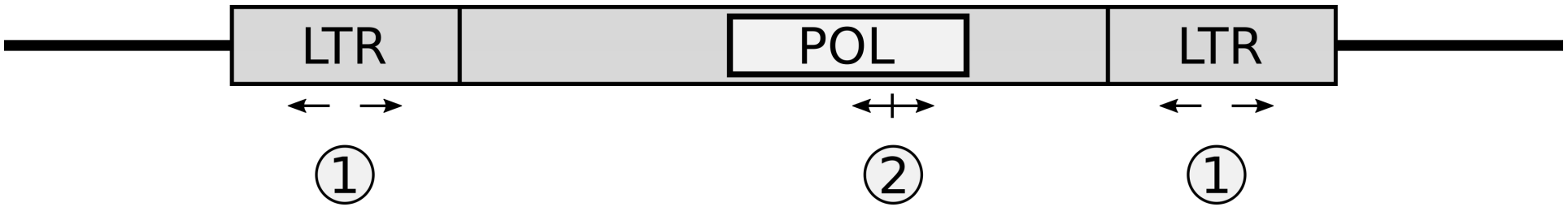

Figure S1. SIP Primer binding sites on KoRV provirus

Genomic DNA was first fragmented and circularized. Following, DNA was amplified using one of two primer sets in the inverse orientation. Primer binding sites were either to the duplicated LTR regions on the provirus (Primer Set 1) or to the polymerase gene (pol) (Primer Set 2), which was designed end-to-end. Notably, inverse PCR amplifications of the LTR and pol genes were performed in separate PCR reactions and subsequently pooled together for library construction and PacBio sequencing.

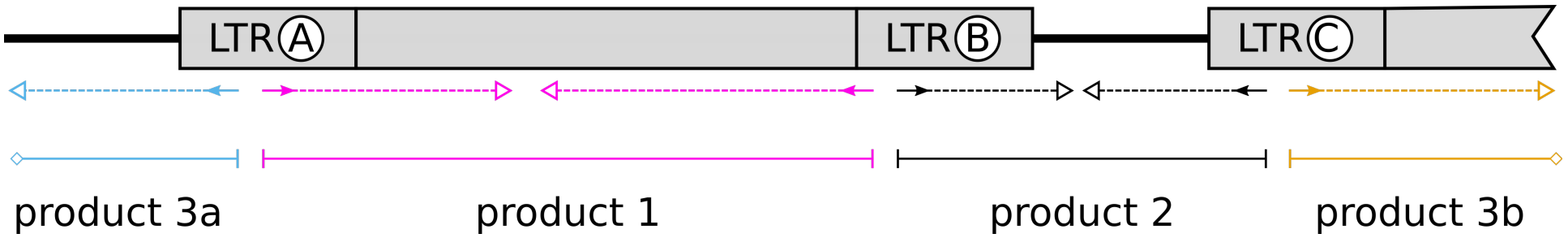

Figure S2. Hypothetical standard PCR amplification in inverse PCR reactions

Linear DNA, i.e. DNA that has not been circularized, can hypothetically be amplified by standard PCR by primers in inverse orientation. This figure represents a full copy of a KoRV provirus and another partial KoRV provirus downstream to display possible standard PCR amplification. For simplicity, putative standard PCR products are represented only with the primers. Provided a fragmented linear DNA molecule spans the length of the 5´ and 3´ LTR primer binding sites on a provirus; exponential PCR amplification can occur from a primer located on LTR A to a primer located in LTR B (product 1). Similarly, provided that two LTR's from two different proviruses are in close proximity to each and on the same DNA molecule, linear DNA can be amplified exponentially from a primer located on LTR B to a primer located on LTR C (product 2). Linear DNA can also be amplified using non-exponential single primer PCR amplification. This is represented from both a primer binding to LTR A and amplifying toward the integration site (product 3a), and separately, a primer binding to LTR C and amplifying towards the internal parts of the provirus (product 3b).

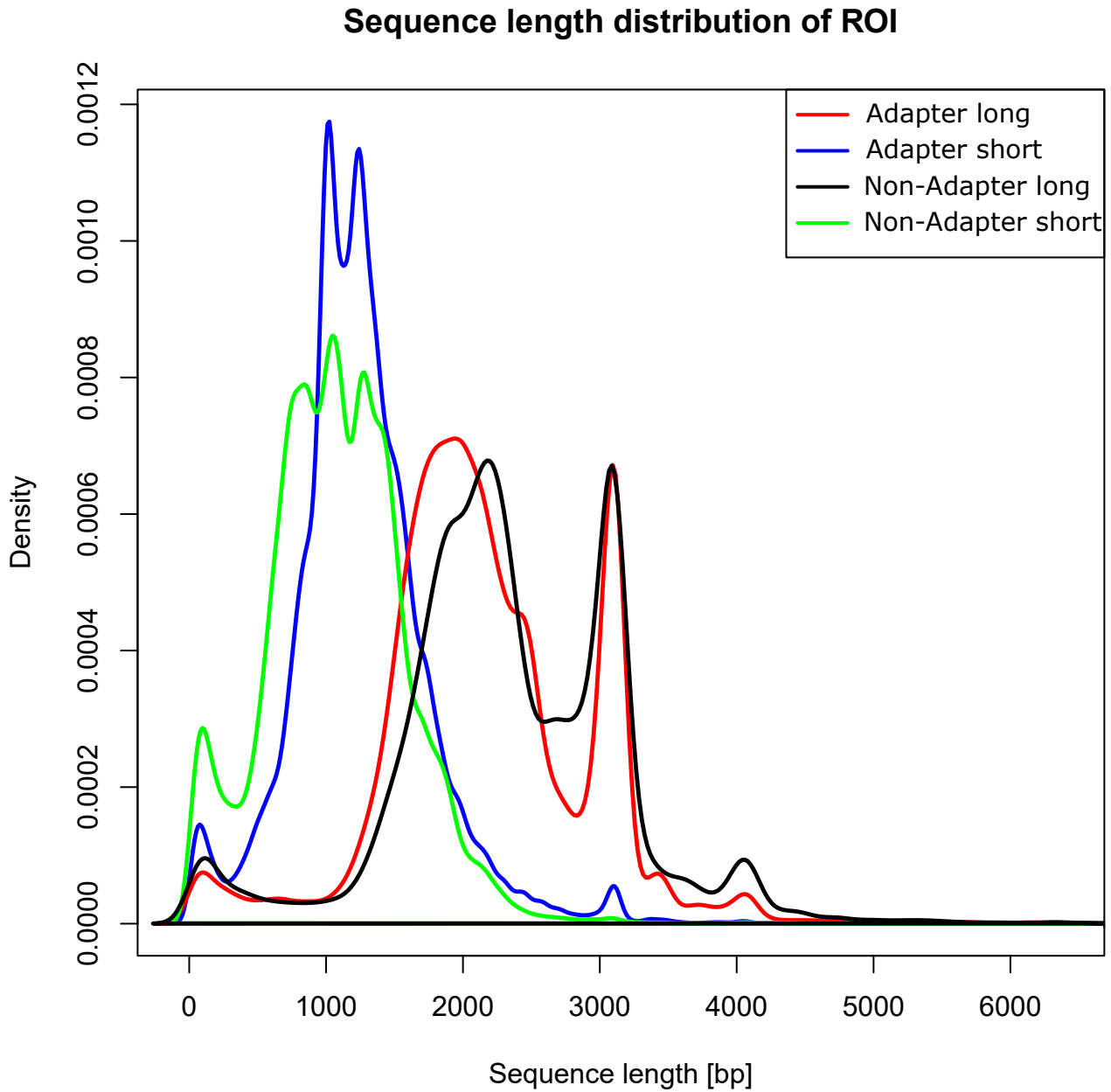

Figure S3. Sequence length distribution of the Reads of Insert from four standard SIP datasets.