous concerns. I therefore support its rapid publication in EMBO Reports.

Referee #2:

This revised manuscript was an enjoyable read. The data were excellent and well presented, with the correct level of interpretation and discussion. The results from this work will be of considerable interest to the molecular parasitology community and should stimulate further research into the intriguing chromosome biology of Leishmania. I would recommend publication. One caveat: the manuscript would benefit from some final input to tidy up minor issues of English usage.

Referee #3:

The revisions performed by the authors of "Identification of the centromeres of Leishmania major: revealing the hidden pieces" are satisfactory. Publication of this manuscript will move the field of unconventional kinetochore forward.

2nd Revision - Authors' response 12 August 2017

Please find the final revised version of the manuscript. We have performed all the modifications that were asked; find below the point-by-point report:

Please have the manuscript revised by a native English speaker (as also indicated by referee #2).

DONE: Cameron Ross MacPherson, who is a native English speaker, has edited the manuscript.

For a Short Report, results and discussion must be combined into one section (Results and Discussion). Please do that for your manuscript text.

DONE

Please shorten the abstract to below 176 words and provide it written in present tense.

DONE

The supplementary material could be presented differently. As you have only 4 main figures, you could show some of these data as expanded view items. Expanded View items will be displayed in the main HTML of the paper in a collapsible format. You can submit up to 5 images or tables as Expanded View. Please follow the nomenclature Figure/Table EV1, Figure/Table EV2 etc. The figure legend for these should be included in the main manuscript document file in a section called Expanded View Figure Legends after the main Figure Legends section, and these items need to be uploaded as single files.

DONE

All additional supplementary material should be supplied as a single PDF labeled Appendix. The Appendix includes a table of content on the first page, all figures and tables and their legends. Currently, you have uploaded 4 tables as Tables S2-S5 (but marked as data sets). If these are supplementary tables, please move them to the Appendix. If these are data sets, please rename them, and put a legend on the first tab of the xls files. Or should some of these be shown as EV tables? The Appendix currently contains many items not labeled as Appendix Table or Appendix Figure (e.g. Supplemental material on LmjF and Supplemental material on retrotransposons). All these items need a nomenclature and a call out in the manuscript text. If these are figures or tables, please name them accordingly. If these are datasets (maybe all the sequence info?), please upload them as datasets as indicated above, and remove them from the Appendix.

DONE: The appendix was thoroughly modified

Finally, after these changes, please update the callouts in the manuscript text, using our style, i.e. Figure/Table EVx, Appendix Figure/Table Sx, Dataset Sx, and delete the text about supplementary files on pages 21-22 of the manuscript.

DONE

It is not clear (to non-experts) what the grey scale inserts in Figures 1 and 4 are. Please describe these in the legend.

DONE

In 1C, there seem to be two grey scale images overlapping. Please clearly separate them, and explain what they are.

DONE

Please also use the same format for all figure legends.

DONE

The text in Fig 2 and 3A is small and hard to read at 100%. Please use bigger fonts and arrange the figures in a more comprehensive way. Please refer to our guidelines for figure preparation.

DONE

3rd Editorial Decision 15 August 2017

Thank you for the submission of your revised manuscript to our editorial offices. Before we can proceed with formal acceptance, I have some final editorial requests [detailed in letter to authors].

3rd Revision - Authors' response 15 August 2017

The authors made the requested editorial changes and submitted their revised manuscript.

4th Editorial Decision -- Acceptance 28 August 2017

I am very pleased to accept your manuscript for publication in the next available issue of EMBO reports. Thank you for your contribution to our journal.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Yvon Sterkers
Journal Submitted to: EMBO reports
Manuscript Number: EMBOR-2017-44216

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com	Antibodypedia
http://1degreebio.org	1DegreeBio
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo	ARRIVE Guidelines
http://grants.nih.gov/grants/olaw/olaw.htm	NIH Guidelines in animal use
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm	MRC Guidelines on animal use
http://ClinicalTrials.gov	Clinical Trial registration
http://www.consort-statement.org	CONSORT Flow Diagram
http://www.consort-statement.org/checklists/view/32-consort/66-title	CONSORT Check List
http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun	REMARK Reporting Guidelines (marker prognostic studies)
http://datadryad.org	Dryad
http://figshare.com	Figshare
http://www.ncbi.nlm.nih.gov/gap	dbGAP
http://www.ebi.ac.uk/ega	EGA
http://biomodels.net/	Biomodels Database
http://biomodels.net/miriam/	MIRIAM Guidelines
http://jjj.biochem.sun.ac.za	JWS Online
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html	Biosecurity Documents from NIH
http://www.selectagents.gov/	List of Select Agents

B- Statistics and general methods

Please fill out these boxes (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	Deep sequencing was applied to INPUT and ChIP samples11/07/2017 Peaks were detected using MACS2 and signal was measured as the fold-change between ChIP and input using custom Python scripts.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	Not applicable
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	Not applicable
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	Not applicable
For animal studies, include a statement about randomization even if no randomization was used.	Not applicable
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	Not applicable
4.b. For animal studies, include a statement about blinding even if no blinding was done	Not applicable
5. For every figure, are statistical tests justified as appropriate?	Not applicable
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Not applicable
Is there an estimate of variation within each group of data?	Not applicable
Is the variance similar between the groups that are being statistically compared?	Not applicable

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	Antibodies used for ChIP-seq were from abcam (ab290), were ChIP-grade antibodies and according to the manufacturer has been referenced in 1126 publications, listed here: http://www.abcam.com/gfp-antibody-chip-grade-ab290-references.html)
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	L. major strain Friedlin promastigotes (MHOM/IL/81/Friedlin) comes from the CRB-leishmania, a culture collection belonging to the World Data Centre for Microorganisms de la World Federation for Culture Collection under the N° WDCM 879 Assays found no mycoplasma DNA in the cultures

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	Not applicable
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	Not applicable
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	Not applicable

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	Not applicable
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	Not applicable
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	Not applicable
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	Not applicable
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	Not applicable

16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	Not applicable
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	Not applicable

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	Done: Data Availability NGS sequence data are deposited in the European Nucleotide Archive (ENA) database under accession number PRJEB21722 and accessible through the following link: http://www.ebi.ac.uk/ena/data/view/PRJEB21722.
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	NGS sequence data are deposited in the European Nucleotide Archive (ENA) database under accession number PRJEB21722 and accessible through the following link: http://www.ebi.ac.uk/ena/data/view/PRJEB21722.
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	Not applicable
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	Not applicable

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	The study based on divergent eukaryoteparasite do not fall under use research restrictions