Systematic functional analysis of *Leishmania* protein kinases identifies regulators of differentiation or survival

Baker N.[1#], Catta-Preta C.M.C.[1#], Neish R.[1], Sadlova, J.[2], Powell B.[3], Alves-Ferreira E.V.C.[1], Geoghegan V.[1], Carnielli, J.B.T.[1], Newling K.[1], Hughes, C[1]., Vojtkova, B[2]. Anand, J.[1], Mihut, A.[1], Walrad, P.B.[1], Wilson, L.G.[4], Pitchford, J.W.[1,3], Volf, P.[2] and Mottram J.C. [1*]

[1] York Biomedical Research Institute and Department of Biology, University of York, Wentworth Way, Heslington, York YO10 5DD, UK

[2] Department of Parasitology, Faculty of Science, Charles University, Prague, Czech Republic

[3] Department of Mathematics, University of York, Wentworth Way, Heslington, York YO10 5DD, UK

[4] York Biomedical Research Institute and Department of Physics, University of York, Wentworth Way, Heslington, York YO10 5DD, UK

# Authors equally contributed to the work.
[*] Correspondence: jeremy.mottram@york.ac.uk

Summary

Differentiation between distinct stages is fundamental for the life cycle of intracellular protozoan parasites and for transmission between hosts, requiring stringent spatial and temporal regulation. Here we applied kinome-wide gene deletion and gene tagging in *Leishmania mexicana* promastigotes to define protein kinases with life cycle transition roles. Whilst 162 were dispensable, 44 protein kinase genes were refractory to deletion in promastigotes and are likely core genes required for parasite replication. Phenotyping of pooled gene deletion mutants using bar-seq and projection pursuit clustering revealed functional phenotypic groups of protein kinases involved in differentiation from metacyclic promastigote to amastigote, growth and survival in macrophages and mice, colonisation of the sand fly and motility. This unbiased interrogation of protein kinase function in *Leishmania* allows targeted investigation of organelle-associated signalling pathways required for successful intracellular parasitism.

Keywords
Leishmania, Protein kinase, CRISPR-Cas9, Differentiation

Introduction

The life cycles of many eukaryotic microbial pathogens are intricate due the requirement to transition between insect and mammalian hosts and cell type differentiation is central to their ability to adapt to different environments [1]. Some parasitic protozoa undergo cell cycle arrest in response to parasite-derived autocrine signals in their host. These arrested cells can undergo differentiation in response to environmental cues to produce a distinct cell type that can proliferate. Little is known about the molecular mechanisms that underpin these events, although recent advances have been made in the malaria parasite *Plasmodium* [2–4] and the African trypanosome *Trypanosoma brucei* implicating protein kinase signalling pathways [5,6]. As models to investigate life cycle progression *Plasmodium* is relevant for its mosquito transmission and intra-erythrocytic replication, whilst *Trypanosoma brucei* is pertinent for its transmission via the tsetse fly and extracellular replication. *Leishmania,* a kinetoplastid parasitic protozoa that is the causative agent of leishmaniasis, also has a complex life cycle with a requirement to replicate and differentiate in both mammalian and sand fly (Diptera: Psychodidae) hosts. The parasite is transmitted during the vector blood meal on an infected vertebrate with the ingestion of infected cells or amastigote forms that have reached the extracellular milieu in the mammalian host. In the vector midgut, the protozoa are enclosed in the peritrophic matrix, in contact with the blood to be digested, and differentiate into procyclic promastigotes. After blood digestion, the parasites attach to midgut epithelium, replicate, differentiate and migrate to reach the anterior part of the thoracic midgut and colonize the stomodeal valve. This localisation of parasites is a prerequisite to successful transmission [7]. The infective stage is the metacyclic promastigote, adapted to migration, transmission and invasion [8,9]. After entering the mammalian host *Leishmania* metacyclic promastigotes are taken up by phagocytes, where they survive and proliferate in the acidic environment of the parasitophorous vacuole. *In vitro* systems are available for most life cycle stages of *Leishmania*, although these models do not completely reflect the natural environment and may not activate some of the intrinsic pathways required for normal differentiation.

Phosphorylation-mediated signal transduction in eukaryotes is a key regulator of protein function and likely plays a pivotal role in *Leishmania* differentiation considering the changes reported for protein phosphorylation in different life cycle stages [10]. *Leishmania* has 195 eukaryotic protein kinases (ePKs) divided in six groups, classified according to catalytic domain conservation (CMGC, AGC, CAMK, STE, NEK, CK1), and an additional group here classified as 'Others', which contains diverse important eukaryotic cell signal transduction functions such as Aurora, Polo and AMP kinases. Atypical protein kinases (aPKs) lacking a conserved ePK structure but otherwise

catalytically active are present in *Leishmania* that harbors, for example, 11 phosphatidylinositol 3' kinase-related kinases (PIKK) resembling lipid kinases [11]. The kinome of *Leishmania* is greatly reduced in comparison to the human equivalent, lacking tyrosine kinases (TK) and tyrosine kinase-like (TKL) groups, similar to *T. brucei* and *Plasmodium* [12,13]. In contrast, CMGC, STE and NEK groups are significantly expanded in *Leishmania* relative to humans.

A number of studies have explored the role of individual *Leishmania* protein kinases in promastigote survival, metacyclic differentiation and amastigote replication, as well as their role in infectivity. Less than 10% of the kinome, mostly from the CMGC group, have been investigated using genetic and chemical approaches [14–31]. There are 17 mitogen activated protein kinases (MAPK) in *Leishmania* [32] and they were the focus of most of those studies, which highlighted the roles of MPK9 and MPK3 on flagellum maintenance, the latter being phosphorylated by MKK1 [15,16,24]. Another example is MPK2, previously demonstrated to modulate aquaglyceroporin 1 (AQP1) and amino acid transporter (APP3), proteins involved in drug resistance and the arginine depletion response, respectively, and therefore intra-macrophage amastigote survival [21,22]. Amastigote growth suppression and persistence was observed in TOR3 and MAPK4 null mutants [18,25], MPK10 has been implicated in amastigote differentiation, and CRK1 and CRK3 in cell cycle control and amastigote survival [14,27,33]. Only DYRK1 was shown to be essential for differentiation to metacyclic promastigotes [31], and there is a gap of knowledge on the role of protein kinases in differentiation from promastigotes to amastigotes and survival in the sand fly.

Here, we systematically tagged protein kinases with mNeonGreen fluorescent protein for localisation studies, and attempted to create individual null mutants for *Leishmania mexicana* ePKs and PIKK using CRISPR-Cas9. We performed pooled infections and assessed the relative proportion of barcodes representing 159 *Leishmania* protein kinase mutants to establish their importance in survival, differentiation and ultimately infection success *in vitro* and *in vivo* in the invertebrate and vertebrate hosts.

Results

Generation of gene deletion mutants identify non-essential protein kinases in promastigotes

Leishmania mexicana eukaryotic protein kinases (ePK) and PIKK family atypical protein kinases (aPKs) gene sequences were retrieved from TriTrypDB after searching for orthologues of *L. major* and *L. infantum* genes that had been identified previously [11] and by searching for protein kinase Pfam

domains. We investigated a total of 204 *L. mexicana* protein kinases, of which 193 were ePKs, distributed into 8 groups, and 11 of which were aPK PIKKs (Fig. 1a; Supplementary Table 1). We also investigated the regulatory subunits of AMPK (AMPK□ and AMPKγ). The *Leishmania* kinome encodes for orthologs and paralogs across a selection of trypanosomatids. 174 protein kinases have orthologues in trypanosomes and *Leishmania*, whilst 17 protein kinases are only found in the *Leishmaniinae* (termed *Leishmaniinae* unique kinases (LUKs). Only one protein kinase, LmxM.03.0780 was found in both *Leishmania* and *Endotrypanum*, but not *Leptomonas* or *Crithidia* (Fig. 1b).

The requirement for these protein kinases in promastigote survival was assessed by the attempted generation of gene deletion lines using CRISPR-Cas9. Briefly, two repair templates containing drug resistance markers (*BSD* and *PUR*) were amplified using long primers with 30bp homology arms containing a unique 12 bp barcode and two guides [34]. The repair templates and guides were transfected into *L. mexicana* procyclic promastigotes expressing Cas9 and T7 RNA polymerase (Fig. 2a). PCR amplification was used to determine the presence of the coding sequence (CDS) and integration of the drug resistance cassette (Fig. 2b, Supplementary Data 1). PCR confirmation of an exemplar gene deletion mutant, Δ*25.0853*, showed a PCR product for integration of the drug resistance cassette in the specific locus, and an absence of the CDS (Fig 2bi). In contrast, when a gene deletion mutant could not be generated, for example with *PKAC1,* either no populations were recovered after transfection, or cell lines had both a CDS and an integration PCR product (designated *PKAc1:: PKAC1,* Fig. 2bii). This may be due to gene or chromosome duplication, which is a common mechanism by which *Leishmania* retains an essential gene following attempts at generating gene deletion mutants [35,36]. Whole genome sequencing of 17 mutant cell lines (15 gene deletion mutants and 2 that retained an additional CDS) in comparison with Cas9T7 confirmed loss of reads over the CDS of interest and presence of resistance markers in this location (Fig. 2c). Of note, the copy number of some other protein kinase genes changed in the gene deletion mutants. For example, Δ*06.1220* has duplicated its genome and in Δ*20.1340* the copy number of *LmxM.29.0600* and *LmxM.07.0900* has increased; gene copy number changes could influence expression of one protein kinase as a way to compensate for loss of another. If a gene deletion mutant was unable to be identified after three transfection attempts, it was classed as 'required' with a 1 star quality classification [37]. Forty three protein kinases and AMPKγ fell into this category (Fig. 1a, 1c).

Gene deletion mutants were successfully produced for 161 protein kinases, equating to 79% of the kinome, and the AMPK□ regulatory subunit. Only 2/22 (9%) NEKs were found to be refractory to

gene deletion compared to 12/48 (25%) CMCGs and 4/11 AGCs (36%) (Fig.1c). Gene deletion mutants were obtained for all 11 pseudokinases that lack all three of the key conserved residues thought to be important for catalytic activity (Supplementary Fig. 1) [11], demonstrating no essential requirement in promastigotes. A further 40 protein kinases were found to lack either one or two of these residues and gene deletion mutants could be generated for 33/40 (82.5%). As previous work showed that it is possible to mutate the lysine residue and retain some ATP binding activity [38], we have classified these protein kinases as putatively active.

The essentiality of protein kinases was considered in relation to orthology across 10 trypanosomatid species. Of the 174 protein kinases found in all 10 of these species, 21% were found to be required in promastigotes. In contrast 41% of the protein kinases found only in the *Leishmaniinae* group (LUKs) were required (Fig. 1b). This suggests that *Leishmania* promastigotes require these LUKs specifically for their life cycle adaptations. Of note, two of the eleven non-essential pseudokinases are also LUKs.

Our gene deletion attempts indicate that 43 of the protein kinases are essential in promastigotes (Table 1). For additional validation, we generated gene deletions for both endogenous alleles of *LmxM.04.0650* whilst providing an extrachromosomal copy of the gene. Δ*04.0650* [*6xmyc-04.0650*] (Fig. 2d) retained the episome for at least seven generations without drug pressure. In some instances genes were refractory to the anticipated genetic manipulation and in the case of *CLK1* (*LmxM.09.0400*) and *CLK2* (*LmxM.09.0410*) this was found to be due to incorrect assembly of the reference genome. A recent nanopore genome allowed design of new primers and the successful generation of Δ*clk1* and Δ*clk2* gene deletion mutants, suggesting that disruption to one of the two *CLK* genes can be compensated for by the other (Supplementary Fig. 2).

Localisation of protein kinases

We generated 199 N- or C-terminal mNeonGreen tagged protein kinases for localisation studies in procyclic promastigotes (Fig. 3a). We used the atlas of *Leishmania* cellular landmarks [39] to localise proteins to the cytoplasm (48.5%), basal body (11.6%), nucleus (10.1%), endomembrane (7.1%), flagellum (6.5%), lysosome (6.5%), flagellar pocket (3%), pellicular membrane (2.5%), cytoplasmic organelles (2%) and mitochondrion (1%) (Fig. 3b,c, Table 1, Supplementary Data 1 and Supplementary Table 1). The fluorescence signal for some protein kinases varied during the cell cycle, for example PKAC1 (LmxM.34.4010) accumulates in the nucleus in early G1/S and at late G2/M in the cytoplasm and flagellum. This also holds true for MPK6 (LmxM.31.3250) which is

localised within the cytoplasm and flagellum in G1 but is located on the pellicular membrane during S phase. In other cases protein kinases were found in more than one localisation, for example LmxM.14.1070 was found in the basal body, flagellar pocket, lysosome and cytoplasm (Supplementary Data 2).

Protein kinases required for differentiation and survival in the mammalian host

In order to ascertain the role of protein kinases in differentiation across various life cycle stages, mutants were pooled and subjected to *Leishmania* life cycle progression. This pool (Pool 1) contained 159 barcoded mutants that included 152 protein kinase gene deletion mutants, five protein kinase mutants in which gene deletion mutants could not be generated ("required"), Δ*ampk* and one non-protein kinase control, metacaspase (Δ*mca*), which was included to provide comparator data with a previous peptidase bar-seq screen [36,40]. We tested the ability of mutants to transition through the *Leishmania* life cycle as procyclic promastigotes (PRO), metacyclic promastigotes (META), axenic amastigotes (AXA), amastigotes in macrophages (inMAC) and amastigotes in the footpads of mice (FP), taking samples at various time points for bar-seq analysis[36] (Fig 4a). The relative growth rate of each mutant was determined by counting barcodes represented in each time point and calculating the proportion of each mutant within the population. Log-proportions were plotted as trajectories across time points for each experiment. Outputs from experimental arms 1-3 were plotted for direct comparison against one another for six example genes that had no loss of fitness, an increased relative fitness or a decreased relative fitness (Fig. 4b). The eukaryotic initiation factor kinase mutant, Δ*eif2ak3* showed a strong increase in representation for axenic amastigotes, macrophage infection and infection of mice, but may not represent gain-of-fitness due to decrease in the relative abundance of mutants that have a loss of fitness. The mitogen-activated protein kinase kinase 3 (Δ*mpkk3*) and cdc2-related kinase 10 (Δ*crk10*) mutants maintained an average level of representation across all timepoints in all three experiments. This was in contrast to the mitogen-activated protein kinase gene deletion mutants Δ*mpk1* and Δ*mpk2,* which both demonstrated a loss of fitness in both axenic and intracellular amastigote stages, a phenotype that has been reported previously [21,24,41]. Our results agree with those findings and also showed that loss of fitness is seen in the mouse footpad infection. The zinc finger kinase mutant (Δ*zfk*) also displayed a loss of fitness across the three experimental arms, but with an increased representation upon enrichment of metacyclic promastigotes.

We examined each barcode proportion profile individually (Supplementary Data 3 c,f,i), but also opted to use an unbiased clustering algorithm to sort the mutants into groups with similar phenotypes, taking all time points into account. We used the projection pursuit method [42] to calculate the differences between the trajectories of logged-proportions, looking at both the first and second differences between each time point within the series, for each of the mutant strains. The projection pursuit clustering algorithm finds a 2-dimensional projection of the phenotype data in which clusters are as distinct as possible. This involves iterating between projection and clustering steps until the algorithm converges. The result is a 2-dimensional map in which whole phenotypes are represented by individual points. Clusters of points on this map were then scrutinized and checked for biological plausibility in a way that would be impractical in a higher-dimensional setting. This allowed qualitative similarities and differences between trajectories to be assessed in an unbiased way. After examining multiple solutions, we set our parameters to produce six clusters for each of the experimental arms (EA1-EA3) (Supplementary Table 2). Axenic amastigote trajectories allowed for identification of mutants where differentiation and replication were impaired, without interference by host factors (Supplementary Data 3). Macrophage infections *in vitro* were used for identification of mutants for genes important during infection, either differentiation and adaptation to the acidic environment of the phagolysosome or replication inside it. The last experimental setup for *in vivo* infections took into account the mouse immune system, able to eliminate mutants less able to survive in this environment. The mouse footpad infection experiment, examining the first differences in the series, was therefore chosen as an example of the cluster analyses (Fig. 4c). Cluster 1 (green) contains 12 trajectories for mutants that exhibited an increased representation during the mouse infection time points compared to others, indicating a relative gain of fitness in the population. Cluster 2 (red) contains trajectories with an average representation throughout promastigote stages and small increase between the metacyclic enrichment and three-week footpad infection stage. Cluster 3 (sky blue) contains the largest group of protein kinase mutants with fairly flat trajectories, decreasing slightly on average from the point of metacyclic enrichment. Cluster 4 (purple) contains protein kinase mutants that on average exhibit fairly flat trajectories throughout the promastigote stages, demonstrating average growth, with a decrease in fitness upon footpad infection. Cluster 5 (blue) contains mutants that show a significant increase in representation during metacyclic enrichment followed by a sharp decrease during footpad infection time points. Cluster 6 (grey) contains protein kinase mutants with trajectories which show a general loss of representation over time with a slight recovery during metacyclic enrichment. Each of the individual kinase trajectory examples (Fig 4b, Supplementary Data 3) are found in separate clusters and provide an example of the data for the axenic amastigotes and macrophage infections. The percentage barcode

representations for each of the mutants are also shown in a heat map format for each of the clusters, depicting the trends for individual mutants. PCA analysis supports the relationship between the different clusters (Fig. 4d, Supplementary Data 3).

Protein kinases important for differentiation from promastigotes to amastigotes

We took advantage of our three experimental arms for cross validation of mutants showing loss of fitness in amastigote infection, differentiation and/or replication. Clusters that identified protein kinases important for the amastigote stages were compared across the three arms for a first difference output from the projection pursuit analysis. Twenty nine protein kinase mutants were identified as important for amastigotes across at least two of the experimental arms (EA1-3) (Fig. 5a, Supplementary Table 2). These included MPK1 and MPK2 and six other protein kinase mutants that had been identified previously as being important for *in vivo* infection of *T. brucei* [43], as well as TOR3, previously identified as required for macrophage infections [25]. A further 15 protein kinases were identified with loss of fitness only in the mouse footpad infection (Supplementary Table 2), suggesting that these mutants can differentiate into amastigotes successfully and infect macrophages but cannot survive in the harsh *in vivo* environment.

We had identified previously a STE protein kinase in *T. brucei*, RDK1, as a repressor of bloodstream form differentiation [12] and the *L. mexicana* syntenic ortholog fell into cluster 3 (Fig. 4b). We analysed the *L. mexicana* Cas9 T7 progenitor line and Δ*rdk1* for life cycle progression and found that Δ*rdk1* had the same proliferation profile as the progenitor promastigotes and axenic amastigotes in culture (Supplementary Fig. 3a, b). In contrast, Δ*rdk1* had a loss of fitness of ~40% after 72 h in the *in vitro* macrophage assay and a loss of fitness in mice that was complemented with the add-back, Δ*rdk1::RDK1* (Supplementary Fig. 3c, d). These analyses confirmed the bar-seq results, providing confidence and validation of the pooled library data.

Protein kinases required for survival in the sand fly.

We next addressed the ability of gene deficient mutants to survive in the sand fly vector *Lutzomyia longipalpis*. We put together a new pool of gene deletion mutants (Pool 2), which were separated into three sub-pools based on their relative growth rates of slow, average or fast. Each sub-pool of ~50 procyclic promastigote mutants had four additional control lines [44], where a barcode was added to the ribosomal locus (Supplementary Table 1, Supplementary Fig. 4). These controls had been used for a previous bar-seq experiment [44] and provided a benchmark to compare sand fly infection experiments. *L. longipalpis* sand flies were allowed to feed on blood containing the three sub-pools,

then bar-seq analyses were carried out on DNA samples from day 0, 1, 5 and 8 post blood meal (Supplementary Fig. 4, Supplementary Table 2). Four mutants (Δ*mpk9*, Δ*31.1020*, Δ*19.1610* and Δ*19.0360*), had a significant loss of representation after 5 days of infection (p = <0.05), whilst fifteen protein kinase mutants had a significant loss of representation after 8 days of infection (p = <0.05) (Fig 5b, Supplementary Table 2). The four controls maintained an average ratio throughout, as expected [44] (Fig 5b, Supplementary Table 2).

The two *in vivo* screens provided a list of protein kinases important for successful infection of mice and colonisation of the sand fly midgut. Seven protein kinases were identified as being required in both screens, indicating that they play fundamental roles and are required for differentiation and/or survival in both the insect vector and mammalian host (Fig. 5b). This list includes MPK1, MPK2, CRK7 and LmxM.06.0640, a MAP3K. Eight protein kinases, including MPK9, ATM and PI4K were identified as important for colonisation of the sand fly only, as these cluster with average trajectories in the amastigote study (Fig. 5c). Sixteen protein kinases were identified as being required for differentiation and/or survival in the mammalian host only (Fig. 5d), as these show no significant decrease in representation by day 8 post blood meal in the sand fly. These include CMGC family protein kinases MAPK10, MAPK15 and CRK8, as well as PKAC3, PK4 and TOR3.

Protein kinases involved in flagellum function.

We hypothesised that protein kinase mutants that had defective motility as promastigotes would have a loss of fitness in the sand fly [44], but not in the mouse. To test this we initially screened Pool 1 for mutants that had defective motility in a transwell assay. We identified 9 mutants that were significantly different (p<0.05) between the upper and lower chambers after 4 hours equilibration (Supplementary Table 2). We validated the screen with four mutants, Δ*ulk4*, Δ*stk36*, Δ*02.0570* and Δ*29.0600*, which had cells with significantly shortened cell body: flagellum ratio and short flagella observed by microscopy (Fig. 6a, Supplementary Fig. 5, Supplementary Table 2). These mutants also had severely impaired motility, as assessed by 2D tracking (Fig. 6b, Supplementary Movie 1). The motile cells with normal length flagella present in these populations of Δ*ulk4* (~8%), Δ*stk36* (~3%) showed slow swim rates compared to the parental line (Supplementary Fig. 5). Re-expression of *ULK4* in Δ*ulk4* and *STK36* in Δ*stk36* restored the flagellum length to that of the parental line (Supplementary Fig. 6). ULK4 and STK36 share not only similar domain architecture with an N-terminal kinase domain and a number of ARM domains (Fig. 6c) but both are located at the flagellum and flagellar tip (Fig. 6d, Supplementary Fig. 6). Subsequently, we repeated the motility screen with Pool 2 and identified 20 protein kinases that were significantly different between the

transwell chambers ($p<0.05$), 6 of which had been identified in Pool 1 (Δ*stk36*, Δ*ulk4*, Δ*29.0600*, Δ*36.1520*, Δ*02.0570* and Δ*07.0900*) and one of which, Δ*mpk3*, had previously been reported to have very short flagella and reduced motility [16].

LmxM.02.0570 is one of only five *Leishmania* protein kinases with a predicted transmembrane domain and it was localised to both the lysosome and flagellum, a location found for just two other protein kinases in our dataset (LmxM.26.1730 and LmxM.36.1530). The trajectories for mouse infection for Δ*stk36* and Δ*ulk4* follow a similar trend showing increased representation after enrichment of metacyclic promastigotes and decreasing representation during mouse infection and hence group with protein kinases important for amastigote differentiation and survival in clusters 5 and 6 respectively (Fig. 6e). LmxM.29.0600 and LmxM.02.0570 on the other hand are found in clusters 1 and 2 respectively for the mouse footpad infection due to the former exhibiting a small increase in fitness and the latter a general but less significant decrease. Interestingly, STK36 and ULK4 were also identified as important protein kinases for survival in the sand fly, whereas MPK3, LmxM.29.0600 and LmxM.02.0570 were not (Fig. 6f). This suggests that these proteins are fundamental to infection for a reason independent of their short flagellum and provides evidence that not all mutants with flagella defects as promastigotes are hindered in colonisation of the sand fly.

Discussion

Protein kinases are fundamental to orchestrating the cellular signalling required for *Leishmania* survival throughout the life cycle. Here we have taken a holistic approach to exploring the role of each protein kinase in the parasite's life cycle, first assessing whether each gene is dispensable for growth as promastigotes and then using pooled library bar-seq phenotyping to assess the ability of gene deletion mutants to progress through the life cycle. We also determined the location of each protein kinase in promastigote stage cells, allowing generation of hypotheses for potential signalling pathways associated with specific organelles. Gene deletion mutants were successfully generated for 161 protein kinases in promastigotes, equating to 79% of the kinome, indicating a high level of functional redundancy. We were unable to generate gene deletion mutants for 43 protein kinases, which are likely to be essential for parasite proliferation. No technical difficulties could be identified with the CRISPR system, as the same guide RNAs and selection methods were used to successfully generate nNeonGreen tagged cell lines. This result is consistent with the level of essentiality for the kinome also seen using RNAi in bloodstream form trypanosomes [12]. In agreement with previous studies, we were unable to generate gene deletion mutants for six protein kinases previously characterised as being essential for promastigotes (CRK1, CRK3, CRK12, MPK4, TOR1 and TOR2)

[17,25,33,45,46]. We were however able to successfully generate a gene deletion mutant for LmxM.01.0750 and LmxM.32.1710, previously published as essential in *L. major* and *L. infantum* promastigotes respectively [20,26] and failed to isolate a gene deletion mutant for DYRK1, despite this being achieved in *L. infantum* [31]. These may be due to different gene deletion approaches taken allowing time for compensatory mutations or differences between the species. When we consider essentiality within the different protein kinase families, the AGC and CMGC kinases host a high proportion of essential genes. This is perhaps anticipated given that AGC protein kinases mediate diverse and vital cellular functions in eukaryotes and mutations in AGC protein kinases are associated with human disease such as cancer [47]. In addition, CMGC protein kinases include the cyclin dependent kinases and MAP kinases, which are fundamental to cell cycle progression [48]. In comparison the NEK protein kinases, which have an expanded family in *Leishmania* [11] are relatively dispensable in *Leishmania* promastigotes with 18/20 gene deletion mutants successfully generated. We noted that protein kinases located in some organelles were more likely to have essential functions than in others, for example, 9/20 (45%) in the nucleus were found to be essential; many of these are CDKs (CRK9, CRK11, CRK12) or other protein kinases likely to be involved in essential nuclear functions such as cell cycle control and transcription (KKT2, KKT3, AUK1, TLK, PKAC1, LmxM.24.0670)[12,49,50]. In contrast, as the flagellum is not essential for promastigote growth *in vitro* [16,44], it is not surprising that only 2/13 (15%) protein kinases found in the flagellum are essential. The basal body is another organelle whose correct duplication and segregation is crucial for cell division [51,52] yet, surprisingly, only 2/23 (9%) of protein kinases localised to the basal body are essential (PLK and LmxM.14.0830).

In order to pool our library of protein kinase gene deletion mutants, and test requirements through life cycle stages, we utilised bar-seq, a powerful technique previously used for genome-wide screens in multiple human cell types, as well as parasites such as *Toxoplasma* and *Plasmodium* [53,54,55]. This approach has now been applied to a small group of *Leishmania* flagellar proteins [44], ubiquitination pathway enzymes [36,40] and the protein kinases of this study. Whilst the plot of barcode proportions as trajectories across experimental time points provided information on life cycle progression for individual mutants, projection pursuit cluster analyses provided a non-biased way to group the experimental outcomes into clusters of mutants with comparable trends. The largest proportion of mutants were clustered with flat trajectories, indicating no change in fitness and whilst it may be that many of these protein kinases are not important for differentiation or survival, other possible explanations include redundancy of protein kinases regulating the same pathways or their roles in other mechanisms are not related to differentiation, replication and survival. For example, Δ*clk1* and

Δ*clk2* promastigotes are viable, however, double *CLK1*/*CLK2* gene deletion mutants could not be isolated, indicating that they might have redundant functions in regulating the kinetochore, as has been described in trypanosomes [49,56]. Five protein kinases have signal peptides and two of the casein kinase family, CK1.4 and CK1.2 are secreted [29,57] and potential virulence factors, however, they are essential in promastigotes and therefore were not present in the pooled libraries. The loss of fitness phenotype for gene deletion mutants of secreted protein kinases may also be masked if they are secreted by other mutants in the pool, still influencing the host-parasite interaction and immune response.

Successful infection of the mouse was the most challenging phenotype assessed for the gene deletion mutants, as the cells required the ability to differentiate from promastigote to amastigote, infect and survive within the macrophage, as well as withstand the host immune response. Using cluster analysis we identified 29 genes required across at least two of the differentiation specific experimental arms and a further 15 only required for infection of the mouse. It is worth noting that the screen does not distinguish between those mutants that have a significant loss of fitness during differentiation itself and those that differentiate normally and then have a loss of fitness growing as amastigotes. Using these data protein kinases were identified that are involved in extrinsic pathways activated in the parasite by external factors (e.g. from macrophage or sand fly) or intrinsic pathways activated by internal factors (e.g. state of the cell cycle), independent of external factors. For example, CRK4 is required for infection of the mouse, but is not required for axenic amastigote growth, suggesting an extrinsic role. On the other hand, MPK10 clusters as important across all three experimental arms implying an intrinsic requirement, supported by research showing it to have stage-specific activity important in promastigote and amastigote differentiation [14]. A careful consideration is required when interpreting the data from enriched metacyclic promastigote cells, where a cluster of protein kinases was identified that have an increased trajectory, yet has a loss of fitness during infection of macrophages or mice. Whilst some mutants may become *bona fide* metacyclic promastigotes yet not infect macrophages, it is also possible that enrichment occurred for repressors of differentiation mutants, which have made an early transition towards amastigote morphology and alternative life cycle stages such as haptomonads. An example is *RDK1,* a regulator of amastigote to promastigote differentiation, and a homologue of *T. brucei* RDK1, which has previously been reported to be involved in differentiation of the bloodstream to procyclic form [12]. Additionally, a mutant that had an impaired proliferation rate as a promastigote would not reach stationary phase by day 7, leading to fewer infective metacyclic forms present in the pooled culture. AMPKα is an example of this, with a consistent loss of fitness throughout the procyclic promastigote

stages and metacyclic enrichment, leading to poor representation in macrophage and mouse infections. Phenotyping of selected individual mutants, rather than a pool, will allow more detailed analyses.

Fifteen protein kinases were identified as being required for successful colonisation of the sand fly midgut. Development of *Leishmania* in the sand fly is far more complex than that observed in in vitro tissue culture and involves sequential differentiation from procyclic promastigote to the nectomonad form and then to the leptomonad form. At day 1 post-infection procyclic forms are numerous within the bloodmeal surrounded by the peritrophic matrix and by day 3 long nectomonads prevail; these escape from the ruptured peritrophic sac and attach to the midgut epithelium to prevent defecation with the remains of the blood meal [58]. In the next few days, leptomonad promastigotes colonize the anterior part of the midgut and differentiate into non-dividing metacyclic promastigotes in preparation for transmission to a mammalian host. The presence of these highly infective metacyclic parasites, together with colonization of the stomodeal valve, is typical for fully developed late-stage infections at day 8 [9]. Further investigation will be required to determine the exact roles that these fifteen protein kinases play a role in these differentiation, migration and establishment events. Of note, seven of the fifteen protein kinases required for establishment of infection in the sand fly were also required in the amastigote stage, suggesting a stage-independent role in growth and differentiation. LmxM.25.1560, a cytoplasmic CMGC kinase, is notable in being both a likely pseudokinase and the only *Leishmaniinae* unique kinase (LUK) where the mutant had a phenotype. Eight protein kinases, including the aPK ATM (LmxM.02.0120), recently associated with a novel DNA damage-induced checkpoint in *Leishmania* [59], were identified as dispensable for full life cycle progression in the mouse but found to be important for development in the sand fly.

The transwell screens identified a number of protein kinase mutants with potential motility defects, including Δ*mpk3,* but surprisingly not Δ*mkk1;* both of these mutants had previously been reported to have reduced motility [15,16]. We focused on four mutants that had both a short flagellum and severely defective motility, allowing a comparison between the phenotypes *in vitro* and in the sand fly. Motility is required for survival in the sand fly [44] and indeed we found that Δ*ulk4* and Δ*stk36* had loss of fitness for sand fly colonisation, as anticipated. In contrast, Δ*29.0600* and Δ*02.0570* were non-motile yet were able to colonise the sand fly and did not diminish even after defecation of the blood meal. Further investigation will be required to determine if the short flagellum provides sufficient binding capacity for attachment to epithelium and colonisation of the midgut. The data on ULK4 and STK36 provide an example of how localisation and life cycle progression data can be

used for signalling pathway hypothesis generation. ULK4 and STK36 have very similar infection profiles, localisation and their orthologues were also identified in a flagella attachment proteome in *T. brucei* [60] providing further evidence that they act together in *Leishmania* and other trypanosomatids.

In summary, we have identified fifteen protein kinases that are required for colonisation of the sand fly, eight of which play an intrinsic role and seven of which are extrinsic protein kinases required for survival in the sand fly only. We also identified twenty nine protein kinases that are required for successful differentiation from metacyclic promastigote to amastigote and survival in macrophages and the mouse (Fig. 7). These include 5 protein kinases potentially involved in MAP kinase signalling (one of which, MPK2 plays a role in the arginine depletion response [21,22]), 2 cyclin-dependent kinases, 2 casein kinase II and 13 protein kinases that have not previously been studied. Thus, a large databank of information is now available for each of the protein kinases that make up the *Leishmania* kinome, providing an opportunity to decipher signalling pathways and link protein kinases involved in regulating important aspects of the parasite's biology such as replication, differentiation and responses to stress. Perhaps more importantly, these data present an opportunity to characterise and focus experimental effort on those pathways which are divergent to humans, and which potentially can be targeted in new drug and vaccine therapies.

Methods

Cell culture *Leishmania mexicana* (MNYC/BZ/62/M379) promastigotes were routinely cultivated in HOMEM medium (modified Eagle's medium) supplemented with 10% (v/v) heat-inactivated fetal calf serum (HIFCS – Gibco) and 1% Penicillin/Streptomycin solution (Sigma-Aldrich) at 25°C. For differentiation assays promastigotes were grown in Grace's insect medium (Sigma-Aldrich) with 10% HIFCS (Gibco), 1% Penicillin/Streptomycin solution (Sigma-Aldrich) and 1% BME vitamins (Sigma), pH5.5 at 25°C. Parasites were subcultured at 1.5 x 10^5 parasites mL^{-1}, reaching mid-log phase at 72 h with approximately 5 x 10^6 parasites mL^{-1}, and stationary-phase at 168 h with approximately 2 x 10^7 parasites mL^{-1}. Axenic amastigote differentiation and proliferation was achieved by incubating log-phase promastigotes in Schneider's *Drosophila* Medium (Sigma-Aldrich) supplemented with 150 μg mL^{-1} of Hemin and 20% HIFCS at the concentration of 1 x 10^6 parasites mL^{-1} for 120 h at 35°C, 5% CO_2. Axenic amastigotes were subcultured by transferring cells into fresh medium at the same initial concentration. Bone marrow-derived macrophages (BMDM) were differentiated using committed myeloid progenitors isolated from BALB/c mice and incubated with macrophage colony-stimulating factor secreted by L929 cells as previously described [61]. Macrophages were infected with metacyclic promastigotes isolated by centrifugation over Ficoll 400 (Sigma-Aldrich) gradient [62] and quantified on a hemocytometer. Metacyclic promastigotes were incubated with BMDM (1:1) for 4 h and washed before incubation at 35°C and 5% CO_2 for 12h and 72 h in RPMI supplemented with 10% HIFCS and 1% Penicillin/Streptomycin solution.

Generation of the CRISPR-Cas9 mutants and bar-seq screens The protein kinase gene deletion library and N- or C-terminal tagged cell lines were generated and validated as described [36] using http://leishgedit.net/ to design primers in the *L. mexicana* Cas9 T7 line [34]. Primers used to generate sgRNA expression cassettes and donor cassettes, as well as the ones used for diagnostic PCRs are available in Supplemental table 1.

Life cycle phenotyping of gene deletion mutants *in vitro* and *in vivo* was carried out as described [36] but with the following modifications. Mid-log phase promastigote mutants (n=159, Pool 1) were grown separately, mixed in equal proportions to give six independent replicates (4 x 10^4 mL^{-1} for each cell line), and then placed in Grace's media supplemented with 20% of FCS, 1% Penicillin/Streptomycin solution and 1% BME vitamins, pH5.5 at 25°C. Samples were collected from each of the six replicates at 0 h, 24 h, 48 h, 72 h and 168 h. Axenic amastigotes were differentiated from 168 h promastigote cultures and grown for 120 h. Metacyclic promastigotes from six independent 168 h promastigote cultures were purified from Ficoll 400 gradients and samples were collected for time point 0 h. These purified metacyclic promastigotes were used for infection of BMDM at a 1:1 ratio (samples collected at 12 h and 72 h) and 12 mice were also infected with 10^6 parasites in the left footpad (two animals for each replicate, and time points collected at 3 and 6 weeks). DNA was extracted using DNeasy Blood & Tissue (Quiagen) following manufacturer's instructions for animal cells (promastigotes, axenic amastigotes and macrophage infections) or animal tissue (mice lesions).

For *Lutzomyia longipalpis* infection, promastigote growth data collected in experimental arms 1-3 at 72 h was ranked according to barcode representation and split into Fast, Average and Slow groups accordingly (Supplementary Table 1). Three sub-pools (slow, ave and fast) of protein kinase mutants plus controls were mixed in equal proportions using mid-log procyclic promastigotes (n = 57, 57 and 56, respectively). For this experiment (Pool 2), three heterozygous protein kinase mutants were removed and an additional five protein kinase gene deletion mutants that had not been included in the mammalian infection screen were added (Δ*03.0210*, Δ*29.3580*, Δ*29.3050*, Δ*20.1340* and Δ*36.2630*; Supplementary Table 1). The pool final density was counted and 50 ml cultures were set up at 2 x 10^5 cells ml^{-1} in HOMEM media. At 72 h, cells were washed three times in PBS and resuspended in defibrinated heat-inactivated sheep blood at a concentration of 2 x 10^7 promastigotes mL^{-1} [63]. The *L. longipalpis* colony was maintained at 26°C and high humidity on 50% sucrose solution and a 12 h light/12 h dark photoperiod, as described previously [64]. The three sub-pools were each fed to a cage of 200 female flies between 3-5 d old, through a chick skin membrane. It was assumed that each fly could take ~300 of each mutant in the sub-pool. DNA samples were prepared from the pools prior to sand fly feeding (72 h after pooling, referred to as Time 0) and also from duplicate sets of 25 whole sand flies taken at 1, 5 and 8 d post blood meal. DNA was extracted using a Tissue Lysis Buffer and PCR template preparation kit (Roche).

Preparation of DNA, PCR and sequencing of barcodes was performed as described previously [36]. Amplification of the barcodes was carried out using F primer OL8684 (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAgatgatgattacTAATACGACTCACTATA ACTGGAAG) and R primer OL11799 (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACACTCGTTTTCATCCGgcag). Q5 polymerase was used to amplify the barcodes in a 20 ul reaction containing 100 ng of DNA, with 28 cycles of amplification, 60°C annealing temperature and 10 s extension time. Nextera adapters were then added to the resulting DNA before being analysed on an Illumina HiSeq 3000. A Python script was used to search each Illumina read for the 12bp sequence preceding each barcode (Supplemental code 1). Total counts for each unique barcode are provided in Supplemental table 2.

Animal work All experiments were conducted according to the Animals (Scientific Procedures) Act of 1986, United Kingdom, and had approval from the University of York Animal Welfare and Ethical Review Body (AWERB) committee. To assess virulence of the *Leishmania mexicana* Cas9 T7 progenitor line, Δ*rdk1* mutant and an add-back Δ*rdk1::RDK1* groups of 5 female BALB/c mice (4-6 weeks) were infected subcutaneously at the left footpad with 1 x 10^6 purified metacyclic promastigotes. Animals were culled after 6 weeks and the parasite burden determined with a limiting dilution assay essentially as described previously [65]. Footpads were digested with 4 mg ml^{-1} collagenase D for 2 h at 37°C before dissociation through 70 µm cell strainers (BD Biosciences). Homogenates were resuspended in HOMEM supplemented with 20% FCS and serial dilutions (2-fold) performed in 96-well clear flat-bottom plates. Dilutions were performed in duplicate, distributed in at least 4 plates and incubated for 7-10 days at 25°C. Wells were analysed for the presence of parasites, and number of parasites calculated by multiplying the dilution factors. Results are shown as total parasites per tissue. Two-tailed unpaired *t*-test was performed as a statistical analysis.

Bar-seq statistical analyses For the purposes of clustering, we characterised mutant phenotypes in terms of the changes over time of the logged-proportion of each mutant subpopulation relative to the total population. For each mutant these changes constitute a time series to which we apply a differencing operator that removes the time series' average heights and slopes so that attention is instead focused on their curvature. The differenced times series are then treated as high-dimensional Euclidean vectors. These vectors are clustered using a Projection Pursuit Clustering algorithm, adapted from [42] and coded up in R (R: The R Project for Statistical Computing, n.d.). The algorithm, for which pseudo-code is provided in Supplementary Equation 1, involves iteratively projecting and re-clustering the vectors. This method is used to find a plottable two dimensional space in which the phenotype vectors are maximally separated. Once projected into this space, proposed clustering solutions for the vectors can be scrutinised and checked against prior knowledge of genetic characteristics. Such scrutinising led to the decision to choose a 6-cluster solution from the algorithm, whose output is illustrated in Supplementary Data 3.

Whole genome sequencing of protein kinase mutants Sequence reads generated from an Illumina HiSeq 3000 were trimmed with Cutadapt version 2.5 [66] and mapped to the *L. mexicana* Cas9 T7 genome assembly [67] and resistance cassettes with BWA-MEM algorithm as part of bwa version 0.7.17 [68]. Mosdepth 0.2.6 [69] was used to calculate the median coverage of each chromosome and scaffold in the assembly, as well as the mean coverage of each gene of interest. The ratio of gene coverage to chromosome coverage was used to assess gene deletion.

DNA cloning Expression of protein kinases either for facilitated gene deletion mutant or addback generation was achieved by using the pNUS system [70] modified for expression with a C-terminal 6xMYC fusion. Gibson Assembly (NEBuilder® HiFi DNA Assembly) was used. The pNUS backbone plasmid was digested using AvrII and gel extracted (Qiagen). The required CDS was amplified from *L. mexicana* genomic DNA using primers designed on NEBuilder assembly online tool, with a 5' 20 nt overlap with the pNUS backbone. Fragments were purified (Qiagen) and Gibson Assembly was performed as suggested by the manufacturer and transformed into *E. coli* DH5α chemo competent.

Live cell imaging Endogenous tagged cell lines in log-phase were centrifuged at 1000 x *g* for 10 min, washed in phosphate saline buffer pH 7.2 (PBS) and re-suspended in 10 µg mL^{-1} of Hoechst 33342 (Thermo Scientific) before incubation for 15 min at room temperature protected from light. To remove Hoescht 33342 the cells were washed with PBS, re-suspended and immobilised in 50 µL of ice-chilled CyGEL™ (Biostatus). Samples were transferred to slides, covered with a coverslip and

immediately imaged using a Zeiss LSM880 confocal microscope, with 488 and 405 nm lasers, at room temperature. Images were processed using Zen Black (Zeiss).

Western blotting Cells were re-suspended in NuPAGE sample buffer supplemented with 50 mM of dithiothreitol (DTT) and loaded on 4-12% NuPAGE Bis-Tris Protein Gels and electrophoresed at 180V for 1 h. Gels were transferred to Nitrocellulose membranes using wet transfer system XCell II Blot Module (Invitrogen) at a constant amperage of 350 mA for 2 h. Membranes were blocked with 5% skimmed milk in Tris Buffered saline-Tween for 1 h before incubation with primary antibody anti-myc tag monoclonal clone 4A6 (Millipore) diluted to final concentration of 200 ng mL^{-1} in 3% skimmed milk in TBST for 1h at room temperature. After washing membranes proceeded to incubation with secondary antibody anti-mouse IgG (H+L) conjugated to HRP (Molecular Probes) at 300 ng mL^{-1} in 3% skimmed milk in TBST for 1h at room temperature. Membranes were washed, incubated with Clarity Max substrate (BioRad) and developed in ChemiDoc MP (BioRad).

Motility screen and assays Six replicates were prepared for Pool 1. Cultures were grown in HOMEM media for 12 h in 20ml culture. A 2ml culture was added to the bottom of each well of a 8µm transwell plate (Costar #3428 24mm Diameter insert, 8.0 µm Pore size 6 well plate) and additional media was added to the top of each well before incubating at 25°C for 4 h to allow motile cells to cross the membrane and equilibrate on both sides. The culture was collected from the top and bottom of each well. Cells were pelleted, DNA extracted and bar-seq analyses completed. Statistical analyses included two-tailed unpaired t- test and Discovery using the Two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli, with Q = 50%. Each row was analysed individually, without assuming a consistent SD. Outputs with a p value <0.005 were chosen for further phenotypic characterisation. This experiment was repeated with Pool 2, using 2 replicate wells.

Motility mutants were grown individually to a mid-log culture. Cells were fixed in a final concentration of 1% paraformaldehyde (Alfa Aesar 16% W/V #43368) for 15 min. Fixed cells were washed in PBS by pelleting at 1000 x g for 1 min. The remaining pellet was gently re-suspended in 20 □l PBS. 10 µl of suspension was added to a glass slide (Thermo Superfrost plus #J1800AMNZ) with 5 µl of mounting solution and coverslip. Cells were imaged on an inverted Zeiss AxioObserver with a 63x lens. Multiple FOV images were taken for each line. Cell morphology measurements were made using Zen-2.6 (blue-edition). For consistency, cell width was always measured across the nucleus, while flagellum length was measured from the tip to the edge of the cell body. Statistical analyses included two-tailed unpaired parametric t- test with Welch's correction. Each row was analysed individually, without assuming a consistent SD.

For tracking, cells were loaded into glass sample chambers measuring approximately 5mm × 20mm × 0.5 mm constructed from glass slides and UV-curing glue. The samples were imaged under dark field illumination on a Nikon Eclipse E600 upright microscope using a ×10 magnification objective lens with a numerical aperture of 0.3. Five video microscopy sequences were acquired from different areas of the sample chamber for each mutant using a CMOS camera (Mikrotron MC-1362). Movies were acquired at 50 Hz for 40 s, at a resolution of 1024 × 1024 pixels, corresponding to a field of view of approximately 1.4 mm× 1.4 mm within which cells were tracked. Minimum images were constructed, composed of the lowest values at each pixel in each video sequence. A rolling ball filter with a radius of 30 pixels was applied to each minimum image before using them to correct the video sequences and remove static background features. The corrected movies were then thresholded to highlight cells moving by either swimming or Brownian motion. The cell coordinates in each frame were extracted, and these coordinates were assembled into cell tracks using custom-written software routines [71]. To smooth short-time fluctuations associated with camera noise, the cell tracks were fitted using piecewise cubic splines. This allowed reliably extracted instantaneous cell velocities.

This technique also allows separation of motile and non-motile cells based on their mean-squared displacements. The spline-smoothed trajectory of a non-motile particle has a mean-squared displacement that scales with time, whereas a swimming cell's mean squared displacement scales with time squared. By setting a threshold for the slope of each cell's mean-squared displacement at short times, and a threshold for the total mean squared displacement after 2 s, motile and non-motile particles were effectively separated and the motile fraction estimated. The instantaneous speed of each swimming cell was averaged over the course of the cell track, and a distribution of average swimming speeds made.

Author Contribution. J.C.M. conceived the study. N.B., C.C.M.C-P., R.N., E.V.C.A.F., V.G., J.B.T.C made the kinome libraries. N.B., C.C.M.C-P., R.N. A.J. carried out the bar-seq screens. B.P. carried out statistical analysis. N.B., C.C.M.C-P., B.P., K.N., J.W.P analysed data. J.S., B.V., P.V did sand fly experiments. P.B.W and L.G.W did motility analysis. N.B., C.C.M.C-P. R.N.,V.G., A.M. did phenotype analysis. N.B, C.C.M.C-P., J.C.M prepared and wrote the manuscript. All authors reviewed, edited and approved the paper. J.C.M. obtained funding.

Acknowledgments. This work was supported by the Wellcome Trust (200807) to JCM and the EPSRC (EP/N014731/1) to LGW. The research has, in part, received funding from the Research Council United Kingdom Grand Challenges Research Funder under grant agreement 'A Global Network for Neglected Tropical Diseases' grant number MR/P027989/1. JS, BV and PV were supported by ERD Funds, project CePaViP (CZ.02.1.01/0.0/0.0/16_019/0000759). This project has also received resources funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 731060 (Infravec2). We thank Peter Ashton, Sally James, Lesley Gilbert, Karen Hogg, Graeme Park, Karen Hodgkinson and Sarah Forrester for technical advice and support within the Bioscience Technology Facility, University of York.

Declaration of Interests. The authors declare no conflict of interest

Data availability. Whole genome sequencing data for *L. mexicana* Cas9T7 and gene deletion mutants are available from the European Nucleotide Archive under study accession PRJEB27113. All other data generated or analysed during this study are included in this published article (and its supplementary information files).

Code Availability. All code used during this study is included in this published article (and its supplementary information files).

Bibliography

1. Matthews, K. R. Controlling and coordinating development in vector-transmitted parasites. *Science* **331**, 1149–1153 (2011).
2. Tewari, R. *et al.* The systematic functional analysis of Plasmodium protein kinases identifies essential regulators of mosquito transmission. *Cell Host Microbe* **8**, 377–387 (2010).
3. Solyakov, L. *et al.* Global kinomic and phospho-proteomic analyses of the human malaria parasite Plasmodium falciparum. *Nat. Commun.* **2**, 565 (2011).

4. Stanway, R. R. *et al.* Genome-Scale Identification of Essential Metabolic Processes for Targeting the Plasmodium Liver Stage. *Cell* **179**, 1112–1128.e26 (2019).

5. Mony, B. M. *et al.* Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. *Nature* **505**, 681–685 (2014).

6. Cayla, M., McDonald, L., MacGregor, P. & Matthews, K. An atypical DYRK kinase connects quorum-sensing with posttranscriptional gene regulation in Trypanosoma brucei. *Elife* **9**, (2020).

7. Volf, P., Hajmova, M., Sadlova, J. & Votypka, J. Blocked stomodeal valve of the insect vector: similar mechanism of transmission in two trypanosomatid models. *Int. J. Parasitol.* **34**, 1221–1227 (2004).

8. Walters, L. L. Leishmania Differentiation in Natural and Unnatural Sand Fly Hosts 1. *J. Eukaryot. Microbiol.* **40**, 196–206 (1993).

9. Dostálová, A. & Volf, P. Leishmania development in sand flies: parasite-vector interactions overview. *Parasit. Vectors* **5**, 276 (2012).

10. Tsigankov, P., Gherardini, P. F., Helmer-Citterich, M., Späth, G. F. & Zilberstein, D. Phosphoproteomic analysis of differentiating Leishmania parasites reveals a unique stage-specific phosphorylation motif. *J. Proteome Res.* **12**, 3405–3412 (2013).

11. Parsons, M., Worthey, E. A., Ward, P. N. & Mottram, J. C. Comparative analysis of the kinomes of three pathogenic trypanosomatids: Leishmania major, Trypanosoma brucei and Trypanosoma cruzi. *BMC Genomics* **6**, 127 (2005).

12. Jones, N. G. *et al.* Regulators of Trypanosoma brucei cell cycle progression and differentiation identified using a kinome-wide RNAi screen. *PLoS Pathog.* **10**, e1003886 (2014).

13. Ward, P., Equinet, L., Packer, J. & Doerig, C. Protein kinases of the human malaria parasite Plasmodium falciparum: the kinome of a divergent eukaryote. *BMC Genomics* **5**, 79 (2004).

14. Cayla, M. *et al.* Transgenic analysis of the Leishmania MAP kinase MPK10 reveals an auto-inhibitory mechanism crucial for stage-regulated activity and parasite viability. *PLoS Pathog.* **10**, e1004347 (2014).

15. Wiese, M., Kuhn, D. & Grünfelder, C. G. Protein kinase involved in flagellar-length control. *Eukaryot. Cell* **2**, 769–777 (2003).

16. Erdmann, M., Scholz, A., Melzer, I. M., Schmetz, C. & Wiese, M. Interacting protein kinases involved in the regulation of flagellar length. *Mol. Biol. Cell* **17**, 2035–2045 (2006).

17. Dacher, M. *et al.* Probing druggability and biological function of essential proteins in Leishmania combining facilitated null mutant and plasmid shuffle analyses. *Mol. Microbiol.* **93**, 146–166 (2014).
18. Wang, Q., Melzer, I. M., Kruse, M., Sander-Juelch, C. & Wiese, M. LmxMPK4, a mitogen-activated protein (MAP) kinase homologue essential for promastigotes and amastigotes of Leishmania mexicana. *Kinetoplastid Biol. Dis.* **4**, 6 (2005).
19. Martin, J. L. *et al.* Metabolic reprogramming during purine stress in the protozoan pathogen Leishmania donovani. *PLoS Pathog.* **10**, e1003938 (2014).
20. Martínez-Calvillo, S., Stuart, K. & Myler, P. J. Ploidy changes associated with disruption of two adjacent genes on Leishmania major chromosome 1. *Int. J. Parasitol.* **35**, 419–429 (2005).
21. Mandal, G. *et al.* Modulation of Leishmania major aquaglyceroporin activity by a mitogen-activated protein kinase. *Mol. Microbiol.* **85**, 1204–1218 (2012).
22. Goldman-Pinkovich, A. *et al.* An Arginine Deprivation Response Pathway Is Induced in Leishmania during Macrophage Invasion. *PLoS Pathog.* **12**, e1005494 (2016).
23. Kuhn, D. & Wiese, M. LmxPK4, a mitogen-activated protein kinase kinase homologue of Leishmania mexicana with a potential role in parasite differentiation. *Mol. Microbiol.* **56**, 1169–1182 (2005).
24. Bengs, F., Scholz, A., Kuhn, D. & Wiese, M. LmxMPK9, a mitogen-activated protein kinase homologue affects flagellar length in Leishmania mexicana. *Mol. Microbiol.* **55**, 1606–1615 (2005).
25. Madeira da Silva, L. & Beverley, S. M. Expansion of the target of rapamycin (TOR) kinase family and function in Leishmania shows that TOR3 is required for acidocalcisome biogenesis and animal infectivity. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11965–11970 (2010).
26. Bhattacharya, A. *et al.* Coupling chemical mutagenesis to next generation sequencing for the identification of drug resistance mutations in Leishmania. *Nat. Commun.* **10**, 5627 (2019).
27. Duncan, S. M. *et al.* Conditional gene deletion with DiCre demonstrates an essential role for CRK3 in Leishmania mexicana cell cycle regulation. *Mol. Microbiol.* **100**, 931–944 (2016).
28. Martel, D., Beneke, T., Gluenz, E., Späth, G. F. & Rachidi, N. Characterisation of Casein Kinase 1.1 in Leishmania donovani Using the CRISPR Cas9 Toolkit. *Biomed Res. Int.* **2017**, 4635605 (2017).
29. Rachidi, N. *et al.* Pharmacological assessment defines Leishmania donovani casein kinase 1 as a drug target and reveals important functions in parasite viability and intracellular infection. *Antimicrob. Agents*

Chemother. **58**, 1501–1515 (2014).

30. Agron, P. G., Reed, S. L. & Engel, J. N. An essential, putative MEK kinase of Leishmania major. *Mol. Biochem. Parasitol.* **142**, 121–125 (2005).
31. Rocha, V. P. C. *et al.* Leishmania dual-specificity tyrosine-regulated kinase 1 (DYRK1) is required for sustaining Leishmania stationary phase phenotype. *Mol. Microbiol.* **113**, 983–1002 (2020).
32. Wiese, M. Leishmania MAP kinases--familiar proteins in an unusual context. *Int. J. Parasitol.* **37**, 1053–1062 (2007).
33. Mottram, J. C., McCready, B. P., Brown, K. G. & Grant, K. M. Gene disruptions indicate an essential function for the LmmCRK1 cdc2-related kinase of Leishmania mexicana. *Mol. Microbiol.* **22**, 573–583 (1996).
34. Beneke, T. *et al.* A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *R Soc Open Sci* **4**, 170095 (2017).
35. Cruz, A., Coburn, C. M. & Beverley, S. M. Double targeted gene replacement for creating null mutants. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 7170–7174 (1991).
36. Damianou, A. *et al.* Essential roles for deubiquitination in Leishmania life cycle progression. *PLoS Pathog.* **16**, e1008455 (2020).
37. Jones, N. G., Catta-Preta, C. M. C., Lima, A. P. C. A. & Mottram, J. C. Genetically Validated Drug Targets in Leishmania: Current Knowledge and Future Prospects. *ACS Infect Dis* **4**, 467–477 (2018).
38. Carrera, A. C., Alexandrov, K. & Roberts, T. M. The conserved lysine of the catalytic domain of protein kinases is actively involved in the phosphotransfer reaction and not required for anchoring ATP. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 442–446 (1993).
39. Halliday, C. *et al.* Cellular landmarks of Trypanosoma brucei and Leishmania mexicana. *Mol. Biochem. Parasitol.* **230**, 24–36 (2019).
40. Burge, R. J., Damianou, A., Wilkinson, A. J., Rodenko, B. & Mottram, J. C. Leishmania differentiation requires ubiquitin conjugation mediated by a UBC2-UEV1 E2 complex. 2020.07.09.195024 (2020) doi:10.1101/2020.07.09.195024.
41. Wiese, M. A mitogen□activated protein (MAP) kinase homologue of Leishmania mexicana is essential for parasite survival in the infected host. *EMBO J.* **17**, 2619–2628 (1998).

42. Bolton, R. J. & Krzanowski, W. J. Projection Pursuit Clustering for Exploratory Data Analysis. *J. Comput. Graph. Stat.* **12**, 121–142 (2003).

43. Fernandez-Cortes, F. *et al.* RNAi screening identifies Trypanosoma brucei stress response protein kinases required for survival in the mouse. *Sci. Rep.* **7**, 6156 (2017).

44. Beneke, T. *et al.* Genetic dissection of a Leishmania flagellar proteome demonstrates requirement for directional motility in sand fly infections. *PLoS Pathog.* **15**, e1007828 (2019).

45. Hassan, P., Fergusson, D., Grant, K. M. & Mottram, J. C. The CRK3 protein kinase is essential for cell cycle progression of Leishmania mexicana. *Mol. Biochem. Parasitol.* **113**, 189–198 (2001).

46. Wyllie, S. *et al.* Cyclin-dependent kinase 12 is a drug target for visceral leishmaniasis. *Nature* **560**, 192–197 (2018).

47. Pearce, L. R., Komander, D. & Alessi, D. R. The nuts and bolts of AGC protein kinases. *Nat. Rev. Mol. Cell Biol.* **11**, 9–22 (2010).

48. Harashima, H., Dissmeyer, N. & Schnittger, A. Cell cycle control across the eukaryotic kingdom. *Trends Cell Biol.* **23**, 345–356 (2013).

49. Akiyoshi, B. & Gull, K. Discovery of unconventional kinetochores in kinetoplastids. *Cell* **156**, 1247–1258 (2014).

50. Li, Z., Gourguechon, S. & Wang, C. C. Tousled-like kinase in a microbial eukaryote regulates spindle assembly and S-phase progression by interacting with Aurora kinase and chromatin assembly factors. *J. Cell Sci.* **120**, 3883–3894 (2007).

51. Ambit, A., Woods, K. L., Cull, B., Coombs, G. H. & Mottram, J. C. Morphological events during the cell cycle of Leishmania major. *Eukaryot. Cell* **10**, 1429–1438 (2011).

52. Wheeler, R. J., Gluenz, E. & Gull, K. The cell cycle of Leishmania: morphogenetic events and their implications for parasite biology. *Mol. Microbiol.* **79**, 647–662 (2011).

53. Bushell, E. *et al.* Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. *Cell* **170**, 260–272.e8 (2017).

54. Wincott, C. J. *et al.* The host brain is permissive to colonization by Toxoplasma gondii. 2020.08.06.239822 (2020) doi:10.1101/2020.08.06.239822.

55. Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**,

511–516 (2019).

56. Saldivia, M. *et al.* Targeting the trypanosome kinetochore with CLK1 protein kinase inhibitors. *Nat Microbiol* (2020) doi:10.1038/s41564-020-0745-6.

57. Dan-Goor, M., Nasereddin, A., Jaber, H. & Jaffe, C. L. Identification of a secreted casein kinase 1 in Leishmania donovani: effect of protein over expression on parasite growth and virulence. *PLoS One* **8**, e79287 (2013).

58. Sádlová, J. & Volf, P. Peritrophic matrix of Phlebotomus duboscqi and its kinetics during Leishmania major development. *Cell Tissue Res.* **337**, 313–325 (2009).

59. Zhou, Q., Pham, K. T. M., Hu, H., Kurasawa, Y. & Li, Z. A kinetochore-based ATM/ATR-independent DNA damage checkpoint maintains genomic integrity in trypanosomes. *Nucleic Acids Res.* **47**, 7973–7988 (2019).

60. Varga, V., Moreira-Leite, F., Portman, N. & Gull, K. Protein diversity in discrete structures at the distal tip of the trypanosome flagellum. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E6546–E6555 (2017).

61. Weischenfeldt, J. & Porse, B. Bone Marrow-Derived Macrophages (BMM): Isolation and Applications. *CSH Protoc.* **2008**, db.prot5080 (2008).

62. Späth, G. F. & Beverley, S. M. A lipophosphoglycan-independent method for isolation of infective Leishmania metacyclic promastigotes by density gradient centrifugation. *Exp. Parasitol.* **99**, 97–103 (2001).

63. Pruzinova, K. *et al.* Comparison of Bloodmeal Digestion and the Peritrophic Matrix in Four Sand Fly Species Differing in Susceptibility to Leishmania donovani. *PLoS One* **10**, e0128203 (2015).

64. Volf, P. & Volfova, V. Establishment and maintenance of sand fly colonies. *J. Vector Ecol.* **36 Suppl 1**, S1–9 (2011).

65. Castanys-Muñoz, E., Brown, E., Coombs, G. H. & Mottram, J. C. Leishmania mexicana metacaspase is a negative regulator of amastigote proliferation in mammalian cells. *Cell Death Dis.* **3**, e385 (2012).

66. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

67. Beneke, T. & Gluenz, E. Bar-seq strategies for the LeishGEdit toolbox. *Mol. Biochem. Parasitol.* 111295 (2020).

68. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).

69. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

70. Tetaud, E., Lecuix, I., Sheldrake, T., Baltz, T. & Fairlamb, A. H. A new expression vector for Crithidia fasciculata and Leishmania. *Mol. Biochem. Parasitol.* **120**, 195–204 (2002).

71. Kühn, M. J. *et al.* Spatial arrangement of several flagellins within bacterial flagella improves motility in different environments. *Nat. Commun.* **9**, 5369 (2018).

72. Dean, S., Sunter, J. D. & Wheeler, R. J. TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. *Trends Parasitol.* **33**, 80–82 (2017).

73. Martínez de Iturrate, P. *et al.* Towards discovery of new leishmanicidal scaffolds able to inhibit GSK-3. *J. Enzyme Inhib. Med. Chem.* **35**, 199–210 (2020).

74. Chhajer, R. *et al.* Leishmania donovani Aurora kinase: A promising therapeutic target against visceral leishmaniasis. *Biochimica et Biophysica Acta (BBA) - General Subjects* vol. 1860 1973–1988 (2016).

Figure Legends

Fig. 1 *L. mexicana* kinome and conservation in other pathogenic trypanosomatids. (a) Representation of eukaryotic (ePK) and phosphatidylinositol 3' kinase-related kinases (PIKK) protein kinase families, separated by colours defined by primary protein sequence conservation. Image for illustrative purpose only with no phylogenetic significance. Red dots represent protein kinases for which gene deletion mutants could not be generated and are therefore potentially essential genes in procyclic promastigotes. (b) Phylogenetic tree of trypanosomatids with the number of orthologous protein kinases unique to each node indicated by blue circles and the percentage of potentially essential orthologous protein kinases unique to each node indicated within a red rectangle. (c) Pie-charts show dispensable (non-essential) and required (potentially essential) protein kinases in procyclic promastigotes, separated into families.

Fig. 2 Generation of the *L. mexicana* kinome gene deletion library. (a) Schematic representation of gene editing by CRISPR-Cas9 in the *L. mexicana* progenitor cell line showing integration of repair cassettes containing 30nt homology sites, a unique barcode and antibiotic resistance genes for puromycin (*PAC*) and blasticidin (*BSD*). Diagnostic PCRs for the presence of the CDS with Oligo 1 and Oligo 2, (depicted in blue), correct integration of the repair cassettes using Oligo 3 and Oligo 4 (depicted in red), and the gene in the endogenous locus with Oligo 5 and Oligo 6 (depicted in green). (b) Diagnostic PCR for (i) *LmxM.25.0853* and (ii) *PKAC1* (*LmxM34.4010*) using primer pairs as above in *L. mexicana* Cas9T7 (WT) and Populations A and B (c) Heat map indicating the presence or absence of target genes after whole genome sequencing. The ratio of gene coverage to chromosome coverage was used as a measure of gene copy number in selected mutants. A value of 1 represents two alleles of a given gene. (d) Diagnostic PCR for facilitated gene deletion mutant *LmxM.04.0650*

using Oligo 1 and 2 for the CDS, Oligo 5 and 6 for the endogenous locus and Oligo 3 and 4 for correct integration. Western blot using anti-myc confirmed episomal expression of LmxM.04.0650 at the anticipated size of 140 kDa.

Fig. 3 Endogenous tagging of *L. mexicana* protein kinases. (a) Schematic representation of process for N-terminal tagging. Transfection of the sgRNA (PCR1) directs the double stranded break immediately upstream of the target CDS and the repair template (PCR2) contains 30nt at 5'and 3' of the fragment for homologous recombination on the target locus to insert the fluorescent tag and an antibiotic-resistance gene. (b) Bar graph representing the localisation of dispensable and required protein kinases for each organelle (c) Pictorial illustration of the 10 localisation groups. An exemplar live cell image of protein kinases tagged with mNeonGreen for each localisation. Procyclic promastigotes were incubated with Hoechst 33342 for DNA labelling (pink). Scale bar: 5 µm.

Fig. 4 Identification of protein kinases involved in *L. mexicana* differentiation and infection. (a) Schematic illustration of the bar-seq screen design with three experimental arms (EA1, blue; EA2 , black; EA3, red) depicting investigation of differentiation to amastigote stages and investigation of protein kinase mutants in sand fly infection (EA4, green). A pool of gene deletion mutants were generated in procyclic promastigotes (PRO), which were grown in Graces Media at pH 5.5 for 168 h. Cells were then either diluted into amastigote media at 35°C and grown as axenic amastigotes (AXA) for 5 d or enriched for metacyclic promastigotes. Metacyclic promastigotes (META) were used to either infect macrophages (inMAC) or inoculate mouse footpads (FP). DNA was taken at the indicated time points and the unique barcodes amplified by PCR to apply bar-seq analyses. (b) Proportion of barcodes across 8 time points, as indicated in A. The trajectories of six exemplar protein kinase mutants have been plotted for the three experimental arms. Values are mean +/- S.D., N = 6. (c) Projection pursuit cluster analysis was applied to the trajectories from each experimental arm, grouped into six clusters, each consisting of trajectories with a similar trend. Only clusters resulting from the mouse footpad infection are shown. The image for each cluster shows gene trajectories overlaid with average trends in bold. Trajectories plotted using logged % barcode representation data, normalised to time 0. Heat maps below show % barcode representation data and depict the trend for each individual gene. Gene IDs in red (MCA, MPK1 and MPK2) have documented phenotypes in *Leishmania* and serve to benchmark the dataset. (d) The relationship between the clusters is shown on a two-dimensional PCA plot. Colours match the clusters in C.

Fig. 5 Identification of protein kinases involved in colonisation of *Lutzomyia longipalpis*. (a) Heat map of 29 protein kinase mutants that cluster in groups important for the amastigote stage in two of the three experiments using both 1 and 2 difference clustering. In bold are the genes that have been identified as important for mouse infection in a previous trypanosome study [43]. (b) Summary showing the protein kinases important for sand fly infection. Control plot (top) shows the relative barcode representation for four control lines with a barcode inserted in the ribosomal locus at three timepoints post blood meal ((PBM), Day 1, Day 5 and Day 8). Protein kinases important for *L. longipalpis* infection plot (bottom) shows the 15 mutants with significant loss of representation by Day 8 identified using a multiple *t*-test (n=2). 7 protein kinases identified as important for both sand fly infectivity and for survival in amastigote stages indicated in bold. (c) Heat map showing the amastigote stage data for the 8 protein kinase mutants that are important for sand fly infection only. (d) Plot showing the infectivity of the protein kinase mutants important for survival of amastigotes only.

Fig. 6 Identification of protein kinases involved in motility. (a) Bar chart showing flagellar length: cell body length ratio for 5 protein kinase deficient mutants identified with significant loss of representation at the top of the transwell, indicating a motility defect. Error bars indicate standard

deviation (n = >50). Mutants ordered according to cell body width measurement. Widest cells shown to the right. Mutants with significant difference in flagella:cell body ratio when compared to Cas9T7 were identified using a multiple *t*-test (b) Graph showing the fraction of the population for each promastigote protein kinase mutant progressively swimming. The motile fraction of cells in the control (*Cas9T7,* n=2532) and four mutant strains: Δ*ulk4* (n=2936), Δ*stk36* (n=4896), Δ29.0600 (n=2623), and Δ02.0570 (n=3532). The error bars represent standard error of the mean/95% confidence intervals and the data points denote results from individual technical replicates. (c) Schematic showing the protein structures of the four protein kinase mutants identified in the motility screen. (d) Localisation of the proteins tagged with mNeonGreen. (e) Bar-seq trajectories for the four protein kinase mutants for Pool 1, EA3 (f) and normalised bar-seq reads for these four protein kinase mutants during development in the sand fly.

Fig. 7 Schematic representation of *L. mexicana* differentiation cycle. Fifteen protein kinases were identified as required for colonisation of the sandfly. Twenty nine protein kinases were identified as required for survival as amastigotes in vivo and in vitro.

Supplementary Information

Supplementary Table 1 Kinome data overview and primers. Tab 1: Overview of data on required/dispensable and primary localisation. Tab 2: Primers used for endogenous tagging of protein kinases. Tab 3: Summary of localisation data and comparison with TrypTag [72], presence of a signal peptide, Mitoprot score, presence of a nuclear localisation signal (NLS), presence in the flagellar proteome [44]. Tab 4: Primers used for gene deletion and predicted size of PCR products for ORF and Integration. Tab 5: Orthology of protein kinases, as shown in TritrypDB, in C*rithidia fasciculata* strain Cf-Cl, *Endotrypanum monterogeii* strain LV88, *Leptomonas seymouri* ATCC 30220, *Blechomonas ayalai* B08-376, *Trypanosoma grayi* ANR4, *Trypanosoma cruzi*, *Trypanosoma vivax, Trypanosoma brucei*, *Paratrypanosoma confusum* CUL13.

Supplementary Table 2 Kinome wide library outputs. Tab 1: Axa_InMac_Mouse data Pool 1; bar-seq data for protein kinases required for differentiation and survival in culture and in the mammalian host. Tab 2: Sand fly data Pool 2; bar-seq data for protein kinases required for colonisation of the sand fly. Tab 3: Motility data Pool 1; bar-seq data for protein kinases required for motility. Tab 4: Motility data Pool 2; bar-seq data for protein kinases required for motility. Tab 5: Axa_InMac_Mouse data analysis; cluster analysis for data in Pool 1. Tab 6: Sandfly analysis; analysis of data from Pool 2. Tab 7: Motility analysis; protein kinases with significant loss of motility from Pool 1 and Pool 2 and cell phenotype measurements.

Supplementary Equation 1 Projection pursuit cluster analysis methodology and script.

Supplementary Code 1 Python script to count the occurrence of barcode sequences in FASTA files generated in a BarSeq experiment

Supplementary Data 1 Diagnostic PCRs to test for gene deletion mutants. The DNA prepared from populations (Pop) recovered after transfection was used to test for the presence/ or absence of the protein kinase coding sequence (CDS) and integration (INT) of the resistance cassette at the correct locus. Exemplar gel explained in Fig 1b. All primers used listed in Supplementary Table 1.

Supplementary Data 2 Localisation of mNeonGreen tagged protein kinases. An example image is shown for each protein kinase in the G1 phase of the cell cycle (1K 1N). Additional images are

included for protein kinases that are stage specific or for which N- and C- terminal tagging differ. Images are ordered by the protein kinase family. Scale bar: 2 µm. Descriptors are based on those described in [39].

Supplementary Data 3 Heat maps and cluster analysis for amastigotes. Data for the three experimental arms (EA1, 2 and 3 (see Fig 4a)) sorted via cluster analysis. Heat maps show % abundance of barcodes. Cluster trajectories plotted using the logged % of barcodes, normalised to time 0 (relative abundance (Log10)). (a) Axenic amastigote data clustered by 1 difference (b) Individual plots for axenic amastigote data (coloured according to 1 difference clustering). (c) Macrophage infection data clustered by 1 difference. (d) Individual plots for macrophage infection data (coloured according to 1 difference clustering). (e) Mouse footpad infection data clustered by 1 difference. (f) Individual plots for mouse footpad infection data (coloured according to 1 difference clustering). Procyclic promastigotes (PRO), axenic amastigotes (AXA), metacyclic promastigotes (META), infected macrophages (inMAC) and mouse footpads (FP).

Supplementary Fig. 1 Putative pseudokinases of *L. mexicana*. (a) Schematic of a conserved protein kinase structure. The highly conserved lysine and two aspartic acid residues required for an active kinase are highlighted. (b) List of all protein kinases missing one or more of these residues. Boxed K, D, D letters with diagonal line indicate missing residues. The 11 pseudokinases are indicated in bold text. Green star denotes protein kinases that are specific to the *Leishmaniinae* family. Solid red circle indicates that the protein kinase is required for promastigote survival.

Supplementary Fig. 2 Redundancy of *CLK1* and *CLK2*. **(**a) Gene deletion (circle) for *CLK1* (LmxM.09.0400) and *CLK2* (LmxM.09.0410). Attempts were made to delete *CLK1*, *CLK2* or both *CLK1* and *CLK2*. Successful deletions are indicated after antibiotic selection. (b) Diagnostic PCRs are shown for gene deletion mutants. WT; *L. mexicana* M379, T7 Cas9; *L. mexicana* cell line expressing Cas9-T7 in populations (Pop) and clones (C). (c) Integrative genomics viewer of the illumina whole genome sequencing reads from Δ*clk1* and Δ*clk2* mutants are aligned against the reference genome of the Cas9 T7 progenitor cell line generated by Nanopore sequence. Reads absent at the N-terminal end of the *CLK1* and *CLK2* indicate a gene deletion mutant as the remainder of the CDSs are identical between both genes. The multiple alignments are indicated by the coloured white reads (mapping quality zero).

Supplementary Fig. 3 Characterisation of *RDK1* (a) Proliferation of promastigotes comparing *L. mexicana* Cas9 T7 progenitor line and Δ*rdk1* clones 1 and 2. (b) Proliferation of axenic amastigotes comparing *L. mexicana* Cas9 T7 progenitor line, Δ*rdk1* and Δ*rdk1::RDK1*. (c) Bone-marrow macrophage infections using purified metacyclic promastigotes of *L. mexicana* Cas9 T7 progenitor line and Δ*rdk1* clones 1 and 2 analysed for total number of amastigotes infecting 200 macrophages (MΦ) at early (12 h) and late (72 h) time points. (d) Parasite burden determined by limiting dilution of 5 BALB/c footpad lesions infected with *L. mexicana* Cas9 T7 progenitor line, Δ*rdk1* mutant and an add-back Δ*rdk1::RDK1* for six weeks. Values plotted with the mean and 95% SEM interval. Statistical analysis are two-tailed unpaired student *t*-test. Data are mean of 2 independent experiments.

Supplementary Fig. 4 Fitness plots for protein kinase mutants during sand fly colonisation. Data are shown for Day 1 (Black), 5 (Dark grey) and 8 (light grey). Top plot shows mutants included in the slow pool, middle plot indicates average growth pool and the bottom plot indicates fast growth pool. Fitness was calculated using normalised reads at each time point against the normalised reads at time point 0. Mutants contributing to less than 0.2% of the reads on the day of the infected blood meal are highlighted in grey and were excluded from the loss of fitness analyses. Mutants showing a

significant loss of fitness are indicated with a *. Corresponding raw data can be found in Supplementary Table 2.

Supplementary Fig. 5 Speed histograms for protein kinase mutants. These panels show the distribution of swimming speeds in the parental line (Cas9T7, n=1149) and four protein kinase mutants: Δ*ulk4* (n=57), Δ*stk36* (n=144), Δ*29.0600* (n=6), Δ*02.0570* (n=145). The instantaneous swimming speed of each cell was averaged across the duration of its swimming track to produce one count in the histograms.

Supplementary Fig. 6 Protein kinases involved in flagellum function. (a) Phenotypic analysis measurements for protein kinase mutants important for motility. (b) Additional images supporting the exempla protein kinase localisations in Fig. 6d and short flagellum null mutants in Fig. 6b (c) Phase images of Δ*ulk4* and Δ*stk36* mutants (top) and Δ*ulk4::ULK4 and* Δ*stk36:STK36* add-back mutants (bottom).

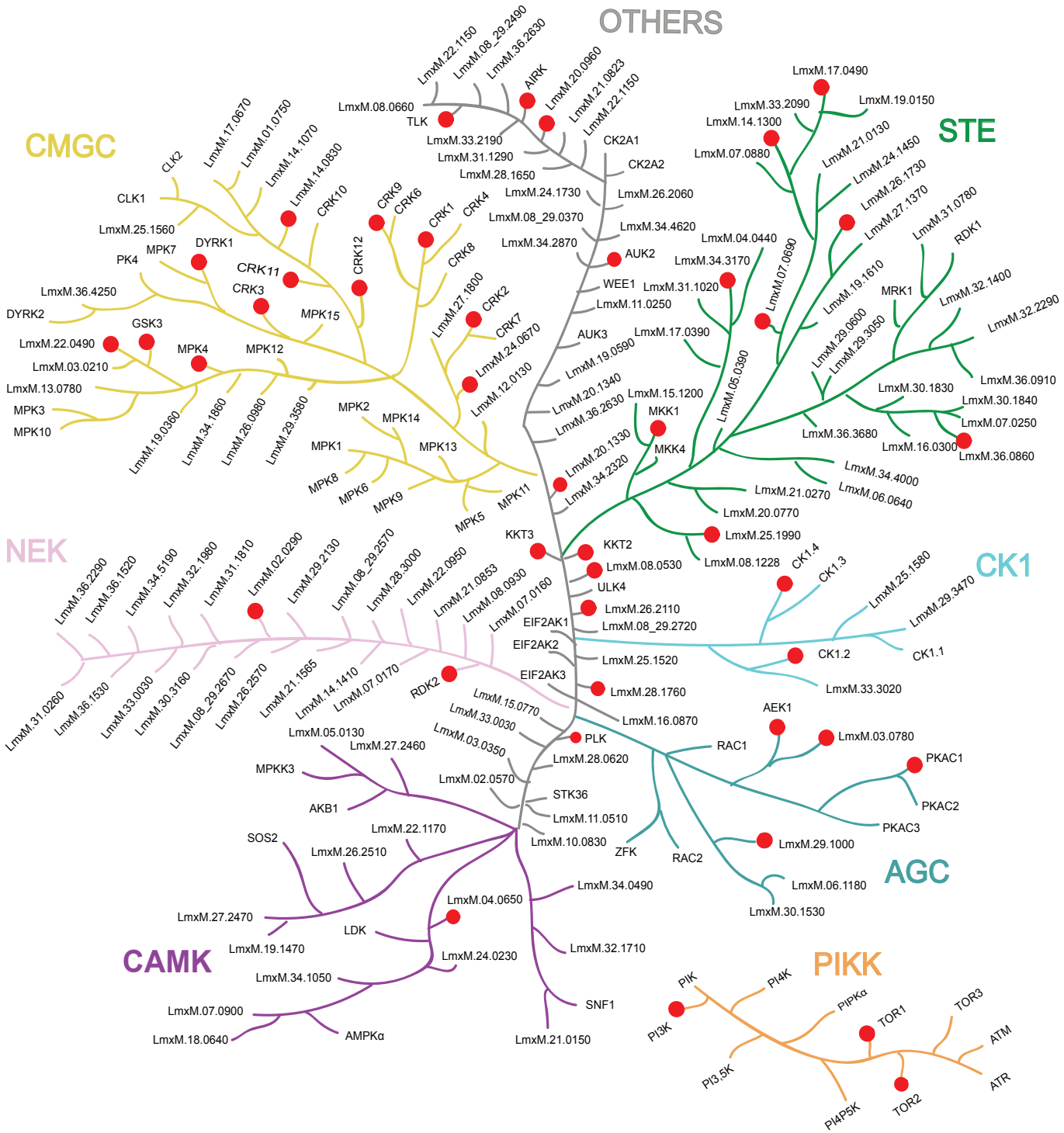
Supplementary Movie 1 Motility of protein kinase mutants. These movies show the effect of mutations on the flagellar beat of *L. mexicana*. The movies were acquired at a frame rate of 500 Hz using a phase-contrast imaging and a 60x magnification microscope objective. The scale is the same for all panels, and the movie is played back at 1/5 th of the true speed so the high-speed flagellar motion can be more easily seen.

Table 1. *L. mexicana* protein kinase genes for which gene deletion mutants could not be generated in procyclic promastigotes and are likely to be essential. Includes previous literature about gene essentiality.

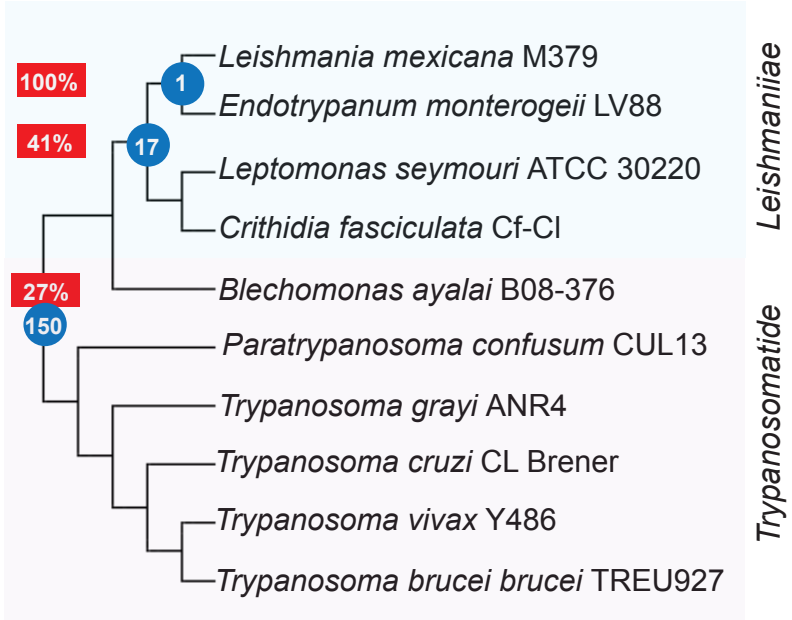
Gene ID	Name	Group/Family	Primary Localisation	Previous literature
LmxM.25.2340	AEK1	AGC	cytoplasm	
LmxM.29.1000		AGC	cytoplasm	
LmxM.34.4010	PKAC1	AGC/PKA	nucleus	
LmxM.03.0780		AGC/RSK	cytoplasm	
LmxM.04.0650		CAMK/CAMKL	cytoplasm	
LmxM.27.1780	CK1.4	CK1/CK1	flagellar pocket	
LmxM.34.1010	CK1.2	CK1/CK1	cytoplasm	29
LmxM.24.0670		CMGC	nucleus	
LmxM.05.0550	CRK2	CMGC/CDK	cytoplasm	
LmxM.09.0310	CRK12	CMGC/CDK	nucleus and kinetoplast	46
LmxM.21.1080	CRK1	CMGC/CDK	mitochondrion	33
LmxM.27.1940	CRK9	CMGC/CDK	nucleus	
LmxM.29.1780	CRK11	CMGC/CDK	nucleus	
LmxM.36.0550	CRK3	CMGC/CDK	cytoplasm	27,45
LmxM.14.0830		CMGC/DYRK	basal body	
LmxM.15.0180	DYRK1	CMGC/DYRK	lysosome	31
LmxM.18.0270	GSK3	CMGC/GSK	pellicular membrane	73
LmxM.22.0490	GSKA	CMGC	no signal	
LmxM.19.1440	MPK4	CMGC/MAPK	lysosome	17

LmxM.24.2320	MKK4	STE	endomembrane	
LmxM.25.1990		STE	endomembrane	
LmxM.34.3170		STE	cytoplasm	
LmxM.07.0690		STE	cytoplasm	
LmxM.14.1300		STE	flagellum	
LmxM.17.0490		STE	flagellum	
LmxM.27.1370		STE	endomembrane	
LmxM.36.0860	MKK5	STE	cytoplasm	
LmxM.02.0290		Other/NEK	cytoplasm	
LmxM.30.2960	RDK2	Other/NEK	cytoplasm	
LmxM.26.2110		Orphan	lysosome	
LmxM.34.4050	KKT3	Orphan	nucleus	
LmxM.36.5350	KKT2	Orphan	nucleus	
LmxM.20.1330		Other/unique	endomembrane	
LmxM.28.0520	AUK1/AIRK	Other/AUR	nucleus	[74]
LmxM.08_29.1330	AUK2	Other/AUR	cytoplasm	
LmxM.17.0790	PLK	Other/PLK	basal body	
LmxM.30.2860	TLK	Other/TLK	nucleus	
LmxM.28.1760		Other/VPS15	not attempted	
LmxM.08.0530		Other/unique	endomembrane	
LmxM.20.0960		Other/unique	endomembrane	
LmxM.24.2010	PI3K	Atypical	cytoplasm	
LmxM.36.6320	TOR1	Atypical	cytoplasm	[25]
LmxM.33.4530	TOR2	Atypical	cytoplasmic organelles	[25]

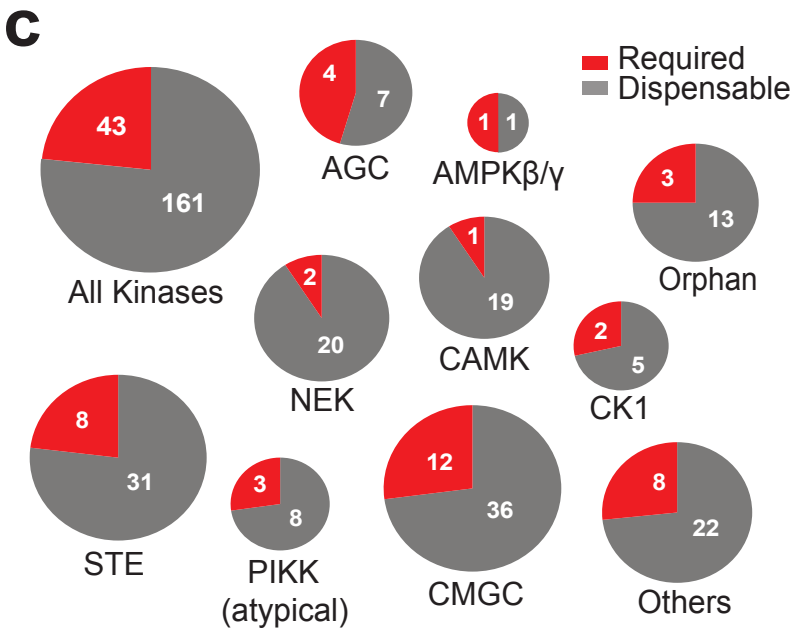
1. a *Leishmania mexicana* Kinome

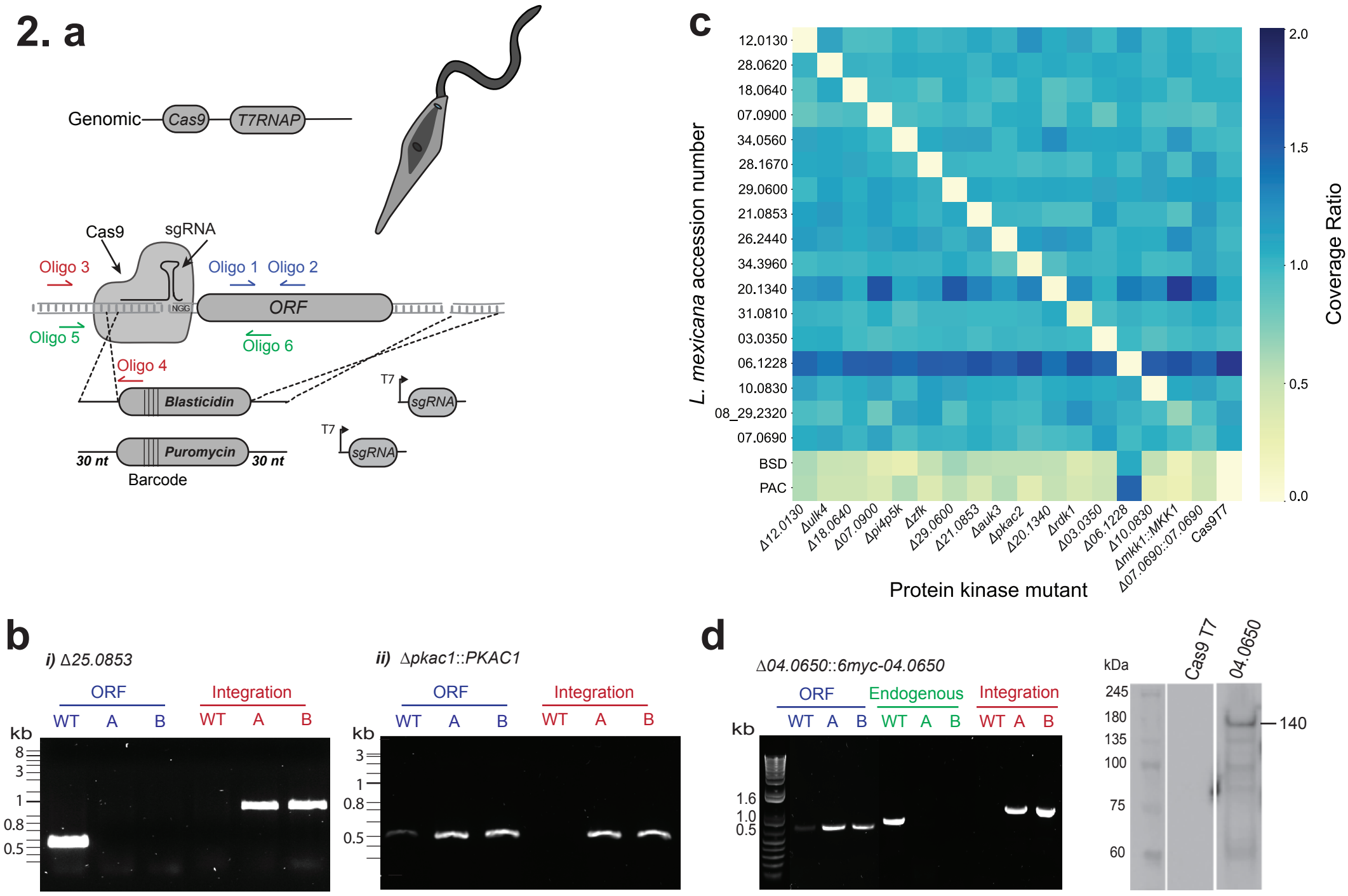


b

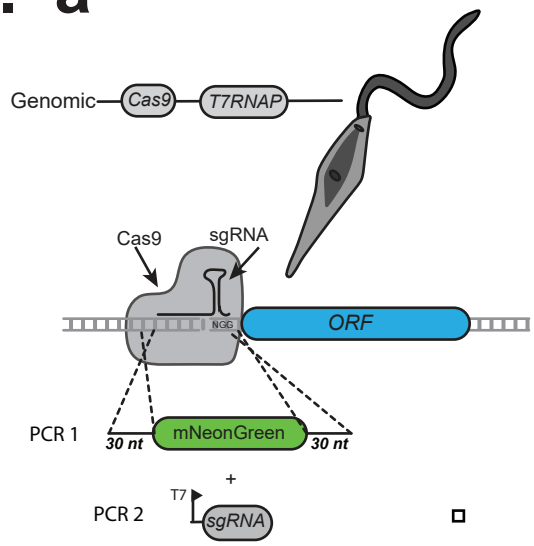


c

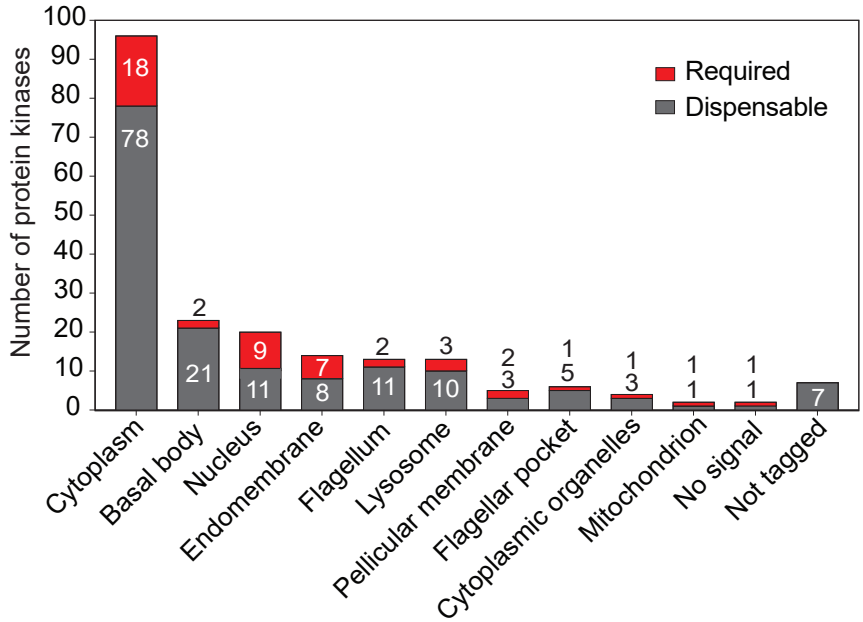



2. a
Genomic
Cas9
T7RNAP
Cas9
sgRNA
Oligo 3
Oligo 1
Oligo 2
NGG
ORF
Oligo 5
Oligo 6
Oligo 4
Blasticidin
T7
sgRNA
T7
sgRNA
Puromycin
30 nt
30 nt
Barcode
c
L. mexicana accession number
12.0130
28.0620
18.0640
07.0900
34.0560
28.1670
29.0600
21.0853
26.2440
34.3960
20.1340
31.0810
03.0350
06.1228
10.0830
08_29.2320
07.0690
BSD
PAC
Δ12.0130
Δulk4
Δ18.0640
Δ07.0900
Δpi4p5k
Δzfk
Δ29.0600
Δ21.0853
Δauk3
Δpkac2
Δ20.1340
Δrdk1
Δ03.0350
Δ06.1228
Δ10.0830
Δmkk1::MKK1
Δ07.0690::07.0690
Cas9T7
Protein kinase mutant
Coverage Ratio
2.0
1.5
1.0
0.5
0.0
b
i) Δ25.0853
ORF
Integration
WT A B WT A B
kb
8
3
1
0.8
0.5
ii) Δpkac1::PKAC1
ORF
Integration
WT A B WT A B
kb
3
1
0.8
0.5
d
Δ04.0650::6myc-04.0650
ORF
Endogenous
Integration
WT A B WT A B WT A B
kb
1.6
1.0
0.5
kDa
Cas9 T7
04.0650
245
180
135
100
75
60
140

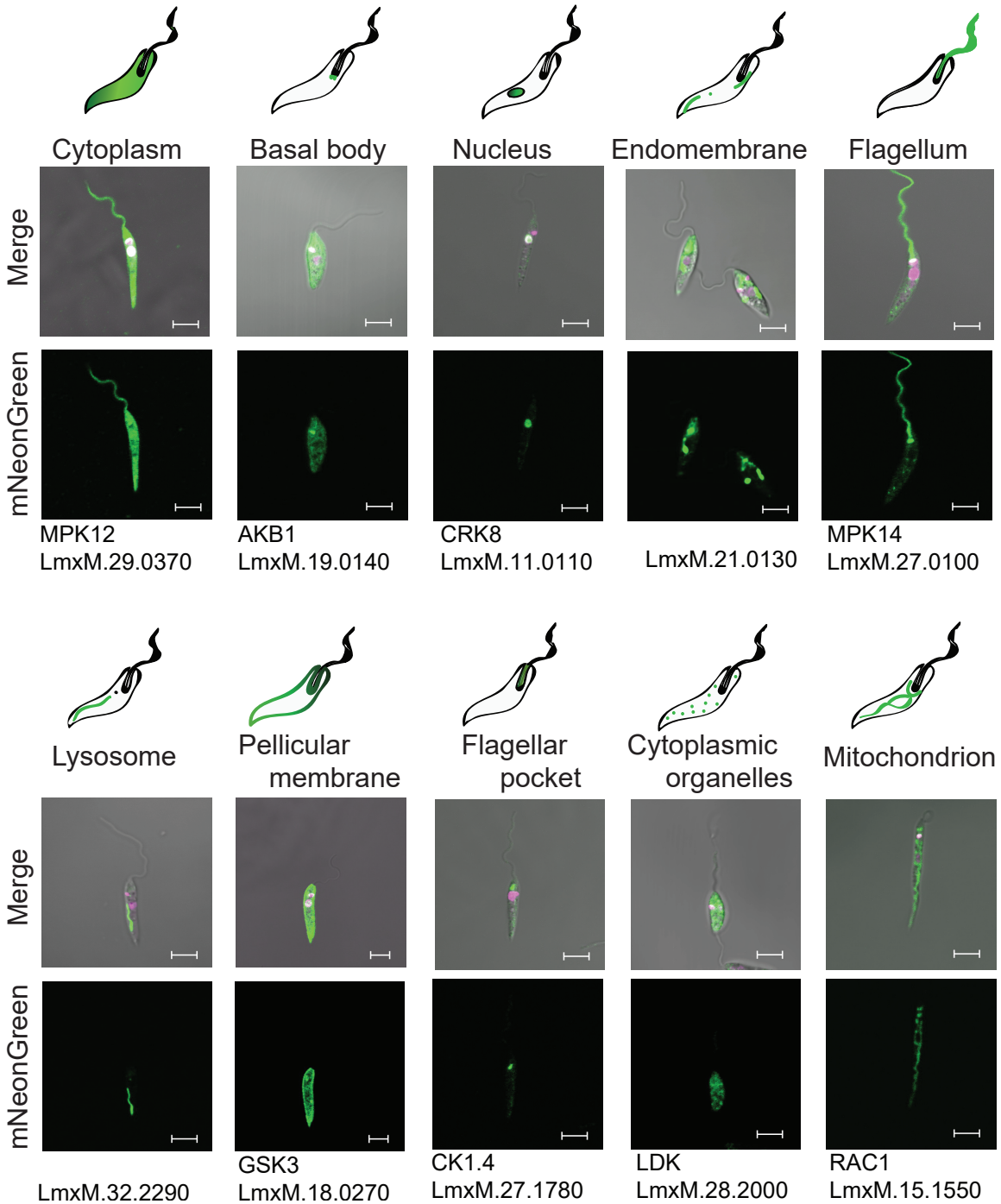
3. a

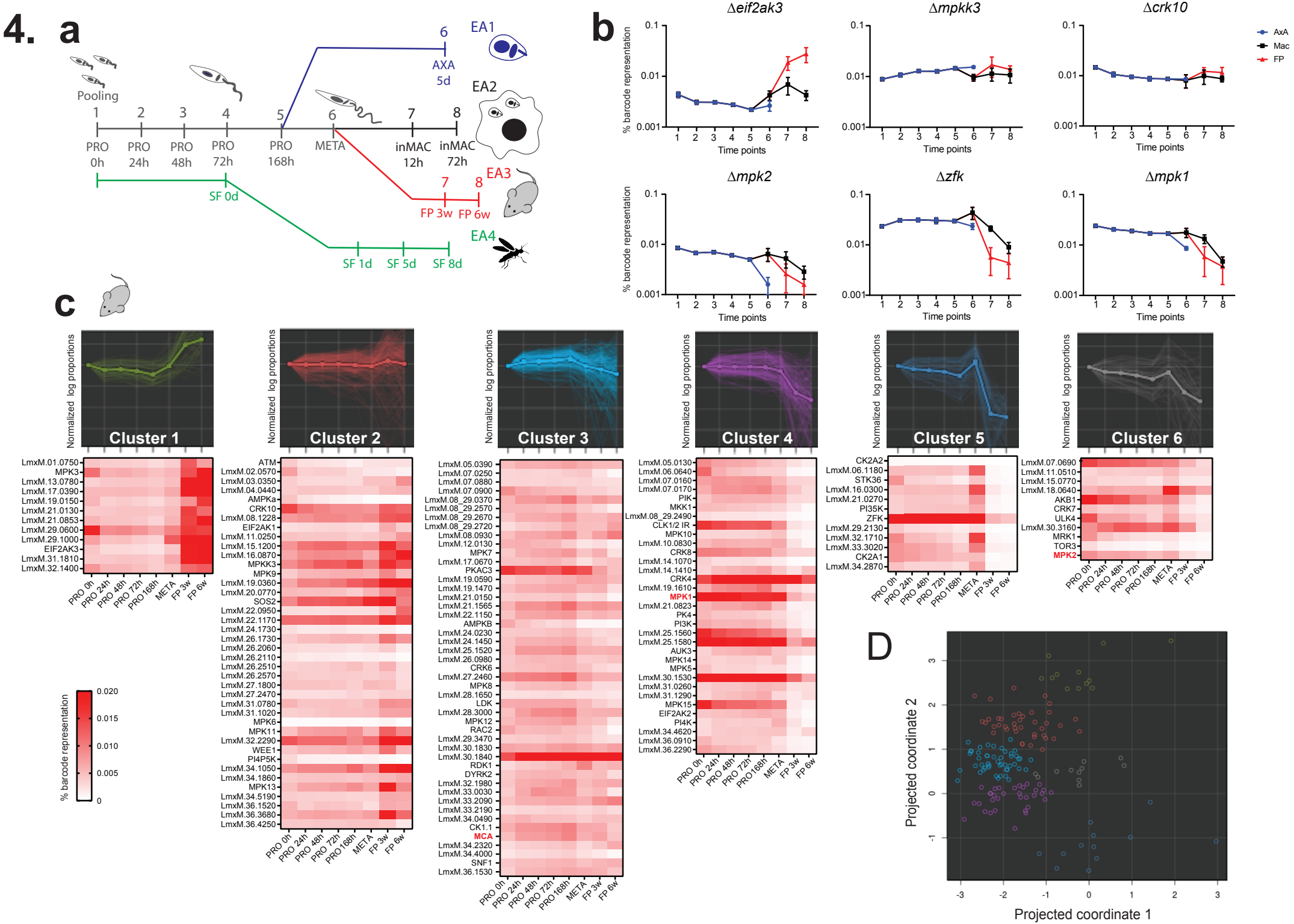

Genomic
Cas9
T7RNAP
Cas9
sgRNA
NGG
ORF
PCR 1
30 nt
mNeonGreen
30 nt
+
PCR 2
T7
sgRNA

b

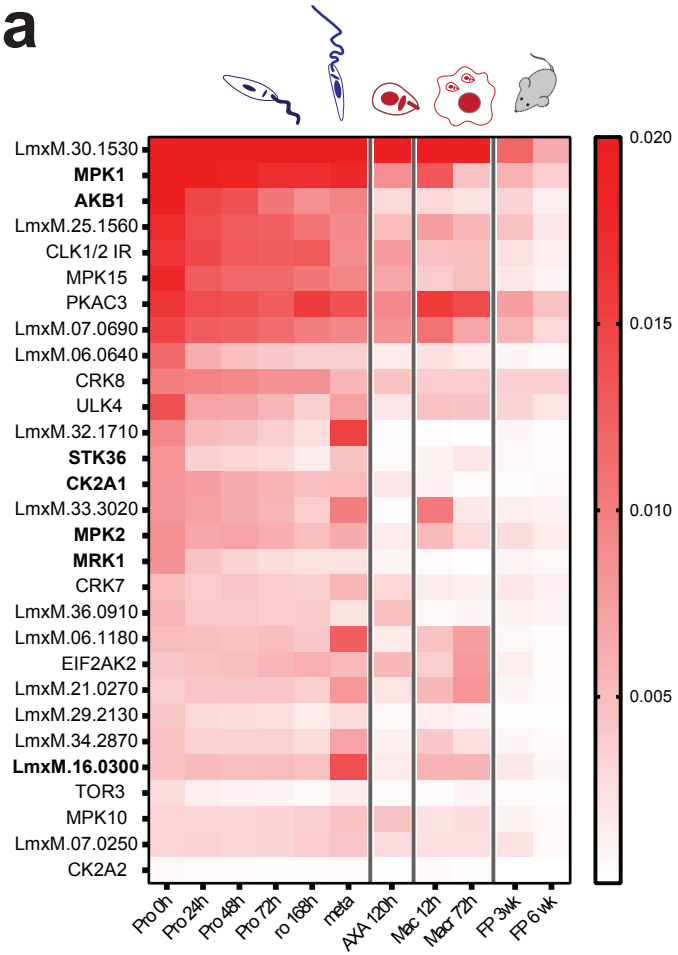

Number of protein kinases
100
90
80
70
60
50
40
30
20
10
0
Required
Dispensable
18
78
2
21
9
11
7
8
2
11
3
10
2
3
1
5
1
3
1
1
1
1
7
Cytoplasm
Basal body
Nucleus
Endomembrane
Flagellum
Lysosome
Pellicular membrane
Flagellar pocket
Cytoplasmic organelles
Mitochondrion
No signal
Not tagged

c

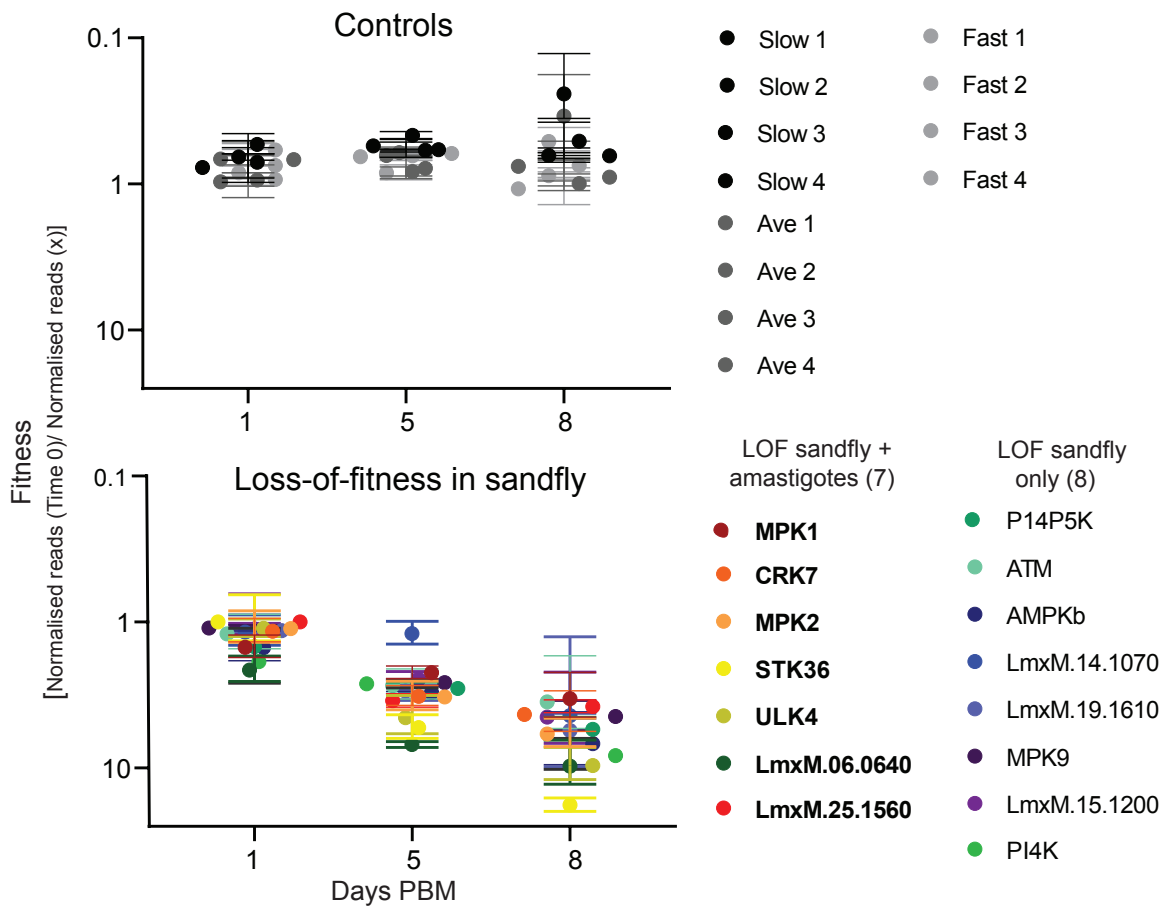

Cytoplasm
Basal body
Nucleus
Endomembrane
Flagellum
Merge
mNeonGreen
MPK12
LmxM.29.0370
AKB1
LmxM.19.0140
CRK8
LmxM.11.0110
LmxM.21.0130
MPK14
LmxM.27.0100
Lysosome
Pellicular membrane
Flagellar pocket
Cytoplasmic organelles
Mitochondrion
Merge
mNeonGreen
LmxM.32.2290
GSK3
LmxM.18.0270
CK1.4
LmxM.27.1780
LDK
LmxM.28.2000
RAC1
LmxM.15.1550


4.
a
Pooling
1
2
3
4
5
6
7
8
PRO 0h
PRO 24h
PRO 48h
PRO 72h
PRO 168h
META
inMAC 12h
inMAC 72h
AXA 5d
EA1
EA2
EA3
EA4
FP 3w
FP 6w
SF 0d
SF 1d
SF 5d
SF 8d
b
Δeif2ak3
Δmpkk3
Δcrk10
Δmpk2
Δzfk
Δmpk1
% barcode representation
Time points
AxA
Mac
FP
c
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Normalized log proportions
% barcode representation
D
Projected coordinate 1
Projected coordinate 2

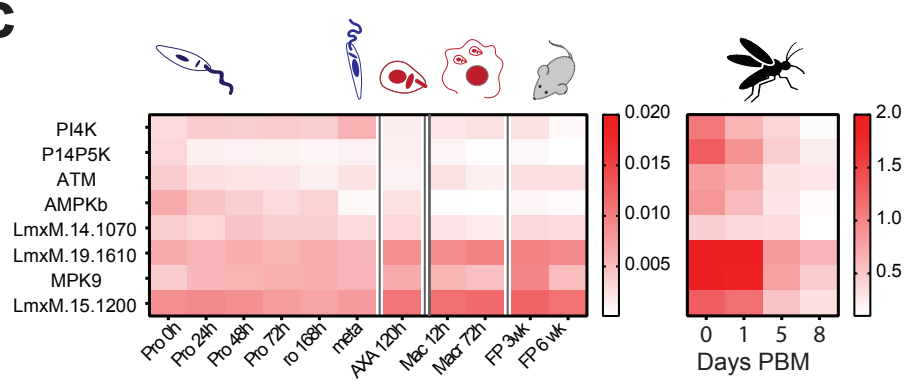
5. a


LmxM.30.1530
MPK1
AKB1
LmxM.25.1560
CLK1/2 IR
MPK15
PKAC3
LmxM.07.0690
LmxM.06.0640
CRK8
ULK4
LmxM.32.1710
STK36
CK2A1
LmxM.33.3020
MPK2
MRK1
CRK7
LmxM.36.0910
LmxM.06.1180
EIF2AK2
LmxM.21.0270
LmxM.29.2130
LmxM.34.2870
LmxM.16.0300
TOR3
MPK10
LmxM.07.0250
CK2A2
Pro 0h
Pro 24h
Pro 48h
Pro 72h
ro 168h
meta
AXA 120h
Mac 12h
Macr 72h
FP 3wk
FP 6 wk
0.020
0.015
0.010
0.005

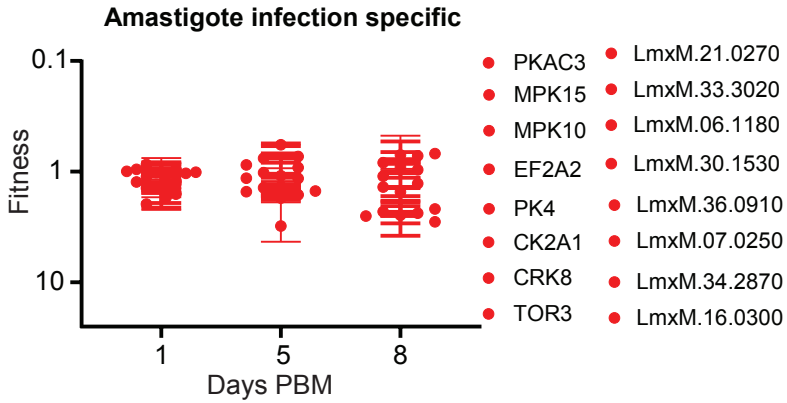
b

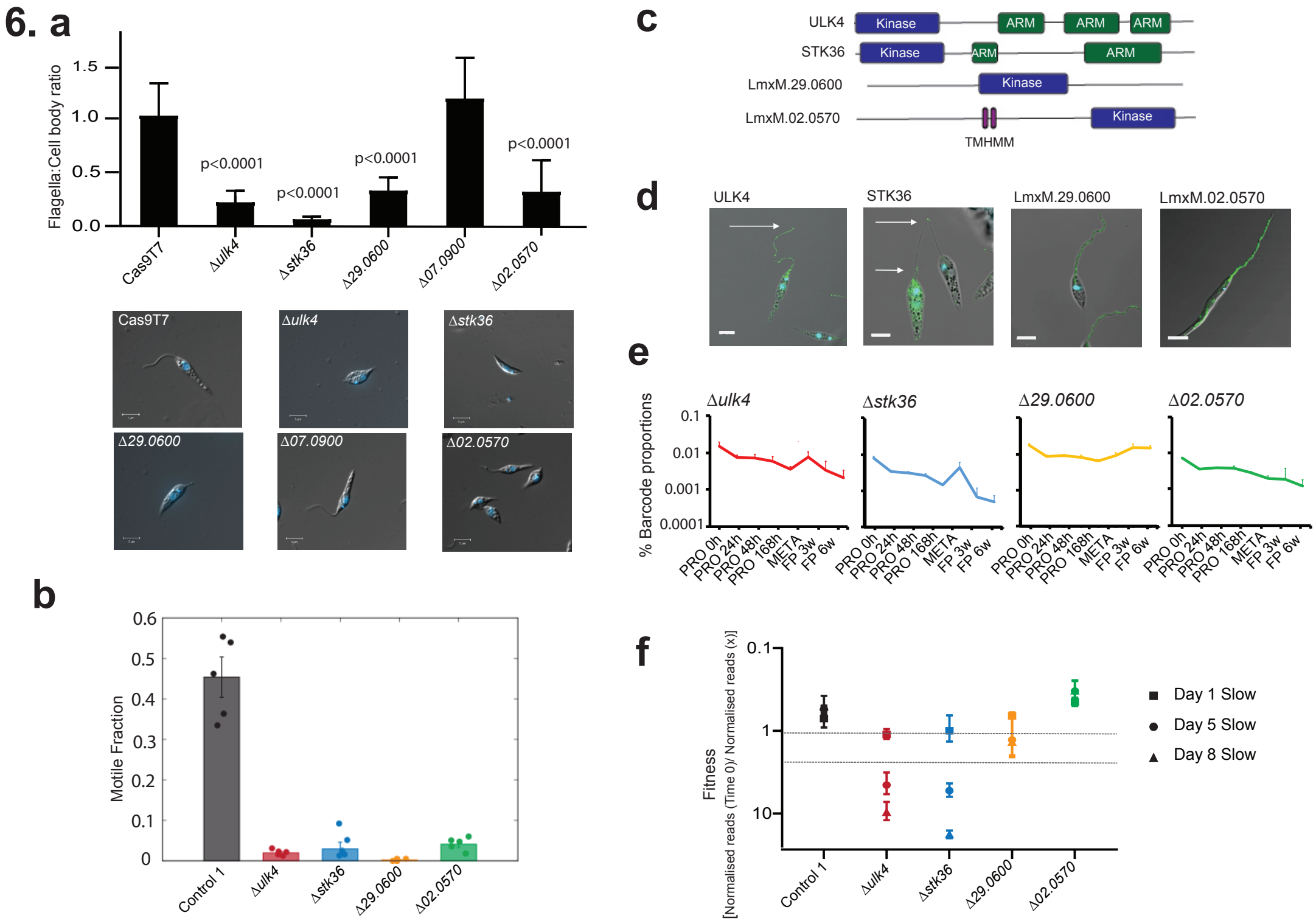

Controls
Loss-of-fitness in sandfly
Fitness
[Normalised reads (Time 0)/ Normalised reads (x)]
0.1
1
10
1
5
8
Days PBM
Slow 1
Slow 2
Slow 3
Slow 4
Ave 1
Ave 2
Ave 3
Ave 4
Fast 1
Fast 2
Fast 3
Fast 4
LOF sandfly + amastigotes (7)
MPK1
CRK7
MPK2
STK36
ULK4
LmxM.06.0640
LmxM.25.1560
LOF sandfly only (8)
P14P5K
ATM
AMPKb
LmxM.14.1070
LmxM.19.1610
MPK9
LmxM.15.1200
PI4K

c

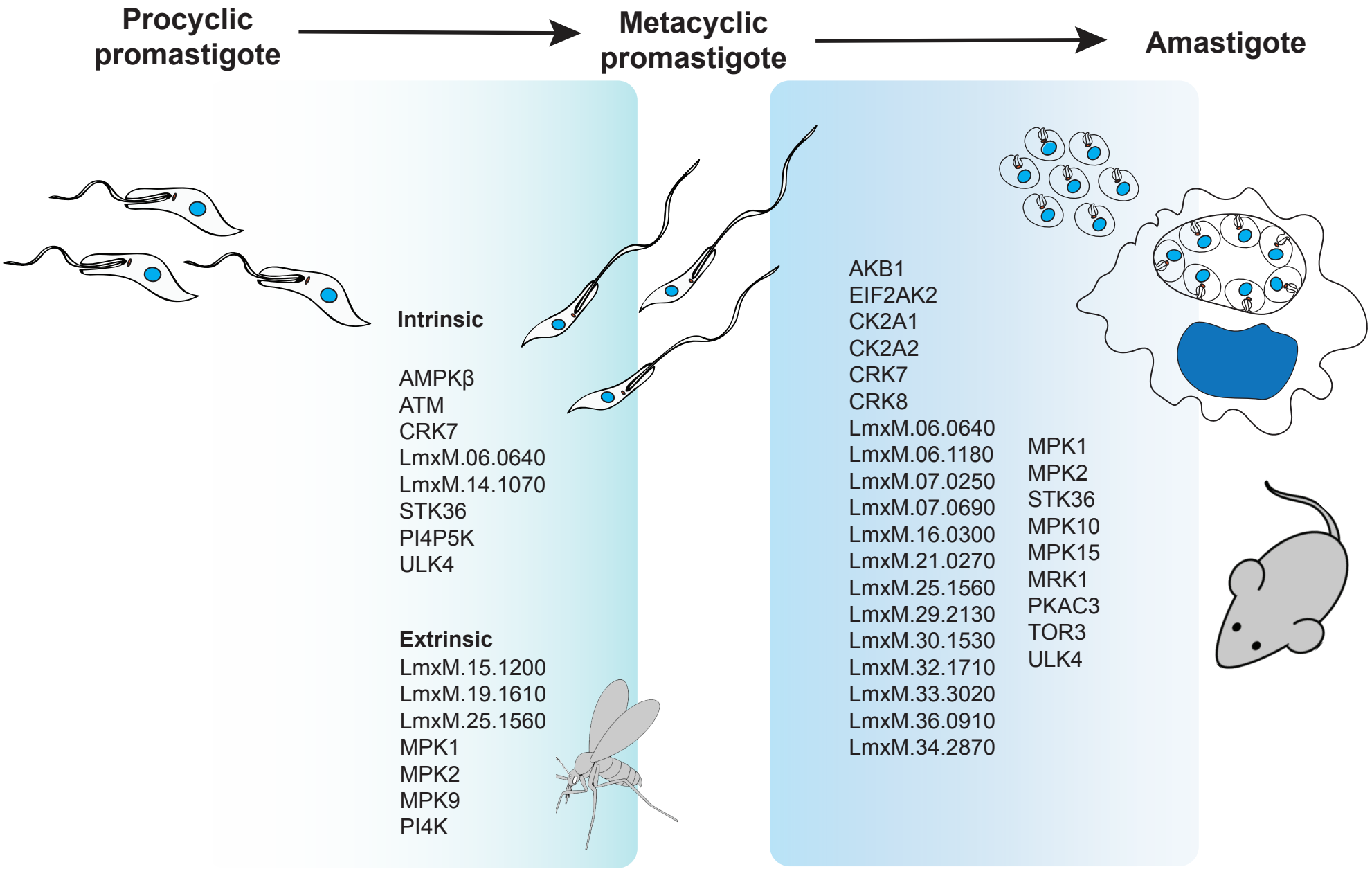

PI4K
P14P5K
ATM
AMPKb
LmxM.14.1070
LmxM.19.1610
MPK9
LmxM.15.1200
Pro 0h
Pro 24h
Pro 48h
Pro 72h
ro 168h
meta
AXA 120h
Mac 12h
Macr 72h
FP 3wk
FP 6 wk
0.020
0.015
0.010
0.005
2.0
1.5
1.0
0.5
0
1
5
8
Days PBM

d

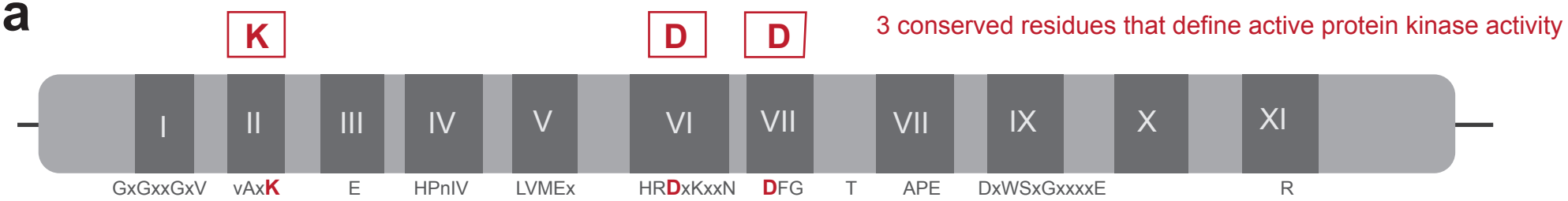

Amastigote infection specific
Fitness
0.1
1
10
1
5
8
Days PBM
PKAC3
MPK15
MPK10
EF2A2
PK4
CK2A1
CRK8
TOR3
LmxM.21.0270
LmxM.33.3020
LmxM.06.1180
LmxM.30.1530
LmxM.36.0910
LmxM.07.0250
LmxM.34.2870
LmxM.16.0300


6. a
Flagella:Cell body ratio
1.5
1.0
0.5
0.0
p<0.0001
p<0.0001
p<0.0001
p<0.0001
Cas9T7
Δulk4
Δstk36
Δ29.0600
Δ07.0900
Δ02.0570
Cas9T7
Δulk4
Δstk36
Δ29.0600
Δ07.0900
Δ02.0570
b
Motile Fraction
0.6
0.5
0.4
0.3
0.2
0.1
0
Control 1
Δulk4
Δstk36
Δ29.0600
Δ02.0570
c
ULK4
Kinase
ARM
ARM
ARM
STK36
Kinase
ARM
ARM
LmxM.29.0600
Kinase
LmxM.02.0570
Kinase
TMHMM
d
ULK4
STK36
LmxM.29.0600
LmxM.02.0570
e
% Barcode proportions
Δulk4
Δstk36
Δ29.0600
Δ02.0570
0.1
0.01
0.001
0.0001
PRO 0h
PRO 24h
PRO 48h
PRO 168h
META
FP 3w
FP 6w
f
Fitness
[Normalised reads (Time 0)/ Normalised reads (x)]
0.1
1
10
Control 1
Δulk4
Δstk36
Δ29.0600
Δ02.0570
Day 1 Slow
Day 5 Slow
Day 8 Slow

7.


Procyclic promastigote
Metacyclic promastigote
Amastigote
Intrinsic
AMPKβ
ATM
CRK7
LmxM.06.0640
LmxM.14.1070
STK36
PI4P5K
ULK4
Extrinsic
LmxM.15.1200
LmxM.19.1610
LmxM.25.1560
MPK1
MPK2
MPK9
PI4K
AKB1
EIF2AK2
CK2A1
CK2A2
CRK7
CRK8
LmxM.06.0640
LmxM.06.1180
LmxM.07.0250
LmxM.07.0690
LmxM.16.0300
LmxM.21.0270
LmxM.25.1560
LmxM.29.2130
LmxM.30.1530
LmxM.32.1710
LmxM.33.3020
LmxM.36.0910
LmxM.34.2870
MPK1
MPK2
STK36
MPK10
MPK15
MRK1
PKAC3
TOR3
ULK4

S1 a



b

Leishmaniinae specific

Required kinase

Pseudo kinase - missing all three residues

K̸ D̸ D̸

1. LmxM.21.0150 (CAMK)
2. LmxM.22.1170 (CAMK)
3. LmxM.01.0750 (CMGC)
4. LmxM.30.1830 (STE)
5. LmxM.30.1840 (STE)
6. LmxM.34.4620 (Pseudo orphan)
7. LmxM.08.0660 (Pseudo orphan)
8 LmxM.26.2060 (Pseudo orphan)
9. LmxM.28.1650 (Pseudo orphan)
10. LmxM.34.2870 (Other)
11. LmxM.22.1150 (Other)

Kinases missing 1 or more important resisdues

K̸ D̸ D

12. LmxM.17.0670 (CMGC)

K̸ D D̸

13. LmxM.04.0650 (CAMK)
14. LmxM.26.2110 (Orphan)
15. LmxM.20.1340 (Other)

K D D̸

16. MPK15 (CMGC)
17. LmxM.36.4250 (CMGC)
18. LmxM.19.1610 (STE)
19. LmxM.16.0870 (Orphan)
20. LmxM.20.1330 (Other)
21. LmxM.15.0770 (Other)

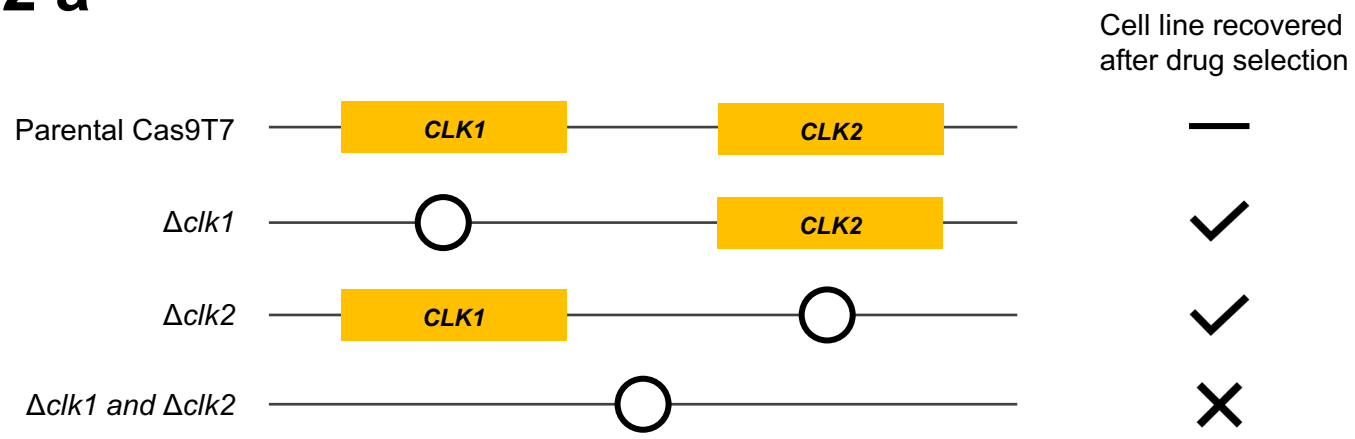
K D̸ D

22. LmxM29.1000 (AGC)
23. LmxM.33.2019 (Pseudo orphan)
24. AUK3 (Other)

K̸ D D

25. LmxM.34.0490 (CAMK)
26. CK1.4 (CK1)
27. LmxM.25.1580 (CK1)
28. LmxM.33.3020 (CK1)
29. CRK10 (CMGC)
30. CRK11 (CMGC)
31. LmxM.25.1560 (CMGC)
32. LmxM.19.0360 (CMGC)
33. LmxM.22.0490 (CMGC)
34. LmxM.03.0210 (CMGC)
35. LmxM.13.0780 (CMGC)
36. LmxM.12.0130 (CMGC)
37. LmxM.31.1020 (STE)
38. LmxM.04.0440 (STE)
39. LmxM.14.1300 (STE)
40. LmxM.32.1400 (STE)
41. LmxM.07.0250 (STE)
42. LmxM.36.2290 (Other)
43. LmxM.02.0570 (Orphan)
44. LmxM.03.0350 (Orphan)
45. LmxM.10.0830 (Orphan)
46. LmxM.31.1290 (Pseudo-orphan)
47. LmxM.11.0230 (Other)
48. ULK4 (Other)
49. LmxM.28.1760 (Other)
50. LmxM.21.0823 (Other)
51. LmxM.36.2630 (Other)

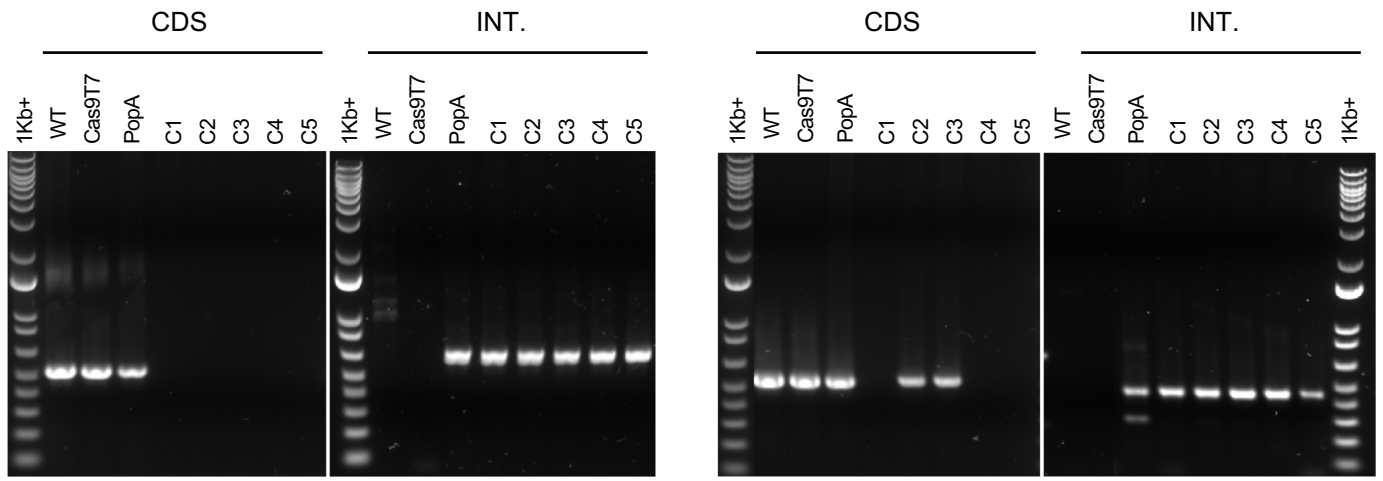
S2 a



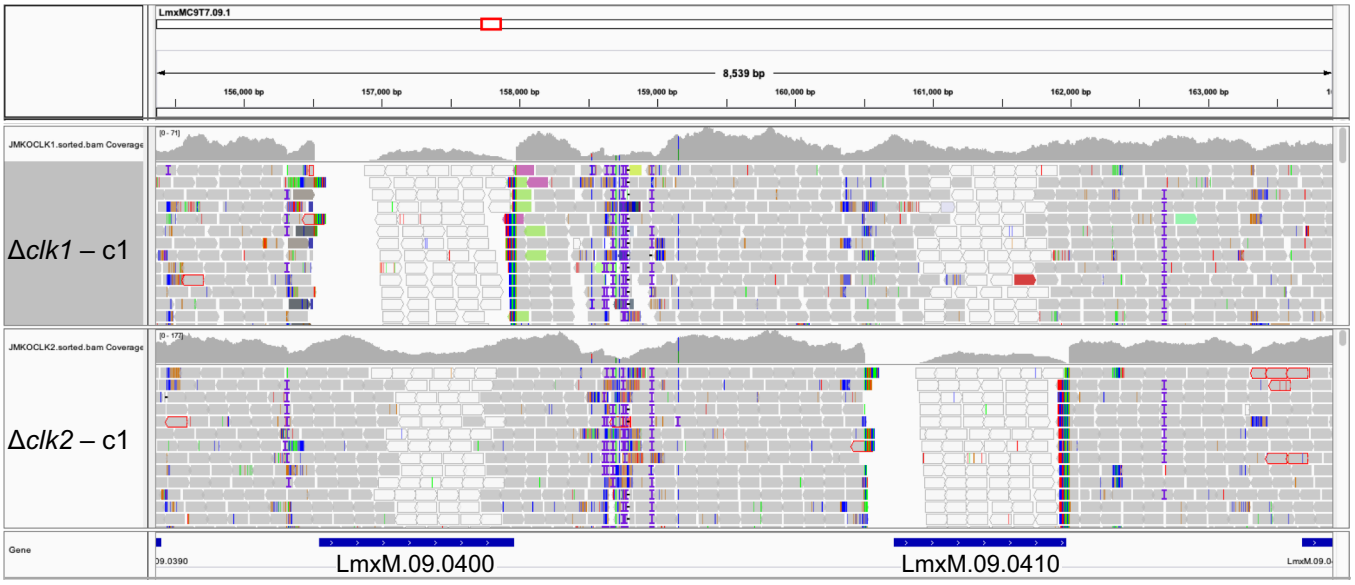
b

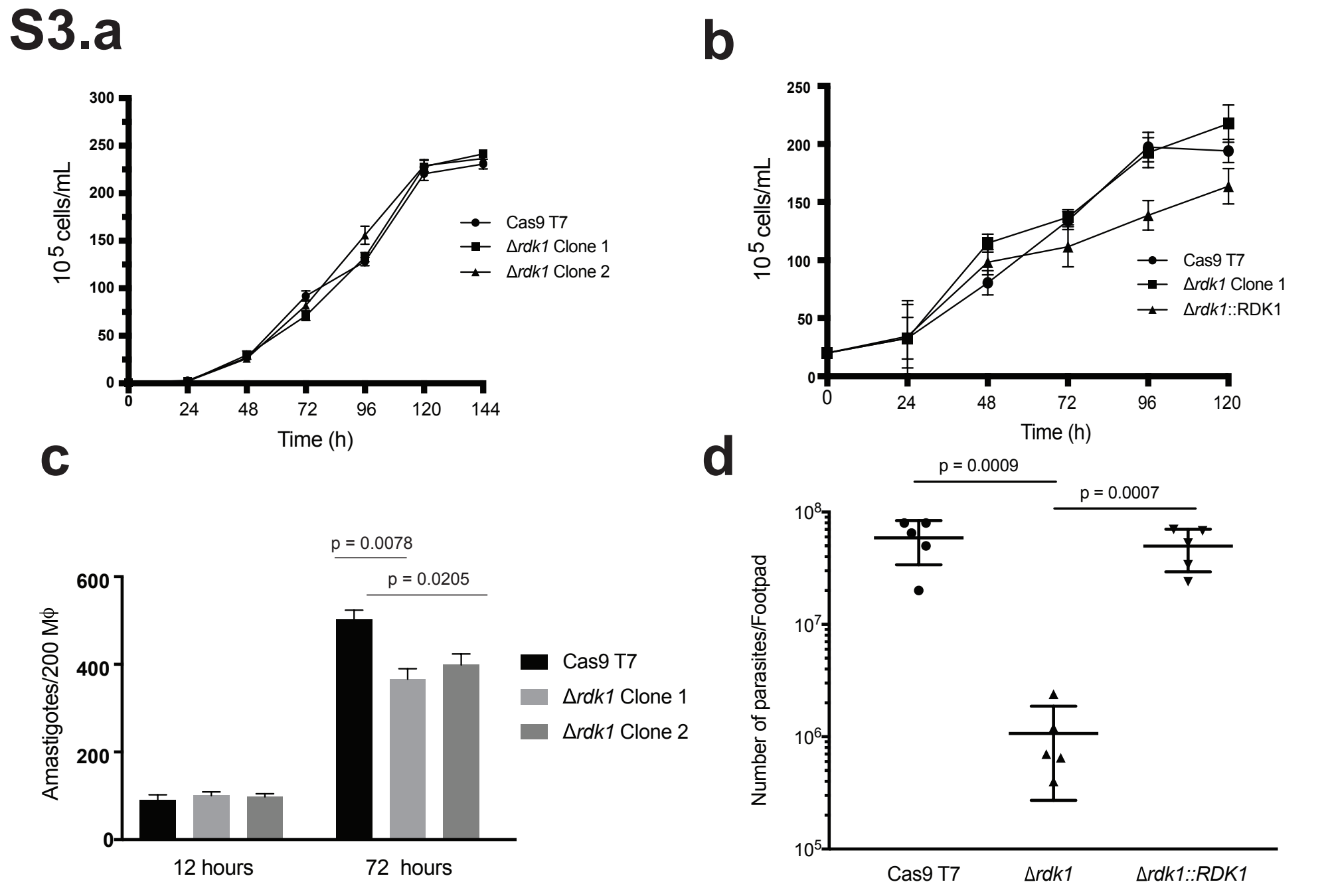
LmxM.09.0400 [CLK1]

LmxM.09.0410 [CLK2]



c

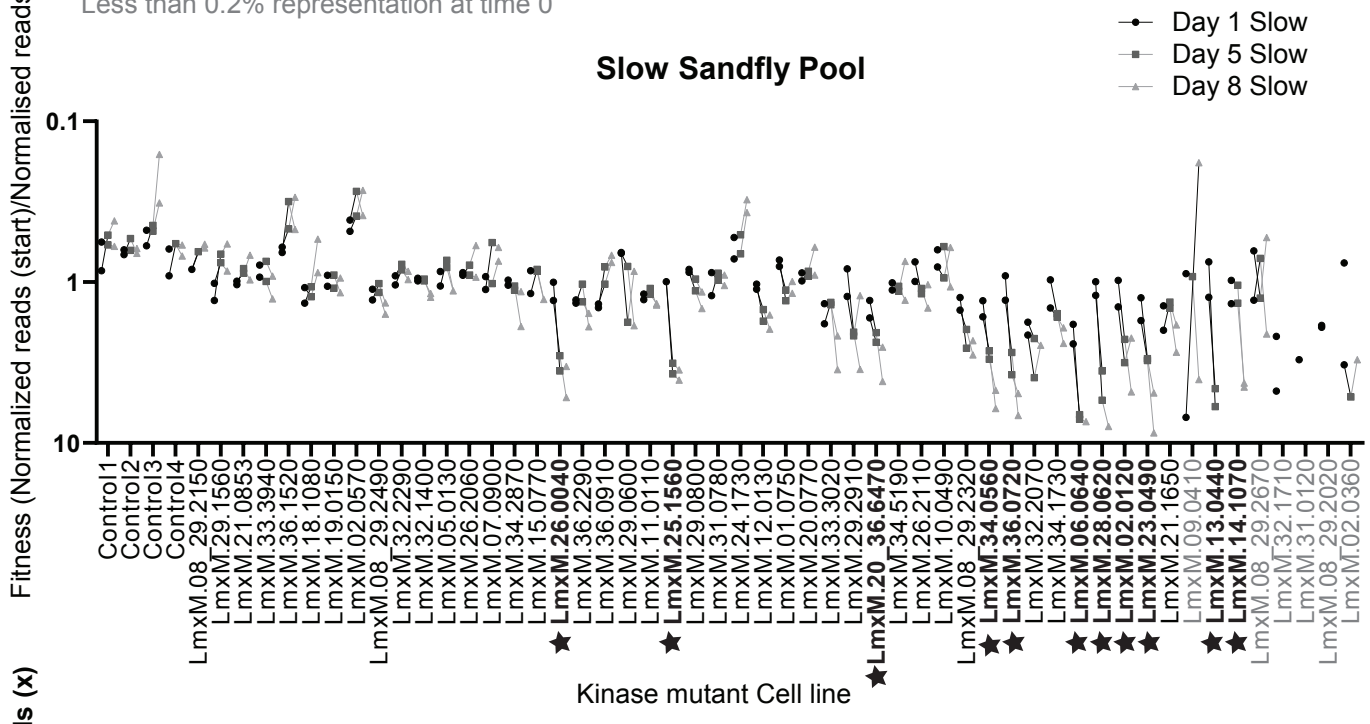


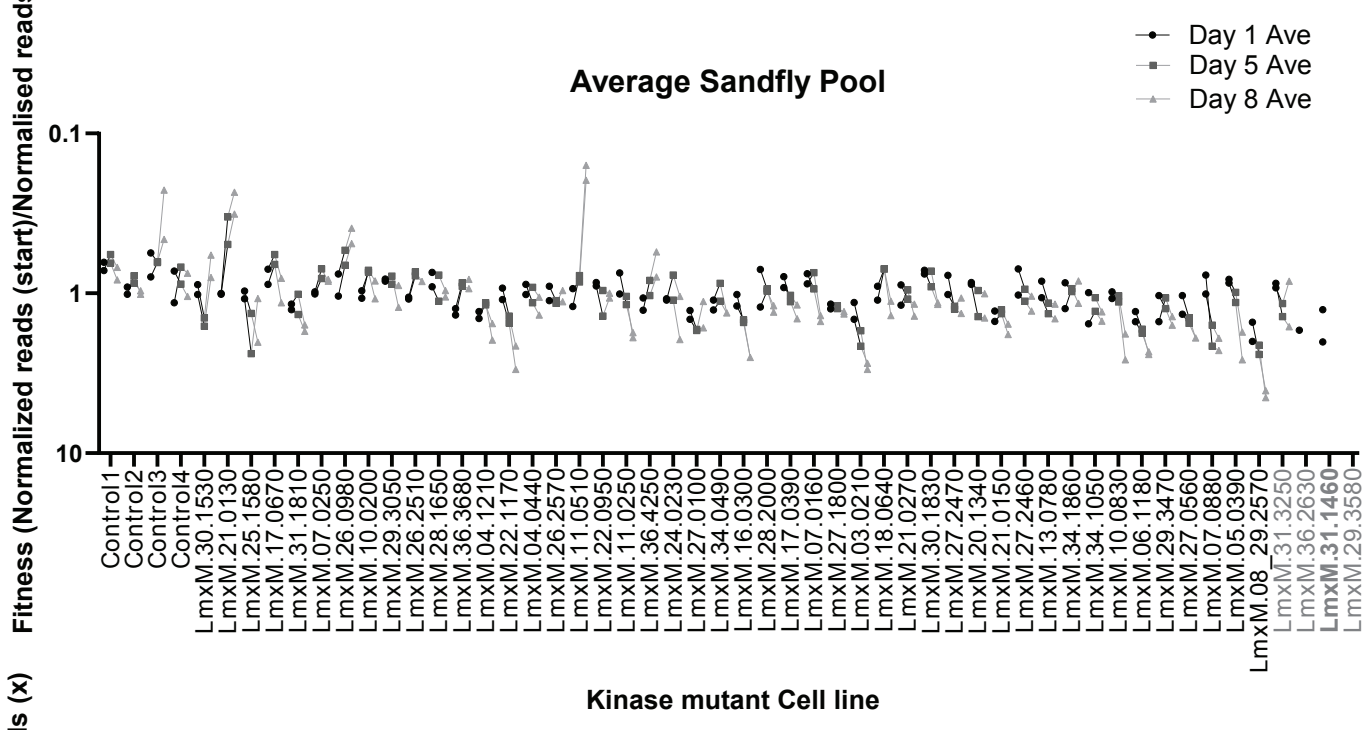

S3.a
300
250
200
150
100
50
0
10^5 cells/mL
0
24
48
72
96
120
144
Time (h)
Cas9 T7
Δrdk1 Clone 1
Δrdk1 Clone 2
b
250
200
150
100
50
0
10^5 cells/mL
0
24
48
72
96
120
Time (h)
Cas9 T7
Δrdk1 Clone 1
Δrdk1::RDK1
c
p = 0.0078
p = 0.0205
600
400
200
0
Amastigotes/200 MΦ
12 hours
72 hours
Cas9 T7
Δrdk1 Clone 1
Δrdk1 Clone 2
d
p = 0.0009
p = 0.0007
10^8
10^7
10^6
10^5
Number of parasites/Footpad
Cas9 T7
Δrdk1
Δrdk1::RDK1

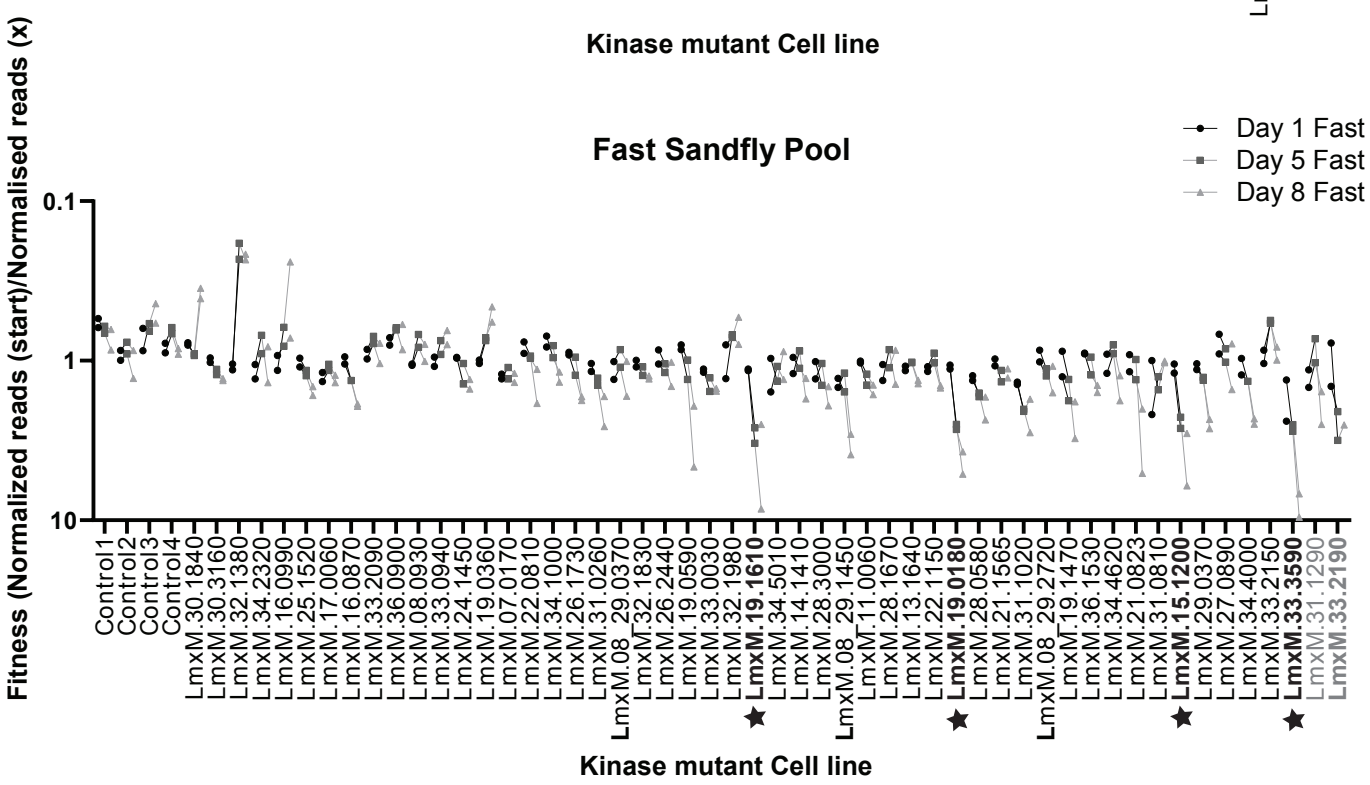
S4.

Significant decrease in fitness ★

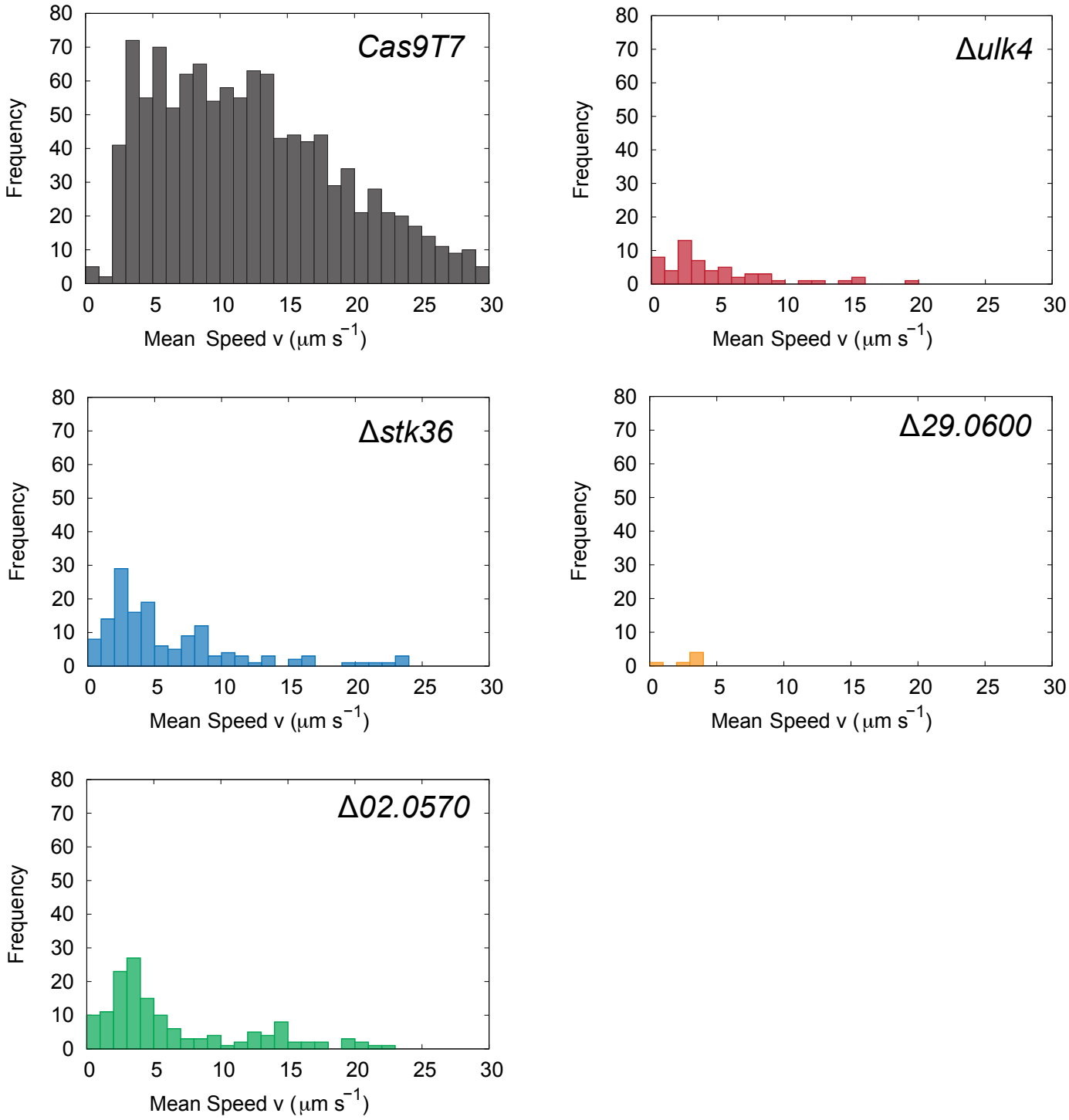
Less than 0.2% representation at time 0







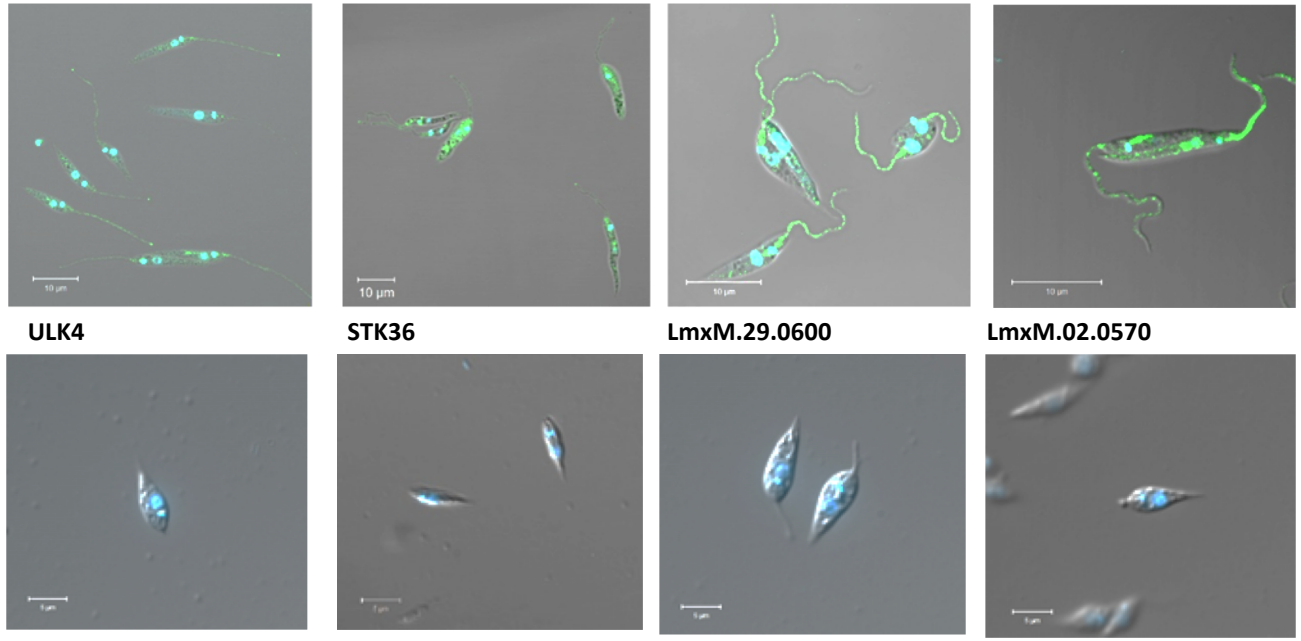
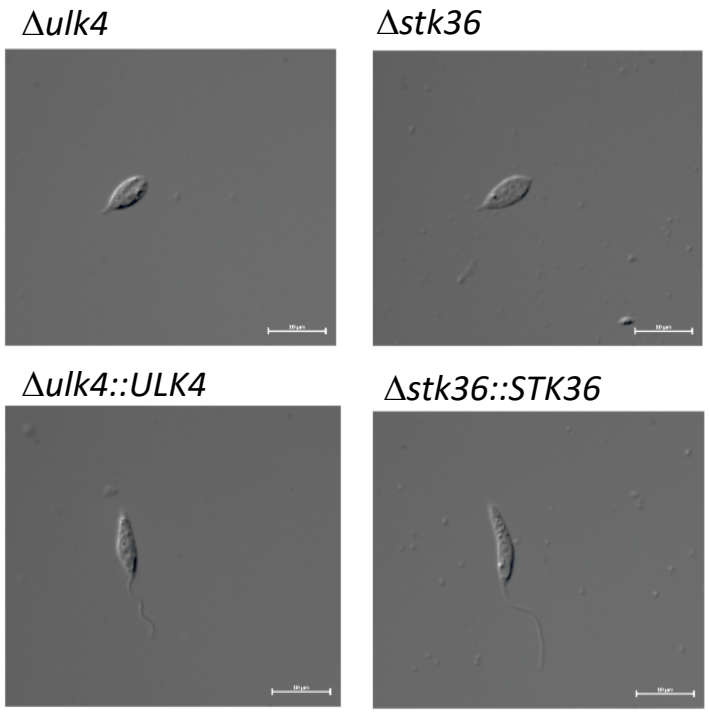
S5.


Cas9T7
Δulk4
Δstk36
Δ29.0600
Δ02.0570
Frequency
Mean Speed v (μm s^{-1})
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25
30

S6 a.

Accession	Gene Name	Mean Flagella: Body ratio (n=50)	S.D. of Flagella: Body ratio	Mean cell body width (µm)(n = 50)	S.D. of cell body width	Phenotype	Exp. identified	Significant LOF in sand fly?
Cas9T7		1.023	0.302	1.978	0.329			No
LmxM.28.0620	*ULK4*	0.218	0.109	1.875	0.427	Short flagellum	Pool 1 and 2	Yes
LmxM.13.0440	*STK36*	0.061	0.027	2.287	0.347	Short flagellum	Pool 1 and 2	Yes
LmxM.29.0600		0.327	0.126	2.36	0.337	Short flagellum	Pool 1 and 2	No
LmxM.07.0900		1.182	0.380	3.276	0.746	Paralysed	Pool 1 and 2	No
LmxM.02.0570		0.316	0.296	3.472	0.428	Short flagellum	Pool 1 and 2	No
LmxM.36.1520		0.914	0.378	4.838	0.779	Uncoordinated	Pool 1	No
LmxM.29.1000		1.901	0.490	2.249	0.421	Long flagellum, paralysed	Pool1	Not in Pool 2

b



c