Causes and Effects of Loss of Classical Nonhomologous End Joining Pathway in Parasitic Eukaryotes

Anna Nenarokova,[a,b] Kristína Záhonová,[a,c] Marija Krasilnikova,[d] Ondřej Gahura,[a] Richard McCulloch,[d] Alena Zíková,[a,b] Vyacheslav Yurchenko,[e,f] Julius Lukeš[a,b]

[a]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic
[b]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic
[c]Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Prague, Czech Republic
[d]Wellcome Centre for Molecular Parasitology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, Scotland
[e]Martsinovsky Institute of Medical Parasitology, Sechenov University, Moscow, Russia
[f]Life Science Research Centre and Institute of Environmental Technologies, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

ABSTRACT We report frequent losses of components of the classical nonhomologous end joining pathway (C-NHEJ), one of the main eukaryotic tools for end joining repair of DNA double-strand breaks, in several lineages of parasitic protists. Moreover, we have identified a single lineage among trypanosomatid flagellates that has lost Ku70 and Ku80, the core C-NHEJ components, and accumulated numerous insertions in many protein-coding genes. We propose a correlation between these two phenomena and discuss the possible impact of the C-NHEJ loss on genome evolution and transition to the parasitic lifestyle.

IMPORTANCE Parasites tend to evolve small and compact genomes, generally endowed with a high mutation rate, compared with those of their free-living relatives. However, the mechanisms by which they achieve these features, independently in unrelated lineages, remain largely unknown. We argue that the loss of the classical nonhomologous end joining pathway components may be one of the crucial steps responsible for characteristic features of parasite genomes.

KEYWORDS DNA repair, genome size, parasite

While DNA integrity and genome stability are crucial for all living organisms, they are permanently challenged by various factors causing DNA damage. The most deleterious DNA lesions are double-strand breaks (DSBs), since accurate repair of one strand using the other one as a template, as occurs in other types of DNA damage, is not possible in this case. To fix such an extreme type of damage, cells have evolved repair mechanisms known as homologous recombination (HR) and nonhomologous end joining (NHEJ).

HR, which relies on the presence of a homologous intact template, starts with 5′-to-3′ resection at the DSB, producing 3′ overhangs usually longer than 100 nucleotides. At least one of the single strand ends invades the homologous region of an intact chromosome, preferentially the sister chromatid (1). This strand invasion of single-stranded DNA into a template sequence produces a displacement loop (D-loop) and is mediated by recombinases of the RecA/Rad51/RadA family, found in all three domains of life (2). Upon invasion, the free 3′ end of the strand is then extended by DNA polymerase(s). Subsequent steps diverge into one of the three pathways with various mutagenic potentials: (i) the double Holliday junction (dHJ) pathway engages both ends of the DSB and can lead to sequence crossover between the broken and intact molecules, (ii) synthesis-dependent strand annealing initially involves only one

DSB end, and (iii) break-induced replication employs only one end of the break and can copy many kilobases from the donor sequence (3–5).

In contrast to HR, NHEJ repairs a DSB by religating the broken ends without engaging an unbroken homologous template. It is divided into two main types, classical (C-NHEJ) and alternative (A-NHEJ) NHEJ. Unlike A-NHEJ, C-NHEJ has no enzymatic overlap with HR and in mammals is directed by five core components: Ku70/Ku80 heterodimer (Ku), DNA-dependent protein kinase catalytic subunit (DNA-PKcs), DNA ligase IV (Lig4), and the XRCC4 and XLF proteins (6–8). The Ku heterodimer first recognizes and binds a DSB in a sequence-independent manner, preventing extensive DSB end resection and serving as a scaffold on which other components of the C-NHEJ machinery are subsequently assembled.

Ku recruits DNA-PKcs, with which it forms a stable complex, tethers the broken DNA ends, and blocks access of other proteins. The lesion is processed, and DNA ends are sealed by the Lig4-XRCC4-XLF complex. Depending on the type of DNA end (overhang or blunt end), other factors (such as the endonuclease Artemis and DNA polymerases) and processes (end resection and DNA synthesis) may also be involved in this repair mechanism (6, 7).

The C-NHEJ machinery is conserved from bacteria to higher eukaryotes, although the levels of conservation of its components differ. In eukaryotes, the Ku heterodimer and Lig4 represent its core. Other components are less conserved and may even be absent. While retained in animals (9, 10), DNA-PKcs is absent in the yeast *Saccharomyces cerevisiae*, in which its roles are carried out by the MRX complex (11). Whether the absence of DNA-PKcs results in a reduced use of C-NHEJ is unclear, though yeasts certainly use HR as the main mechanism for DSB repair (12). Bacterial C-NHEJ employs a reduced enzymatic machinery, which comprises a Ku homodimer, homologous to eukaryotic Ku70 and Ku80, and a DNA ligase often fused to other functional domains (13–16). C-NHEJ in Archaea also utilizes a Ku homodimer, but with a different DNA ligase, DNA polymerase, and phosphodiesterase, all of which nonetheless appear closely related to their bacterial homologues (17).

Although the C-NHEJ pathway is often considered more error-prone than the HR pathway, this view has been challenged recently by emerging evidence that the latter can often be erroneous as well, especially in large and repetitive genomes (3, 18), whereas the C-NHEJ is often robust and accurate (19). However, such fidelity does not apply to the A-NHEJ pathways, named microhomology-mediated end joining (MMEJ) and single-strand annealing (SSA). Both are frequently associated with deletions, since they rely on short regions of homology around a DSB, revealed by more extensive DSB processing than in the case of C-NHEJ. The SSA pathway is independent of Rad51 but operates by annealing 25- to 400-bp-long stretches of high sequence homology in a Rad52-dependent reaction, suggesting at least some functional overlap with the HR machinery (3–5). Since such long stretches of homology are relatively rare, SSA normally generates large deletions around the DSB and is often associated with tandem repeats. MMEJ also results in deletions (20), but the shorter lengths of homology needed for strand annealing, allied to the reaction's tolerance of mismatches, ensure that deletions are normally less extensive. However, the same substrate requirements also imply that MMEJ can cause translocations, as well as complex deletions/insertions, where insertions are usually 2- to 30-bp-long, reiterating either adjacent or distant sequences (21, 22).

In metazoans, MMEJ is facilitated by poly(ADP-ribose) polymerase 1 (23), while DSB recognition requires additional proteins. Six- to 20-bp-long microhomologies are used to allow annealing around the processed DSB (24, 25), the overhangs are cleaved off, and single-stranded gaps are filled in and ligated by DNA ligases I and/or III (26, 27). Another key component of metazoan MMEJ is DNA polymerase theta (Pol θ), which possesses both polymerase and helicase domains, tethers DSB ends, anneals the broken ends at microhomology sites, and synthesizes DNA in template-dependent and -independent manners to allow DSB religation (21, 28–31). Despite this central role in
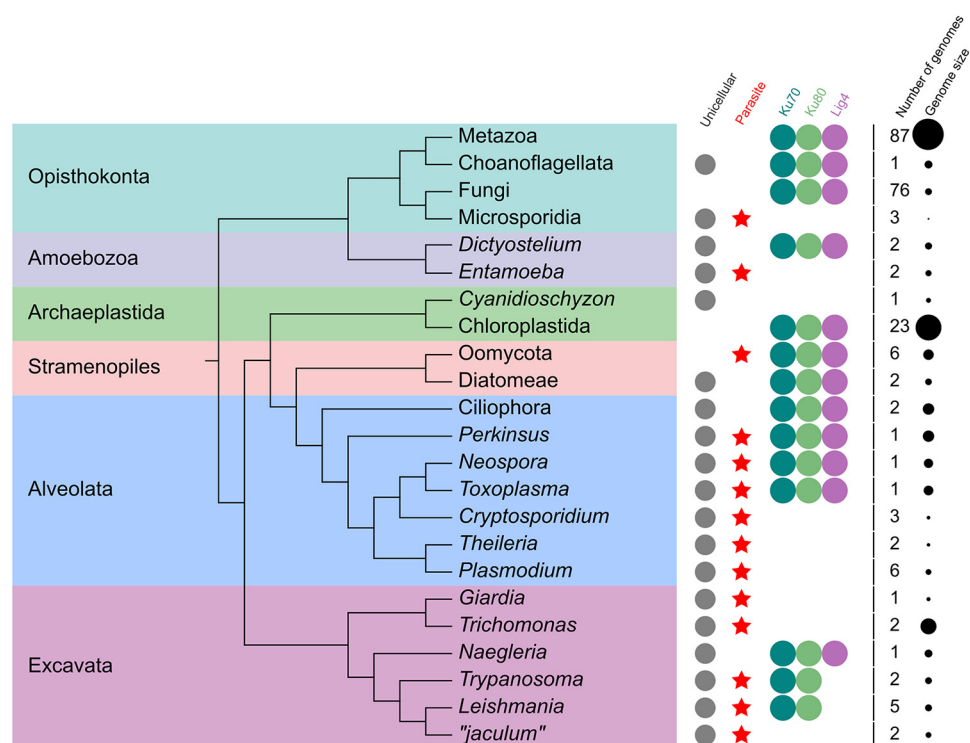
FIG 1 Distribution of main C-NHEJ components across eukaryotes. Median genome size is represented as black circles of corresponding size.

MMEJ, Pol θ is not present in all organisms. For example, yeasts employ other polymerases for this purpose (32).

The HR pathway predominates in the S and G_2 phases of the cell cycle, when newly replicated, homologous sister chromatids are present. In contrast, Ku-dependent C-NHEJ operates during the whole cell cycle, being the major DSB repair mechanism in multicellular eukaryotes (12, 33, 34). Whether MMEJ or SSA is limited to specific parts of the cell cycle is unclear.

Parasites tend to lose C-NHEJ. Perhaps because C-NHEJ is not the sole mechanism of end joining in eukaryotes, the pathway has been lost in several lineages (32, 35, 36). Prominent among the organisms lacking C-NHEJ are parasites. The absence of C-NHEJ components has been documented for the human parasitic protists *Trypanosoma* spp. (37), *Plasmodium* spp. (38), and *Encephalitozoon cuniculi* (39). Experimental analysis of DSB repair has shown that only A-NHEJ and not C-NHEJ is used in at least two of these genera (40–45).

To understand the phylogenetic distribution of C-NHEJ across eukaryotes, we searched for the orthologues of Ku70, Ku80, and Lig4, since these are the main widely conserved factors (Fig. 1).

From 230 eukaryotic genomes present in the EggNOG database (the genome of *Aspergillus oryzae*, in which Ku70 and Ku80 were artificially deleted to make HR more effective, was not included), 181, 26, and 3 genomes encoded all three, two, and one component, respectively, and in 20 genomes, all three components were missing (Table S1). The analysis revealed an overall trend of parasitic protists to lack the C-NHEJ pathway. For example, C-NHEJ is lost in microsporidia and *Entamoeba* spp., yet it is retained in free-living fungi (46) and *Dictyostelium* spp. that form their sister clades, respectively. Nonetheless, this rule is not without exceptions. Among apicomplexan parasites, all C-NHEJ components were retained in the genera *Toxoplasma* and *Neospora* yet lost in *Plasmodium*, *Cryptosporidium*, and *Theileria*. Moreover, C-NHEJ is absent in the red alga *Cyanidioschyzon merolae*, the only known free-living protist

lacking it (Fig. 1). Such a sporadic absence of C-NHEJ is most readily explained by multiple independent losses during eukaryotic evolution.

Why parasites? Two important questions arise from the observation that multiple eukaryotic lineages have discarded C-NHEJ. What processes and forces triggered the loss of such an important DNA repair pathway? What consequences might it have for genome stability and structure?

It has been suggested that the distribution of C-NHEJ in bacteria is connected with their life cycle, with the pathway present in species with a prolonged stationary phase (47, 48), during which there is no available sister chromatid to perform HR. This is also consistent with the observed predominance of C-NHEJ in the haploid cells of eukaryotes, as well as in the G_1 or G_0 phase of the cell cycle, when HR cannot be implemented and the cell has to rely on the nonhomologous DSB repair pathways (49, 50). Vice versa, the organisms that divide often and spend long time in the diploid state tend to rely on HR and lose C-NHEJ.

Alternatively, the loss of C-NHEJ may be triggered by an attempt to limit or even eradicate transposons that rely on it for their movement (51). Finally, the patchy distribution of different DSB repair pathways may reflect their relative impact on genome changes. For example, C-NHEJ can be mutagenic, contributing to sequence diversity during maturation of vertebrate immune genes (52). Consequently, the balance between the beneficial and detrimental aspects of C-NHEJ-associated mutagenesis (53) may dictate the need for its loss, facilitating use of the more faithful HR. However, the absence of C-NHEJ also results in a higher dependence on the A-NHEJ pathway, as appears to be the case during DSB repair in trypanosomatids and other organisms without C-NHEJ (40–45, 54, 55). Such prominence of A-NHEJ may become useful because of additional functions that C-NHEJ cannot perform, such as enhanced genome rearrangement, due to the reliance of A-NHEJ on annealing short, imperfect regions of homology. However, at least in the case of trypanosomatids, the extensive synteny of the *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania* genomes (56) argues against the function of A-NHEJ in genome rearrangements, although we cannot exclude its reclusive role in localized genome variation, such as in multigene families (57–59).

Instead, loss of C-NHEJ can be better correlated with reduced genome size. For instance, the chordate *Oikopleura* (54), the red alga *Cyanidioschyzon* (60), and the prokaryote *Mycobacterium leprae* (61) have undergone a process of genome compaction and, unlike their relatives, notably lack C-NHEJ. Similarly, the size range from 8 to 23 Mb of the C-NHEJ-lacking genomes of the apicomplexans *Theileria parva* (62), *Cryptosporidium* spp. (63), and *Plasmodium* spp. (64) is significantly smaller than the 80-Mb genome of the related *Toxoplasma gondii* (65) (Fig. 1). The loss of C-NHEJ and subsequent gradual compaction of the genome were also observed in the evolution of microsporidians (46, 66). Importantly, Deng and colleagues associated the genome compaction in *Oikopleura* with the loss of C-NHEJ machinery (54). Consistent with this suggestion, our comparative analysis of eukaryotic genomes lacking and containing C-NHEJ machinery revealed a mean size of 29.2 Mb for the former and 667.9 Mb for the latter, a remarkable difference of >20 times ($P = 1.0 \times 10^{-8}$). While this cannot be the sole explanation of size differences, since the ~165-Mb genome of *Trichomonas vaginalis* (67) also lacks C-NHEJ machinery (although its close relative *Trichomonas tenax* has a genome of only 46 Mb [68]), it is highly plausible that when genome streamlining is advantageous, C-NHEJ tends to be discarded, either due to its dispensability or because this step further accelerates sequence loss.

Selective pressure makes parasites fast, concise, and economic, preferably exceeding their hosts in these parameters. Moreover, compared with their free-living relatives, parasites typically have smaller and streamlined genomes and are more susceptible to gene loss. All this is beneficial, since smaller genomes allow parasites to multiply faster and with lower metabolic costs (69, 70). In this context, we posit that the observed multiple independent losses of the C-NHEJ components in parasitic

lineages provide evidence that loss of this DSB repair mechanism leads to genome compaction and, in turn, provides parasites with a number of selective advantages detailed below.

At a DSB, the Ku heterodimer binds promptly to the broken DNA ends (71), protecting them from further degradation and resection by nucleases, which would lead to deleterious deletions (72). In the absence of C-NHEJ, the organism uses A-NHEJ pathways, such as MMEJ and SSA, which inevitably triggers sequence deletions (20). Moreover, the HR-based break-induced replication and SSA pathways can also produce deletions at the breakpoint flanks (73, 74). Thus, following the loss of C-NHEJ, a eukaryotic genome undergoes chromosome aberrations, including deletions and translocations, leading to loss of genetic material and consequent genome shrinkage (75–77). For instance, it has been experimentally demonstrated that A-NHEJ causes novel indel mutations in *Oikopleura*, and this process was implicated in the mechanism of genome shrinkage (54).

We may speculate about the potential mechanisms behind the genome shrinkage. Keeling and Slamovits considered two principal ways leading to the shrinkage of a genome, which are not mutually exclusive: reduction and compaction (78). Reduction is a process of elimination of some functional elements, such as protein-coding genes, whereas compaction is a process of rearranging the existing functional elements in a denser way, for instance, by removing the parts of the noncoding sequences. Both processes operate in the eukaryote genomes: they can occur together or separately. The smallest known nuclear genomes are those of parasitic microsporidia (2.5 Mb) and nucleomorphs (0.373 Mb). They represent extreme cases of both processes, having the highest gene density and the smallest number of genes among eukaryotes (78).

The physical mechanism of genome shrinkage is the loss of whole chromosomes (aneuploidy) or their parts (deletion mutations). Aneuploidy occurs due to the erroneous cell division when the chromosomes do not distribute correctly between the daughter cells. Large deletions originate as a result of DSB without rejoining, translocation of mobile elements, or erroneous, unequal, and ectopic recombination, such as between repeated regions. It is probable that this recombination is more likely to occur in the noncoding parts of genome, which have more repeated elements than protein-coding sequences, causing genome compaction (79). Small deletions occur as a result of DNA polymerase errors, such as slippering on repeats (80). Comparative studies of various animal genomes showed that on the level of small (<400-bp) indels, deletions prevail over insertions both in the protein-coding genes (81) and in the noncoding sequences (82), which may also lead to gradual loss of genetic material.

Still, we cannot exclude the possibility that loss of the C-NHEJ pathway is not the cause but rather the consequence of genome shrinkage. Even though HR occurs in mammals, C-NHEJ acts as their main DSB repair pathway (12, 33, 34). However, eukaryotes with smaller genomes and functional C-NHEJ, such as *S. cerevisiae*, preferably employ HR for DSB repair (12). There is at least one reason for C-NHEJ being the main DSB repair pathway in large eukaryotic genomes. The search for a homologous sequence during HR occurs across the entire genome, raising the risk of invading homologous ectopic sequences, which is especially high given the abundance of almost identical retrotransposon repeats in such genomes (3, 18, 83). In contrast, HR may be the mechanism of choice in small, nonrepetitive genomes, such as those of most bacteria and some unicellular eukaryotes, including parasites. The dependence of HR on the presence of homologous chromatids implies that during haploid cell cycle stages, organisms without C-NHEJ must rely on other repair pathways, such as MMEJ and/or SSA. However, as mentioned above, these pathways are highly error-prone, with a tendency to generate indel mutations (20, 75, 84–86). While deleterious for free-living eukaryotes, this sloppiness in repair mechanisms may be beneficial for parasites. By depending on these mutagenic pathways, they increase their mutation rate, thus benefiting in the arms race with the host's immune system (69, 70).

The nonrandom loss of the Ku proteins in parasitic lineages might be also associated with function(s) of the heterodimer in telomere maintenance. Ku is known to protect

telomeres from abnormal fusions and has an inhibitory effect on the recombination of normal telomeres. The Ku heterodimer also controls telomere length by recruiting telomerase and is involved in the telomere silencing effect (87–89). Furthermore, chromosomal ends and adjacent subtelomeric regions are of particular importance for parasites, as this is where factors involved in host cell interaction and immune escape mechanisms are frequently located (90, 91). Genes specifying variant surface antigens that allow parasites to evade the hosts' immune response are often found in the (sub)telomeric regions. Such surface variation systems are known for *Plasmodium* and *Babesia* spp. (64, 92), *T. brucei* (93, 94), and the fungus *Pneumocystis carinii* (95). Similar strategies have also been described for several prokaryotic pathogens, such as *Neisseria* spp. (96), *Haemophilus influenzae* (97), and *Borrelia* spp. (98). Importantly, variation of these polymorphic and fast-evolving surface proteins is promoted by DSBs, at least in the case of *T. brucei* (99). In the (sub)telomeric regions of *P. falciparum*, antigenic variation occurs via homologous and ectopic recombination (100–102), which is inhibited by Ku in the organisms that have it (10, 103). In this regard, the retention of Ku in *T. brucei* and other trypanosomatids, in the absence of other C-NHEJ components, is a notable anomaly.

Why is Ku retained in trypanosomatids? The human parasites *Trypanosoma* and *Leishmania* (Trypanosomatida, Kinetoplastida) retain Ku70 and Ku80 (104, 105) but have lost Lig4. This is an unusual combination, since other organisms lacking Lig4 usually also do not possess the Ku proteins (Fig. 1). Recently, we have sequenced and annotated the genomes of two unnamed insect flagellates belonging to the "*jaculum*" clade, a novel trypanosomatid lineage (106, 107); the raw sequencing data and the draft assembly are available at NCBI (www.ncbi.nlm.nih.gov) under BioProject PRJNA543408. Their genome sizes are 19.8 Mb and 24.9 Mb in the draft genome assemblies, and the numbers of predicted proteins are 6,163 and 7,571, correspondingly. Unexpectedly, unlike for other trypanosomatids, both Ku genes were ablated from these genomes, proving that the Ku heterodimer is not indispensable for these organisms. Interestingly, a detailed inspection of the genomes of both "*jaculum*" species revealed a high frequency of specific insertions in protein-coding genes, while deletions were rare (Fig. 2; see also Fig. S1 in the supplemental material). Since "*jaculum*" is not a basal trypanosomatid clade, but rather one from the crown (106, 107), and the insertions are specific for this group, the most parsimonious scenario is that the insertions appeared *de novo* in the common ancestor of "*jaculum*."

Insertions were present in the majority of examined coding sequences, although they were underrepresented or completely absent from the most conserved genes, such as ribosomal proteins and glycolytic enzymes (Table S2). In 247 analyzed alignments in the two "*jaculum*" species, inserted sequences constituted 14.9% and 17.4% of the alignments, whereas in *T. brucei* only 8.9% of the alignment were represented by insertions ($P_1 = 4.3 \times 10^{-11}$; $P_2 = 1.4 \times 10^{-13}$) (Table S2). We compared the amino acid compositions of insertions and sequences without insertions, and we found that some amino acids were overrepresented or underrepresented in the inserted sequences; however, this pattern was similar in all the analyzed species (Table S3). Mass spectrometry confirmed that the insertions were indeed retained in mature proteins (Fig. 2 and Fig. S1).

Next, we investigated whether the observed insertions are neutral with respect to the function of the affected proteins. For that purpose, we mapped the insertions in selected conserved "*jaculum*" proteins on experimentally determined structures of their orthologues in *T. brucei* (Fig. 3). The inspected insertions either formed terminal extensions or were located to the external loops, but they never occurred in regions involved in ligand binding, ion coordination, or interaction with other molecules. This observation is fully consistent with the hypothesis that all insertions are functionally neutral.

We propose that the observed features are a consequence of the loss of the Ku heterodimer. Moreover, our data suggest an additional, so far unexplored, role(s) of Ku
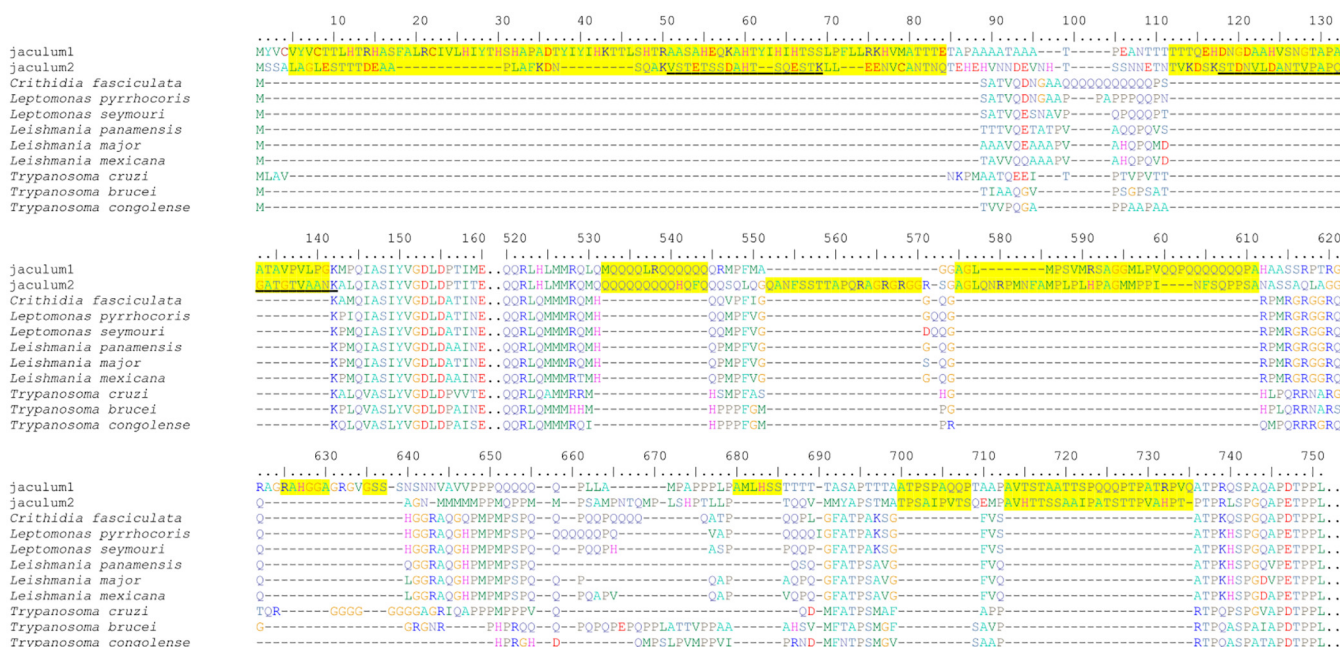
FIG 2 Multiple insertions are present in "*jaculum*" proteins. The N-terminal part of the poly(A)-binding protein alignment of chosen trypanosomatids is shown (full-length alignment is available in Fig. S1). Insertions present in "*jaculum*" proteins are highlighted by yellow background. Peptides identified by mass spectrometry are underlined in black. Two dots represent regions of the sequence alignment that are conserved among the species and were omitted for simplicity.

in trypanosomatid parasites. In all examined species, with the sole exception of the "*jaculum*" lineage, Lig4 is absent but both Ku70 and Ku80 are retained (Fig. 1). Data available from *Trypanosoma cruzi*, *T. brucei*, and *Leishmania* spp. indicate that the Ku heterodimer does not participate in C-NHEJ and that in the corresponding genomes DSBs are predominantly repaired via HR and MMEJ (37, 43–45, 108). However, it is possible that the Ku70/80 complex plays a role in DSB repair even without its partner Lig4, because it may act as "first aid," binding within seconds to the disrupted DNA ends (71), holding them together and protecting them from further damage until the slower HR or A-NHEJ proteins come to serve. Such a role may be important in *Leishmania* spp. and *T. brucei*, in which pronounced levels of genome rearrangements are observed, either genome-wide or in the subtelomeric region for immune evasion, and might involve DNA DSBs (109, 110). Alternatively, Ku70 and Ku80 are involved in other DNA repair pathways, such as base excision and DNA alkylation repair (111), although a role for Ku in these processes has so far not been examined in trypanosomatids. Moreover, together with the MRN complex, the Ku heterodimer may serve as a signaling molecule, modulating activity of the ATM kinase, which phosphorylates other factors and initiates a signaling cascade in the DNA damage response pathway (10). Again, the function of the ATM kinase has not yet been scrutinized in trypanosomatids. Finally, the Ku proteins play an important role in telomere maintenance (104, 105, 112). Data obtained from the analysis of the "*jaculum*"genomes may shed light on the genome-wide roles of these conserved and multifunctional proteins not only in trypanosomatids but also in other eukaryotes.

Taking the alternative end joining pathways into consideration may give us a hint regarding the origin of the insertions that are prominent in "*jaculum*." In metazoan MMEJ, DNA polymerase θ uses only one to four complementary nucleotides to initiate polymerization, frequently producing short templated and nontemplated insertions (113, 114), reminiscent of those pervading the "*jaculum*" genome. We consider as highly plausible a hypothesis that in the "*jaculum*" trypanosomatids, the insertions may result from the erroneous A-NHEJ and HR repair processes, unconstrained by the Ku proteins. Similarly, in tunicate *Oikopleura dioica*, which lacks Ku70/80 and other components of C-NHEJ, DSB repair by A-NHEJ results in acquisition of multiple novel insertions (54).
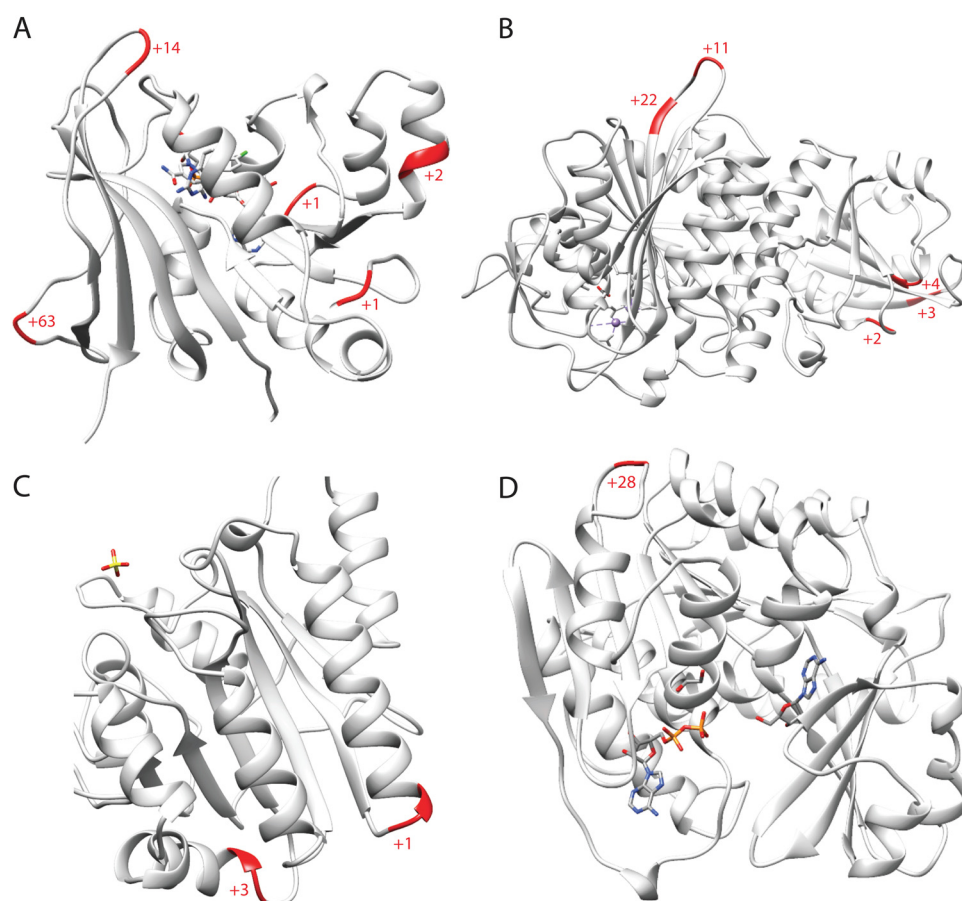
FIG 3 Mapping of insertions in the *"jaculum"* proteins onto structures of dihydrofolate reductase in complex with pyrimethamine (118) (A), leucyl aminopeptidase (119) (B), the phosphatase domain of phosphoglycerate mutase (120) (C), and adenosine kinase in complex with adenosine and AMPPNP (121) (D) from *T. brucei*. The positions and lengths of insertions in the *"jaculum"* proteins are shown in red.

An interesting question is why the observed insertions in "*jaculum*" and other trypanosomatids were significantly prevalent over deletions (Fig. 2 and Fig. S1). It is known that insertions in protein-coding sequences are usually several times more frequent than deletions, apparently because the latter are generally more deleterious and more susceptible to purifying selection (115). We also noticed that amino acids are predominantly altered in the flanking regions of the insertions and may represent remnants of the deletions, rendering these parts of the alignment to be inaccurately aligned. Moreover, the lengths of the inserted region are often variable in different species, which may be explained by consequent insertions and deletions (Fig. 2 and Fig. S1).

A comparably high incidence of indel mutations, accompanied by loss of all main C-NHEJ components, has been reported for the causative agent of human malaria, *Plasmodium falciparum* (42) (Fig. 1). In this protist, the occurrence of indels is over 10-fold higher than that of base substitutions (116). It is therefore worth pointing out that in most other organisms, base substitutions are much more frequent than indels. For example, the substitution-to-indel ratios are approximately 10:1 in primates and 20:1 in bacteria (117). While *P. falciparum* is known to be a highly polymorphic and fast-evolving parasite (116), these features are so far not associated with the absence of C-NHEJ. The above-described circumstantial evidence makes the putative connection between the DNA repair pathways and the unique features of the *Plasmodium* genomes worth exploring.

Concluding remarks. We have found that the C-NHEJ pathway, which is a highly conserved key eukaryotic DNA repair pathway, has been independently lost multiple

times in several parasitic protist lineages. We provide several alternative explanations for these seemingly nonrandom losses. Moreover, we raise the question of whether parasites benefit from this repair mechanism or, unlike their free-living kin, try to free themselves from its constraints.

SUPPLEMENTAL MATERIAL

ACKNOWLEDGMENTS

REFERENCES

1. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol 18:495–506. https://doi.org/10.1038/nrm.2017.48.

2. Jain S, Sugawara N, Haber JE. 2016. Role of double-strand break end-tethering during gene conversion in *Saccharomyces cerevisiae*. PLoS Genet 12:e1005976. https://doi.org/10.1371/journal.pgen.1005976.

3. Rodgers K, McVey M. 2016. Error-prone repair of DNA double-strand breaks. J Cell Physiol 231:15–24. https://doi.org/10.1002/jcp.25053.

4. Haber JE. 2018. DNA repair: the search for homology. Bioessays 40:1700229. https://doi.org/10.1002/bies.201700229.

5. Kramara J, Osia B, Malkova A. 2018. Break-induced replication: the where, the why, and the how. Trends Genet 34:518–531. https://doi.org/10.1016/j.tig.2018.04.002.

6. Waters CA, Strande NT, Wyatt DW, Pryor JM, Ramsden DA. 2014. Nonhomologous end joining: a good solution for bad ends. DNA Repair (Amst) 17:39–51. https://doi.org/10.1016/j.dnarep.2014.02.008.

7. Williams GJ, Hammel M, Radhakrishnan SK, Ramsden D, Lees-Miller SP, Tainer JA. 2014. Structural insights into NHEJ: building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time. DNA Repair (Amst) 17:110–120. https://doi.org/10.1016/j.dnarep.2014.02.009.

8. Her J, Bunting SF. 2018. How cells ensure correct repair of DNA double-strand breaks. J Biol Chem 293:10502–10511. https://doi.org/10.1074/jbc.TM118.000371.

9. Doré AS, Drake AC, Brewerton SC, Blundell TL. 2004. Identification of DNA-PK in the arthropods: evidence for the ancient ancestry of verte-brate non-homologous end-joining. DNA Repair (Amst) 3:33–41. https://doi.org/10.1016/j.dnarep.2003.09.003.

10. Fell VL, Schild-Poulter C. 2015. The Ku heterodimer: function in DNA repair and beyond. Mutat Res Rev Mutat Res 763:15–29. https://doi.org/10.1016/j.mrrev.2014.06.002.

11. Chen L, Trujillo K, Ramos W, Sung P, Tomkinson AE. 2001. Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. Mol Cell 8:1105–1115. https://doi.org/10.1016/S1097-2765(01)00388-4.

12. Emerson CH, Bertuch AA. 2016. Consider the workhorse: nonhomologous end-joining in budding yeast. Biochem Cell Biol 94:396–406. https://doi.org/10.1139/bcb-2016-0001.

13. Aravind L, Koonin EV. 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res 11:1365–1374. https://doi.org/10.1101/gr.181001.

14. Della M, Palmbos PL, Tseng H-M, Tonkin LM, Daley JM, Topper LM, Pitcher RS, Tomkinson AE, Wilson TE, Doherty AJ. 2004. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. Science 306:683–685. https://doi.org/10.1126/science.1099824.

15. Shuman S, Glickman MS. 2007. Bacterial DNA repair by non-homologous end joining. Nat Rev Microbiol 5:852–861. https://doi.org/10.1038/nrmicro1768.

16. Pitcher RS, Brissett NC, Doherty AJ. 2007. Nonhomologous end-joining in bacteria: a microbial perspective. Annu Rev Microbiol 61:259–282. https://doi.org/10.1146/annurev.micro.61.080706.093354.

17. Bartlett EJ, Brissett NC, Doherty AJ. 2013. Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. Proc Natl Acad Sci U S A 110:E1984–E1991. https://doi.org/10.1073/pnas.1302616110.

18. Malkova A, Haber JE. 2012. Mutations arising during repair of chromosome breaks. Annu Rev Genet 46:455–473. https://doi.org/10.1146/annurev-genet-110711-155547.

19. Bétermier M, Bertrand P, Lopez BS. 2014. Is non-homologous end-joining really an inherently error-prone process? PLoS Genet 10:e1004086. https://doi.org/10.1371/journal.pgen.1004086.

20. Seol J-H, Shim EY, Lee SE. 2018. Microhomology-mediated end joining: good, bad and ugly. Mutat Res 809:81–87. https://doi.org/10.1016/j.mrfmmm.2017.07.002.

21. Koole W, van Schendel R, Karambelas AE, van Heteren JT, Okihara KL, Tijsterman M. 2014. A polymerase theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. Nat Commun 5:3216. https://doi.org/10.1038/ncomms4216.

22. Wood RD, Doublié S. 2016. DNA polymerase θ (POLQ), double-strand break repair, and cancer. DNA Repair (Amst) 44:22–32. https://doi.org/10.1016/j.dnarep.2016.05.003.

23. Robert I, Dantzer F, Reina-San-Martin B. 2009. Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. J Exp Med 206:1047–1056. https://doi.org/10.1084/jem.20082468.

24. Rass E, Grabarz A, Plo I, Gautier J, Bertrand P, Lopez BS. 2009. Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. Nat Struct Mol Biol 16:819–824. https://doi.org/10.1038/nsmb.1641.

25. Lee-Theilen M, Matthews AJ, Kelly D, Zheng S, Chaudhuri J. 2011. CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. Nat Struct Mol Biol 18:75–80. https://doi.org/10.1038/nsmb.1942.

26. Ma J-L, Kim EM, Haber JE, Lee SE. 2003. Yeast Mre11 and Rad1 proteins

define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. Mol Cell Biol 23:8820–8828. https://doi.org/10.1128/mcb.23.23.8820-8828.2003.

27. Paul K, Wang M, Mladenov E, Bencsik-Theilen A, Bednar T, Wu W, Arakawa H, Iliakis G. 2013. DNA ligases I and III cooperate in alternative non-homologous end-joining in vertebrates. PLoS One 8:e59505. https://doi.org/10.1371/journal.pone.0059505.

28. He P, Yang W. 2018. Template and primer requirements for DNA Pol θ-mediated end joining. Proc Natl Acad Sci U S A 115:7747–7752. https://doi.org/10.1073/pnas.1807329115.

29. Chan SH, Yu AM, McVey M. 2010. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. PLoS Genet 6:e1001005. https://doi.org/10.1371/journal.pgen.1001005.

30. Yu AM, McVey M. 2010. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res 38:5706–5717. https://doi.org/10.1093/nar/gkq379.

31. Roerink SF, van Schendel R, Tijsterman M, Schendel R, Tijsterman M. 2014. Polymerase theta-mediated end joining of replication-associated DNA breaks in C. elegans. Genome Res 24:954–962. https://doi.org/10.1101/gr.170431.113.

32. Sfeir A, Symington LS. 2015. Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? Trends Biochem Sci 40:701–714. https://doi.org/10.1016/j.tibs.2015.08.006.

33. Pannunzio NR, Watanabe G, Lieber MR. 2018. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J Biol Chem 293:10512–10523. https://doi.org/10.1074/jbc.TM117.000374.

34. Beucher A, Birraux J, Tchouandong L, Barton O, Shibata A, Conrad S, Goodarzi AA, Krempler A, Jeggo PA, Löbrich M. 2009. ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. EMBO J 28:3413–3427. https://doi.org/10.1038/emboj.2009.276.

35. Deriano L, Roth DB. 2013. Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu Rev Genet 47:433–455. https://doi.org/10.1146/annurev-genet-110711-155540.

36. Sallmyr A, Tomkinson AE. 2018. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. J Biol Chem 293:10536–10549. https://doi.org/10.1074/jbc.TM117.000375.

37. Burton P, McBride DJ, Wilkes JM, Barry JD, McCulloch R. 2007. Ku heterodimer-independent end joining in Trypanosoma brucei cell extracts relies upon sequence microhomology. Eukaryot Cell 6:1773–1781. https://doi.org/10.1128/EC.00212-07.

38. Lee AH, Symington LS, Fidock DA. 2014. DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum. Microbiol Mol Biol Rev 78:469–486. https://doi.org/10.1128/MMBR.00059-13.

39. Gill EE, Fast NM. 2007. Stripped-down DNA repair in a highly reduced parasite. BMC Mol Biol 8:24. https://doi.org/10.1186/1471-2199-8-24.

40. Conway C, Proudfoot C, Burton P, Barry JD, McCulloch R. 2002. Two pathways of homologous recombination in Trypanosoma brucei. Mol Microbiol 45:1687–1700. https://doi.org/10.1046/j.1365-2958.2002.03122.x.

41. Glover L, McCulloch R, Horn D. 2008. Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. Nucleic Acids Res 36:2608–2618. https://doi.org/10.1093/nar/gkn104.

42. Kirkman LA, Lawrence EA, Deitsch KW. 2014. Malaria parasites utilize both homologous recombination and alternative end joining pathways to maintain genome integrity. Nucleic Acids Res 42:370–379. https://doi.org/10.1093/nar/gkt881.

43. Peng D, Kurup SP, Yao PY, Minning TA, Tarleton RL. 2014. CRISPR-Cas9-mediated single-gene and gene family disruption in Trypanosoma cruzi. mBio 6:e02097-14. https://doi.org/10.1128/mBio.02097-14.

44. Zhang W-W, Matlashewski G. 2015. CRISPR-Cas9-mediated genome editing in Leishmania donovani. mBio 6:e00861-15. https://doi.org/10.1128/mBio.00861-15.

45. Zhang W-W, Lypaczewski P, Matlashewski G. 2017. Optimized CRISPR-Cas9 genome editing for Leishmania and its use to target a multigene family, induce chromosomal translocation, and study DNA break repair mechanisms. mSphere 2:e00340-16. https://doi.org/10.1128/mSphere.00340-16.

46. Galindo LJ, Torruella G, Moreira D, Timpano H, Paskerova G, Smirnov A, Nassonova E, López-García P. 2018. Evolutionary genomics of Metchnikovella incurvata (Metchnikovellidae): an early branching microspo-

ridium. Genome Biol Evol 10:2736–2748. https://doi.org/10.1093/gbe/evy205.

47. Bowater R, Doherty AJ. 2006. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet 2:e8. https://doi.org/10.1371/journal.pgen.0020008.

48. Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M, Devine SK, Day JP, Wilkinson A, d'Adda di Fagagna F, Devine KM, Bowater RP, Jeggo PA, Jackson SP, Doherty AJ. 2002. Identification of a DNA non-homologous end-joining complex in bacteria. Science 297:1686–1689. https://doi.org/10.1126/science.1074584.

49. Heidenreich E, Novotny R, Kneidinger B, Holzmann V, Wintersberger U. 2003. Non-homologous end joining as an important mutagenic process in cell cycle-arrested cells. EMBO J 22:2274–2283. https://doi.org/10.1093/emboj/cdg203.

50. Karathanasis E, Wilson TE. 2002. Enhancement of Saccharomyces cerevisiae end-joining efficiency by cell growth stage but not by impairment of recombination. Genetics 161:1015–1027.

51. Ivics Z, Izsvák Z. 2015. Sleeping Beauty transposition. Microbiol Spectr 3:853–874.

52. Chaudhuri J, Alt FW. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol 4:541–552. https://doi.org/10.1038/nri1395.

53. Matic I, Taddei F, Radman M. 2004. Survival versus maintenance of genetic stability: a conflict of priorities during stress. Res Microbiol 155:337–341. https://doi.org/10.1016/j.resmic.2004.01.010.

54. Deng W, Henriet S, Chourrout D. 2018. Prevalence of mutation-prone microhomology-mediated end joining in a chordate lacking the c-NHEJ DNA repair pathway. Curr Biol 28:3337–3341.e4. https://doi.org/10.1016/j.cub.2018.08.048.

55. Chayot R, Montagne B, Mazel D, Ricchetti M. 2010. An end-joining repair mechanism in Escherichia coli. Proc Natl Acad Sci U S A 107:2141–2146. https://doi.org/10.1073/pnas.0906355107.

56. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran A-N, Wortman JR, Alsmark UCM, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N. 2005. Comparative genomics of trypanosomatid parasitic protozoa. Science 309:404–409. https://doi.org/10.1126/science.1112181.

57. Weatherly DB, Peng D, Tarleton RL. 2016. Recombination-driven generation of the largest pathogen repository of antigen variants in the protozoan Trypanosoma cruzi. BMC Genomics 17:729. https://doi.org/10.1186/s12864-016-3037-z.

58. Jackson AP, Berry A, Aslett M, Allison HC, Burton P, Vavrova-Anderson J, Brown R, Browne H, Corton N, Hauser H, Gamble J, Gilderthorp R, Marcello L, McQuillan J, Otto TD, Quail MA, Sanders MJ, van Tonder A, Ginger ML, Field MC, Barry JD, Hertz-Fowler C, Berriman M. 2012. Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. Proc Natl Acad Sci U S A 109:3416–3421. https://doi.org/10.1073/pnas.1117313109.

59. Flegontov P, Butenko A, Firsov S, Kraeva N, Eliáš M, Field MC, Filatov D, Flegontova O, Gerasimov ES, Hlaváčová J, Ishemgulova A, Jackson AP, Kelly S, Kostygov AY, Logacheva MD, Maslov DA, Opperdoes FR, O'Reilly A, Sádlová J, Ševčíková T, Venkatesh D, Vlček Č, Volf P, Votýpka J, Záhonová K, Yurchenko V, Lukeš J. 2016. Genome of Leptomonas pyrrhocoris: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of Leishmania. Sci Rep 6:23704. https://doi.org/10.1038/srep23704.

60. Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Nishida K, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii Y, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimizu N, Sugano S, Sato N, Nozaki H, Ogasawara N, Kohara Y, Kuroiwa T. 2004. Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. Nature 428:653–657. https://doi.org/10.1038/nature02398.

61. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honoré N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean

J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG. 2001. Massive gene decay in the leprosy bacillus. Nature 409: 1007–1011. https://doi.org/10.1038/35059006.

62. Gardner MJ, Bishop R, Shah T, De Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M, Wilson RJM, Sato S, Ralph SA, Mann DJ, Xiong Z, Shallom SJ, Weidman J, Jiang L, Lynn J, Weaver B, Shoaibi A, Domingo AR, Wasawo D, Crabtree J, Wortman JR, Haas B, Angiuoli SV, Creasy TH, Lu C, Suh B, Silva JC, Utterback TR, Feldblyum TV, Pertea M, Allen J, Nierman WC, Taracha ELN, Salzberg SL, White OR, Fitzhugh HA, Morzaria S, Venter JC, Fraser CM, Nene V. 2005. Genome sequence of Theileria parva, a bovine pathogen that transforms lymphocytes. Science 309:134–137. https://doi.org/10.1126/science.1110439.

63. Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, Pearson WR, Dear PH, Bankier AT, Peterson DL, Abrahamsen MS, Kapur V, Tzipori S, Buck GA. 2004. The genome of Cryptosporidium hominis. Nature 431:1107–1112. https://doi.org/10.1038/nature02977.

64. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan M-S, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin D, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498–511. https://doi.org/10.1038/nature01097.

65. Kissinger JC, Gajria B, Li L, Paulsen IT, Roos DS. 2003. ToxoDB: accessing the Toxoplasma gondii genome. Nucleic Acids Res 31:234–236. https://doi.org/10.1093/nar/gkg072.

66. Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivarès CP. 2001. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature 414:450–453. https://doi.org/10.1038/35106579.

67. Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao QQ, Wortman JR, Bidwell SL, Alsmark UCM, Besteiro S, Sicheritz-Ponten T, Noel CJ, Dacks JB, Foster PG, Simillion C, Van de Peer Y, Miranda-Saavedra D, Barton GJ, Westrop GD, Müller S, Dessi D, Fiori PL, Ren Q, Paulsen I, Zhang H, Bastida-Corcuera FD, Simoes-Barbosa A, Brown MT, Hayes RD, Mukherjee M, Okumura CY, Schneider R, Smith AJ, Vanacova S, Villalvazo M, Haas BJ, Pertea M, Feldblyum TV, Utterback TR, Shu CLC-L, Osoegawa K, de Jong PJ, Hrdy I, Horvathova L, Zubacova Z, Dolezal P, Malik S-BSB, Logsdon JM, Henze K, Gupta A, Wang CC, Dunne RL, Upcroft JA, Upcroft P, White O, Salzberg SL, Tang P, Chiu CHC-H, Lee YSY-S, Embley TM, Coombs GH, Mottram JC, Tachezy J, Fraser-Liggett CM, Johnson PJ. 2007. Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science 315:207–212. https://doi.org/10.1126/science.1132894.

68. Benabdelkader S, Andreani J, Gillet A, Terrer E, Pignoly M, Chaudet H, Aboudharam G, La Scola B. 2019. Specific clones of Trichomonas tenax are associated with periodontitis. PLoS One 14:e0213338. https://doi.org/10.1371/journal.pone.0213338.

69. Cavalier-Smith T. 2005. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Ann Bot 95:147–175. https://doi.org/10.1093/aob/mci010.

70. Poulin R, Randhawa H. 2015. Evolution of parasitism along convergent lines: from ecology to genomics. Parasitology 142:S6–S15. https://doi.org/10.1017/S0031182013001674.

71. Mari P-O, Florea BI, Persengiev SP, Verkaik NS, Bruggenwirth HT, Modesti M, Giglia-Mari G, Bezstarosti K, Demmers JAA, Luider TM, Houtsmuller AB, van Gent DC. 2006. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. Proc Natl Acad Sci U S A 103:18597–18602. https://doi.org/10.1073/pnas.0609061103.

72. Mimitou EP, Symington LS. 2010. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. EMBO J 29:3358–3369. https://doi.org/10.1038/emboj.2010.193.

73. Bhargava R, Sandhu M, Muk S, Lee G, Vaidehi N, Stark JM. 2018. C-NHEJ without indels is robust and requires synergistic function of distinct XLF

74. Malkova A, Ira G. 2013. Break-induced replication: functions and molecular mechanism. Curr Opin Genet Dev 23:271–279. https://doi.org/10.1016/j.gde.2013.05.007.

75. Boulton SJ, Jackson SP. 1996. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. EMBO J 15:5093–5103. https://doi.org/10.1002/j.1460-2075.1996.tb00890.x.

76. Difilippantonio MJ, Zhu J, Chen HT, Meffre E, Nussenzweig MC, Max EE, Ried T, Nussenzweig A. 2000. DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. Nature 404: 510–514. https://doi.org/10.1038/35006670.

77. Ferguson DO, Sekiguchi JM, Chang S, Frank KM, Gao Y, DePinho RA, Alt FW. 2000. The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations. Proc Natl Acad Sci U S A 97:6630–6633. https://doi.org/10.1073/pnas.110152897.

78. Keeling PJ, Slamovits CH. 2005. Causes and effects of nuclear genome reduction. Curr Opin Genet Dev 15:601–608. https://doi.org/10.1016/j.gde.2005.09.003.

79. Vinogradov AE. 2004. Evolution of genome size: multilevel selection, mutation bias or dynamical chaos? Curr Opin Genet Dev 14:620–626. https://doi.org/10.1016/j.gde.2004.09.007.

80. Garcia-Diaz M, Kunkel TA. 2006. Mechanism of a genetic glissando: structural biology of indel mutations. Trends Biochem Sci 31:206–214. https://doi.org/10.1016/j.tibs.2006.02.004.

81. de Jong WW, Rydén L. 1981. Causes of more frequent deletions than insertions in mutations and protein evolution. Nature 290:157–159. https://doi.org/10.1038/290157a0.

82. Graur D, Shuali Y, Li WH. 1989. Deletions in processed pseudogenes accumulate faster in rodents than in humans. J Mol Evol 28:279–285. https://doi.org/10.1007/BF02103423.

83. Guirouilh-Barbat J, Lambert S, Bertrand P, Lopez BS. 2014. Is homologous recombination really an error-free process? Front Genet 5:175. https://doi.org/10.3389/fgene.2014.00175.

84. Bahmed K, Nitiss KC, Nitiss JL. 2010. Yeast Tdp1 regulates the fidelity of nonhomologous end joining. Proc Natl Acad Sci U S A 107:4057–4062. https://doi.org/10.1073/pnas.0909917107.

85. Daley JM, Wilson TE. 2005. Rejoining of DNA double-strand breaks as a function of overhang length. Mol Cell Biol 25:896–906. https://doi.org/10.1128/MCB.25.3.896-906.2005.

86. Moore JK, Haber JE. 1996. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Mol Cell Biol 16:2164–2173. https://doi.org/10.1128/mcb.16.5.2164.

87. Boulton SJ, Jackson SP. 1996. Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. Nucleic Acids Res 24:4639–4648. https://doi.org/10.1093/nar/24.23.4639.

88. Chico L, Ciudad T, Hsu M, Lue NF, Larriba G. 2011. The Candida albicans Ku70 modulates telomere length and structure by regulating both telomerase and recombination. PLoS One 6:e23732. https://doi.org/10.1371/journal.pone.0023732.

89. d'Adda di Fagagna F, Hande MP, Tong WM, Roth D, Lansdorp PM, Wang ZQ, Jackson SP. 2001. Effects of DNA nonhomologous end-joining factors on telomere length and chromosomal stability in mammalian cells. Curr Biol 11:1192–1196. https://doi.org/10.1016/S0960-9822(01)00328-1.

90. Barry JD, Ginger ML, Burton P, McCulloch R. 2003. Why are parasite contingency genes often associated with telomeres? Int J Parasitol 33:29–45. https://doi.org/10.1016/S0020-7519(02)00247-3.

91. Merrick CJ, Duraisingh MT. 2006. Heterochromatin-mediated control of virulence gene expression. Mol Microbiol 62:612–620. https://doi.org/10.1111/j.1365-2958.2006.05397.x.

92. Jackson AP, Otto TD, Darby A, Ramaprasad A, Xia D, Echaide IE, Farber M, Gahlot S, Gamble J, Gupta D, Gupta Y, Jackson L, Malandrin L, Malas TB, Moussa E, Nair M, Reid AJ, Sanders M, Sharma J, Tracey A, Quail MA, Weir W, Wastling JM, Hall N, Willadsen P, Lingelbach K, Shiels B, Tait A, Berriman M, Allred DR, Pain A. 2014. The evolutionary dynamics of variant antigen genes in Babesia reveal a history of genomic innovation underlying host-parasite interaction. Nucleic Acids Res 42:7113–7131. https://doi.org/10.1093/nar/gku322.

93. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bar-

tholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Böhme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UCM, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth T-J, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DMA, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream M-A, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CMR, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM. 2005. The genome of the African trypanosome *Trypanosoma brucei*. Science 309:416–422. https://doi.org/10.1126/science.1112642.

94. Navarro M, Cross GA, Wirtz E. 1999. *Trypanosoma brucei* variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. EMBO J 18:2265–2272. https://doi.org/10.1093/emboj/18.8.2265.

95. Stringer JR, Keely SP. 2001. Genetics of surface antigen expression in *Pneumocystis carinii*. Infect Immun 69:627–639. https://doi.org/10.1128/IAI.69.2.627-639.2001.

96. Meyer TF, Mlawer N, So M. 1982. Pilus expression in *Neisseria gonorrhoeae* involves chromosomal rearrangement. Cell 30:45–52. https://doi.org/10.1016/0092-8674(82)90010-1.

97. Maskell DJ, Szabo MJ, Butler PD, Williams AE, Moxon ER. 1992. Molecular biology of phase-variable lipopolysaccharide biosynthesis by *Haemophilus influenzae*. J Infect Dis 165:S90–S92. https://doi.org/10.1093/infdis/165-Supplement_1-S90.

98. Plasterk RHA, Simon MI, Barbour AG. 1985. Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium *Borrelia hermsii*. Nature 318:257–263. https://doi.org/10.1038/318257a0.

99. Li B. 2015. DNA double-strand breaks and telomeres play important roles in *Trypanosoma brucei* antigenic variation. Eukaryot Cell 14:196–205. https://doi.org/10.1128/EC.00207-14.

100. Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U, Wellems TE, Scherf A. 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. Nature 407:1018–1022. https://doi.org/10.1038/35039531.

101. Taylor HM, Kyes SA, Newbold CI. 2000. Var gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. Mol Biochem Parasitol 110:391–397. https://doi.org/10.1016/S0166-6851(00)00286-3.

102. Calhoun SF, Reed J, Alexander N, Mason CE, Deitsch KW, Kirkman LA. 2017. Chromosome end repair and genome stability in *Plasmodium falciparum*. mBio 8:e00547-17. https://doi.org/10.1128/mBio.00547-17.

103. Celli GB, Denchi EL, de Lange T. 2006. Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. Nat Cell Biol 8:885–890. https://doi.org/10.1038/ncb1444.

104. Conway C, McCulloch R, Ginger ML, Robinson NP, Browitt A, Barry JD. 2002. Ku is important for telomere maintenance, but not for differential expression of telomeric VSG genes, in African trypanosomes. J Biol Chem 277:21269–21277. https://doi.org/10.1074/jbc.M200550200.

105. Janzen CJ, Lander F, Dreesen O, Cross G. 2004. Telomere length regulation and transcriptional silencing in KU80-deficient *Trypanosoma brucei*. Nucleic Acids Res 32:6575–6584. https://doi.org/10.1093/nar/gkh991.

106. Maslov DA, Votýpka J, Yurchenko V, Lukeš J. 2013. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. Trends Parasitol 29:43–52. https://doi.org/10.1016/j.pt.2012.11.001.

107. Záhonová K, Kostygov AY, Ševčíková T, Yurchenko V, Eliáš M. 2016. An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. Curr Biol 26:2364–2369. https://doi.org/10.1016/j.cub.2016.06.064.

108. Genois M-M, Paquet ER, Laffitte M-C, Maity R, Rodrigue A, Ouellette M, Masson J-Y. 2014. DNA repair pathways in trypanosomatids: from DNA repair to drug resistance. Microbiol Mol Biol Rev 78:40–73. https://doi.org/10.1128/MMBR.00045-13.

109. Laffitte M-C, Leprohon P, Papadopoulou B, Ouellette M. 2016. Plasticity of the *Leishmania* genome leading to gene copy number variations and drug resistance. F1000Res 5:2350. https://doi.org/10.12688/f1000research.9218.1.

110. da Silva MS, Hovel-Miner GA, Briggs EM, Elias MC, McCulloch R. 2018. Evaluation of mechanisms that may generate DNA lesions triggering antigenic variation in African trypanosomes. PLoS Pathog 14:e1007321. https://doi.org/10.1371/journal.ppat.1007321.

111. Li H, Marple T, Hasty P. 2013. Ku80-deleted cells are defective at base excision repair. Mutat Res 745–746:16–25. https://doi.org/10.1016/j.mrfmmm.2013.03.010.

112. Reis H, Schwebs M, Dietz S, Janzen CJ, Butter F. 2018. TelAP1 links telomere complexes with developmental expression site silencing in African trypanosomes. Nucleic Acids Res 46:2820–2833. https://doi.org/10.1093/nar/gky028.

113. van Schendel R, van Heteren J, Welten R, Tijsterman M. 2016. Genomic scars generated by polymerase theta reveal the versatile mechanism of alternative end-joining. PLoS Genet 12:e1006368. https://doi.org/10.1371/journal.pgen.1006368.

114. Kent T, Mateos-Gomez PA, Sfeir A, Pomerantz RT. 2016. Polymerase θ is a robust terminal transferase that oscillates between three different mechanisms during end-joining. Elife 5:e13740. https://doi.org/10.7554/eLife.13740.

115. Ajawatanawong P, Baldauf SL. 2013. Evolution of protein indels in plants, animals and fungi. BMC Evol Biol 13:140. https://doi.org/10.1186/1471-2148-13-140.

116. Hamilton WL, Claessens A, Otto TD, Kekre M, Fairhurst RM, Rayner JC, Kwiatkowski D. 2017. Extreme mutation bias and high AT content in *Plasmodium falciparum*. Nucleic Acids Res 45:1889–1901. https://doi.org/10.1093/nar/gkw1259.

117. Chen JQ, Wu Y, Yang H, Bergelson J, Kreitman M, Tian D. 2009. Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. Mol Biol Evol 26:1523–1531. https://doi.org/10.1093/molbev/msp063.

118. Vanichtanankul J, Taweechai S, Yuvaniyama J, Vilaivan T, Chitnumsub P, Kamchonwongpaisan S, Yuthavong Y. 2011. Trypanosomal dihydrofolate reductase reveals natural antifolate resistance. ACS Chem Biol 6:905–911. https://doi.org/10.1021/cb200124r.

119. Timm J, Valente M, García-Caballero D, Wilson KS, González-Pacanowska D. 2017. Structural characterization of acidic M17 leucine aminopeptidases from the TriTryps and evaluation of their role in nutrient starvation in *Trypanosoma brucei*. mSphere 2:e00226-17. https://doi.org/10.1128/mSphere.00226-17.

120. Mercaldi GF, Pereira HM, Cordeiro AT, Michels PAM, Thiemann OH. 2012. Structural role of the active-site metal in the conformation of *Trypanosoma brucei* phosphoglycerate mutase. FEBS J 279:2012–2021. https://doi.org/10.1111/j.1742-4658.2012.08586.x.

121. Timm J, González-Pacanowska D, Wilson KS. 2014. Structures of adenosine kinase from *Trypanosoma brucei brucei*. Acta Crystallogr F Struct Biol Commun 70:34–39. https://doi.org/10.1107/S2053230X13033621.