Supporting Figure Legends and information

Figure S1. *STRIP assay*

SV40 replication templates were incubated with whole cell extracts (HeLa, LoVo, or 46BR derivatives) and SV40 Large T-antigen. After 4 hours, reactions were stopped and DNAs purified by phenol/chloroform extraction. Unreplicated material was digested with *Dpn*I, leaving behind *Dpn*I-resistant material, which had been fully replicated by human cell extracts. Replication products were transformed into bacteria (DH5α) at DNA:bacteria ratios that permit less than one plasmid per bacteria [67], with only *Dpn*I-resistant material giving rise to colonies. Plasmid DNA was isolated from colonies, the repeat-containing fragment was released by restriction digest, and digests were run on 4% polyacrylamide gels. Figure has been modified from [26].

Figure S2. *Sample set of pDM79E plasmid DNAs derived from colonies containing single or double repeat containing fragments*

A. Acrylamide gel shows representative repeat-containing fragment(s) released from pDM79E plasmid after replication with HeLa (left) or LoVo (right) cell extract. Arrows indicate the presence of double repeat-containing fragments.

B. Table shows the percentage of different types of doublets obtained with LoVo replication of pDM79E plasmid.

Supplementary text. *Double-repeat colonies are not due to tandem repeat tracts, catenated circular dimers, or recombination intermediates.*

The presence of double repeat-containing fragments within a single colony might be due to any of four scenarios: 1) each colony may have a single plasmid that recombined to harbor two tandem repeat-containing fragments; 2) the persistence of linked newly-replicated circular daughter molecules with different repeat lengths; 3) unprocessed recombination intermediates between two plasmids containing different repeat lengths; or 4) SV40 replication products are slipped-DNA heteroduplexes.

Not tandem repeats in the same plasmid: Each double-repeat colony may have a single plasmid that recombined to harbor two tandem repeat-containing fragments. This is unlikely since tandem duplication of CAG/CTG repeat containing fragments requires an initial template that contains at least two tandem copies [112]. To demonstrate that the distinct-sized fragments were derived from non-concatemeric plasmids, DNA isolated from the mixed colonies was re-transformed into bacteria – if individual colonies from this secondary transformation each displayed a single repeat size, this would indicate that there were two separate plasmid populations found in the initial mixed colony (Supplementary Figure S3A). This test was performed on mixed colony plasmid DNA derived from SV40 replication using LoVo cell extracts, and the secondary colonies each harbored plasmid DNA with a single repeat length (of either one or the other repeat tract length found in the initial mixed colony) (Supplementary Figure S3B). This finding further confirms that the presence of two repeat lengths in any given mixed colony occurs due to the propagation of a slipped heteroduplex formed during SV40 replication in LoVo extracts.

Not circular catenated plasmid dimers: A second scenario that could produce two different repeat lengths per bacterial colony is the persistence of catenated circular dimers (linked daughter molecules) with different repeat lengths after in vitro SV40 replication (Figure S4A). In this situation, the daughter

products of SV40 replication would not have been unlinked, and upon transformation into bacteria each those colonies would amplify the two differently sized plasmids from replication, leading to the appearance of two repeat-containing fragments. To test this, excess human topoisomerase II alpha (TopoII) was added to LoVo extracts during in vitro replication since this should ensure decatenation of newly replicated plasmids [113]. Sufficient amount of TopoII was added, using amounts that could resolve the many catenated forms present in kinetoplast DNAs (Figure S4B). Upon inclusion of TopoII in the SV40 LoVo extract replication reactions a decrease in the number of double fragments was not seen (41.8% with TopoII versus 44.0% without), indicating that the LoVo replicated material was likely monomeric (Figure S4, LoVo + TopoII). This strongly rules-out the possibility that the double repeat colonies were the result of catenated dimers.

Not recombination intermediates in the same plasmid: To assess whether or not recombination intermediates were responsible for our double repeat containing fragments, we performed 2D gel analysis on DNA replicated in both HeLa and LoVo. (Analysis was performed on pDM79E-2 and pDM79H-2 because these plasmids had an additional fragment inserted to facilitate detection of recombination intermediates or paused replication forks by 2D gel analysis. The proportion of double repeat-containing fragments after LoVo replication compared to HeLa replication of pDM79E-2, and their absence for pDM79H-2, was consistent with the values found for pDM79E and pDM79H.) While slipped-DNA formation would not be expected to alter migration in a 2D gel, the presence of recombination intermediates would be apparent. Replication products isolated from HeLa or LoVo cells were assessed for repeat lengths in bacteria, as above, to determine if DNAs replicated in human cells also yielded double-repeat containing fragments following bacterial transformation. Replication in HeLa cells yielded 6.25% double repeat-containing fragments (6 out of 96 colonies). In contrast, significantly more double repeat-containing fragments (25%) were found in DNAs replicated in LoVo cells (24 out of 96, $p = 0.00035$). These DNAs replicated in human cells were then assessed for the presence of recombination intermediates on 2D gels. Figure S5 shows that there are no differences in the replication products between HeLa and LoVo for each replication direction. This indicates that the double repeat containing fragments are not due to recombination intermediates, but rather are more likely due to slip-outs that had escaped repair in the MMR-deficient LoVo cell extract.

Are slipped-heteroduplexes processed by bacteria?: Since in vitro replication by human cell extracts only permits a single-round of replication [57], when heteroduplexes are formed at least one of the strands should contain the template starting repeat length. Provided bacteria only faithfully replicate each of the parental strands of the slipped-heteroduplex, we would expect mixed colonies to contain one set of plasmids harboring the starting length and the other with a mutation (expansion or contraction). Alternatively, bacteria might repair the transformed slipped-heteroduplexes using either of the two strands as a template, giving rise to colonies with plasmids having only one repeat length (starting length or mutant). To determine whether bacteria would faithfully replicate each of the parental strands of a slipped-heteroduplex, repair it to either of its parental strands, or process it to create lengths differing from either parental strand, we transformed bacteria with either pre-formed circular heteroduplex DNA containing a single repeat slip-out (48 CTG repeats paired with 47 CAG repeats), or with fully-duplexed DNA containing 48 repeats in both strands as a control, then carried out STRIP analysis (Figure S6). For the fully-paired 48/48 plasmid most colonies harbored a single repeat length within the starting length range of 46-49 repeats (>81%). Only 19% show double repeat fragments, where 15.7% contained the starting size and an expansion or contraction, and 3% were two contractions. In contrast, transformation of the pre-formed 47/48 heteroduplex yielded only mixed colonies at a level significantly greater than the fully-duplexed plasmid ($p = 2.7 \text{ X } 10\text{-}8$). Surprisingly, the majority (>76%) had lengths distinct from either parental strand, indicating that bacteria were producing two plasmids but in an unfaithful manner (Figure S6). These results confirm that the mixed colonies arising from LoVo-replicated material are in fact due to slipped-DNAs, and reveal the utility of bacterial transformation as a means to identify the presence of slipped-heteroduplexes.

It is presently unclear how bacteria are processing the slipped-heteroduplexes and what factors might be involved. While a comprehensive study of potential factors that process slipped-heteroduplexes is beyond the scope of the current study, our initial findings argue against a contribution of either GATC methylation status or of bacterial MutS in slipped-heteroduplex processing. The DpnI-resistant LoVo replicated products would be hemimethylated at the Adenosine of GATC sites (the strand discrimination signal for base-base mismatch repair in E. coli). The 47/48 heteroduplex and DpnI-resistant LoVo-replicated material are both hemi-methylated at GATC sites. Since transformation of both yielded many mixed colonies with repeat lengths distinct from either of the strands of the input heteroduplex, this suggests that differences in dam methylation are not contributing, albeit several groups have demonstrated the existence of a bacterial heteroduplex repair system that is independent of methylation [70,72,114].

Transformation of a pre-formed slipped-heteroduplex into a MutS-defective strain yielded the similar results compared to its parental isogenic MutS-proficient strain. For bacterial transformation of large heteroduplexes, the repair outcome was independent of bacterial MMR status, except in cases of co-repair of an adjacent base-base mismatch [68,72,73]. As briefly covered below, others have observed MMR-independent processing of large heteroduplexes, but the factors involved are unknown. While a definitive assessment of a wide-range of bacterial strains defective in various DNA repair defects is beyond the scope of the current study, future research in this avenue may prove interesting.

Transformation of bacterial cells with circular heteroduplex molecules (plasmids or bacteriophage with base-base mismatches, insertion/deletion heteroduplexes of 1 to several hundred excess non-repetitive nucleotides, as well as multiple mismatched regions) has previously been demonstrated to produce mixed colonies with two plasmids, each derived from either of the input parental strands [68,69,70,71,72,73]. These studies demonstrated that the products of transformations could be classified according to whether they contained two plasmids derived from both strands of the original heteroduplex (mixed colonies) or plasmids corresponding to only one of the strands of the input heteroduplex (pure colonies). Only in very rare instances were products other than sequence equivalents of input parental strands detected (whether mixed or pure) [72,73]. To this degree our slipped-DNA heteroduplexes appear to be unique in that the mixed colonies yielded high levels of plasmids with lengths distinct from either strand of the input heteroduplex – indicating that alterations are arising on both strands. While we are unable to determine repeat size changes using our assay due to this bacterial processing, the STRIP assay does give an accurate representation of the number of slipped-heteroduplex molecules present before transformation (Figure S6). The ability of bacteria to produce mixed colonies following transformation of a slipped-heteroduplex, regardless of size, provides a useful tool to detect the presence of slipped-heteroduplexes amongst primate replication products.

Figure S3. *Re-transformation of plasmids isolated from a mixed colony*

A. Schematic of experiment, see text above.

B. Analysis of SV40-replicated DNA by LoVo extracts. Starting template length, S, the primary colonies arising from transformed replicated material harboring two repeat tract lengths/colony. The re-transformed material, lanes 1-5, revealing either of the two lengths in the primary colony – indicates that the primary colonies actually harbored two distinct plasmids, rather than a single plasmid with two lengths or two linked plasmids.

Figure S4. *TopoII decatenation*

A. Schematic showing the formation of bacterial colonies containing two (double) repeat-containing fragments after transformation by absence of decatenation of two plasmids.

B. To ensure that exogenous TopoII is active in the presence of whole cell extracts, its activity was tested on the heavily-catenated kinetoplast DNA (kDNA) in the presence or absence of LoVo cell extract. Lane 1: untreated kDNA. Lane 2: kDNA + TopoII (kDNA has been decatenated based on presence of both relaxed circular and nicked circular forms). Lane 3: kDNA + TopoII + LoVo extract (kDNA is still decatenated in presence of LoVo extract). Lane 4: kDNA + *Xho*I (linearized kDNA)

C: Graph showing the percentage of double fragment containing colonies in LoVo replicated pDM79E plasmid and LoVo replicated pDM79E plasmid treated with TopoII. Actual numbers are shown at bottom and significance level indicated above graphs ($\chi 2$ test).

Figure S5. *2D gel analysis of replication products*

To ensure that aberrant recombination was not occurring between replication templates, 2D gel analysis of replication intermediates was performed.

A. Replication templates. pDM79E-2 and pDM79H-2 are derivatives of pDM79E/H with an insert between the repeat and the SV40-ori (blue circle) making the distance between the two sites ~800 nucleotides. Replication of the pDM79E-2 template in the LoVo cell line still yielded the double repeat-containing fragments in the STRIP assay, and these products remained negligible after HeLa replication. Probe for 2D gels is the XmnI/AlwNI fragment (darker).

B. 2D gel analysis of pDM79E-2 and pDM79H-2 replication from HeLa and LoVo cells. pDM79E-2 was digested with *Avr*II/*Xmn*I for the analysis. pDM79H-2 was digested with *Alw*NI/*Avr*II.

Figure S6. *Heteroduplex DNA in DH5a*

To ensure that DH5α bacteria were not repairing slipped-heteroduplexes before replicating the DNA, fully-duplexed DNA (left - 48 repeat duplex) and slipped-heteroduplex DNA (right - 47/48 repeats) were transformed into the bacteria and then the STRIP protocol was carried out. Multiple repeat-containing fragments are seen in colonies from heteroduplex DNA transformation as shown in the table below – they are not correctly repaired before replication.