Salthouse, Holly Rebecca (2025) *Investigating the role of 70 bp repeats and VSGs in formation of DNA double strand breaks in Trypanosoma brucei bloodstream expression sites.* MSc(R) thesis.

Investigating the role of 70 bp repeats and VSGs in formation of DNA double strand breaks in *Trypanosoma brucei* bloodstream expression sites

Holly Rebecca Salthouse
BSc (Hons)

Submitted to the University of Glasgow in fulfilment of

the requirement for the Degree of Master of Science

(Research)

School of Infection and Immunity,
College of Medical, Veterinary and Life Sciences
University of Glasgow

September 2024

Abstract

Trypanosoma brucei are protozoan parasites, which evade the immune system by antigenic variation with the use of a variant surface glycoprotein (VSG). A double strand DNA break (DSB) has been proposed as the first step in VSG switching and it has previously been demonstrated that when DNA breaks arise in the active expression site, this leads to accumulation of RNA-DNA hybrids, R-loops, however, there is still uncertainty of how and where the double stranded breaks form. Mapping of DNA breaks in the expression site by BLISS has implicated two key features in the formation of the breaks: the *VSG* 3' end where the break is found and in the 70 bp repeats. This project aimed to examine the relative contribution of both the *VSG* 3' end and the 70 bp repeats to the formation of a DNA break and VSG switching. To address this, bloodstream form *T. brucei* cell lines with alterations in the active bloodstream expression site (BES) were generated using CRISPR/Cas9. Two regions of 70 bp repeats were successfully removed and a distinct *VSG* sequence (*VSG121*) was relocated to two other locations within the active BES. In addition, a plasmid containing the *VSG* of the active expression site in this cell line (*VSG221*) was successfully generated. To confirm alterations to the active BES PCR screening, VSG immunofluorescence, and Oxford Nanopore sequencing were performed. It was identified by Oxford Nanopore sequencing that when the large section of 70 bp repeats was removed, the smaller section of 70 bp repeats expanded. In future work, these cell lines will allow for investigation of R-loops and double strand breaks by comparison of DRIP-seq and INDUCE-seq/BLISS data with the natural organisation of the active BES and so, give an insight into whether the 70 bp repeats and/or *VSG* sequence drive R-loop and DNA DSB formation and if they are necessary for break formation linked to VSG switching.

Table of Contents

List of Tables

List of Figures

List of Appendices

Acknowledgments

I would like to thank my supervisors, Professor Richard McCulloch and Dr Jane Munday, for proposing the project, giving me the opportunity to carry out this research, and their guidance throughout.

From not being able to properly use a pipette on my first day to completing experiments alone, thank you to Jane who took the time to teach and support me through many lab techniques and experiments. Thank you for all your advice, answering all my questions, and helping me through the final stages of writing up. I couldn't have done this without you!

Thank you to everyone in the McCulloch lab who provided me with support, help with practical lab work, and made my time in the lab enjoyable. Thank you to Craig Lapsley and Marija Krasilnikova for their help with Nanopore sequencing. A special thanks to Grace Gill whose positivity and endless support in the lab helped me through.

Finally, thank you to all my friends and family for their support throughout.

Authors declaration

I declare that the work presented in this thesis is my own work, except where reference is made to the contribution of others. This work has not been submitted for any other degree at the University of Glasgow or any other academic institution.

Holly Salthouse

September 2024

Abbreviations

BES	Bloodstream expression site
Bp	Base pair
BSD	Blasticidin
BSF	Bloodstream form
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
DRIP	DNA RNA hybrid immunoprecipitation
DSB	Double stranded break
ES	Expression site
ESAG	Expression site associated gene
FBS	Foetal bovine serum
gDNA	Genomic DNA
HYG	Hygromycin
IFA	Immunofluorescence assay
kb	kilobase
KO	Knockout
LB	Luria Bertami medium
NEO	Neomycin
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
rcf	Relative centrifugal force
RNA	Ribonucleic acid
rpm	Revolutions per minute
sgRNA	Single guide RNA
VSG	Variant surface glycoprotein
WB	Western blot
WT	Wild type
γH2A	phosphorylated histone H2A on Thr130

1. Introduction

1.1 African trypanosomiasis

Human African trypanosomiasis (HAT), often referred to as sleeping sickness, is a vector-borne parasitic disease caused by infection with the single-cell protozoan parasite *Trypanosoma brucei*. HAT is a neglected tropical disease endemic in 36 countries of sub-Saharan Africa, which places 55 million people at risk (WHO, 2023). *T. brucei* is transmitted through the bite of an infected tsetse fly (*Glossina spp.*), but can also be transmitted through mother to child infection, contact with contaminated needles, or through sexual contact.

HAT progresses in two stages with stage one, the haemolymphatic stage, involving trypanosomes multiplying in the subcutaneous tissue, blood, and lymph, which can result in symptoms such as headaches, joint pains, and an intermittent fever that can last from a day to a week and corresponds to the level of parasitaemia fluctuation (Amin *et al*., 2010). Stage two, the meningoencephalitic stage, is characterised by the invasion of trypanosomes into the central nervous system (Amin *et al*., 2010). In stage two, patients experience sleep cycle disturbance due to a dysregulation of the circadian rhythm, which controls the sleep and wake cycle and resulted in the name sleeping sickness (Bentivoglio and Bertini, 2018; Rijo-Ferreira *et al*., 2018). Patients also experience neurological symptoms such as confusion, encephalitis, and tremors. Without treatment, the second stage is normally fatal (Bonnet *et al.,* 2015).

The causative agents of HAT are endemic in distinct areas and cause distinct disease patterns with *T. b. gambiense* prevalent in Western and Central Africa, and *T. b. rhodesiense* prevalent in Eastern and Southern Africa (Hollingshead and Bermudez, 2024). *T. b. gambiense* accounts for 97% of reported cases of HAT and causes chronic HAT, which is characterised by a long asymptomatic incubation period that can last from months to years (Kuepfer *et al.,* 2011). When symptoms of infection with *T. b. gambiense* emerge, the central nervous system is usually already affected (Kuepfer *et al.,* 2011). In comparison, *T. b. rhodesiense* causes acute HAT, which progresses quickly with patients having neurological involvement within weeks. Some patients may develop multiorgan failure or septic shock as quickly as within a few days of developing a fever (Kuepfer *et al*., 2011). A chancre (a persistent swelling) can appear at the site of the tsetse fly bite of *T. b. rhodesiense* infections but is rarely seen in *T. b. gambiense* infections (Kuepfer *et al*., 2011).

Early diagnosis of infection is essential to avoid HAT progressing into the neurological stage requiring complex treatment (Bonnet *et al.,* 2015*).* Antibody and parasite detection is required for successful diagnosis and may differ depending on the *T. brucei* subspecies (Álvarez-Rodríguez *et al*., 2022). Current diagnostic tests rely on detection of host antibodies, which react with the parasite's antigens (Álvarez-Rodríguez *et al*., 2022). Diagnostic tests include the Card-Agglutination Trypanosomiasis Test (CATT), which is a serological test and is the standard approach for initial population screening, and several other rapid diagnostic tests for HAT have been developed, including lateral flow immunochromatographic assays (Bonnet *et al.,* 2015; Álvarez-Rodríguez *et al*., 2022). Antibody-based screening tests are extremely sensitive and specific, but have a low positive predictive value, which requires further confirmation of parasites in body fluids, such as by PCR, before treatment (Álvarez-Rodríguez *et al*., 2022). Accurate staging of the disease is also critical for treatment (Amin *et al*., 2010). A new CRISPR-based molecular detection toolkit, SHERLOCK4HAT, has been developed and could be used as a highly specific and sensitive tool in diagnosis (Sima *et al.,* 2022). SHERLOCK4HAT has the potential to be used as an adaptable diagnostic method, for mass screening, and as an epidemiological surveillance tool and can discriminate between subspecies (Sima *et al.,* 2022). Drugs used in the early stage of infection, Pentamidine and Suramin, do not cross the blood-brain barrier efficiently so are not appropriate for when infection has progressed to the second stage (Bonnet *et al.,* 2015*;* Amin *et al*., 2010). Second stage treatments, Melarsoprol and Eflornithine, are more toxic and so should be avoided in the first stage to limit unnecessary toxicity and side effects (Amin *et al*., 2010). These drugs can cause significant adverse effects, so patients require clinical surveillance during treatment (Bonnet *et al.,* 2015*).*

Some *Trypanosoma* (sub)species are unable to infect humans due to the trypanolytic protein apolipoprotein-L1 and two trypanosome lytic factors (TLF1 and TLF2), which are present in human serum and provide an innate resistance that can prevent human infections (Ponte-Sucre, 2016). These species, *T. brucei brucei, T. congolense and T. vivax,* infect animals, causing animal African trypanosomiasis (AAT), also known as Nagana disease. This disease can affect a range of animals such as cattle, horses, sheep, pigs, camels, and goats (Desquesnes *et al.,* 2022). AAT can cause severe disease with symptoms such as fever, oedema, and paralysis. As the illness progresses animals become weaker and unable to work, which is a significant issue for livestock production in sub-Saharan Africa (Steverding, 2008).

1.2 Life cycle of *T. brucei*

The life cycle of *T. brucei* involves molecular and morphological changes as it adapts to the mammalian host or compartments within the tsetse fly (Figure 1-1) (Matthews, 2005; CDC, 2019). When an infected tsetse fly takes a blood meal from a mammalian host, metacyclic trypomastigotes in the tsetse's saliva are injected into the subcutaneous tissue of the host (Oethinger and Campbell, 2010). Metacyclic trypomastigotes are a cell-cycle arrested life cycle stage, which expresses a metacyclic variant surface glycoprotein coat comprised of a single VSG type (Ramey-Butler *et al.,* 2015). Expression of the mVSG is controlled through monoallelic expression, which is complex due to the multiple parasite stages within the tsetse fly salivary glands (Hutchinson *et al.,* 2021). The VEX complex is essential for monoallelic expression in BSFs (see 1.3), and recent data suggests a role for VEX1 in metacyclic forms as well, with modulation of *VEX1* expression leading to dysregulation of *VSG* expression in metacyclic forms (Tihon *et al.,* 2022). The cell-cycle arrested metacyclic trypomastigotes are then activated, re-entering the cell cycle, and differentiating into proliferative trypanosomes (Reuter *et al*., 2023). The skin is a reservoir tissue, and trypanosomes are often found in the skin of asymptomatic patients when they are not detectable in the blood (Reuter *et al*., 2023).

The metacyclic trypomastigotes enter the lymphatic system and into the bloodstream where they differentiate into replicating long-slender bloodstream forms (BSFs). BSFs establish infection and escape the immune response using variant surface glycoprotein (VSG) switching (see below). BSFs are carried to other body sites in the blood, lymph, and cerebrospinal fluid and replicate extracellularly by binary fission (Moreno *et al.,* 2019). A significant population of extravascular *T. brucei* exist within the skin and are a reservoir of infection (Capewell *et al.,* 2016). BSFs are foundin the interstitium of tissues including within interstitial fluids in adipose tissue and in abundance in seminiferous tubules of the male testes (Trindade *et al.,* 2016; Claes *et al.,* 2009). Long slender BSFs continue to increase in number until differentiation to short stumpy BSFs, a process triggered via a quorum sensing process (Brigg*s et al.,* 2021). The short stumpy BSFs are a cell-cycle arrested, non-replicating form and the accumulation of this form prolongs host survival and increases the chance of transmission to a new host. The short stumpy BSFs are pre-adapted to survive in the tsetse fly's midgut with characteristics including resistance to acidic and proteolytic stress, activation of mitochondrion, and repositioning of the lysosome (Nolan *et al.,* 2000; Vickerman, 1965; Vanhollebeke *et al.,* 2010). Short stumpy forms can regulate their internal pH to withstand potentially harmful changes in the environment upon ingestion by a tsetse fly (Nolan *et al.,* 2000). Differentiation to the short

stumpy form also coincides with the activation and elaboration of their mitochondrion, and the ability to use the Krebs cycle to maintain motility in the absence of glucose, exhibiting pre-adaption for the metabolic environment within the tsetse midgut (Vickerman, 1965; Dewar *et al.,* 2018). There are also endocytic differences, with the flagellar pocket increasing in size and the lysosome expanding and relocating anterior to the nucleus in short stumpy forms (Brickman *et al.,* 1994; Vanhollebeke *et al.,* 2010).

The obvious distinction between forms is that short stumpy BSFs are in cell-cycle arrest, and this contributes to their capacity for transmission (Shapiro *et al.,* 1984; Silvester *et al.,* 2017). Cells are uniformly arrested in the G0/G1 phase of the cell cycle and re-enter the cell cycle after differentiation to procyclic trypomastigotes in the tsetse fly (Shapiro *et al.,* 1984). This allows for morphological changes, such as organelle repositioning, to be coordinated with re-entry into the cell cycle which is important for completion of the cell cycle of tsetse midgut procyclic forms (Matthews, 2005).

When a tsetse fly takes a blood meal from an infected host, the BSF trypomastigotes are ingested into the midgut of the tsetse fly. Short stumpy forms are induced to differentiate into procyclic trypomastigotes, which are proliferative in the tsetse midgut. They continue to multiply and establish infection and generate a reservoir of infective parasites in the tsetse's midgut (Ooi and Bastin, 2013). A study by Schuster et al. (2021) found that proliferative slender BSFs can complete the life cycle in the tsetse fly *in vivo* and were equally competent for fly passage as short stumpy forms, and so stumpy forms are not required to complete the life cycle. VSG expression is downregulated in the procyclic form and the procyclic acidic repetitive protein, a glycoprotein, replaces the VSG on the surface membrane (Navarro *et al.,* 1999).

After proliferation, procyclic trypomastigotes leave the midgut, move to the salivary gland, and differentiate to epimastigotes (Matthews, 2005). Epimastigotes attach to the salivary gland wall through elaborations of the flagellar membrane and multiply (Matthews, 2005). The parasites then differentiate into metacyclic trypomastigotes, undergoing cell division arrest and re-expressing a VSG coat (Matthews, 2005). They then move to the salivary gland lumen and prepare to inoculate a new host (Moreno *et al.*, 2019).
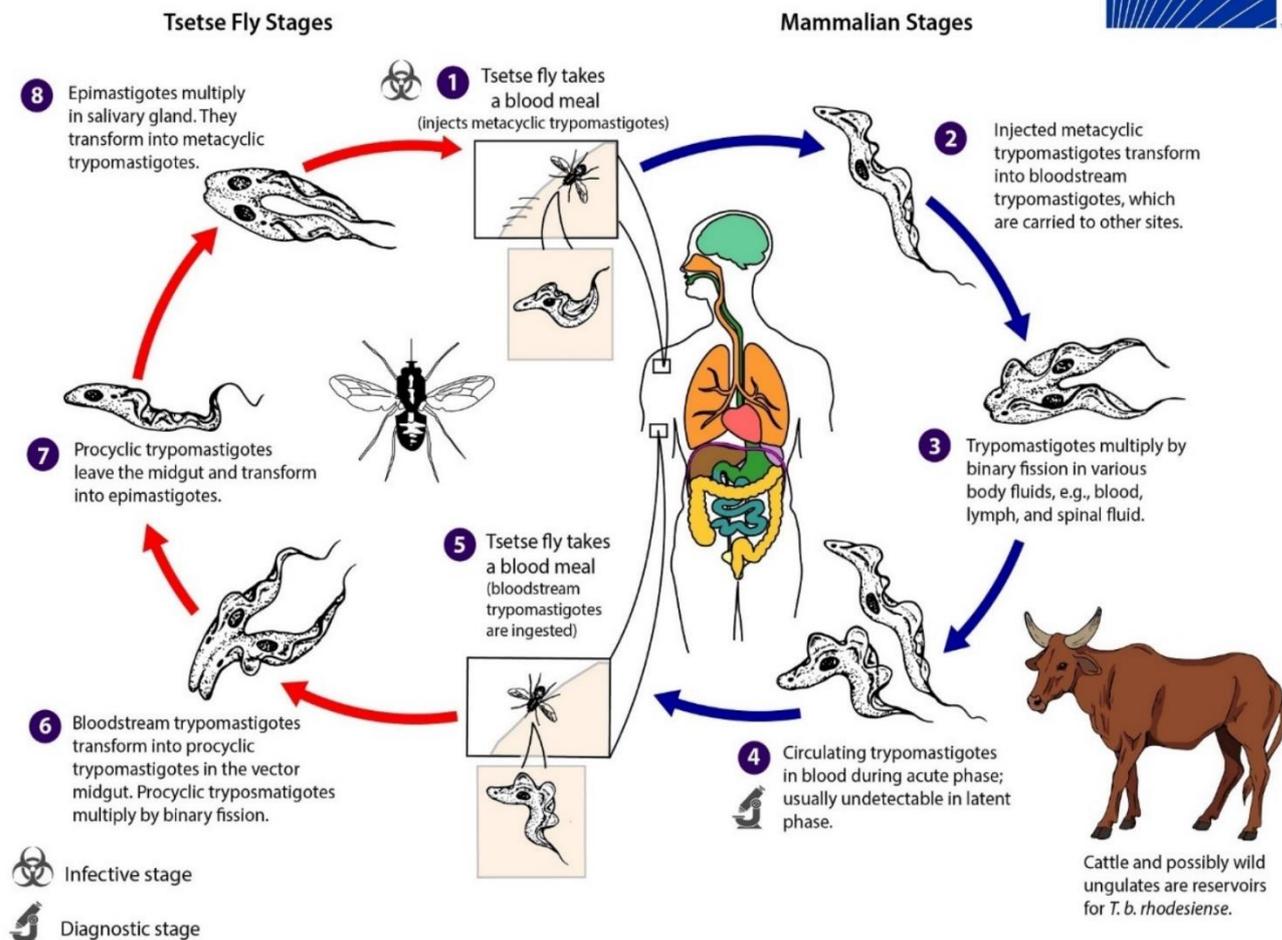
Figure 1-1. Life cycle of *Trypanosoma brucei* through the tsetse fly vector and host stages. Diagrammatic representation of morphologies adopted by the life cycle stages of *T. brucei* between mammalian hosts and the tsetse fly vector indicating when *T. brucei* is replicating and dividing or in cell-cycle arrest. Indicates when *T. brucei* is in an infective stage and diagnostic stage (CDC, 2019).

1.3 Antigenic variation of *T. brucei*

The major surface component of *T. brucei* BSFs is the variant surface glycoprotein (VSG), which is key in evading the host immune response. The VSG is a glycophosphatidylinositol (GPI) anchored glycoprotein and compromises around 95% of the surface proteome, forming a densely packed monolayer, or 'coat' (Mugnier *et al*., 2016; Chattopadhyay *et al*., 2005). The VSG coat has a structural role in protecting other cell surface proteins from effector molecules from the mammalian immune system (Chattopadhyay *et al*., 2005). It extends 12-15 nm from the cell membrane and, by density and steric hindrance, shields invariant surface proteins from

host antibody binding as they are inaccessible (Mugnier *et al*., 2016). The VSG is an essential virulence factor and represents up to 20% of total cell proteins (Sheader *et al.,* 2005; Horn, 2014). The VSG is essential even *in vitro* with inhibition of VSG synthesis by RNAi resulting in cells triggering pre-cytokinesis cell cycle arrest, characterised by stalled cells not reinitiating S phase, to preserve integrity of the VSG (Sheader *et al.,* 2005). Although cell cycle stalled cells do not grow in volume, cells are metabolically active, actively try to maintain lipid homeostasis, and GPI-anchor biosynthesis is unaltered (Sheader *et al.,* 2005; Ooi *et al.,* 2018). Induction of *VSG* RNAi *in vivo* by Sheader et al. (2005) resulted in rapid clearance of trypanosome infection within 12 hours and suggested that even minor damage of the VSG is fatal for *T. brucei*. Blocking synthesis of the VSG at translation level results in dramatic changes to both the number and morphology of secretory structures and organelles, endoplasmic reticulum exit sites and Golgi bodies, but the secretory pathway remains functional (Ooi *et al.,* 2018).

VSGs are 400-500 amino acid proteins and have two domains: the N-terminal and C-terminal domain (Silva Pereira *et al.,* 2022; Schwede *et al*., 2011). The N-terminal domain is larger and consists of 300-400 amino acids (Silva Pereira *et al.,* 2022). The core of the N-terminal domain contains a 10 nm helical coiled-coil, and the domain is elongated and contains highly variable antigenic epitopes (Schwede *et al*., 2011; Silva Pereira *et al.,* 2022). The C-terminal domain is small, <100 amino acids, inaccessible to host antibodies, and links to the plasma membrane via a GPI anchor (Silva Pereira *et al.,* 2022). The C-terminal domain is more diverse and the cysteine distribution more variable (Marcello and Barry, 2007). VSGs are separated into two subfamilies which contain variant and non-variant antigens: a-VSGs and b-VSGs (Silva Pereira *et al.,* 2022). Subfamilies are classed based on the N-terminal and C-terminal types and can be further divided into phylotypes (Silva Pereira *et al.,* 2022). Each subfamily includes VSGs which contain highly related amino acid sequences (Silva Pereira *et al.,* 2022).

VSGs are highly immunogenic, and a host immune response is effectively generated in response to the VSG (Oethinger and Campbell, 2010). Host antibodies specific to the VSG are rapidly produced and are effective at clearing trypanosomes. The host immune response means most trypanosomes are destroyed by antibody, complement, or phagocytosis (Oethinger and Campbell, 2010). However, the immune response takes some time to generate anti-VSG antibodies, and during this time, a subpopulation of trypanosomes evades the response by antigenic variation of the VSG where the trypanosomes switch to a VSG that is structurally similar but antigenically distinct (Burri *et al.,* 2014). The new VSG-expressing population then

expands until host antibodies are developed specific to the new VSG and, once again, most trypanosomes are destroyed and some produce another variation of the VSG, which results in waves of parasitaemia (Stockdale *et al*., 2008). In each peak of parasitaemia, many distinct VSGs are expressed by the population at one time (Hall *et al.,* 2013; Mugnier *et al.,* 2015). The strategy of antigenic variation is successful as it prolongs the life of the host and extends the time trypanosomes reside in the host, enhancing transmission (Stockdale *et al*., 2008).

Antigenic variation is a common method of immune evasion, which has three main requirements for nonviral pathogens (Stockdale *et al*., 2008). The first requirement is that pathogens have a large family of surface antigens which allow for variation. *T. brucei* has a huge family of around 2,000 genes encoding for antigenically distinct VSGs, though it should be noted that around 80% of the *VSG* repertoire consists of pseudogenes or gene fragments (Berriman *et al.,* 2005; Marcello and Barry, 2007; Cross *et al.,* 2014). The VSG archive is distributed across three genomic locations: in telomeric VSG expression sites (ESs), at the telomeres of mini- and intermediate-chromosomes, and in arrays of *VSG*s in subtelomeres of the megabase chromosomes (Hall *et al.,* 2013). *T. brucei* has around 100 minichromosomes, which are 30-150 kb in length and *VSG* genes are found on two thirds of the minichromosome telomeres (Hall *et al.,* 2013). Expression site-associated genes (ESAGs) and ES promoters, which are features of sites of VSG expression in BSF cells (see below), have not been found in minichromosomes, which suggests these molecules may only act as store of silent VSGs, which can only be activated by recombination (Devlin *et al.,* 2016) (see below). Total genomic DNA (gDNA) content varies by up to 33% between different strains of *T. brucei* and this variation is mostly due to differences within the subtelomeric VSG arrays (Melville *et al.,* 1998). Indeed, near complete assembly of the genome of *T. brucei* strain Lister 427 has shown that the subtelomeres of allelic chromosomes differ substantially in *VSG* content (Muller *et al.,* 2018; Cosentino *et al.,* 2021).

The second requirement for antigenic variation is that only one variable surface antigen is expressed by a pathogen at one time, which prevents exhausting the repertoire of antigens to prolong infection (Stockdale *et al.,* 2008). In *T. brucei* transcription of the VSG occurs in genomic telomeric regions, termed expression sites (ES), which allows for one VSG to be expressed at a time (Pays *et al.,* 2001). The expression sites in bloodstream and metacyclic stages are distinct (Pays *et al.,* 2001).

Metacyclic form ES (MES) are telomeric, monocistronic transcription units and contain only the *VSG* gene and a RNA polymerase I promoter, with the promoter found ~5 kb upstream of the *VSG* (Graham *et al.,* 1999). Unlike the *VSG* gene in bloodstream forms, metacyclic VSGs have their own promoters and are only loosely conserved with bloodstream *VSG* promoters (Ginger *et al.,* 2002). Life cycle stage-specific regulation of the MES promoter operates through a transcriptional switch (Graham *et al.,* 1990).

In contrast, BESs are polycistronic transcription units and there are ~15 BES present in *T. brucei* Lister 427. The structure of the BESs are shown in Figure 1-2 (Hertz-Fowler *et al.,* 2008). Each BES has ~90% of the same sequence and the same gene organisation (Li and Zhao, 2021). Only one BES is completely transcribed at any time in a process termed allelic exclusion, meaning only a single VSG is generated in a cell at one time (Navarro *et al*., 1999). The BES comprises a promoter, an array of repetitive sequence known as 70 base-pair (bp) repeats, the VSG, which is always proximal to the telomere, several ESAGs, and sporadic retrotransposons (Silva Pereira *et al.,* 2022). The ES promoter is located 40-60 kb upstream of the *VSG* and transcription from this promoter is driven by RNA Polymerase I (Navarro *et al*., 1999). *T. brucei* Lister 427 MiTat1.2, the strain most frequently examined and used here, predominantly expresses VSG221 from BES1. Of note for the work in this thesis: *VSG121* is found in the silent BES3 (Hertz-Fowler *et al.,* 2008). Flanking the 5' end of the ES promoter is a 20-40 kb of 50 bp repeats with an unknown function (Navarro *et al*., 1999). The region separating the VSG and the 70 bp telomere repeats has been termed the 'co-transposed region', and appears to be essential, for unknown reasons (Davies *et al.,* 1997).
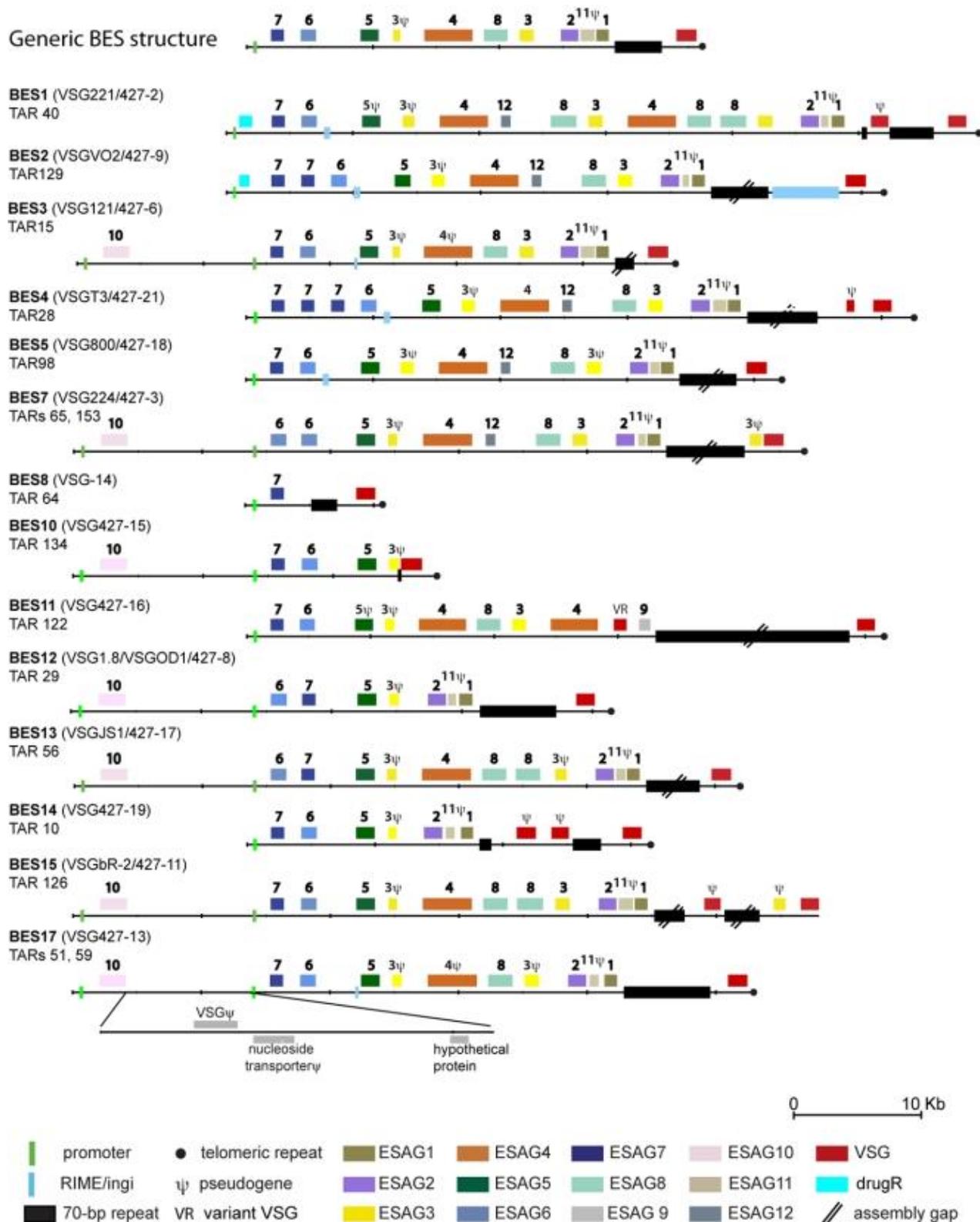
Figure 1-2. Organisation and structure of *T. brucei* Lister 427 bloodstream expression sites. Annotated expression sites are drawn to scale. Diagram indicates the organisation of the promoter, ESAGs, array of 70 bp repeats, and the VSG within different BESs. BESs are aligned at their 5' most *ESAG7* or *ESAG6* sequence (Hertz-Fowler *et al.,* 2008).

A link has been identified between DNA replication timing and transcription of the active BES suggesting replication may drive VSG switching (Devlin *et al*., 2016). The transcribed VSG ES differs to other telomeric sites as it is the only telomeric site which replicates early in the S phase with other silent ESs replicated later (Devlin *et al.,* 2016). In procyclic forms where ES transcription is turned off, all subtelomeric ESs are replicated late (Devlin *et al.,* 2016), This suggests a model in which *T. brucei* use targeted DNA damage resulting from replication-transcription clashes within the BES during antigenic variation (Devlin *et al,* 2016).

The active BES is transcribed in a specialised sub-nuclear structure named the expression site body (ESB), which is associated with an accumulation of RNA polymerase I and is non-nucleolar (Lopez-Escobar *et al*., 2022). Monoallelic VSG expression is dependent on multiple factors and is important for trypanosomes virulence (Li and Zhao, 2021). The ESB contains an enrichment of factors required for the maintenance of monoallelic expression (Lopez-Escobar *et al*., 2022). Factors, which regulate the recessive state of the VSG ESs include ISWI, RAP1, and histone deacetylase 3 (Hughes *et al.,* 2007; Yang *et al.,* 2009; Wang *et al.,* 2010). Factors, which specifically associate with the ESB and regulate monoallelic antigenic variation are the *VSG*-exclusion protein 1 (VEX1), *VSG*-exclusion protein 2 (VEX2), and ESB-specific protein 1 (ESB1) (Glover *et al.,* 2016; Faria *et al.,* 2019; Lopez-Escobar *et al.,* 2022). ESB1 and VEX are involved in controlling *VSG* expression and have distinct roles (Faria *et al.,* 2022). ESB1 is a transcriptional activator, which localises to the ESB and is stage specific, only being expressed in *T. brucei* life cycle stages, which express the VSG (Lopez-Escobar *et al.,* 2022). ESB1 associates with DNA near the active promoter of the *VSG* and regulates *VSG* expression (Lopez-Escobar *et al.,* 2022). ESB1 is also required for the recruitment of some components of the ESB including RNA polymerase I (Lopez-Escobar *et al.,* 2022). A depletion of ESB1 leads to a decrease in transcription of the active ES (Lopez-Escobar *et al.,* 2022). The VEX1 protein binds *VSG* associated chromatin and VEX2, a putative RNA helicase (Faria *et al.,* 2019). VEX1 and VEX2 form a *VSG* exclusion (VEX) complex, which assembles a sub-nuclear domain and controls *VSG* allelic exclusion (Faria *et al*., 2019). A loss of the VEX complex results in a collapse of the exclusion system leading to de-repression of silent VSGs (Faria *et al.,* 2019). VEX2 makes the greatest contribution to monoallelic expression with knockout of VEX2 resulting in simultaneous expression of multiple VSGs on cells (Faria *et al*., 2019). Monoallelic expression can be lost during switching between ES as crosstalk appears to occur during rapid switching between two ES, but multiple ES cannot be active at the same time (Borst *et al.,* 1998). The third and final requirement for antigenic variation is

that pathogens can switch expression of one antigen to another. As variants of the VSG begin to be recognised and cleared by the host immune system, a few parasites will stochastically switch their VSG (Mugnier *et al*., 2015). VSG switching events are transcriptional or DNA recombination-mediated (Li and Zhao, 2021). These appear to be highly distinct pathways.

1.4 Mechanisms of VSG switching

The VSG can be changed by transcriptional, or *in situ,* switching between telomeric ESs in which the active expression site is silenced, and a new ES becomes fully expressed (Figure 1-3d). All other ESs continue to be repressed to maintain monoallelic expression (Vink *et al.,* 2012). Transcriptional switching may be mainly used in early infections and is likely limited in its capacity to drive prolonged infections due to the small number of BESs (Bitter *et al.,* 1998). This is an example of an epigenetic switch with no DNA sequence changes as the transcription machinery is transferred from one ES to another (Cestari and Stuart, 2018). This allows for transcription of both a new VSG and the ESAGs associated with the new BES (Cestari and Stuart, 2018). ESAGs encode for a set of diverse transporters to be able to survive in a range of hosts so transcriptional switching allows access to all ESAGs for the requirements of *T. brucei* (Bitter *et al.,* 1998).

DNA recombination is another pathway used to switch the expressed VSG and overcomes only being able to use VSGs that are present in the BES (Devlin *et al.,* 2016). Recombination is the major pathway for VSG switching and likely drives the large diversity of VSGs seen in infections (McCulloch and Field, 2015). There are at least three VSG recombination mechanisms (Devlin *et al.,* 2016).

The most common recombination pathway is gene conversion (Vink *et al.,* 2012). Gene conversion involves a silent VSG being copied into the active ES (Figure 1-3a). Gene conversion is an important mechanism during infection as it enables access to a large number of silent VSGs, and not merely those that are available at telomeres (Aitcheson *et al.,* 2005). Gene conversion involves the deletion of the VSG in the active BES and replacing this with a VSG sequence, which is copied from the silent archive (Devlin *et al.,* 2016). In early infections, most VSG switches involve gene conversion of intact VSGs (Marcello and Barry, 2007). Donor sequences for recombination can come from minichromosomes, the VSG arrays or the silent ESs. Gene conversion of intact and dissimilar VSGs involves the homology of the 70 bp

repeats, which flank the VSG to give upstream homology and help guide recombination (Marcello and Barry, 2007; Devlin *et al.,* 2016).

Another recombination pathway is reciprocal recombination (or telomere exchange), where telomeres are exchanged between chromosomes by a cross-over, which moves a silent telomeric VSG into the active ES (Figure 1-3c). Unlike gene conversion, this does not lead to sequence loss, as the previously active VSG is moved to the end of the other chromosome in the cross-over reaction (Morrison *et al.,* 2009; McCulloch *et al.,* 2015). Reciprocal recombination is a minor pathway, which is used less often than other recombination pathways (McCulloch *et al.,* 2015).

The third pathway, segmental gene conversion (SGC), involves multiple VSGs, leading to novel 'mosaic' VSGs that, themselves, are not found in the genome but arise from complex, multigene recombination reactions (Figure 1-3b). SGC is mainly used in late infection and allows the activation of VSG pseudogenes (Marcello and Barry, 2007; Hall *et al.,* 2013). As with other homologous recombination pathways, SGC relies on sequence similarity. However, the homology required for SGC, rather than flanking sequences, requires sequence homology within the VSG open reading frames, on at least one end of the involved VSG gene (McCulloch *et al.,* 2015). For SGC, it is possible that the VSG sequence donors used to form a mosaic VSG are preferred if they share sequence similarity (Hall *et al*., 2013). The formation of mosaic VSGs may be inefficient relative to activation of intact VSGs (Marcello and Barry, 2007). It is assumed that during the formation of a mosaic VSG an assembly intermediate is formed, which may not be capable of expressing an intact VSG, and it has been suggested that this might occur within a silent BES (Barry and McCulloch 2001). SGC expands the number of VSG coats that might be generated beyond the number of genes encoded in the genome, and potentially allows virtually unlimited coat variants to be generated, which may be key to prolonging infections (Barry *et al.,* 2012).
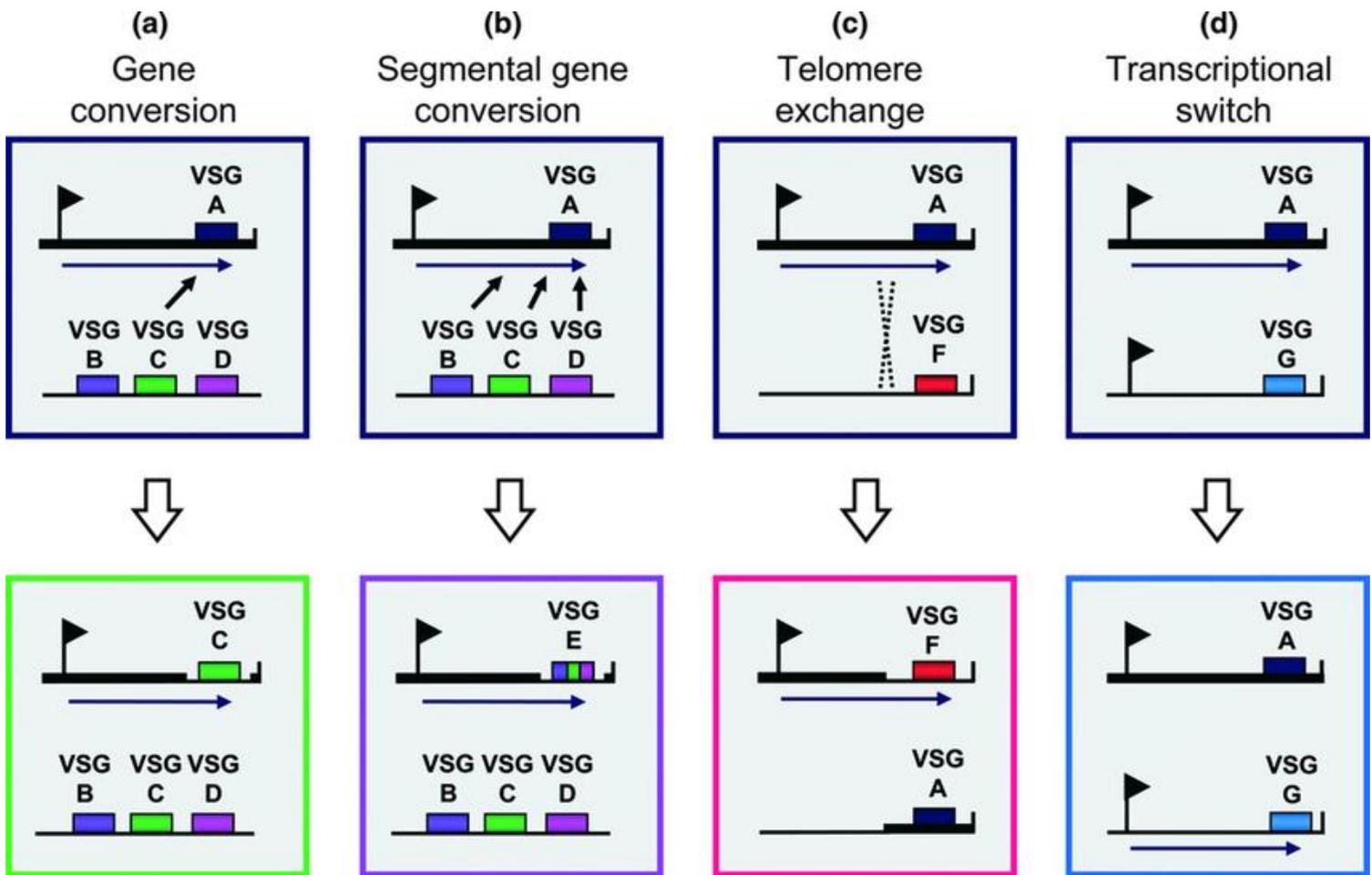
Figure 1-3. Mechanisms of VSG switching in _T. brucei._
a) Gene conversion involves replacing the active VSG (VSG A) with a silent VSG (VSG C) in the active expression site. This results in a loss of VSG A sequence.
b) Segmental gene conversion (SGC) involves the recombination of several VSG genes or pseudogenes (VSG B, C, D) to form a novel mosaic VSG (VSG E) replacing the active VSG (VSG A) in the expression site.
c) Telomere exchange involves the cross-over between telomeric ends replacing the active VSG (VSG A) with a silent one (VSG F).
d) Transcriptional, in situ, switching involves transcription of a new expression site with a new VSG (VSG G) and new ESAGs replacing the previously active VSG (VSG A) and ESAGs (Vink _et al.,_ 2012).

1.5 Repair mechanisms of double-strand DNA breaks

DNA double-strand breaks (DSBs) are toxic and can cause genomic alterations, including chromosome rearrangements, and can lead to cell death (da Silva, 2021). Natural DSBs occur in both active and silent telomeres due to their fragility and can result from metabolic reactions or DNA stressors, including DNA replication fork stalling and ionising irradiation (Mehnert _et al_., 2021; Li, 2015). Due to this, it is essential to repair the DNA damage properly and, as a result, DSB repair mechanisms have evolved. There are two major repair pathways for DSB: non-homologous end joining (NHEJ) and homologous recombination (HR). Trypanosomatids

exploit at least one of these repair pathways for survival within the host, as repair of DSBs has been linked to antigenic variation (da Silva, 2021; Mehnert *et al.*, 2021).

The NHEJ pathway involves the repair of DSBs by directly ligating the break ends together, meaning the reaction does not require a homologous template (Figure 1-4). DNA ends are bound by heterodimers Ku70/80, which shifts the repair mechanism towards NHEJ and promotes binding of DNA-PKcs, creating a binding platform that facilitates ligation of DNA ends together (Ackerson *et al.,* 2021). End joining is typically not conservative and usually results in a sequence loss, with NHEJ typically resulting in a loss of <10 bp either side of the break (Lieber, 2010). In most eukaryotes, NHEJ is active throughout the cell cycle and in proliferating and non-dividing cells, whilst HR is not active in non-dividing cells and likely most active in G2-M phase (Passos-Silva *et al.,* 2010). NHEJ is active in mammalian cells but *T. brucei* lacks detectable coding sequences of key factors, including DNA ligase IV, which is crucial for NHEJ (Burton *et al.,* 2007). Targeted assays investigating the repair of induced DSBs, such as after ISceI-directed DSB formation, did not detect repair by NHEJ in *T. brucei,* which suggests the NHEJ pathway is absent or highly supressed (Glover *et al.,* 2010).

Microhomology-mediated end joining (MMEJ), also known as alternative end joining (alt-EJ), is an alternative repair pathway to NHEJ, and can be detected in *T. brucei* (Conway *et al.,* 2002) (Figure 1-4). The use of MMEJ in *T. brucei* after an induced break is only minor, with ~10% of DNA repair through MMEJ (Glover *et al.,* 2008). MMEJ is more error-prone with loss of sequence occurring as MMEJ requires short stretches of ~2-20 bp in homology for repair (McVey and Lee, 2008; Ackerson *et al.,* 2021). In many cases, the ends of a break are resected to reveal these short microhomologies, which are aligned to guide repair of the break, meaning the sequence between the microhomologies is deleted (Ackerson *et al.,* 2021). Central to MMEJ is a specialised polymerase, DNA polymerase theta (PolQ), which fills ssDNA regions and stabilises the paired intermediates to prevent resection, which would promote HR and can be used to resolve persistent DSBs (Brambati *et al.,* 2021; Sfeir *et al.,* 2024). In other eukaryotes, this activity has been only ascribed to PolQ, but *T. brucei* does not encode a true PolQ with both DNA Pol and helicase domains (Leal *et al.,* 2020). Instead, *T. brucei* encodes discrete DNA Pol and DNA helicase proteins, which may have a PolQ like function in MMEJ (Leal *et al.,* 2020).

The major DSB repair pathway in kinetoplastid parasites is homologous recombination (HR). HR is a universal process, which relies on a template DNA to repair the DSB and is more

efficient than MMEJ (Figure 1-4). HR is dominant in the S and G2/M phases in the cell cycle when sister chromatids or homologous chromosomes are present as a DNA template (Ackerson *et al.,* 2021). HR is important for repair in many cellular processes such as recombination of the VSG, repairing DNA damage, and reinitiating stalled replication forks (Passos-Silva *et al.,* 2010). HR is highly efficient in *T. brucei* with a minimum homology of 42 bp being sufficient in tsetse fly stages and only 24 bp homology needed in BSF and, unlike MMEJ, is impeded by base mismatches during homology search and strand exchange (Barnes and McCulloch, 2007).

When a DSB occurs, it is recognised by a DNA damage sensing MRE11-RAD50-NBS1 (MRN) complex, which recognises the free DNA ends at the DSB (Mehnert *et al.,* 2021). The MRN complex binds and leads to the ends of the DSB being resected (Mehnert *et al.,* 2021). Ataxia-telangiectasia-mutated (ATM) is recruited by the MRN complex and phosphorylates the H2A histone, which recruits factors for response to the DSB (Kuo and Yang, 2008). An exonuclease, EXO1, processes DNA from the resected ends and generates long stretches of single strand DNA (ssDNA) at the ends of the DSB with 3' overhangs (Mehnert *et al.,* 2020). The overhangs are used as a nucleation site for recombinases (Kelso *et al.,* 2017). The replication protein A (RPA), which has high affinity for ssDNA, binds to the overhangs (Wold, 1997). BRCA2 transfers DNA filaments from RPA to RAD51 and RAD51 forms a nucleoprotein filament on the ssDNA overhang (Yang *et al.,* 2005; San Filippo *et al.,* 2008). BRCA2 is essential for error-free repair of DSBs (Yang *et al.,* 2005). Loading of recombinases to the ssDNA limits the rate but is enhanced by recombination mediators including BRCA2, RAD51 paralogs, and RAD52 (Marin *et al.,* 2018). The nucleoprotein filament searches the genome for sequences with homology to the ssDNA overhang to use as a template for repair (Marin *et al.,* 2018). After finding a sequence with homology, the filament invades the duplex DNA, in a process termed strand invasion, and facilitates base pairing of the ssDNA to the complementary template sequence (Kelso *et al.,* 2017). The homologous strand of duplex DNA displaces forming a displacement loop (D-loop) structure between the 3' overhang strand and the homologous chromosome (van der Heijden *et al.,* 2008). A DNA polymerase extends the invading strand by synthesising new DNA to restore the DNA lost at the DSB and DNA ends are ligated together (Kelso *et al.,* 2017). HR accurately repairs DSBs.
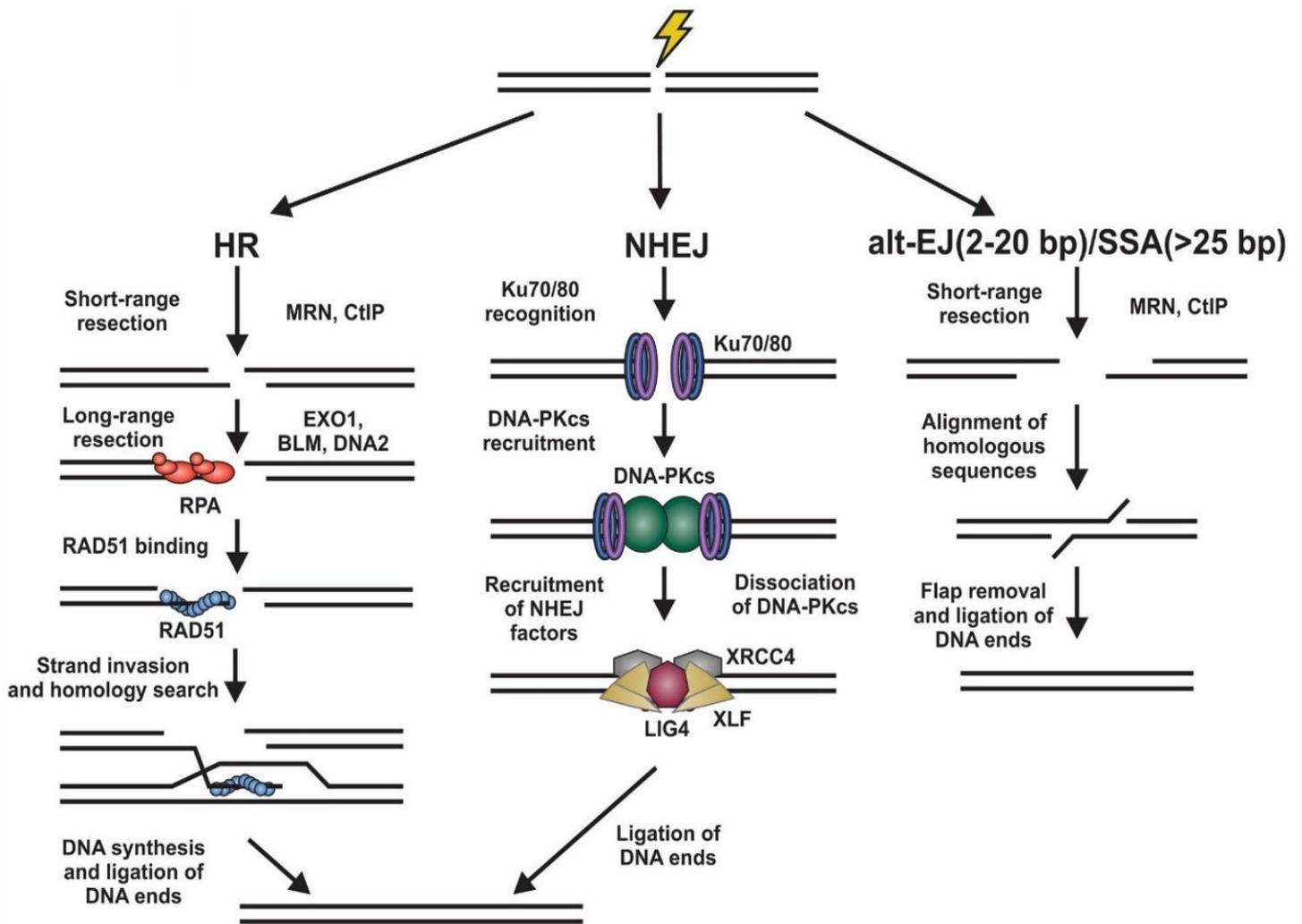
Figure 1-4. DNA double-strand break repair mechanisms.
Homologous recombination (HR) involves resection of break ends to generate ssDNA which is bound by RPA. RAD51 facilitates search for homology as HR requires a donor DNA template for repair. Non-homologous end joining (NHEJ) involves direct ligation of break ends back together. Alternative end joining (Alt-EJ) involves the break being resected to reveal microhomologies, which are aligned and ligated together (Ackerson *et al.,* 2021).

1.6 The role of double-strand DNA breaks in VSG switching

HR has been shown to be closely linked to antigenic variation as mutations to key factors involved with HR affect VSG switching. These key factors include RAD51, RAD51-3 paralogue and the RAD51 interacting protein BRCA2, as loss of any of these proteins impairs VSG switching (McCulloch and Barry, 1999; Proudfoot and McCulloch, 2005; Trenaman *et al.,* 2013). Loss of other DNA repair proteins, RECQ2, TOPO3α or RMI1, results in increased VSG switching (Devlin *et al.,* 2016; Kim and Cross, 2010; Kim and Cross, 2011). Although inactivation of MRE11 in *T. brucei* leads to an impairment in HR, this does not seem to affect

VSG switching efficiency, though it does alter the range of *VSG*s activated after a DSB is targeted to the VSG BES (Robinson *et al.,* 2002; Mehnert *et al.,* 2021). Due to the link between HR and VSG switching, a DSB in the active ES has been long suspected as a first step in VSG switching (Barry, 1997; Borst *et al.,* 1997).

A study by Boothroyd et al. (2009) found that using the yeast endonuclease I-SceI to introduce a DSB into the ES adjacent to the 70 bp repeats upstream of the active VSG increased the rate of *in vitro* VSG switching. The DSB elicited VSG recombination and demonstrated that a DSB can trigger VSG switching (Boothroyd *et al.,* 2009). The increased frequence of VSG switching was not observed when the I-SceI DSB was introduced into the *VSG* pseudogene or in place of the 70 bp repeats (Boothroyd *et al.,* 2009). The location of the DSB also influences VSG switching, as not all DSBs in the ES induce switching with equal efficiency (Glover *et al.,* 2013). Glover et al. (2013) showed that inducing a break at the promoter or telomere does not result in VSG switching, but a break adjacent to the 70 bp repeats almost always triggered switching. These data indicate that a DSB can induce VSG switching and implicate the 70 bp repeats as important mediators of DSB repair and potential sites of break formation (Boothroyd *et al.,* 2009; Glover *et al.,* 2013). However, no native *T. brucei* endonuclease that might act like I-SceI to generate a break has been described, so the route for formation of a VSG switch-initiating lesion or lesions is not addressed by these experiments (Vink *et al.,* 2012). As a result, proposals for the how a DSB in the ES may form have been described.

RNA-DNA hybrids (R-loops) have been suggested to be involved in VSG switching by potentially causing a DSB in the BES (Briggs *et al.,* 2018b) or by forming in response to a DSB (Girasol *et al.,* 2023b). R-loops are a particular form of RNA-DNA hybrid where the RNA base pairs with one strand of the DNA helix and displaces a single strand of DNA forming a three-stranded structure (Nanavaty *et al.,* 2017). R-loops have enhanced stability when compared with dsDNA because of the conformation they adopt (Briggs *et al.,* 2018a). R-loops can be generated by many cellular processes, including transcription, and can lead to genome instability, which may be in part due to the formation of DSBs (Briggs *et al.,* 2018a). Due to the detrimental effects R-loops can have on genome function, *T. brucei* try to limit the formation of R-loops and have activities to resolve them when they form (Briggs *et al.,* 2018a). Amongst these R-loop resolving activities are endonucleases, RNase H1 or RNase H2, which degrade the RNA in the RNA-DNA hybrid (Nanavaty *et al.,* 2017). Previous work has mapped that R-loops localise across the *T. brucei* genome (Briggs *et al.,* 2018a). Loss of *T. brucei* RNase H2 or RNase H1 has been shown to change R-loop distribution resulting in an accumulation of

R-loops and DNA damage across the VSG ES and an increase in VSG switching (Briggs *et al.,* 2018b; Briggs *et al.,* 2019). Loss of RNase H1 results in increased replication-associated damage including in the active VSG ES during DNA replication (Briggs *et al.,* 2018b). R-loops accumulate within the active VSG ES in RNase H1 mutants, with the loss leading to increased transcription from normally silent *VSGs* extending to the VSG surface coat as VSG switching is induced (Briggs *et al.,* 2018b). This work revealed a connection between R-loops in the VSG BESs and VSG switching, which was initially interpreted as the hybrids causing damage and eliciting VSG recombination (Briggs *et al.,* 2018b; Briggs *et al.,* 2019).

Two studies by Girasol et al. (2023) tested the above model by characterising the proteins that interact with RNA-DNA hybrids in *T. brucei*. Many proteins appear to interact with RNA-DNA hybrids, amongst which four proteins were shown to affect VSG switching or monoallelic VSG expression, including two proteins, which were previously associated with HR and VSG recombination: RAD51 and RAD51-3 (Girasol *et al.,* 2023a). The loss of RAD51 resulted in an increase in R-loops identified by performing DNA-RNA hybrid immunoprecipitation (DRIP) with S9.6 antiserum coupled with deep sequencing (DRIP-seq) (Girasol *et al.,* 2023b). This study also identified that R-loops are enriched at the start and ends of polycistronic transcription units (PTUs) with a loss of RAD51 resulting in a change in distribution of R-loops. The loss of RAD51 resulted in a loss of DRIP-seq signal across the 70 bp repeats, with enrichment across the ESAGs, which indicates that RAD51 may be predominantly involved with the formation/stabilisation of R-loops at the 70 bp repeats (Girasol *et al.,* 2023b). Surprisingly, the results also indicated that VSG associated R-loop levels are modulated by RAD51 and RNase H1 in both the active VSG and across the silent VSG archive (Girasol *et al.,* 2023b).

After a link was identified between R-loops and RAD51, Girasol et al. (2023b) mapped putative DNA DSBs by Breaks Labelling In Situ and Sequencing (BLISS). The DRIP-seq signal had pronounced enrichment centring on discrete BLISS signals in WT cells revealing that DNA break sites associate with R-loops across the genome. The loss of RAD51 impairs the formation of R-loops at BLISS mapped breaks, revealing that RAD51 plays a widespread role in recruiting or stabilising R-loops at DNA break sites (Girasol *et al.,* 2023b). To determine if DNA breaks were detected in the VSG ES and if loss of RAD51 impacts the level and distribution of breaks, BLISS reads were mapped to annotated BESs. This revealed BLISS signal was higher in the active BES (BES1 in the strain used) than in silent BESs (Girasol *et al.,* 2023b). Furthermore, the BLISS accumulation was most pronounced around the 3' end of

VSG221 in BES1, and the BLISS signal showed a dramatic increase in RAD51 mutants. In summary, Girasol et al. (2023b) revealed that RAD51 plays a role in repairing DNA breaks localised to the active VSG and that this involves recruiting RNA-DNA hybrids (Figure 1-5).
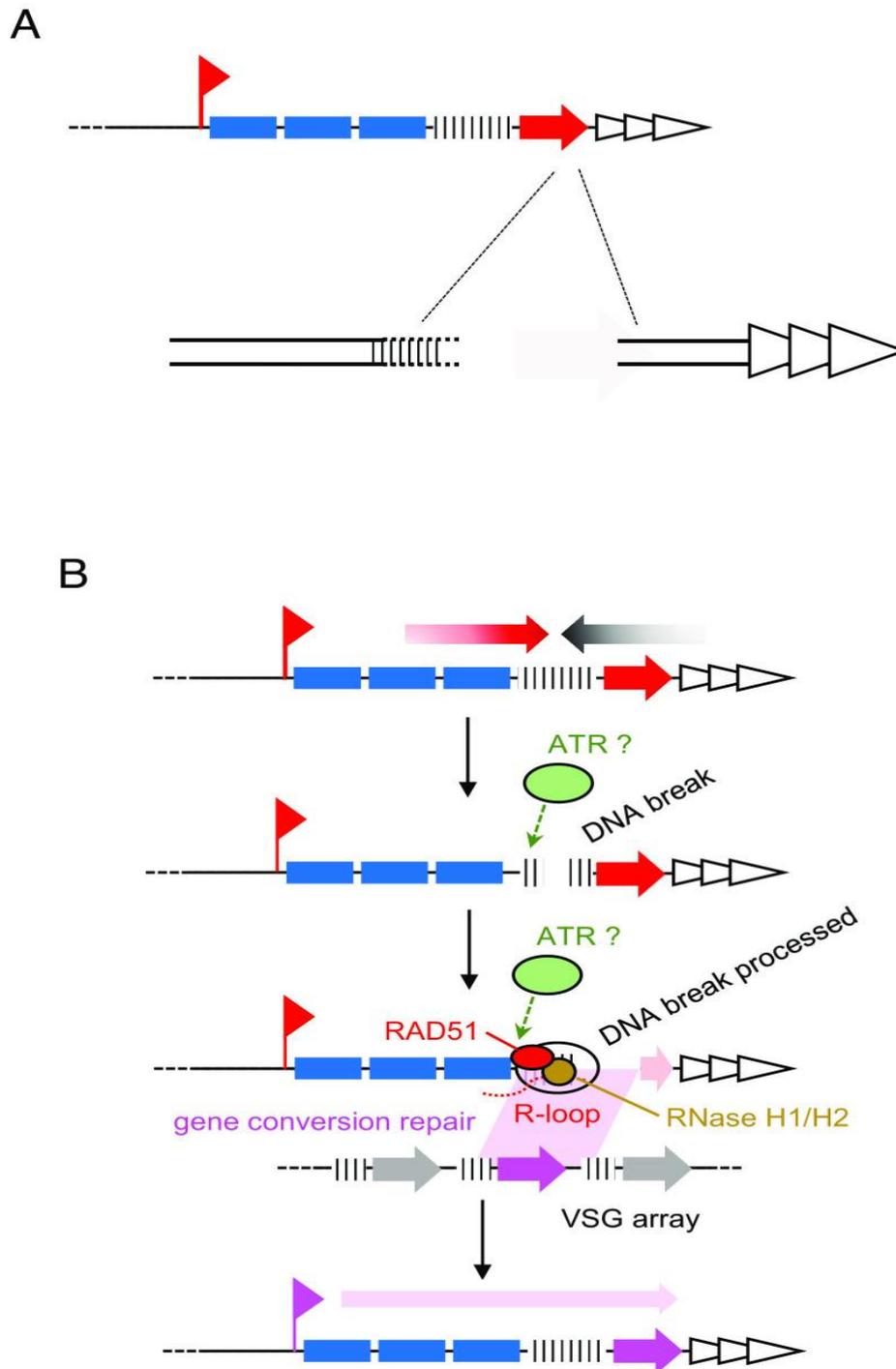
Figure 1-5. Model of R-loops and DNA breaks during _T. brucei_ VSG switching.
A. Diagram of BLISS mapping of predicted DNA DSB in the active BES.
B. Predicted events during VSG recombination. After the break forms in the BES, processing of the DSB occurs and RAD51 is recruited, which directs R-loops to the break. The DSB is then repaired. (Girasol _et al.,_ 2023b).

Data from Girasol et al. (2023b) demonstrated that when DNA breaks arise in the active BES this leads to the RAD51-dependent accumulation of R-loops. However, there is still uncertainty of how and where the DSBs in the BES form (Girasol *et al.,* 2023b). The proposal that DNA breaks originate in the 70 bp repeats and are processed extensively downstream of the *VSG* would match BLISS mapping by Girasol et al. (2023b). The 70 bp repeats could play a role as both the cause of the DNA damage and as a participant in the subsequent repair. However, all these suggestions need to be tested, which was the aim of this project.

1.7 Aims and Approaches

Two key features of the telomere-proximal region of the VSG BES have been implicated in the formation of the BLISS-mapped DNA break: the VSG 3' end, where the downstream end of the break is found; and the 70 bp repeats, where R-loops are enriched and in whose absence repair of DSB-induced VSG switching is impaired. This project seeks to examine how each cis-acting element contributes to break formation and VSG switching. To address this question, this study had two aims, each case asking how alteration of the VSG BES affects R-loop localisation and DNA break formation.

1. What happens when the 70 bp repeats are removed from the active expression site?

2. What happens when the endogenous *VSG* (*VSG221*), or a distinct *VSG* (*VSG121*), are relocated to other locations into the active BES?

Generating cell lines with the above modifications will allow a comparison of DRIP-seq and BLISS with the natural organisation of the active BES, thus testing whether the 70 bp repeats and/or VSG sequence drive R-loop and DNA DSB formation, as well as testing if telomere proximity is necessary for break formation linked to VSG switching.

2. Materials and Methods

2.1 General molecular biology techniques

2.1.1 Primer design

The sequences of genes of interest from *Trypanosoma brucei* Lister 427 genome were obtained from TriTrypDB (http://tritrypdb.org) and viewed on Benchling (https://www.benchling.com/).

For experiments using CRISPR/Cas9, the Eukaryotic Pathogen CRISPR guide RNA/DNA Design Tool (http://grna.ctegd.uga.edu/) was utilised. The "SpCas9: 20nt gRNA, NGG PAM on 3' end" nuclease option and the *T. brucei* Lister427 TriTrypDB-28 genome were selected. The sequence of the genes of interest were entered and sgRNA, which had the lowest number of predicted off target sequences was accepted, with majority of sequences having no off targets. PAM sequences were removed from the primer sequence and a T7 Pol promoter sequence (GAAATTAATACGACTCACTATAGG) added to the 5' end of the primer; and a sequence complementary to a G00 primer (GTTTTAGAGCTAGAAATAGC) was added to the 3' end of the primer. For donor primers, a 30 bp sequence flanking the sgRNA region was selected; with sequences targeting the required section of donor plasmid added to the 3' end. Primers were ordered from Sigma Aldrich and sequences are shown in Appendix 1-3.

2.1.2 Polymerase chain reaction (PCR)

PCR was performed to amplify DNA fragments to be used for CRISPR/Cas9. To amplify donor DNA, the PCR mix (30 µl) contained the following reagents: 10 µl of 5X Phusion® HF buffer (NEB), 1.9 µl 1.87 mM $MgCl_2$, 10 µl 2 µM gene-specific forward primer, 10 µl 2 µM gene-specific reverse primer, 1.5 µl 3% (v/v) DMSO (NEB), 1 µl 10 mM dNTPs (Promega), 1 µl 30 ng pPOTv7 plasmid (Dean *et al.,* 2015; Figure 3-1), and 0.5 µl Phusion® DNA Polymerase (NEB) with 14.1 µl Milli-Q water. The gene-specific primer sequences are outlined in Appendix 1-3.

2.1.3 Agarose gel electrophoresis

For PCR product analysis, agarose gel electrophoresis was performed. A gel containing 1% agarose in 1X TAE buffer (40 mM Tris base, 19 mM acetic acid, and 1 mM EDTA) was prepared. To allow for DNA detection by ultraviolet light, SYBR® Safe DNA Gel Stain (Life Technologies) was added into the melted agarose gel. A 12 µl reaction, 5 µl of PCR product, 5 µl Milli-Q water, and 2 µl of 6X loading dye (50% glycerol, 0.125% bromophenol blue,

0.125% xylene cyanol in 1X TAE), was loaded onto the gel alongside 5 µl of Promega 1 kb DNA ladder. Where GoTaq® Green Master Mix (Promega) was used, SYBR® Safe DNA Gel Stain was not used and 10 µl of reaction was run. Gels were run at 100V for 30-45 minutes in 1X TAE buffer on Mini-Sub® Cell GT Cell tanks (BioRad). For visualisation, gels were imaged using a GelDoc™ XR+ System (BioRad).

2.1.4 Purification of PCR products by ethanol precipitation

To use PCR amplified DNA for transfection, DNA was pooled together and concentrated by ethanol precipitation. For each DNA amplicon to be transfected, a total of 3-10 individual PCR reactions were pooled together in a 1.5 ml tube with 0.1X volume of 3M sodium acetate (pH 5.2) and 2.5X volumes of ice-cold 100% ethanol. The mix was incubated for 30 minutes at -80°C then centrifuged at 14,000 x gfor 15 minutes at 4°C. The DNA pellet was washed with 1 ml of 70% ethanol and centrifugation repeated. If the DNA pellet was difficult to visualise, 2 µl of GlycoBlue™ Coprecipitant (Invitrogen) was added to aid visualisation. The DNA pellet was then air-dried for 5-10 minutes and resuspended in 25 µl Milli-Q water. DNA was stored until transfection at 4°C. DNA concentrations were measured, and 10 µg of DNA used for transfection.

2.1.5 Agarose gel extraction of DNA fragments

Agarose gels were visualised using a UV transilluminator and desired DNA fragments excised from the gel using a sterile razor blade and placed in a 1.5 ml microcentrifuge tube. DNA was extracted from the gel fragment using QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions. DNA was eluted in 30-50 µL of Buffer EB or nuclease-free water.

2.1.6 Measuring DNA concentration

Concentrations were assessed by photo spectrometry using a NanoDrop 2000™ Spectrophotometer (ThermoScientific). 1 µl of Milli-Q water was added to the nanodrop as a reference solution (blank) and then 1 µl used to assess concentration in ng/µl.

2.2 Generation of VSG expression plasmid constructs

2.2.1 Amplification of VSG fragments by PCR

VSG fragments were amplified by PCR. A mix of 10 µl 5X Phusion® GC reaction buffer, 2 µl 10 mM dNTPs (Promega), 2 µl 2µM gene-specific forward primer, 2 µl 2µM gene-specific reverse primer, 0.5 µl Phusion® DNA Polymerase (NEB), 2 µl of *T. brucei* Cas9/T7 wild type

gDNA, and 33.5 µl Milli-Q water. The PCR thermal cycling conditions were 98°C for 5 minutes followed by 35 cycles of 98°C for 15 seconds, single tubes were annealed at differing temperatures between 49-64°C for 20 seconds, and 72°C for 60 seconds followed by a final extension at 72°C for 5 minutes. Reactions were run on a 1% agarose gel for confirmation of VSG amplification (2.1.3).

A-tails were added to the reaction to enable the *VSG* fragment to ligate to pGEM®-T Easy vector (Promega) by briefly spinning the tubes to collect contents then adding 10 µl 5X GoTaq® Green Master Mix (Promega), 2 µl dATPs (Promega), and 0.2 µl GoTaq® (Promega) and incubating at 72°C for 20 mins. The full reaction was run on a 1% agarose gel and desired DNA extracted (2.1.3 and 2.1.4).

2.2.2 DNA ligation

Ligations were set up to ligate the *VSG* fragment into a subcloning vector and subsequently to a final vector. To ligate the gel extracted VSG fragments into the subcloning vector, pGEM®-T Easy (Promega), a 10 µl reaction was set up of 5 µl 2X Ligase Buffer (NEB), 1 µl T4 DNA Ligase (NEB), 0.5 µl pGEM®-T Easy (Promega), and 3.5 µl of >50 ng/µl gel extracted VSG fragment.

On confirmation of the correct insertion of the *VSG* into the subcloning vector by PCR, the subcloning vector + *VSG* were ligated into the final vector – a pPOTv7 backbone with either a neomycin, blasticidin or hygromycin resistance cassette (Dean *et al.,* 2015). A mix of 1 µl of 10X Ligase Buffer (NEB), 1 µl T4 DNA Ligase (NEB), 2 µl of >50 ng/µl gel extracted vector (pPOTv7), and 6 µl of > 50 ng/µl pGEM®-T Easy + VSG fragment was set up.

Two ligations were set up for each ligation – one for each gene (*VSG121* and *VSG221*). Reactions were incubated overnight at 4°C and 5 µl used to transform competent bacteria cells as described in 2.2.3.

2.2.3 Transformation of competent cells

Chemically competent *Escherichia coli* MAX Efficiency® DH5-α ™ cells (Life Technologies) were used for transformation. Aliquots stored at -80°C were thawed on ice and 5 µl of each ligation product was added to 50 µl of bacteria and incubated on ice for 30 minutes. This was followed by a heat shock by incubating at 42°C for 45s and immediately placing on ice for 2 mins. 1 ml of SOC media (2% vegetable peptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM $MgSO_4$, and 20 mM glucose; NEB) was added and incubated at

37°C with shaking for 45 mins. The bacterial suspension was plated on LB agar (10 g/L of Tryptone, 5 g/L of yeast extract, 10 g/L of NaCl, and 10 g/L of agarose) supplemented with 100 µg/ml of ampicillin and incubated at 37°C overnight before being screened (2.2.4).

2.2.4 Colony screening by PCR

Plates were inspected for surviving colonies and 8-16 colonies screened per transformation by PCR. A mix of 4 µl 5X GoTaq® Green Master Mix (Promega), 1 µl 10 mM dNTPs (Promega), 1 µl 10 µM gene-specific forward primer, 1 µl 10 µM gene-specific reverse primer, 0.2 µl GoTaq® DNA Polymerase (Promega) and 13.3 µl Milli-Q water was used for each colony. Each bacterial colony was picked with a sterile 200 µl pipette tip and dipped into 20 µl of PCR mix then streaked on to a new ampicillin-containing agar plate (100 µg/ml). The following PCR thermal cycling conditions were used; 95°C for 3 mins, 35 cycles of 95°C for 30s, 55-60°C for 30s, 72°C for 30s-1m 40 (1 min/kb) and a final extension at 72°C for 5 mins. PCR products were run on a 1% agarose gel to identify the correct insertion of fragment (2.1.3).

Positive colonies were transferred into 5 ml of liquid LB broth containing 100 µg/ml of ampicillin and incubated overnight at 37°C with shaking. The culture was then centrifuged for 10 mins at 800 x g and pellets used for plasmid purification. The QIAprep Spin Miniprep Kit (Qiagen) was used according to the manufacturer's instructions and DNA eluted in 50 µl EB buffer. The concentrations of the plasmids were assessed then sent for sequencing to confirm if sequences were correct (2.1.6 and 2.2.6).

2.2.5 Restriction enzyme digestion

Plasmids were digested with restriction enzymes from NEB according to the manufacturer's instructions. Reactions were set up using the whole miniprep (50 µl), 1 µl of *Kpn*I, 1 µl *Sac*I, and 6 µl of CutSmart® Buffer. The reactions were incubated at 37°C for one hour. The plasmids were separated and extracted from a 1% agarose gel (2.1.3 and 2.1.5). The purified DNA was then used for ligation to final vector (2.2.2).

A test digest was carried out to determine if pGEM®-T Easy + VSG had correctly inserted into the final vector pPOTv7 (Dean *et al.,* 2015). 10 µl of miniprep, 1 µl *Kpn*I (NEB), 1 µl *Sac*I (NEB), 2 µl CutSmart® Buffer (NEB), and 6 µl Milli-Q water were incubated at 37°C for one hour and ran on a 1% agarose gel (2.1.3). If the test digest was correct according to the size of resulting fragments, the plasmid was sent for Sanger sequencing (2.2.6). The generated plasmid

with correct sequence confirmed was then used as a template for PCR amplification of CRISPR/Cas9 donors (2.1.2).

2.2.6 Sanger sequencing

To confirm if generated plasmids or PCR products were correct, Sanger sequencing was carried out according to the sequencer's instructions. For generated plasmids, 4 µl of purified plasmid DNA and 3 µl of 10 µM primer was pipetted into a Mix2Seq bar-coded tube (Eurofins). For PCR products, products were cleaned up using QIAquick PCR Purification Kit (Qiagen), and 6 µl of 10 ng/µl purified PCR product and 6 µl of 5 µM primer were added. Sequence of primers used for sequencing are outlined in Appendix 2. Sequencing was performed by Eurofins Genomics and DNA sequence aligned with reference sequence using Benchling.

2.3 *In vitro* culture of *T. brucei*

2.3.1 Culture maintenance of *T. brucei*

To allow for CRISPR/Cas9 editing in BSF *T. brucei* Lister 427, a puromycin resistant *T. brucei* cell line that constitutively expresses Cas9 and T7 RNA polymerase was used (Girasol *et al.,* 2023a). HMI-9 medium was used to maintain *in vitro* growth. A pre-formulated HMI-9 powder (Gibco) was supplemented with 3% $NaHCO_3$ and 200 µM β-mercaptoethanol and filter sterilized using a 0.22 µM bottle top filter. For use, HMI-9 was supplemented with 10% bovine serum (FBS, Gibco) and 1% penicillin/streptomycin. Media was stored at 4°C for use.

Trypanosomes were maintained in vented cell culture flasks and incubated at 37°C in 5% CO_2. Cells were maintained in 5 ml of media at a density between $1x10^4$ and $2x10^6$ cells/ml. Cell density was determined by counting cells in 10 µl of culture on a Neubauer improved hemocytometer (Marienfeld-Superior) and multiplied by 10^4 to determine number of cells per ml in the culture. Selective media was prepared when needed using concentrations outlined in Table 2-1.

Table 2-1. Concentrations of drugs used for selective media.

Drug	Source	Maintenance	Selection
Puromycin	Invivogen	0.2 µg/mL	Cas9/T7 Polymerase
Blasticidin	Invivogen	10 µg/mL	70bp knockout clones
Neomycin (G418)	Invivogen	2.5 µg/mL	Insertion of VSG121
Hygromycin	Invivogen	5 µg/mL	Insertion of VSG221

2.3.2 Cryopreservation of *T. brucei* cell lines

Cryopreservation was performed to allow long term storage of *T. brucei* cell lines. From mid-logarithmic cultures, a 1:1 ratio of cell culture to freezing medium (30% glycerol in drug-free HMI-9) was added to a 2 ml cryotube giving a final concentration of 15% glycerol. Stabilates were frozen at -80°C wrapped in cotton wool for 24 hours before being transferred to liquid nitrogen for long term storage.

To recover trypanosomes, stabilates were thawed at 37°C in 5% CO_2 and added to 3 ml of drug-free HMI-9. To remove glycerol, stabilates were centrifuged at 1000 x g for 10 minutes, media removed and added to 9 ml of drug-free HMI-9 in a vented cell culture flask. Appropriate selective drugs were added the following day and cells then passaged as normal.

2.3.3 Transfection of BSF *T. brucei* parasites

The TbCas9/T7 cell line was cultured in HMI-9 media and supplemented with 0.2 µg/ml of puromycin. A total of $3x10^7$ cells at a density of $1x10^6$ cells/ml were harvested for transfection by centrifugation at 1000 x g for 10 minutes. The cell pellets were resuspended in 100 µl of transfection buffer (90 mM sodium phosphate, 5 mM potassium chloride, 0.15 mM calcium chloride, 50 mM HEPES, pH 7.3) and transferred to a cuvette containing 10 µg of relevant DNA for transfection (CRISPR PCR products). The cuvette was electroporated in Lonza Amaxa Nucleofector machine using program Z-001. The parasites were transferred into prewarmed HMI-9 medium with puromycin at 0.2 µg/ml. This cell-medium solution was diluted to concentrations of ¼ and 1/20 and 1 ml was distributed into each well on a 24-well plate. The 24-well plates, as well as an undiluted flask, were allowed to recover for 24 hours at 37°C in 5% CO_2.

After recovery, 1 ml of media containing 2X concentration of the appropriate drug (Concentrations shown in Table 2-1) were added to each well and then incubated for 5-7 days. Transformants were identified by examining the 24-well plates under a light microscope and counting the number of wells containing live parasites. 1 ml of surviving clones were added to 5 ml of HMI-9 media with 1X concentration of appropriate drug and incubated for 24-48 hours before gDNA was extracted to confirm if the transfection was successful.

2.3.4 Genomic DNA (gDNA) extraction

To obtain gDNA, cells were harvested by centrifugation at 1,000 x g for 8 minutes at RT after which, gDNA extraction procedure followed manufacturer's instructions for DNeasy Blood

and Tissue DNA extraction kit (Qiagen). The gDNA was eluted in 50 µl buffer, concentration measured and stored at -20°C for future use.

2.4 Characterisation of Clones

2.4.1 Screening clones by PCR

Selected clones were screened by PCR to determine if the transfection was successful. A PCR mix of 4 µl GoTaq® Green Master Mix (Promega), 1 µl 10 mM dNTPs (Promega), 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 0.2 µl GoTaq® DNA Polymerase, 12.8 µl Milli-Q water, and 1 µl of the relevant clones' gDNA was set up. The primer sequences shown in Appendix 2. The PCR thermal cycling conditions used were 95°C for 3 mins, 35 cycles of 95°C for 30s, 55°C for 30s, 72°C for 2 mins 30s followed by a final extension at 72°C for 5 mins. PCR products were then visualised on a 1% agarose gel to confirm if the transfection was successful (2.1.3).

2.4.2 Growth analysis

For cell growth analysis, cell lines were grown at mid-logarithmic phase and new cultures started at a density of $1x10^4$ cells/ml. Cell density was recorded at intervals of 24, 48 and 72 hours.

2.4.3 VSG Immunofluorescence imaging

A total of $1x10^6$ cells were collected by centrifugation at 1,000 x gand fixed in 1% methanol-free formaldehyde for 15 minutes (5 minutes at 37°C then 10 minutes at RT) then washed twice with 1X PBS by centrifugation at 1000 x g for 5 minutes. Pellets were resuspended in 30 µl of ice cold 1% BSA and left to dry overnight at RT. Cells were then rehydrated with PBS for 5 minutes and washed three times with PBS, leaving the PBS on for 5 minutes each wash. Cells were adhered to a slide coated in Poly-L-Lysine solution (Sigma Aldrich). The cells were blocked with 50% FBS in PBS for 15 minutes and washed twice with PBS. The rabbit serotype primary antibody, α-VSG121, was used at a dilution of 1:10,000 in PBS/3% FBS and was added and incubated for 1 hour at RT. The primary antibody was then washed three times with PBS. The goat serotype secondary antibody, α-rat Alexa Fluor® 488 (Molecular Probes®), was used at a dilution of 1:1,000 and added for 1 hour at RT in the dark then washed three times with PBS. Slides were visualised using a Leica DiM8 microscope and images acquired using a Hamamatsu Orca camera, 2k by 2k camera chip. Images were analysed using FIJI software.

2.4.4 Protein Analysis by western blot

2.4.4.1 Cell preparation

A total of 1×10^7 cells were collected at a density of 1×10^6 cells/ml through centrifugation at 1000 x g for 10 minutes. Cell pellets were resuspended in 20 µl of loading dye (50 µl 4X NuPAGE® LDS Sample Buffer (Life Technologies), 5 µl 2% β-mercaptoethanol, 20 µl 5X protease inhibitor, and 25 µl Milli-Q water). Samples were stored at -80°C until use.

2.4.4.2 SDS-PAGE

A 10 µl sample was loaded in each well of a 10-well NuPAGE® 4-12% Bis-Tris Protein Gel (ThermoFisher) alongside 10 µl of Novex® Sharp Protein Standard (Invitrogen) ladder. The gel was run at 180V for 50 minutes in 1X NuPAGE® MOPS SDS Running Buffer (ThermoFisher).

2.4.4.3 Protein transfer

The PVDF membrane (GE Life Sciences) was soaked in methanol to activate and then soaked in 1X transfer buffer for 10 minutes (25 mM Tris pH 8.3, 192 mM glycine, 20% (v/v) methanol) alongside two sponge filter pieces and the NuPAGE® gel. The proteins separated through SDS-PAGE were then transferred to the PVDF membrane in 1X transfer buffer by stacking the sponge, PVDF membrane, NuPAGE® gel, then sponge and ran in position A on Mini Trans-Blot® Cell (Bio-Rad) for 30 minutes. The membrane was incubated with Ponceau S solution (Sigma Aldrich) until bands were visible to confirm transfer was successful.

2.4.4.4 Antibody incubations

The membrane was washed with PBS-T (1X PBS pH 7.4, 0.05% Tween-20) for 5 minutes with shaking. Blocking was carried out by adding 10 ml of blocking buffer (5% milk powder in PBS-T) for 1 hour with shaking at RT. Antibody α-EF1α, mouse serotype, was diluted to 1:40,000 in blocking buffer and added to the membrane and incubated for 1 hour with shaking at RT. The second antibody, α-γH2A, of rabbit serotype was diluted to 1:10,000 in blocking buffer and added to the membrane with incubation for 1 hour at RT. LiCor secondary antibodies (anti-mouse or anti-rabbit) were diluted in blocking buffer at 1:10,000, added to the membrane and incubated for 1 hour with shaking at RT. After each antibody incubation, the membrane was washed three times in PBS-T for 10 minutes. The membrane was imaged using Odyssey® DLx (LI-COR). Images were analysed using FIJI software.

2.4.5 Oxford Nanopore sequencing

A total of $2x10^8$ cells of each clone were collected by centrifugation at 1,000 x g for 10 minutes. The supernatant was removed, and samples stored at -80°C until processing.

Oxford Nanopore sequencing was carried out by Craig Lapsley. DNA was extracted using Monarch® HMW DNA Extraction Kit for Cells & Blood (NEB) and sequenced on a R10 flow cell using an Oxford Nanopore Technologies GridION device. The following method for analysis of the results from Oxford Nanopore sequencing were written and completed by Marija Krasilnikova. For read depth coverage, sequencing results were mapped to the TriTrypDB v.64 *T. brucei brucei* Lister 427 2018 genome using minimap2 (http://tritrypdb.org). The output files were coverted, processed, and indexed into sorted bam files using samtools. To assess the read depth coverage for each sample analysed, bamCoverage (of deepTools) was used applying minimum MAPQ mapping quality cutoff of 1. The following script was used:

```
#!/bin/bash

# Mapping Holly's and Selina's ONT data to b64 427 2018 on the HPCC - WORKS

for sample in $(ls /export/III-
data/mcculloch/2091636k/2023_08_01_Holly_Selina_BES_ONT_runs/fastq_all_concat/*.fas
tq.gz | sed -r 's/.fastq.gz//' | uniq)
do
        minimap2 -ax map-ont /export/III-
data/mcculloch/2091636k/reference_genomes/TriTrypDB-
64_TbruceiLister427_2018_Genome.fasta ${sample}.fastq.gz >
${sample}_427_2018_v64.sam &&
        samtools view -Sbu ${sample}_427_2018_v64.sam -o
${sample}_427_2018_v64.bam &&
        samtools sort ${sample}_427_2018_v64.bam -o
${sample}_sorted_427_2018_v64.bam &&
        rm ${sample}_427_2018_v64.sam &&
        rm ${sample}_427_2018_v64.bam &&
        samtools index ${sample}_sorted_427_2018_v64.bam &&
        bamCoverage -b ${sample}_sorted_427_2018_v64.bam -of bigwig -o
${sample}_sorted_427_2018_v64 &&
        bamCoverage -b ${sample}_sorted_427_2018_v64.bam -of bigwig -o
${sample}_sorted_427_2018_v64_mapQ1 --minMappingQuality 1 &&
        minimap2 -c -x map-ont /export/III-
data/mcculloch/2091636k/reference_genomes/TriTrypDB-
64_TbruceiLister427_2018_Genome.fasta ${sample}.fastq.gz >
${sample}_427_2018_v64.paf
done
```

Sequencing results were also compared to the annotated Oxford Nanopore assembly of the same parent strain using Benchling (https://www.benchling.com/).

2.4.6 INDUCE-seq

2.4.6.1 Cell harvesting and cross-linking

The methodology for INDUCE-seq was adapted from Dobbs et al. (2022). A total of 2×10^8 cells were collected in mid-logarithmic phase and centrifuged at 1000 x g for 10 minutes at 4°C. The supernatant was removed, and pellet resuspended in 17.5 ml pre-warmed media with 2.5 ml of 15% methanol-free formaldehyde. Cells were incubated for 10 minutes with shaking at RT to cross-link then 157 µl of 2M glycine added with shaking for 5 minutes at RT. Cells were then incubated on ice for 5 minutes and centrifuged at 1000 x g for 10 minutes at 4°C. The supernatant was removed, and cells resuspended in 2 ml PBS and stored at 4°C until use.

2.4.6.2 Extraction of genomic DNA (gDNA)

Cells collected were spun at 300 x g for 5 minutes and resuspended in 1 ml of lysis buffer 1 (10 mM Tris-HCl (pH8), 10 mM NaCl, 1 mM EDTA, and 0.2% Triton X-100) and incubated for an hour at 4°C. The centrifugation was repeated and resuspended in 1 ml lysis buffer 2 (10 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, and 0.3% SDS) and incubated for one hour at 37°C. Centrifugation was repeated and pellet washed in 250 µl of 1X CutSmart buffer and repeated for 3 times total. Blunt end repair was carried out using NEB quick blunting kit with 100 µg/ml BSA and incubating for an hour at RT. Washing with 1x CutSmart buffer was repeated a further three times. A-tailing was completed using NEBNext dA-tailing module in a final volume of 50 µl for 30 minutes at 37°C. CutSmart washing was repeated a further three times. It was then incubated for 5 minutes at RT in 1 ml of 1X T4 ligase buffer. The ends were labelled with T4 DNA ligase and 0.4 µM modified P5 adaptor (5'-ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC-3') in a final volume of 20 µl for 16-20 hours.

Following ligation, the excess P5 adaptor was removed by washing 10 times in 250 µl wash buffer (10mM Tris-HCl, 2M NaCl, 2mM EDTA, and 0.5% Triton-X-100) at RT. Washes were carried out by incubating with wash buffer for 2 minutes then spinning for 5 minutes at 300 x g. PBS was used as a final wash.

gDNA was extracted by incubating cells in DNA extraction buffer (10 mM Tris-HCl, 100 mM NaCl, 50 mM EDTA, and 1% SDS) and 1 mg/ml of proteinase K in a final volume of 100 µl

for 5 minutes at 37°C. Cell lysates were transferred into a 1.5 ml Eppendorf RNA/DNA LoBind tubes and incubated for an hour at 65°C with shaking at 800 rpm. The DNA was purified using Zymo Genomic DNA Clean and Concentrator-10 and eluted in 100 µl elution buffer. Samples were stored at 4°C until use.

2.4.6.3 Library preparation

DNA libraries from INDUCE-seq samples were prepared by Craig Lapsley. Qubit DNA HS kit was used to assess DNA yield of 1 µl sample before proceeding to library preparation. DNA was fragmented to 300-500 bp using Qiagen enzymatic kit and SPRI beads were used to select and remove fragments <150 bp. Fragmented and end-repaired DNA was added to ligation reaction of NEBNext® Ultra™ II Ligation Module using 7.5 µM modified half-functional P7 adaptor (5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') and omitting USER enzyme addition. P5 and P7 adaptors were ordered from Sigma Aldrich. The ligated sequencing libraries were then purified using SPRI beads and purification repeated twice more to remove fragments <200 bp. Libraries were then quantified by Qubit. Libraries were pooled and concentrated with a SpeedVac. Sequencing was performed on Illumina NextSeq 2000 using 2x100 bp high output flow cell.

2.4.7 DRIP-seq

2.4.7.1 DNA preparation

A total of $2x10^8$ exponentially growing cells were collected by centrifugation at 1000 x g for 10 minutes and washed once in PBS. The cell pellet was resuspended in 3 ml of lysis buffer (10 mM TrisHCl pH8, 100 mM NaCl, 25 mM EDTA) with 0.1 mg/ml proteinase K and 0.5% SDS. The solution was mixed carefully until cells were totally disrupted then incubated overnight in a 37°C water bath. To purify DNA from the solution, a phenol/chloroform extraction was carried out. A volume of 3 ml of phenol-chloroform-isoamyl alcohol (Sigma) was added and mixedfor 5 minutes, followed by centrifugation at 12,000 x g for 5 minutes to separate the two phases. The extraction was repeated once more and a further time using chloroform.

Ethanol precipitation was then completed by addition of 0.1V of 3M NaAc pH 5.2 and 2V of ice-cold ethanol then mixed gently until a precipitate formed. The solution was then centrifuged at 14,000 x g for 10 minutes at 4°C. The cell pellet was resuspended in 1.5 ml 75% ice cold-ethanol and recentrifuged at 7,000 x g for 10 minutes. The pellet was resuspended in 300 µl ultrapure TE and incubated at 4°C for 8-12 hours to solubilise nucleic acids.

2.4.7.2 DNA RNA hybrid immunoprecipitation (DRIP)

After DNA extraction, the sample was divided into two samples and sample 1 was stored at 4°C as an untreated control. Sample 2 was digested with 20U of RNAse H and incubated overnight in a water bath at 37°C. A phenol/chloroform extraction followed by ethanol precipitation was performed, as described in 2.4.8.1, and the DNA resuspended in 100 µl ultrapure TE buffer.

Both samples were digested with BamHI (GGATCC), NcoI (CCATGG), ApaLI (GCTAGC) and PvuII (CAGCTG). For sample 1, 120U of RNAseOUT was added and 10U RNAse H to sample 2. Samples were incubated overnight in a water bath at 37°C. Samples were extracted once with phenol-chloroform-isoamyl alcohol then precipitated with 0.1V 3M NaAc pH 5.2 and 2V of 100% ethanol. Samples were resuspended in 100 µl ultrapure TE and DNA concentration measured.

AB conjugated beads were prepared using a magnetic rack. For each sample, 400 µl of Dynabeads M-280 Sheep Anti-Mouse IgG was washed twice with wash buffer (1X PBS, 0.2% BSA, 2mM EDTA). The beads were resuspended in 500 µl of wash buffer and 25 µl of S9.6 AB (Millipore MABE1095) added and incubated at 4°C overnight with rotational shaking. Using the magnetic rack, the beads were washed twice with wash buffer. The beads were resuspended in 100 µl of 1X IP buffer (10 mM sodium phosphate pH 6.8, 140 mM NaCl, 0.5% Triton X-100) and stored on ice until use.

A total of 40 µg of each DNA sample was added to 900 µl of 1x IP buffer then added to AB conjugated beads and incubated overnight at 4°C with rotational shaking. The beads were collected using magnetic rack and washed 4 times using 1 ml of IP buffer for 10 minutes in the cold room with overhead shaking. The AB/R-loop complexes were collected with the magnetic rack and the beads resuspended in 200 µl of 1X PBS. PBS was added to increase the volume to 200 µl.

The DNeasy Blood and Tissue DNA extraction kit (Qiagen) was used to extract DNA from isolated complexes. An incubation with proteinase K was carried out in a thermomixer to prevent bead sedimentation. DNA was eluted twice from the column, two elutions of 75 µl, for a total volume of 150 µl.

2.4.7.3 Library preparation

Libraries were prepared as regular genomic DNA libraries using Qiagen FX 96 Library Kit for next-generation Illumina sequencing. DNA was purified using Agencourt AMPure XP beads (Beckman Coulter). Samples were sequenced on an Illumina NextSeq 500 platform.

3. Results

3.1 Generation of *T. brucei* cell lines by CRISPR/Cas9

3.1.1 CRISPR/Cas9 technology

In order to test the roles of the 70 bp repeats and *VSG* sequence in generating the DNA breaks and accumulation of R-loops that are believed to be integral to VSG switching (Girasol, 2023b), it was sought to modify these features in the active BES. The specific changes we made are described below, but the approach to making these changes relied on genome editing by clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR-associated gene 9 (Cas9), adapting the methodology from a study by Beneke et al. (2017). To allow for CRISPR/Cas9 editing in BSF *T. brucei* Lister 427, a puromycin resistant cell line that constitutively expresses Cas9 and T7 RNA polymerase generated by Girasol et al. (2023b) was used. CRISPR/Cas9 editing allowed for attempts to generate cell lines with specific sequence knockouts and insertion, and this approach was chosen due to the likelihood of modification precision.

Single-guide RNAs (sgRNAs) can be designed with sequence-specificity, which directs the Cas9 nuclease to cleave dsDNA at the target locus (Beneke *et al*., 2017), generating a double-stranded DNA break (DSB). The repair of this DSB provides the ability for precise modifications at the site of the break if repair templates are provided. Expression of T7 RNA polymerase is required for *in vivo* transcription of PCR-amplified DNA templates of sgRNA. Oligonucleotides were generated by inputting the genes or sequences of interest into the Eukaryotic Pathogen CRISPR guide RNA/DNA Design Tool (http://grna.ctegd.uga.edu/) to identify relevant sgRNA templates, screening for locus-specificity. The T7 promoter sequence was added to the oligonucleotide to allow for *in vivo* transcription of the sgRNA by T7 RNA polymerase. PCR amplification was used to generate multiple copies of sgRNA by using sense and antisense oligonucleotides.

Donor DNA is introduced as a template for repair of a Cas9 DSB by the *T. brucei* homologous recombination machinery. Primers of a ~30 bp sequence of homology arms on each side of the target locus were used as a donor template for repair of the break to enable precise modifications.

3.1.2 Plasmid structure

The pPOTv7 plasmid was used as a template for PCR amplification of donor DNA (Dean *et al.,* 2015). The pPOTv7 plasmid has coding sequences for a fluorescent tag mNeonGreen (mNG) (tag not used in this study) and two copies of a drug resistance marker to allow use of the plasmid for N-terminal and C-terminal protein tagging, if desired. The pPOTv7 plasmids with the drug resistance marker neomycin (NEO), blasticidin (BSD) or hygromycin (HYG) were used to allow for selection of successful transformants in selective media (Figure 3-1). Designed primers were able to be used across all pPOTv7 plasmids, independent of drug resistance marker.
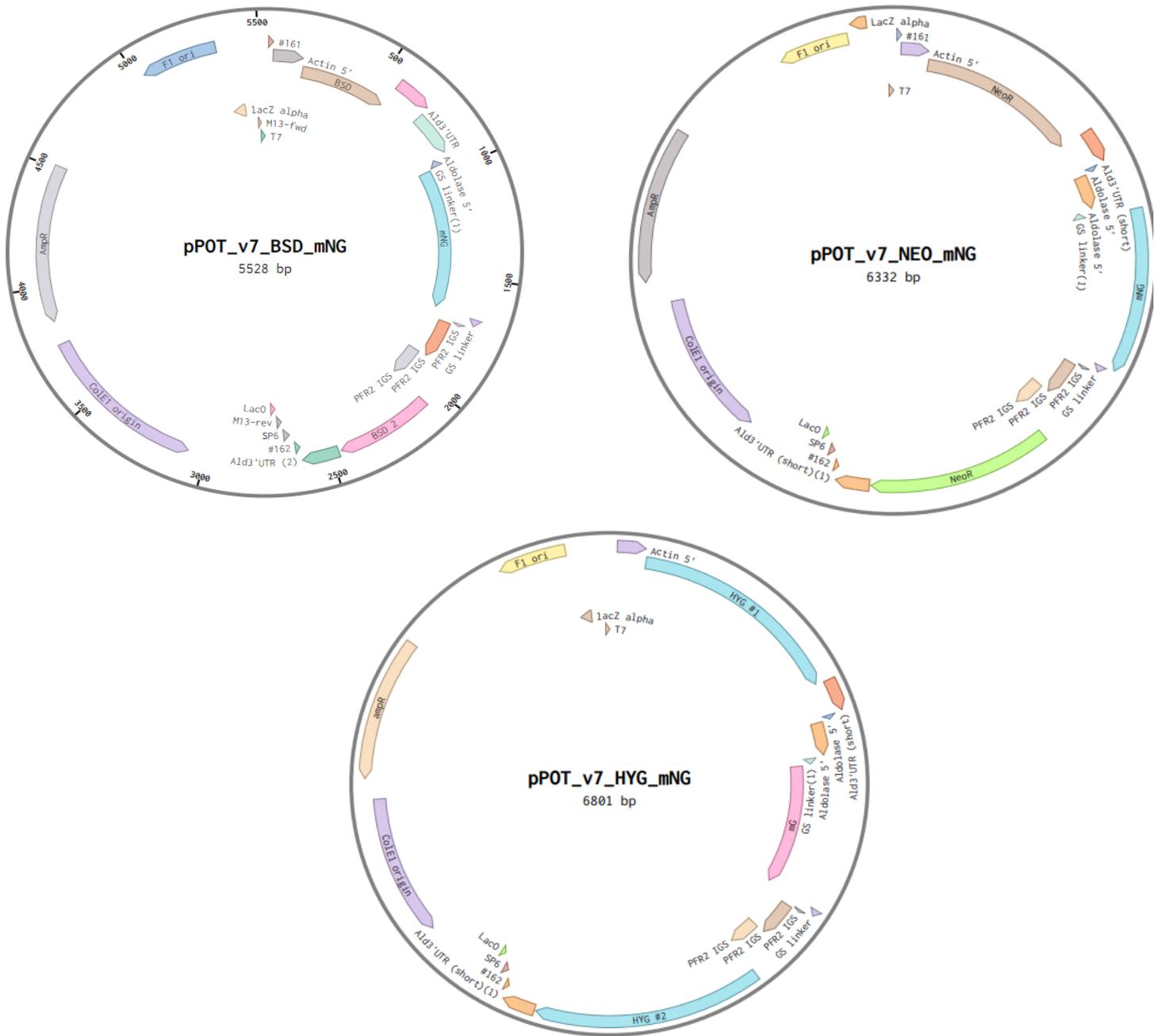
Figure 3-1. Structure of the pPOTv7 plasmid with different drug resistance markers. Plasmid map of pPOTv7 containing either a blasticidin, neomycin, or hygromycin resistance cassette, each plasmid contains two antibiotic genes to allow N- or C-terminal tagging of genes if desired. mNG= mNeonGreen tag**.**

3.2 Generation of *T. brucei* cell lines with the 70 bp repeats removed

3.2.1 Approach for generation of cell lines with the 70 bp repeats removed

To test the role of the 70 bp repeats in BES-focused break formation and its link to R-loop localisation, we sought to ask what happens when the 70 bp repeats are removed from the active BES. The repeats have an unusual structure in the BES (BES1) of the VSG that is transcribed in the strain used in the McCulloch laboratory, in that, they are found upstream of *VSG221* but in two stretches that are separated by a *VSG* pseudogene (Figure 3-2A). As we do not know if either or both of the 70 bp repeats stretches might be important, or indeed if the pseudogene might play a role, our modifications of BES1 involved three separate deletions. The first deletion, deletion 1, should remove the larger, most *VSG221*-proximal section of the repeats (~3.2 kb), deletion 2 was intended to remove the smaller section of repeats upstream of the *VSG* pseudogene (~419 bases), and deletion 3 included both sections of the repeats and a *VSG* pseudogene (~6.5 kb) (Figure 3-2A).

To remove the individual sections of 70 bp repeats or both plus the *VSG* pseudogene by CRISPR/Cas9, primers were designed to generate sgRNA to target5' and 3' regions flanking the repeats, indicated in Figure 3-2A. It was intended that transfection of sgRNA would result in a DSB at either side of the targeted sequences around the 70 bp repeats, which would then be replaced by DSB repair using the introduced donor DNA as a repair template. Donor primers using a ~30 bp sequence were selected at 5' and 3' ends of each deletion and used to PCR up the antibiotic resistance cassette from the pPOTv7 BSD plasmid. The designed sgRNA and donor DNA templates were amplified by PCR (Figure 3-2). All following gel images include a 1 kb ladder with band sizes ranging from 500 bp to 10 kb. Donor DNA was successfully amplified for deletion 2 and 3, with an expected band size of 998 bp visualised on an agarose gel (Figure 3-2B). No band was present for donor 1 and the PCR was repeated (see below). PCR products, to allow transcription of each sgRNA around the three deletions, were successfully amplified and confirmed by fragments at 120 bp visualised on an agarose gel (Figure 3-2C). All PCR amplifications of donor DNA and sgRNA were repeated to provide more DNA to be pooled for transfection.
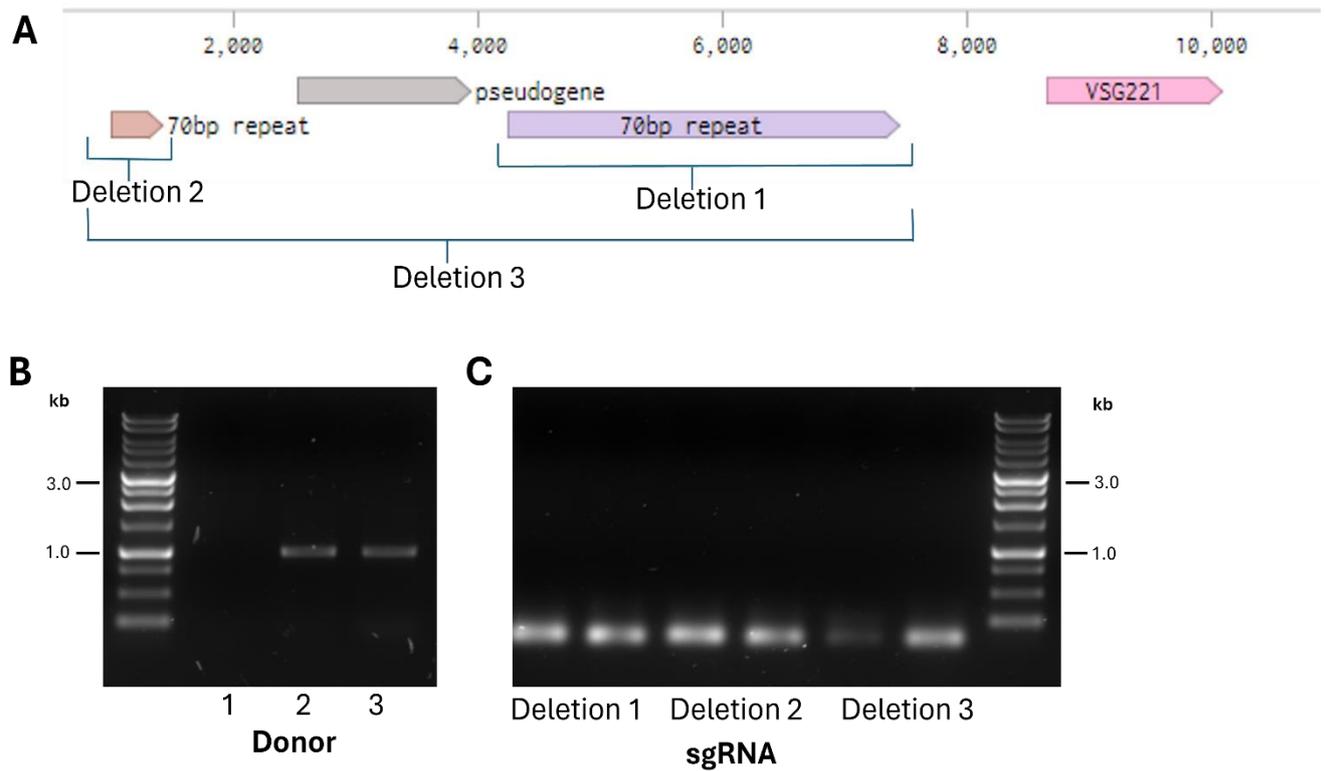
Figure 3-2. Generation of *T. brucei* cell lines with the 70 bp repeats removed by CRISPR/Cas9

A. Approach used for knockout of the 70 bp repeats by CRISPR/Cas9 gene editing. Indicates section of the genome targeted for knock out. Forward and reverse primers were designed at each of these sites.

B. PCR confirmation of amplification of donor DNA amplified from pPOTv7 BSD. Depicts successful amplification of donor DNA for deletion 2 and deletion 3.

C. PCR confirmation of amplification of sgRNA DNA templates for 70 bp knockouts. For each sgRNA DNA template, the first lane is the upstream template followed by the downstream template.

Donor DNA to be used as a repair template for deletion 1 failed to be amplified by PCR using an annealing temperature of 65°C. A gradient PCR using annealing temperatures ranging from 55°C to 70°C was therefore conducted to determine the most efficient annealing temperature (Figure 3-3A). The band intensity increased at higher temperatures with an optimal temperature of 70°C. PCRs were then repeated using this optimised temperature, which successfully amplified deletion 1 donor DNA to be used in transfection (Figure 3-3B).
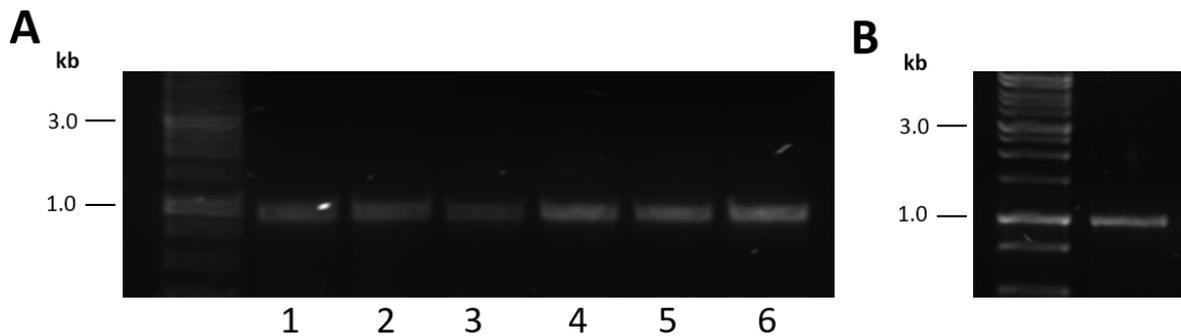
Figure 3-3. Amplification of donor DNA for *T. brucei* 70 bp deletion 1 by PCR.
A. Gradient PCR of donor DNA for 70 bp deletion 1 using an annealing temperature range of 55-70°C. Band intensity increased at higher temperatures with lane 1 using the lowest temperature (55°C) to lane 7 with the highest temperature (70°C).
B. Amplification by PCR of donor DNA for deletion 1 using an annealing temperature of 70°C.

3.2.2 Transfection and confirmation of clones

After successful amplification of PCR products to allow transcription of each sgRNA and donor DNA templates, a total of 3-10 individual PCR reactions were pooled together for each deletion. PCR products were purified by ethanol precipitation and 10 µg of relevant DNA, comprised of the donor DNA and sgRNA DNA templates, used for transfection. BSF *Tb*Cas9 cells were grown to a density of 1×10^6 cells/ml and a total of $1\text{-}3 \times 10^7$ cells used per transfection with DNA transfected through electroporation. Antibiotic resistant clones were selected using 10 µg/ml of BSD.

3.2.2.1 Attempted removal of 70 bp repeats by deletion 1

After multiple failed attempts at transfection, antibiotic resistant cells from the 70 bp repeat deletion 1 transformation were recovered. The surviving cells were not obtained from a well on a diluted 24-well plate but, instead, in an undiluted flask from transfection. The flask was slow growing compared to other attempted transfections, taking 14 days for cells to be visible

under light microscopy, and then taking another 6 days for the culture to grow to a normal density of 1×10^6 cells/ml. Furthermore, it was attempted to replate this flask out into a 96-well plate to obtain clones, which was unsuccessful, and no surviving clones were obtained.

PCR screens were used to test if the deletion 1 region of the 70 bp repeats in BES1 had been successfully removed and for confirmation of insertion of the BSD resistance cassette (Figure 3-4A). The PCR screening for deletion 1 had to be conducted multiple times due to unsuccessful or ambiguous screening. After unsuccessful screening using primers HS18+HS19 and HS19+G27 (Figure 3-4), the PCR conditions were changed in an attempt to optimise the reaction. PCR screening was completed with a longer extension step of 72°C for 4 minutes due to the large product sizes, perhaps explaining why amplification may not have been successful in previous screens.

70 bp deletion 1 gDNA was screened with primers, HS18 + HS19, to screen across the 70 bp deletion 1 region (Figure 3-4B). A band at the expected product size (1149 bp) was seen for deletion 1 cells, suggesting the section of repeats was successfully removed from the ES. A fragment of 3837 bp would be amplified from the WT using HS18 + HS19 however, due to the repetitive nature of the 70 bp repeats, this is unable to be amplified by PCR. A cell line with known BSD resistance was used as a positive control and a cell line with no BSD resistance was used as a negative control for the PCR reactions. No bands were present for the WT, positive control, or negative control. PCR screening was next conducted to test for the presence of the BSD resistance cassette, with the primers G27 and G28 expected to generate a band size at 400 bp if the BSD resistance cassette was present (Figure 3-4C). As expected, the WT cells and negative BSD control cells did not produce a band, as these cell lines do not have BSD resistance. The positive control, gDNA from a *T. brucei* cell line known to have BSD resistance, produced the expected band at 393 bp. However, after multiple attempts, gDNA from the TbCas9 70 bp deletion 1 cells failed to produce a band to confirm BSD integration, which was unexpected as the cell line was continuing to grow in the presence of BSD (Figure 3-4C). A distinct PCR screen was therefore attempted, targeting from the outside of the 70 bp deletion 1 region to the end of the BSD ORF (Figure 3-4D, primers HS18 and G28). This PCR produced an expected band of 624 bp for the deletion 1 cells, and not for the WT or controls. Additionally, screening was carried out using primers from the BSD cassette to the outside of the 70 bp deletion 1 region (Figure 3-4E, primers HS19 + G27). A band at 918 bp was expected for successful insertion of BSD resistance cassette. Unexpectedly, this screen produced no bands for any cell lines.
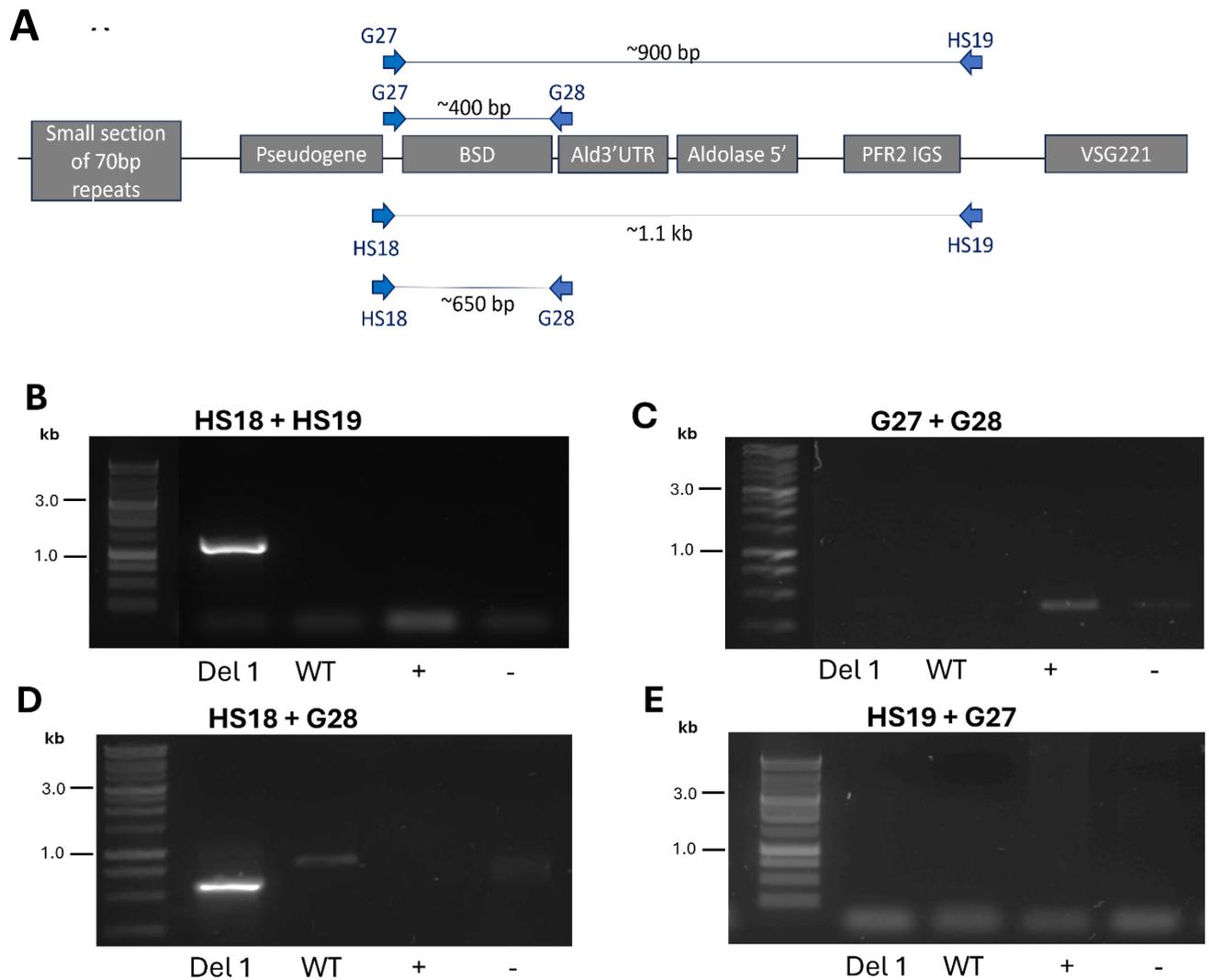
Figure 3-4. Confirmation screening of 70 bp repeats deletion 1 by PCR.
A. PCR confirmation strategy indicating the location of the forward and reverse primers used and the size of the expected fragment.
B-D. PCR confirmation of 70 bp repeats deletion 1 amplified using primers shown in A and visualised on agarose gel. WT band depicts the wildtype TbCas9/T7 cell line. + indicates a BSD resistant positive control. – depicts a negative control cell line not BSD resistant.

Given the difficulty in using PCR to validate the expected changes in BES1, the ~1.1 kb PCR product 70 bp deletion 1 cells with primers HS18 and HS19 was cleaned up using a QIAquick PCR Purification Kit and sent for Sanger sequencing. The sequence confirmed integration of the correct construct with the expected organisation of BSD cassette sequences relative to surrounding BES1 sequences, with no evidence of the 70 bp repeats targeted by this deletion strategy (Appendix 4).

3.2.2.2 Successful removal of 70 bp repeats region 2

In contrast to the difficulties with the deletion 1 strategy, on the first attempt at transfection to remove the deletion 2 section of the 70 bp repeats, seven wells on a 24-well plate were found to have antibiotic resistant cells under light microscopy. From each of the 7 wells, 1 ml was transferred into 5 ml of HMI-9 media in a vented flask with additional BSD.

Screening was next performed by PCR to test if the 70 bp repeat deletion 2 had occurred as expected in any of the seven clones, including successful insertion of the BSD resistance cassette (Figure 3-5A). Each clone was screened across the 70 bp repeat deletion 2 region by PCR and visualised on an agarose gel with a fragment of 1154 bp expected (Figure 3-5B; primers HS20 + HS21). A PCR fragment at the expected size was visible for clones 1, 2, 3, and clone 7, consistent with the expected replacement of the upstream 70 bp repeats by the BSD cassette. A faint band was visible for clone 6. For further confirmation, a screen was completed from the BSD cassette, G27, to the outside of the 70 bp deletion 2 region, HS21 (Figure 3-5C; primers G27 + HS21). The results of this PCR screen were surprising with only 2 of the 4 clones that were detected by the first PCR having visible PCR products. The expected PCR product size was 892 bp and double bands were visible for clones 3 and 7. Given this result, all clones were screened for insertion of the BSD resistance cassette with primers G27 and G28, alongside the WT, positive BSD control, and negative BSD control (Figure 3-5D); here, 6 of the clones (clones 1-5, and clone 7) gave the expected product of 393 bp. A faint band was visible for clone 6. Since all seven clones were surviving in media containing BSD, this may indicate that in at least some cells the resistance marker had integrated elsewhere. A final screen using targeting from the outside of the 70 bp deletion 2 region, HS20, to the end of the BSD ORF, G28, was completed with an expected product size of 655 bp if successful deletion of the repeats and insertion of BSD cassette had occurred (Figure 3-5E). Here, the expected products were visible for clones 1-3, 7, consistent with PCR using HS20 and HS21, and a band was also seen for clone 6 and faint band for clone 4. Clones 2 and 7 were selected for further analysis.
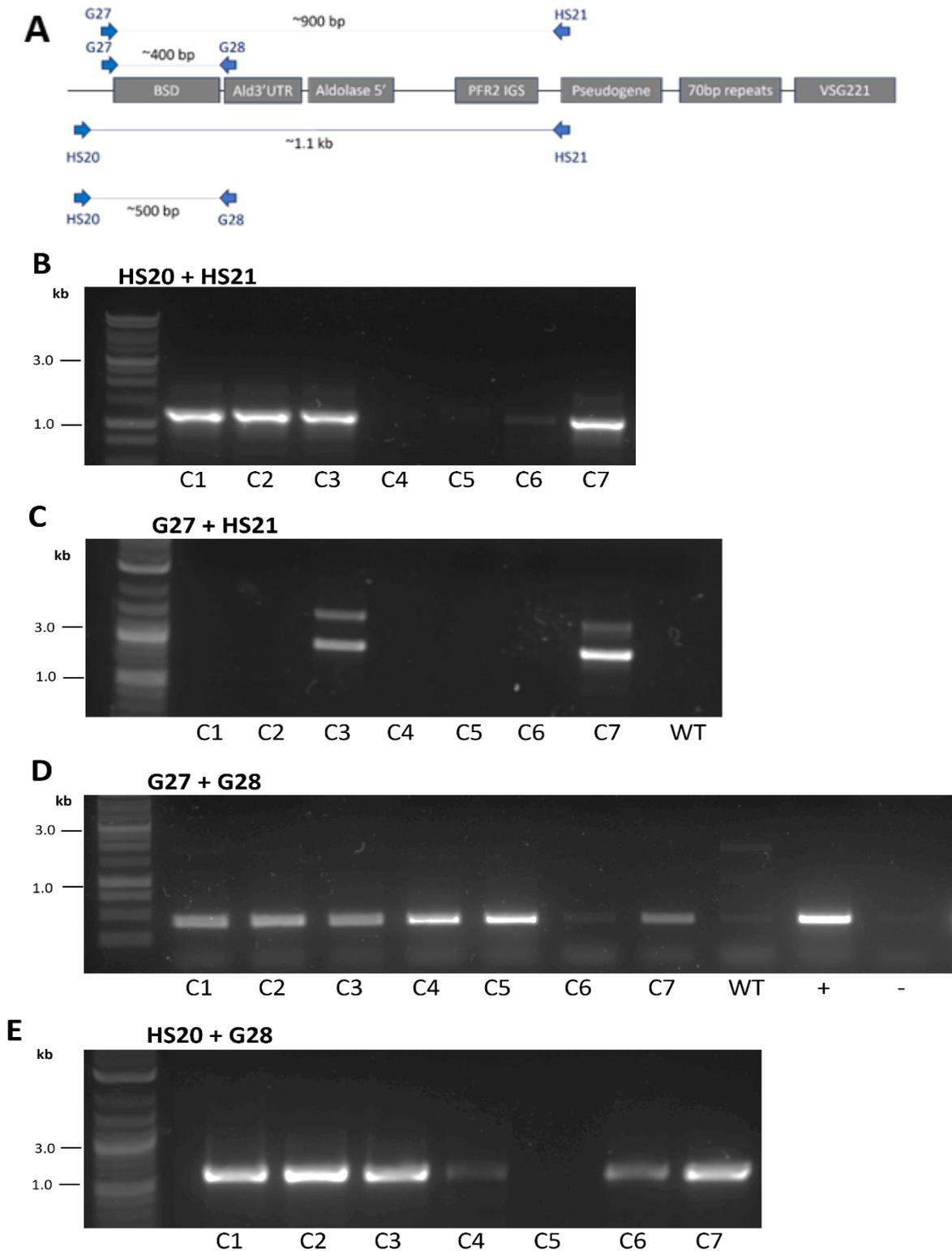
Figure 3-5. Confirmation screening of 70 bp repeat deletion 2 clones by PCR.
A. PCR confirmation strategy indicating the location of the forward and reverse primers used
and the size of the expected fragment.
B-E. PCR confirmation of deletion 2 amplified using priners shown in A. C = clone
numbered 1-7. WT band depicts the wildtype TbCas9/T7 cell line. + indicates a BSD
resistant positive control. – depicts a negative control cell line not BSD resistant.

3.2.2.3 Unsuccessful removal of all 70 bp repeats, deletion 3, from *T. brucei* BES

Three transfections failed to recover any surviving clones for deletion 3 on either diluted 24-well plates or in undiluted flasks. Due to the extended length of time antibiotic-resistant cells took to be visible under light microscopy for deletion 1, the final attempt at transfection for deletion 3 was allowed to recover for up to 17 days, which failed to recover any surviving cells. As no antibiotic cells were recovered, no further tests were possible.

3.3 Generation of *T. brucei* cell lines with an additional *VSG* copy

3.3.1 Approach to generating cell lines with an additional *VSG* copy

To test the role of the *VSG* in BES-focused break formation and its link to R-loop localisation, we sought to ask what happens when an additional *VSG* sequence is relocated in the active BES. To insert an additional copy of a *VSG* gene into BES1 of *T. brucei,* the *VSG* sequence was amplified by PCR then ligated into pGEM®-T Easy to allow subsequent cloning into the pPOTv7. We decided to explore this using the sequences of both the endogenous BES 1 *VSG* (*VSG221)* and a distinct *VSG* (*VSG121,* found in BES3*)* to ask if any effect on DSB or R-loop formation is common to all VSGs or might be due to specific features of *VSG221*. Three sites were chosen in the active BES1 for insertion of the *VSG* sequence: at the promoter of BES1, replacing the 'third' copy of *ESAG8*, and replacing the small section of the 70 bp repeats (deletion 2) (Figure 3-6). Our rationale for this selection was to ask if the *VSG* sequences needed specific context within the BES to act in DSB and R-loop formation.
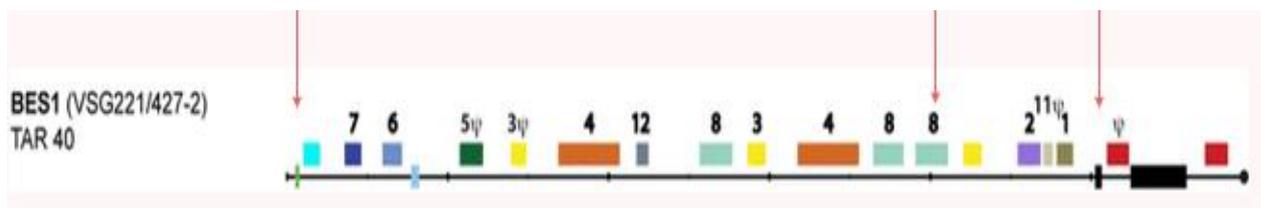
Figure 3-6. Locations in *T. brucei* BES1 selected for insertion of a *VSG* copy. Arrows indicate where insertion of a *VSG* copy was attempted. Image adapted from Hertz-Fowler *et al*., 2008.

3.3.2 Cloning *VSG* sequences into a subcloning vector

The sequences of *VSG121* and *VSG221* were amplified by PCR. A gradient PCR was performed to determine the optimum annealing temperature for the amplification and six tubes were set up for each *VSG* with an annealing temperature range of 49°C to 64°C and then run on a 1%

agarose gel (Figure 3-7). PCR products for *VSG221*, with a fragment size of 1.45 kb, were present at all annealing temperatures but gave stronger bands at higher temperatures. PCR products for *VSG121*, although present, were fainter in comparison to those of *VSG221*. PCR products for *VSG121*, with a fragment size of 1.55 kb, were extremely faint at the 3 lower temperatures but strength of bands improved as the annealing temperature used increased. The gradient PCR identified that a higher temperature was optimal for annealing. The PCR was run using the optimum annealing temperature of 64°C for amplification of *VSG121* and *VSG221* (Figure 3-7). Fragments were produced from the successful amplification of *VSG121* and *VSG221*.
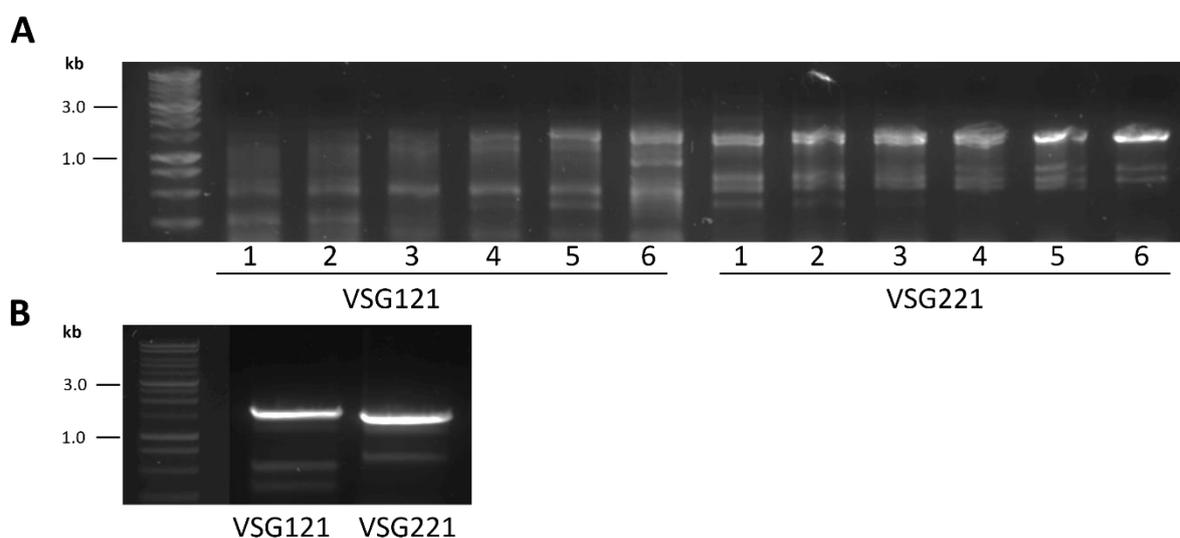
Figure 3-7. Amplification of *VSG* sequences by PCR.
A. Gradient PCR for amplification of *VSG121* and *VSG221* with an annealing temperature range of 49°C to 64°C. Visibility of fragments increased at higher temperatures with lane 1 using the lowest temperature (55°C) to lane 7 with the highest temperature (70°C).
B. Amplification of *VSG* sequences by PCR using the optimal annealing temperature of 64°C.

The successfully amplified *VSG* fragments required the addition of A-tails for the PCR fragment to ligate into the subcloning vector pGEM®-T Easy (Appendix 11). PCR products with added A-tails were run on an agarose gel and the desired fragments, at 1431 bp for *VSG221* and 1590 bp for *VSG121*, were excised from the gel. DNA was extracted from the gel fragments, ligated to the subcloning vector pGEM®-T Easy, and transformed into competent *E. coli* cells. Multiple colonies were screened by PCR for correct insertion of each *VSG* fragment using primers from either side of the insertion site on pGEM®-T Easy (Figure 3-8). PCR products were visualised on an agarose gel and would provide a band of 233 bp if the *VSG* was not successfully inserted and of much greater size, dependent on the *VSG* size, if

correctly inserted: *VSG121,* 1833 bp; and *VSG221*, 1674 bp. Identifying colonies, which had correctly inserted *VSG121* into pGEM®-T Easy required repeating the ligation and transformation multiple times, as the clones recovered and subjected to PCR screening revealed a product of ~200 bp, indicating a failed insertion. This can be seen in pGEM®-T Easy + VSG121 colony 4 (Figure 3-8A). Fragments of the correct size were eventually detected in 7 colonies (Figure 3-8A); colonies 1 and 2 were selected. On visualisation of PCR screening on agarose gel of pGEM®-T Easy + *VSG221* colonies, colonies 2-7 had a fragment of the expected size indicating a correct insertion (Figure 3-8B); colonies 4 and 5 were selected.
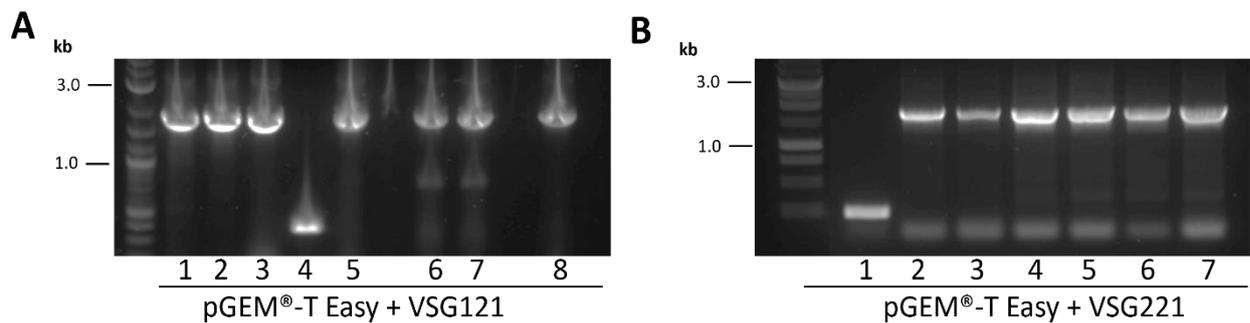
Figure 3-8. Colony screening by PCR for *VSG* insertion into pGEM®-T Easy.
A. Colony screening for correct insertion of *VSG121*. Eight colonies were screened by PCR and seven colonies had a fragment at the expected size.
B. Colony screening for correct insertion of *VSG221*. Seven colonies were screened by PCR and six colonies had a fragment at the expected size.

Sanger sequencing was performed by Eurofins Genomics on pGEM®-T Easy + *VSG121* and pGEM®-T Easy + *VSG221* to identify that the correct gene and relevant restriction enzyme sites were present (Appendix 9 and 10). The full inserts were sequenced to ensure no mutation was introduced by PCR. No alterations from the intended gene sequences were found; by comparison of the Sanger traces to the original copies of *VSG* genes using Benchling software.

3.3.3 Cloning of *VSG* into pPOTv7

3.3.3.1 pPOTv7 BSD

After confirmation of correct *VSG* ORFs and desired restriction sites by Sanger sequencing, pGEM®-T Easy + *VSG* and pPOTv7 BSD were digested using restriction enzymes *Kpn*I and *Sac*I (Figure 3-9). From the pGEM®-T Easy + *VSG221* digestion, the 1431 bp *VSG* fragment was extracted, leaving the 3 kb vector fragment (Figure 3-9A). From pPOTv7 BSD, the vector fragment at 4737 bp was extracted, leaving the band at 790 bp, which is an unwanted fragment

of the mNeonGreen tag being removed from the original pPOTv7 BSD plasmid (Figure 3-9A). From the pGEM®-T Easy + *VSG121* digestion, the 1590 bp *VSG* fragment was extracted, leaving the 3 kb vector band (Figure 3-9B).
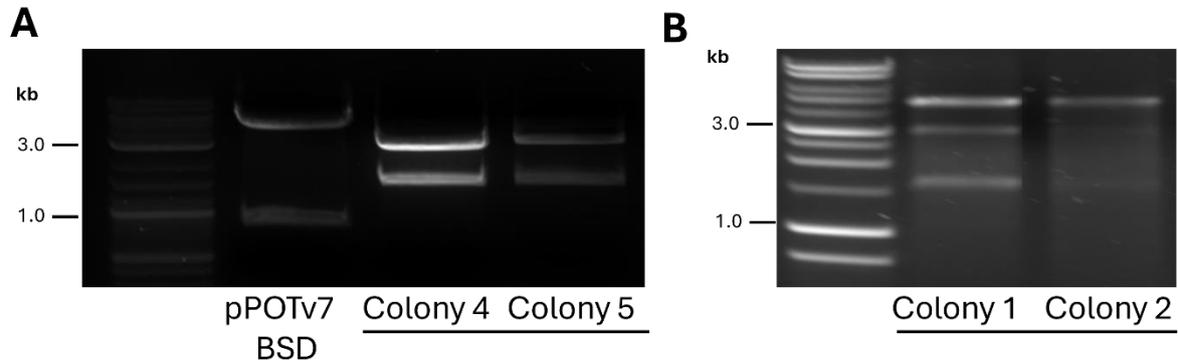
Figure 3-9. Restriction digest of pPOTv7 BSD and pGEM®-T Easy + *VSG* using *Kpn*I and *Sac*I. Visualised on an agarose gel.
A. Restriction digests of pPOTv7 BSD and pGEM®-T Easy + *VSG221* colonies 4 and 5; showing backbone of vectors and relevant DNA fragments after digestion.
B. Restriction digests of pGEM®-T Easy + *VSG121* colonies 1 and 2.

Desired bands were excised from the agarose gel and DNA was purified from the gel. The purified *VSG* fragments were ligated to the pPOTv7 BSD expression vector and transformed into competent *E. coli* cells. Eight colonies for each *VSG* were screened by PCR and visualised on an agarose gel to determine if the *VSG* had correctly inserted into pPOTv7 BSD (Figure 3-10). All eight colonies for insertion of *VSG221* had a fragment of the expected size of 1.45 kb, indicating successful insertion. However, only one colony, colony 7, had a fragment consistent with insertion of *VSG121* (1.55 kb*)*. Selected colonies, indicated in Figure 3-10 with an arrow, were grown in LB broth and plasmid DNA purified.
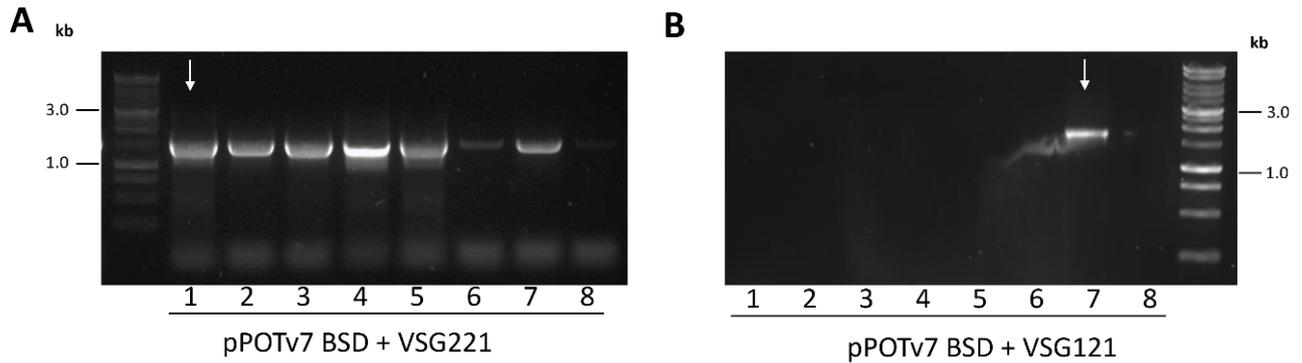
Figure 3-10. Colony screen by PCR of *VSG* insertion into pPOTv7 BSD. Visualised on a 1% agarose gel. Arrow indicates colonies selected for further testing.
A. Colony screen of pPOTv7 BSD + *VSG221*. Seven colonies produced fragments at 1.45 kb.
B. Colony screen of pPOTv7 + *VSG121*. One colony, colony 7, produced a fragment at 1.55 kb.

A test restriction digest was carried out to determine if the *VSGs* had correctly inserted into pPOTv7 BSD of selected colonies. The restriction digest was used to confirm the presence of each *VSG* by excising it from the pPOTv7 BSD backbone using enzymes with restriction sites that flank the sides of the *VSG* insert. The purified DNA was incubated with restriction enzymes, *Kpn*I and *Sac*I, and run on an agarose gel to determine the size of the resulting DNA fragments. The pPOTv7 BSD plasmid backbone produced a size of ~5 kb, with a fragment of 1.55 kb if *VSG121* had correctly inserted and 1.45 kb if *VSG221* had correctly inserted. If the *VSG* insertion was unsuccessful, a fragment of ~800 bp would be expected, which is the mNeonGreen tag from the pPOTv7 construct. The results from two successful tests digest visualised on an agarose gel are shown in Figure 3-11, indicating successful insertion of each *VSG* into pPOTv7 BSD. After confirmation by restriction analysis, the resulting plasmids were sent for Sanger sequencing by Eurofins Genomics. Sequences were confirmed to be correct for pPOTv7 BSD + *VSG221* (Appendix 6) and pPOTv7 BSD + *VSG121* (Appendix 7).
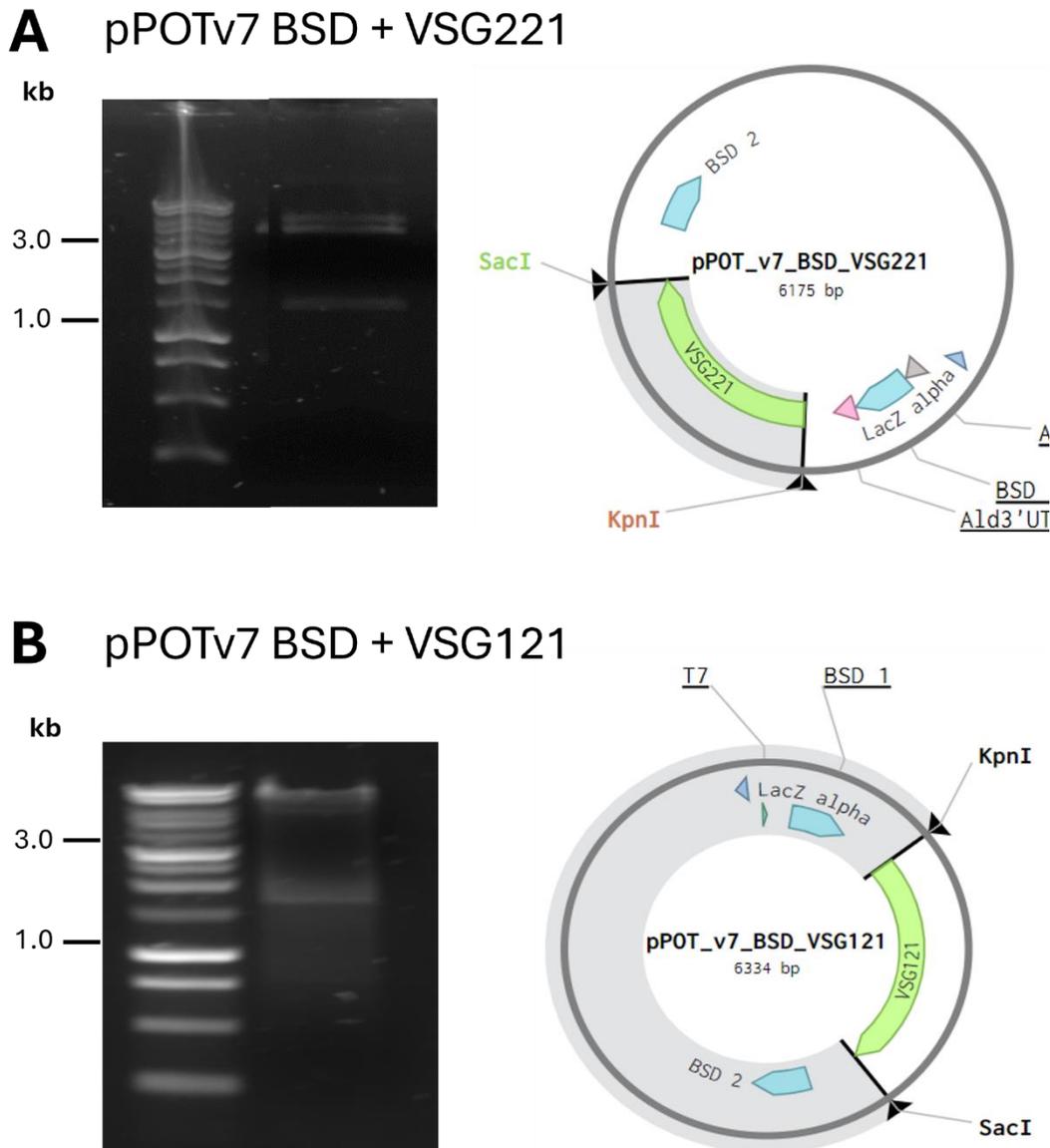
Figure 3-11. Test restriction digest of pPOTv7 BSD + *VSG* using restriction enzymes *Kpn*I and *Sac*I. Agarose gel showing backbone of vectors and relevant DNA fragments after digestion. Plasmid maps of generated vectors indicating the site of the restriction enzymes.

PCR was performed on the pPOTv7 BSD + *VSG* plasmid constructs to add homology flanks that would allow insertion of the *VSGs* into the promoter (HS36; HS37), to replace ESAG8 (HS32; HS33), and to replace the 70 bp repeat deletion 2 region (HS38; HS7). The amplification of these donor DNA templates was visualised on an agarose gel to confirm correct amplification with an expected fragment size of 2527 bp for correct amplification with *VSG221* and 2696 bp of *VSG121* (Figure 3-12). PCR amplification of each of the donor templates produced double bands with the larger band of the expected size, but the lower band was not expected (Figure 3-12A; Figure 3-12B). PCR products, to allow transcription of each sgRNA targeting the three locations, were successfully amplified and confirmed by fragments

at 90 bp visualised on an agarose gel (Figure 3-12C). Ethanol precipitation of the donor DNA which produced double bands and sgRNA resulted in DNA concentrations of 8.5 ng/µl, 8.1 ng/µl and 4.5 ng/µl for *VSG221* and 6.4 ng/µl, 9.6 ng/µl, and 4.5 ng/µl for *VSG121*. Concentrations were too low for DNA to be used for transfection, as 10 µg of DNA is normally required. To achieve higher concentrations, the number of PCR donor amplicons were increased to using 4 donor amplicons with 8 sgRNA per transfection, and ethanol precipitation completed with the addition of GlycoBlue^{TM} Coprecipitant to aid identification of DNA pellets during precipitation to prevent additional DNA loss. These steps were repeated multiple times, failing to significantly increase the DNA concentration.
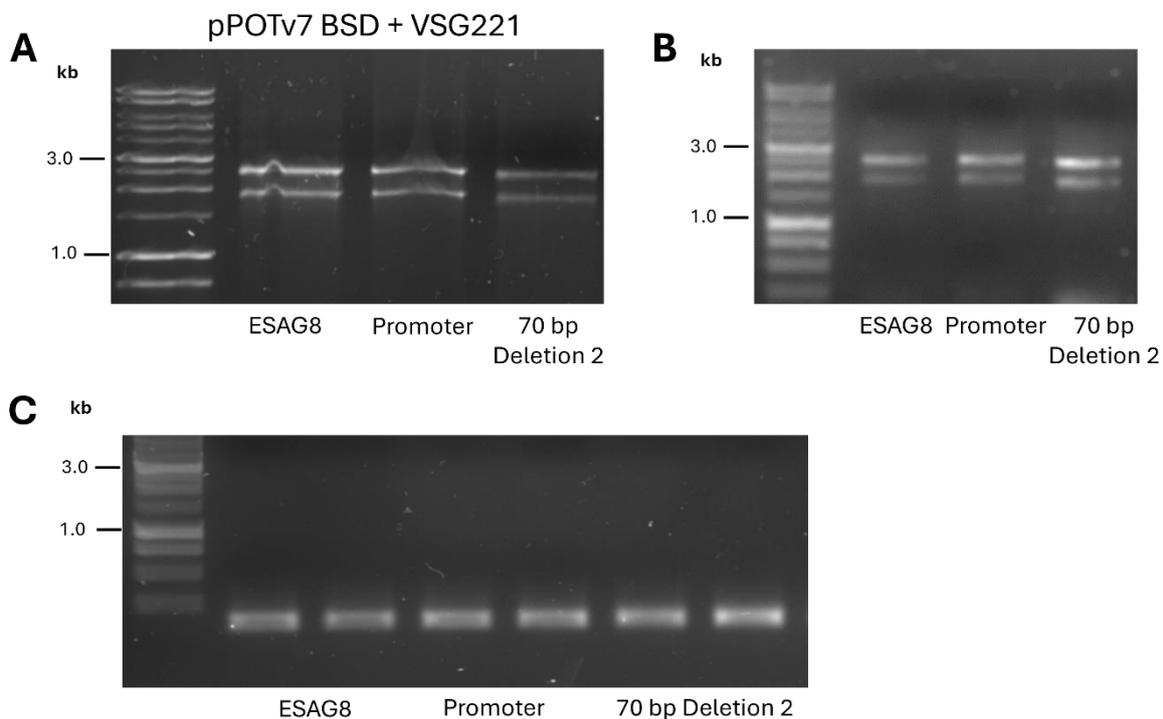
Figure 3-12. Amplification of donor DNA and sgRNA by PCR using pPOTv7 BSD + *VSG* plasmids.
A. Amplification of pPOTv7 BSD + VSG221 donor DNA, with two fragments.
B. Amplification of pPOTv7 BSD + VSG121 donor DNA, with two fragments.
C. Amplification of sgRNA for insertion into the three named locations of BES1, with fragment at 100 bp.

Using the above PCR amplified donor DNA, which produced double bands (Figure 3-12A; Figure 3-12B), gel extraction was completed to excise the top bands of the expected product size (~2.55 kb for *VSG121* and ~2.45 kb for *VSG221*) using a UV transilluminator to visualise the DNA in the agarose gel. This approach was used to try to increase DNA concentrations by only using the desired PCR product and not the additional incorrect product. DNA was

extracted from the gel fragment, pooled with sgRNA, and ethanol precipitation completed. The DNA concentrations ranged from 2.8 ng/µl to 33.0 ng/µl and were too low to be suitable for use in transfection.

Finally, due to the concern that rearrangement of the constructs accounted for the double PCR products, the pPOTv7 BSD + *VSG* colonies were streaked on an ampicillin containing agar plate to achieve a single clean colony. The streaking technique was used to isolate a pure culture from the putative mixed population with the aim that the pure culture would not produce double bands. Eight single colonies were screened by PCR for *VSG* insertion for pPOTv7 BSD + *VSG121* and pPOTv7 BSD + *VSG221*. The colony screens were visualised on agarose gel, and again, both *VSGs* with all donor flanking primers, produced either double bands or the incorrect lower band and were deemed unsuitable for use to amplify donor DNA.

3.3.3.2 pPOTv7 NEO

Due to the consistent generation of double PCR bands and low concentration of DNA when donor DNA was amplified from pPOTv7 BSD, DNA was unable to be used for transfection. To try and circumvent this, the *VSGs* were instead inserted into pPOTv7 NEO, in case the BSD gene was the source of the problems. The pPOTv7 NEO vector, pGEM®-T Easy + *VSG121,* and pGEM®-T Easy + *VSG221* were digested using restriction enzymes *Kpn*I and *Sac*I (Figure 3-13). The vector band at 5541 bp was excised for pPOTv7 NEO. For the *VSG* Figure 3-excised, leaving the pGEM®-T Easy vector fragment ~3 kb and the larger fragment of the linearised pGEM®-T Easy vector without the VSG fully digested out.. DNA was then purified from the excised gel fragments.
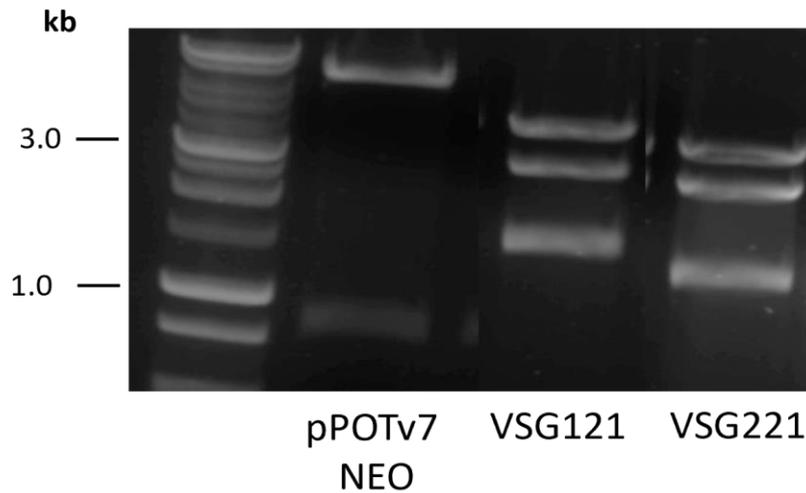
Figure 3-13. Restriction digest of pPOTv7 NEO and pGEM®-T Easy + *VSG* **using** *Kpn*I **and** *Sac*I. Restriction digests of pPOTv7 NEO, pGEM®-T Easy + *VSG121,* and pGEM®-T Easy + *VSG221*; showing backbone of vectors and relevant DNA fragments after digestion.

The purified *VSG* fragments were ligated to the pPOTv7 NEO vector and transformed into competent *E. coli* cells. Eight colonies from each transformation were screened by PCR and visualised on an agarose gel for confirmation of insertion of either *VSG121* or *VSG221* (Figure 3-14). Five colonies appeared to have the correct insertion of *VSG121*, with a fragment size of 1.55 kb, and colonies 3 and 8, indicated with arrows in Figure 3-14A, were selected and grown in LB broth for further testing. No colonies screened by PCR had the correct insertion of *VSG221* as none had a correct fragment size of 1.45 kb; instead, most colonies screened had an incorrect band of ~1 kb, with one colony producing a double band, which was not desired (Figure 3-14B).
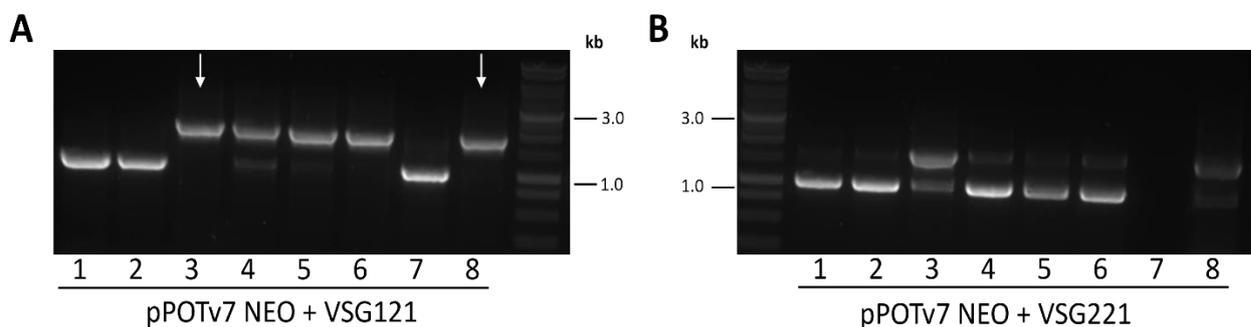
Figure 3-14. Colony screen by PCR for insertion of *VSG* **into pPOTv7 NEO.**
A. Colony screen of 8 colonies for insertion of *VSG121* into pPOTv7 NEO. Arrows indicate colonies with correct band size which were selected to be grown in LB broth.
B. Colony screen of 8 colonies for insertion of *VSG221* into pPOTv7 NEO.

A restriction digest was used to confirm the presence of *VSG121* by excising it from the pPOTv7 NEO backbone using restriction enzymes, *Kpn*I and *Sac*I, from selected colonies. The purified DNA was incubated with restriction enzymes and run on an agarose gel to determine the size of the resulting DNA fragments (Figure 3-15). The pPOTv7 NEO plasmid backbone would produce a size of ~5.5 kb with a fragment of ~1.6 kb if *VSG121* had correctly inserted. There were no bands present for the digest of colony 8, suggesting a failed digest as the band corresponding to the pPOTv7 backbone was not visible. For colony 3, the higher band suggested the presence of undigested plasmid whilst the faint lower band at around ~1.6 kb suggested *VSG121* had correctly inserted. To confirm insertion, pPOTv7 NEO + *VSG121* the plasmid from colony 3 was sent for Sanger sequencing by Eurofins Genomics. Sequencing confirmed correct insertion of *VSG121* (Appendix 5).
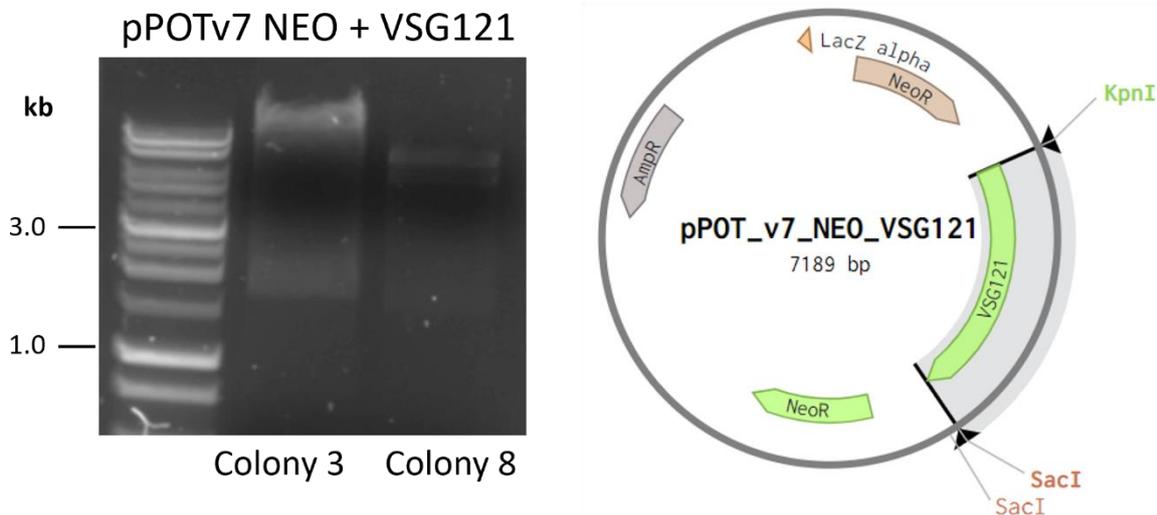
Figure 3-15. Test restriction digest of insertion of *VSG121* into pPOTv7 NEO. Agarose gel showing backbone of vectors and relevant dropouts of DNA fragments after digestion. Plasmid maps of generated vectors indicating the site of the restriction enzymes *Kpn*I and *Sac*I.

The generated plasmid, pPOTv7 NEO + *VSG121*, was used as a DNA template to amplify donors for insertion into the promoter, ESAG8, and 70 bp repeat deletion 2 region in BES1. Donors and primers for amplification of sgRNA were amplified by PCR and visualised on an agarose gel (Figure 3-16). Single bands indicating successful PCR amplifications of each donor were visible (Figure 3-16A). Primers to amplify sgRNA targeting the promoter, ESAG8 and 70 bp repeat deletion 2 were successfully amplified by PCR with a fragment size of ~ 100 bp and pooled with the corresponding donor DNA (Figure 3-16B). Ethanol precipitation was completed to prepare DNA for transfection.
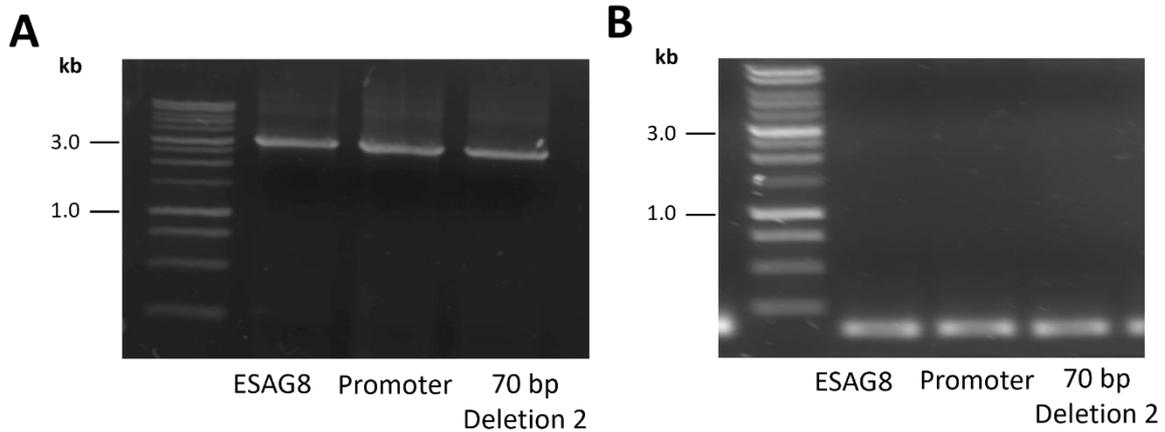
Figure 3-16. Amplification of sgRNA DNA templates and donor DNA fragment by PCR using pPOTv7 NEO + *VSG121* **as template.**
A. Amplification of pPOTv7 NEO + VSG121 donor DNA, with fragment at 3 kb.
B. Amplification of sgRNA DNA templates for insertion into the three named locations of BES1, with fragment at 100 bp.

Additional colonies were screened by PCR for insertion of *VSG221* from the same pPOTv7 NEO + *VSG221* agar plate, which failed to produce any colonies with the correct fragment size when visualised on an agarose gel. The transformation into competent *E. coli* cells was repeated, and additional colonies screened by PCR, which failed to produce any with the correct fragment size. Over 50 colonies in total were screened by PCR with none corresponding to a correct fragment size when visualised on agarose gel.

3.3.3.3 pPOTv7 HYG

Due to the recovery of *VSG221* donors from pPOTv7 BSD, and the insertion of VSG221 into pPOTv7 NEO, being unsuccessful, it was attempted to insert VSG221 into a third vector: pPOTv7 HYG. pPOTv7 HYG was digested using restriction enzymes *Kpn*I and *Sac*I (Figure 3-17). The fragment at 6003 bp was excised for pPOTv7 HYG and DNA purified before being ligated to previously purified pGEM®-T Easy + *VSG221* DNA.
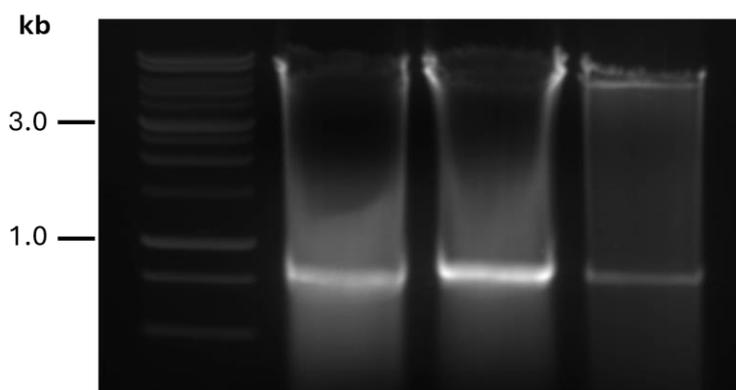
Figure 3-17. Restriction digest of pPOTv7 HYG using *Kpn*I and *Sac*I. Three restriction digests of pPOTv7 HYG visualised on agarose gel; showing backbone of vectors and relevant DNA fragments after digestion.

The ligation was used for transformation into competent *E. coli* cells. Eight colonies were screened by PCR and visualised on an agarose gel for confirmation of insertion of *VSG221* (Figure 3-18). Three colonies had a PCR fragment at the expected product size of 1.45 kb, indicated by arrows on Figure 3-18, and these colonies were grown in LB broth before DNA was extracted.
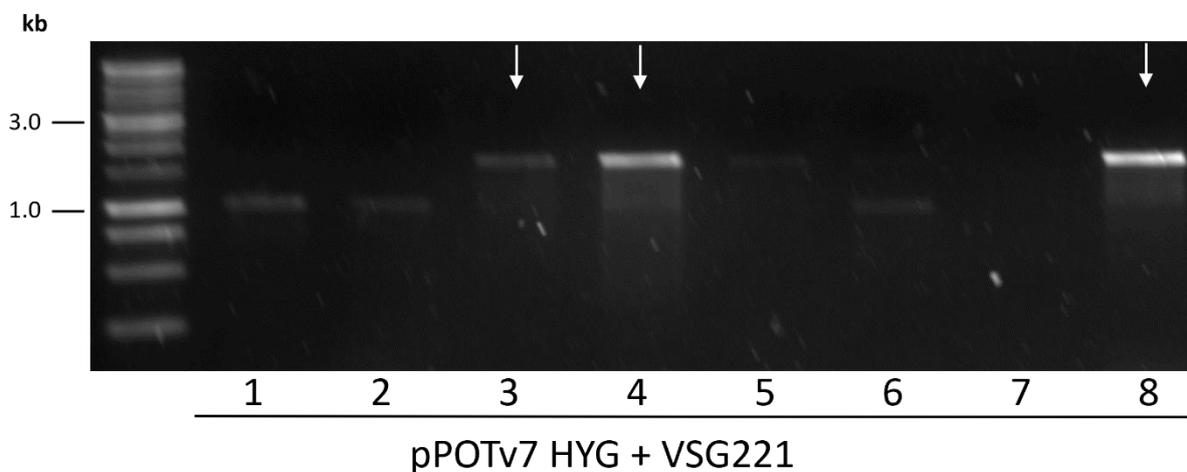
Figure 3-18. Colony screen of *VSG221* insertion into pPOTv7 HYG by PCR. Arrows indicate colonies with a fragment size at 1.45 kb which were selected for further testing.

A test restriction digest was carried out on the selected colonies to determine if *VSG221* had correctly inserted into pPOTv7 HYG. The restriction digest was used to confirm the presence of *VSG221* by excising it from the pPOTv7 HYG backbone. The purified DNA was incubated with restriction enzymes *Kpn*I and *Sac*I and run on an agarose gel to determine the size of the resulting DNA fragments (Figure 3-19). The pPOTv7 HYG plasmid backbone would produce

a size of ~6 kb with a fragment of ~1.5 kb if *VSG221* had correctly inserted. No fragments were present for screening of colony 4 suggesting *VSG221* had not inserted or a failed digestion. Colony 3 and colony 8 both produced fragments of ~6 kb for the pPOTv7 HYG backbone as well as a faint dropout of ~1.5 kb, suggesting correct insertion of *VSG221*. To confirm insertion, pPOTv7 HYG + *VSG221* colony 3 was sent for Sanger sequencing by Eurofins Genomics. Sequencing of colony 3 confirmed insertion of *VSG221*, although part of the 3' sequencing was incorrect, with mixed sequencing found. Due to this, colony 8 was sent for Sanger sequencing, which confirmed correct insertion of *VSG221,* and this was then used as a template for PCR amplification of donor DNA for transfection.
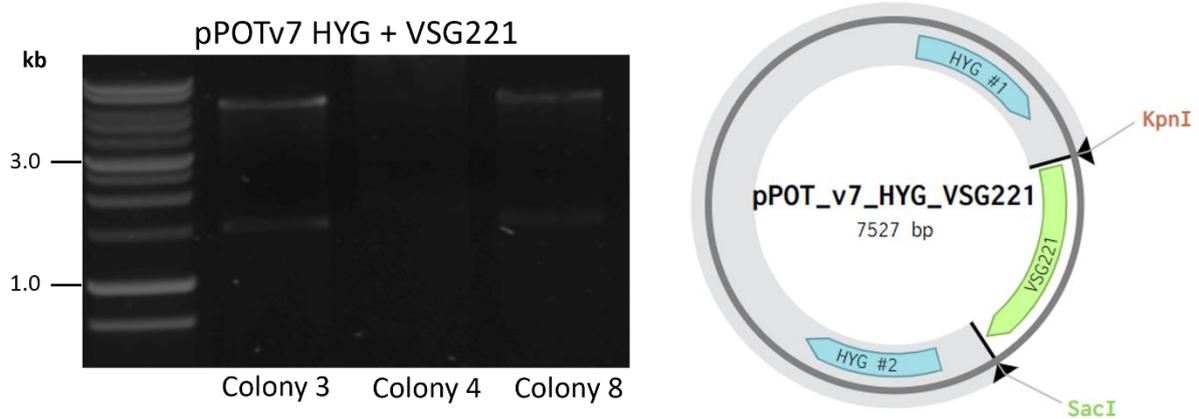
Figure 3-19. Test restriction digest of insertion of *VSG221* into pPOTv7 HYG using *Kpn*I and *Sac*I. Agarose gel showing backbone of vectors and DNA fragments after digestion. Plasmid maps of generated vectors indicating the site of the restriction enzymes *Kpn*I and *Sac*I.

The pPOTv7 HYG + VSG221 plasmid was used as a DNA template to amplify donors for insertion into the promoter, ESAG8, and 70 bp deletion 2 in BES1. Donors were amplified by PCR and visualised on agarose gel (Figure 3-20A). Fragments indicating successful amplifications were present for each donor, with a band at 3409 bp. Primers to generate sgRNA targeting the promoter, ESAG8, and 70 bp repeat deletion 2 were successfully amplified by PCR with a fragment of 90 bp (Figure 3-20B) and pooled with the corresponding donor DNA. Ethanol precipitation was carried out on pooled sgRNA and donors to prepare DNA for

transfection.
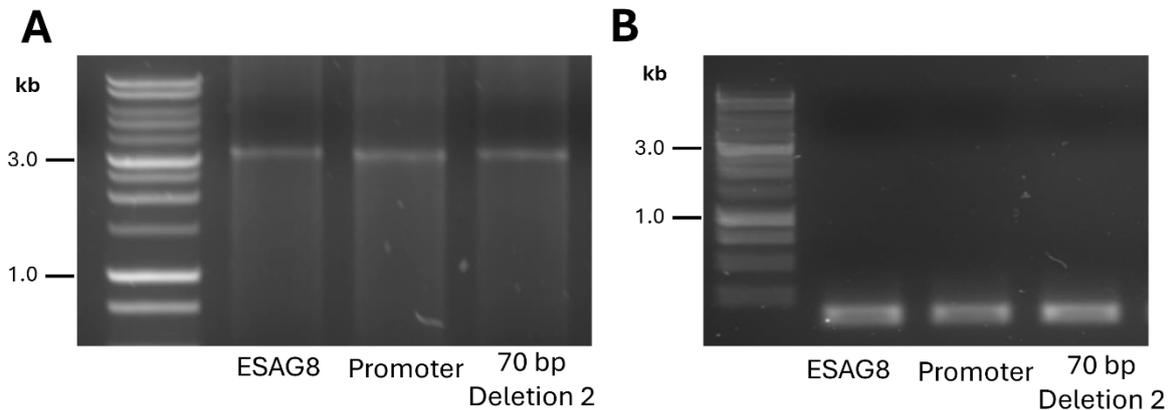
Figure 3-20. Amplification of sgRNA and donor DNA fragment using pPOTv7 HYG + *VSG221* as template.
A. Amplification of pPOTv7 HYG + VSG221 donor DNA, with fragment at 3 kb.
B. Amplification of sgRNA for insertion into the three named locations of BES1, with fragment at 100 bp.

3.3.4 Transfection and confirmation of *VSG* insertion

3.3.4.1 Transfection

After successful amplification of sgRNA and donors containing a DNA template of generated plasmids, either pPOTv7 NEO + *VSG121* or pPOTv7 HYG + *VSG221,* a total of 3-10 PCR reactions were pooled together for each location targeted for *VSG* insertion and purified by ethanol precipitation. After ethanol precipitation of donors and sgRNA for insertion of *VSG121*, concentrations of 6009.1 ng/μl were achieved for targeting ESAG8, 1408.6 ng/μl for the promoter, and 2527.9 ng/μl for insertion at 70 bp repeat deletion 2. For insertion of *VSG221*, a concentration of 1698.2 ng/μl was achieved for targeting ESAG8, 2101.2 ng/μl for the promoter, and 1340.5 ng/μl for insertion at 70 bp repeat deletion 2.BSF *Tb*Cas9 cells were grown to a density of 1×10^6 cells/ml and a total of 1×10^7 cells used per transfection, and 10 μg of DNA transfected through electroporation. Antibiotic resistant clones were then selected using the appropriate selective drug of either 2.5 μg/ml of G418 (Neomycin) for *VSG121* or 5 μg/ml of hygromycin for *VSG221.*

3.3.4.2 Unsuccessful insertion of *VSG121* replacing ESAG8

After addition of neomycin to undiluted flasks and 24-well plates for insertion of *VSG121* into *ESAG8*, no cells survived the first attempted transformation. Transfection of *VSG121* into ESAG8 was attempted multiple further times and each time failed to produce any surviving

clones. This indicates that pPOTv7 NEO + *VSG121* could not be used to integrate VSG121 into ESAG8.

3.3.4.3 Attempted insertion of *VSG121* replacing 70 bp repeat deletion 2

In contrast, neomycin resistant clones were obtained for *VSG121* insertion into 70 bp repeat deletion 2. After transfection, 2 clones from the 1/20 dilution 24-well plate and 8 clones from the 1/4 dilution 24-well plate had survived and were visible under light microscopy. Four clones were selected for further passaging and subsequent testing. Two clones were selected from each dilution, with clones 1 and 2 taken from 1/20 dilution, and clones 3 and 4 from a 1/4 dilution plate. On passaging of the cell line, neomycin was again added to each flask to a concentration of 2.5 µg/ml for continued antibiotic selection. After passaging with addition of neomycin, clones 1 and 2, failed to survive, which perhaps indicates unsuccessful initial antibiotic selection. In contrast, clones 3 and 4 survived multiple passages in the presence of neomycin, showing true neomycin resistance.

To confirm if insertion was successful, gDNA was extracted from clones 3 and 4 which had VSG121 inserted into 70 bp repeat deletion 2. Clones 3 and 4 were screened alongside the WT, and a positive control, which is known to have neomycin resistance. Clones were screened for insertion of *VSG121* and neomycin resistance with primers used for screening indicated in Figure 3-21A. The resulting PCR products were visualised on an agarose gel (Figure 3-21). No band was present for either clone when screened for insertion of *VSG121* using primers HS20 and HS21 (Figure 3-21B). On screening for the neomycin resistance cassette with primers M106 and M107, no bands were present for clone 3 or 4, which is surprising as these were growing in media containing neomycin. No band at the correct size was visible for the positive control but as expected no band was present for the WT, which are not resistant to neomycin (Figure 3-21C)., The forward screening primer for *VSG121* insertion, HS20, used with the reverse neomycin primer, M107, produced a fragment for clone 3 at the expected size ~ 1 kb for successful insertion (Figure 3-21D). No fragments were produced using the forward neomycin screening primer, M106, with the reverse *VSG12*1 screening primer, HS21 (Figure 3-21E).

Despite the recovery of at least 4 clones from the selection, two proved not to be truly neomycin resistant and PCR was only able to confirm the expected replacement of the upstream 70 bp repeat region of BES1 with VSG121, and in only one clone. Taken together, these data indicate

a problem with this approach for insertion, or with the consequences of inserting a second VSG in this location, or with the PCR conditions and primer pairs used giving inconsistent results.
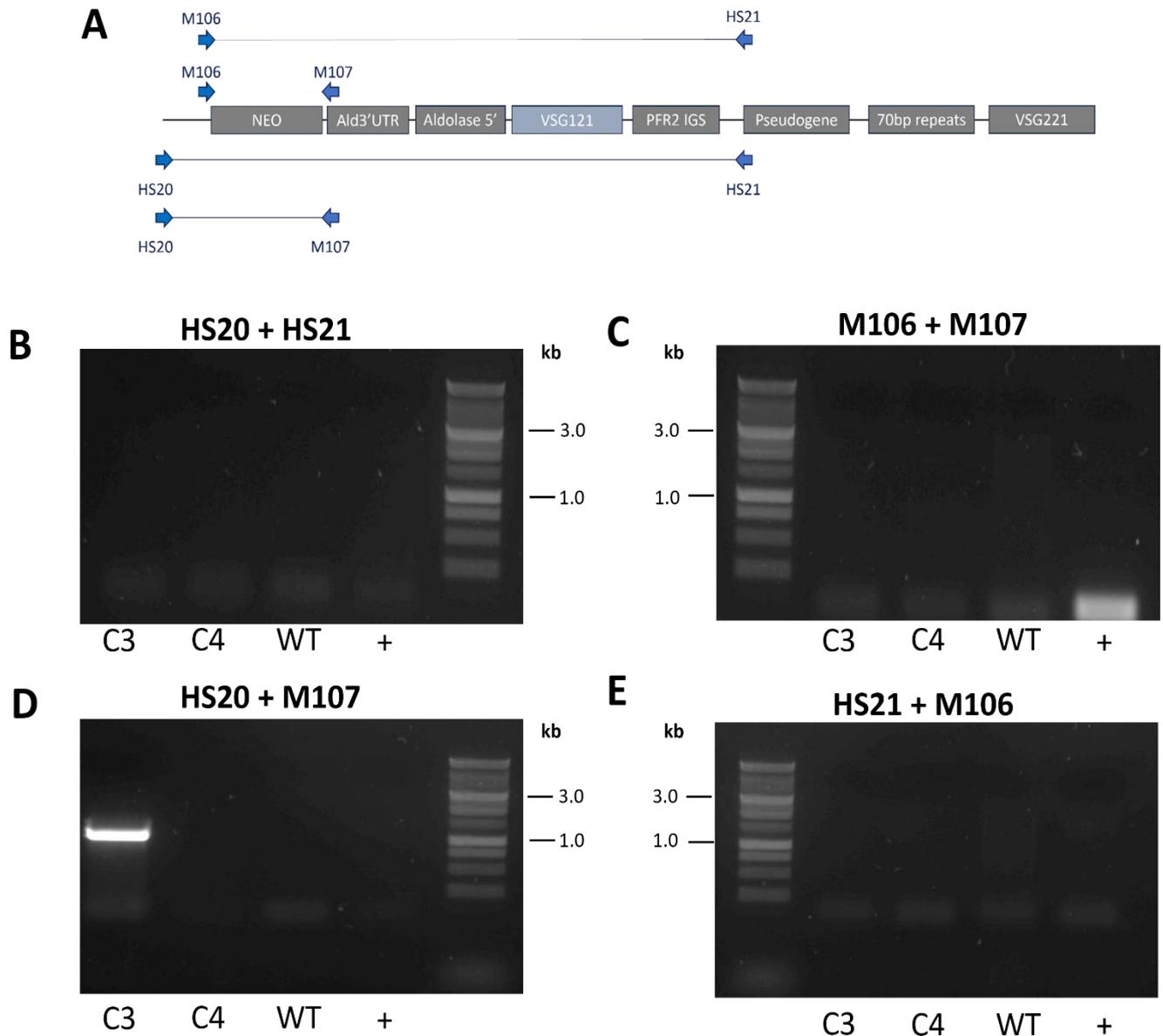
Figure 3-21. Confirmation screening of *VSG121* insertion in 70 bp repeat deletion 2 by PCR.
A. PCR confirmation strategy indicating the location of the forward and reverse primers used and size of the expected fragment.
B-D. PCR analysis of clones amplified using primers shown in A visualised on an agarose gel. WT band depicts the wildtype TbCas9/T7 cell line. + indicates a NEO resistant positive control.

3.3.4.4 Attempted insertion of *VSG121* into promoter region

In contrast to the difficulties with *VSG121* insertion into the 70 bp repeat and ESAG8 regions of BES1 described above, recovery of clones predicted to have integrated *VSG121* into the promoter region of BES1 worked with high efficiency. After transfection, all wells on the 1/4 dilution 24-well plate had surviving clones visible under light microscopy, and an additional 11 wells were growing from the 1/20 dilution plate. A total of five neomycin resistant clones were selected for further testing of *VSG121* insertion into the promoter region of BES1: four clones were selected from the 1/20 dilution plate, and 1 clone from the 1/4 dilution plate, and all cells were passaged. For continued antibiotic selection, 2.5 µg/ml of neomycin was added to the cell lines on passaging. All 5 clones survived passaging, indicating they were neomycin resistant.

To test if *VSG121* had inserted as expected, gDNA was extracted from each clone and confirmatory screening by PCR was conducted and visualised on agarose gels (Figure 3-22). No PCR reactions generated any products, and this was unable to be resolved by repeating the PCRs or by altering the PCR conditions, including by gradient PCR.
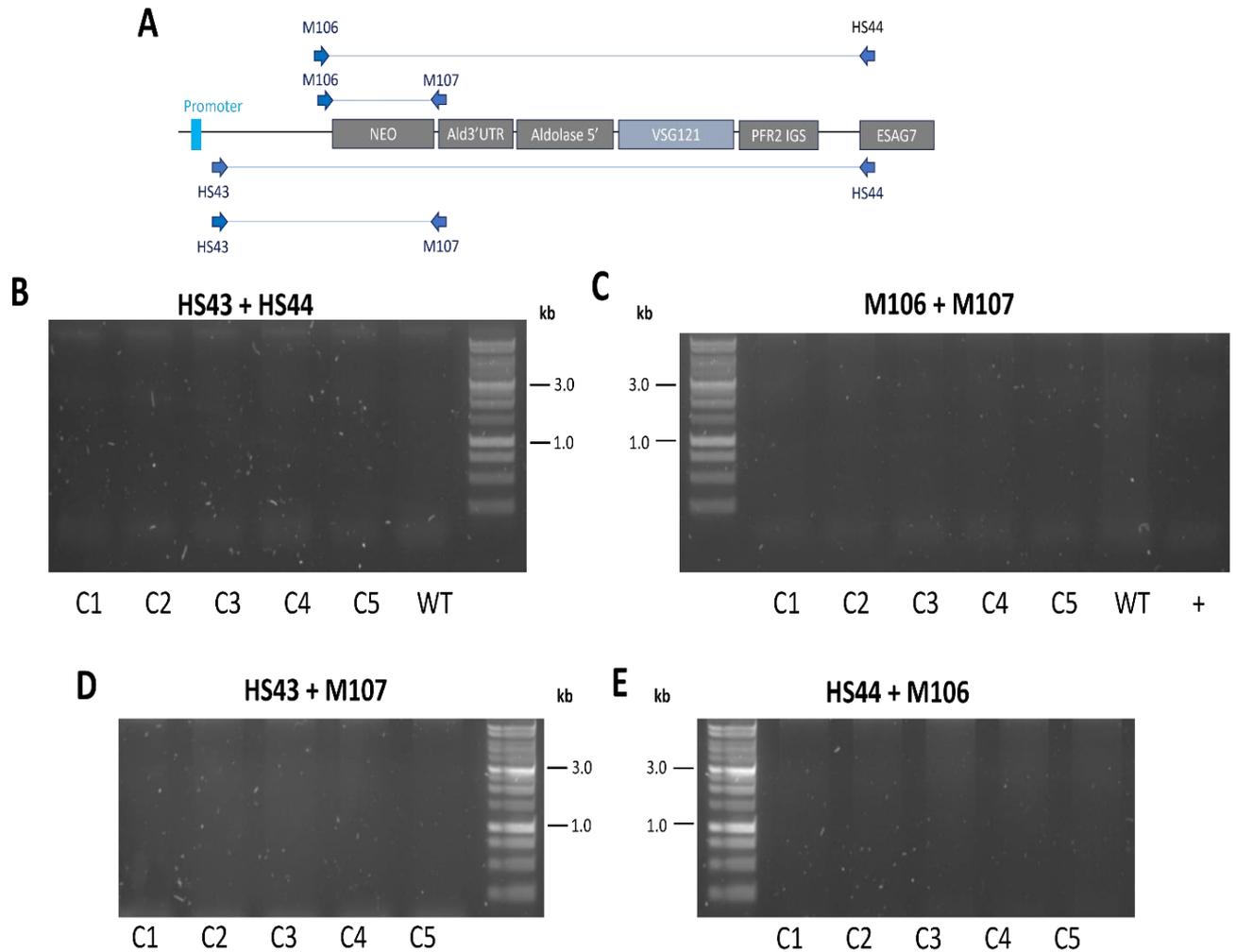
Figure 3- 22. Confirmation screening of *VSG121* insertion at the promoter of BES1 by PCR.
A. PCR confirmation strategy indicating the location of the forward and reverse primers used and size of the expected fragment.
B-D. PCR confirmation of clones amplified using primers shown in A visualised on agarose gel. WT band depicts the wildtype TbCas9/T7 cell line. + indicates a NEO resistant positive control.

3.3.4.5 Unsuccessful transfection of *VSG221*

All transfections attesting to integrate *VSG221* into each location (Promoter, 'third copy' of ESAG8, and the 70 bp repeats region 2) failed to produce any surviving antibiotic-resistant clones.

3.4 Characterisation of *T. brucei* cell lines

3.4.1 VSG Immunofluorescence assay

The above data indicate a number of challenges in manipulating BES1 to integrate the VSGs into non-endogenous locations in the expression site. For *VSG221*, we were unable to recover any transformants, and while antibiotic resistant cells were recovered after attempting to target *VSG121* to the smaller 70 bp repeat region and, especially, the promoter, validation of the integrations by PCR was highly problematic and did not provide evidence for successful integration. Due to these issues, an immunofluorescence assay was performed to ask if it might provide evidence for the expected BES1 modification, since VSG121 should only be expressed in the cells used if it has been inserted into the genome, since BES3 is normally inactive (Girasol *et al*, 2023a). VSG immunofluorescence was therefore performed. In all cases, a total of $1x10^6$ cells were incubated with an α-VSG121 antibody, which is considered specific to VSG121, followed by incubation with a secondary antibody, α-rat Alexa Fluor® 488, which is labelled with a fluorescent dye and binds to the primary antibody. Images were obtained using a Leica DiM8 microscope (Figure 3-23). Five clones with predicted insertion of *VSG121* at the promoter, on the basis of neomycin resistance, were screened for VSG121 expression, with all clones having visible expression of VSG121 (Figure 3-23). Signal levels were not consistent, with clone 1 and clone 4 appearing to have a stronger signal than clones 2, 3 or 5, which were more comparable to the WT cells. Both clones predicted to have insertion of *VSG121* at the 70 bp repeat region 2 also appeared to have a stronger signal when compared to the WT. It was not expected that the WT cells would express VSG121 as this is normally inactive (Girasol *et al*., 2023a), but some immunofluorescence was seen. Based on these data, clones 1 and 4 for putative promoter insertion of VSG121, and clones 3 and 4 for putative insertion of *VSG121* at the 70 bp repeat region were used in further experiments.
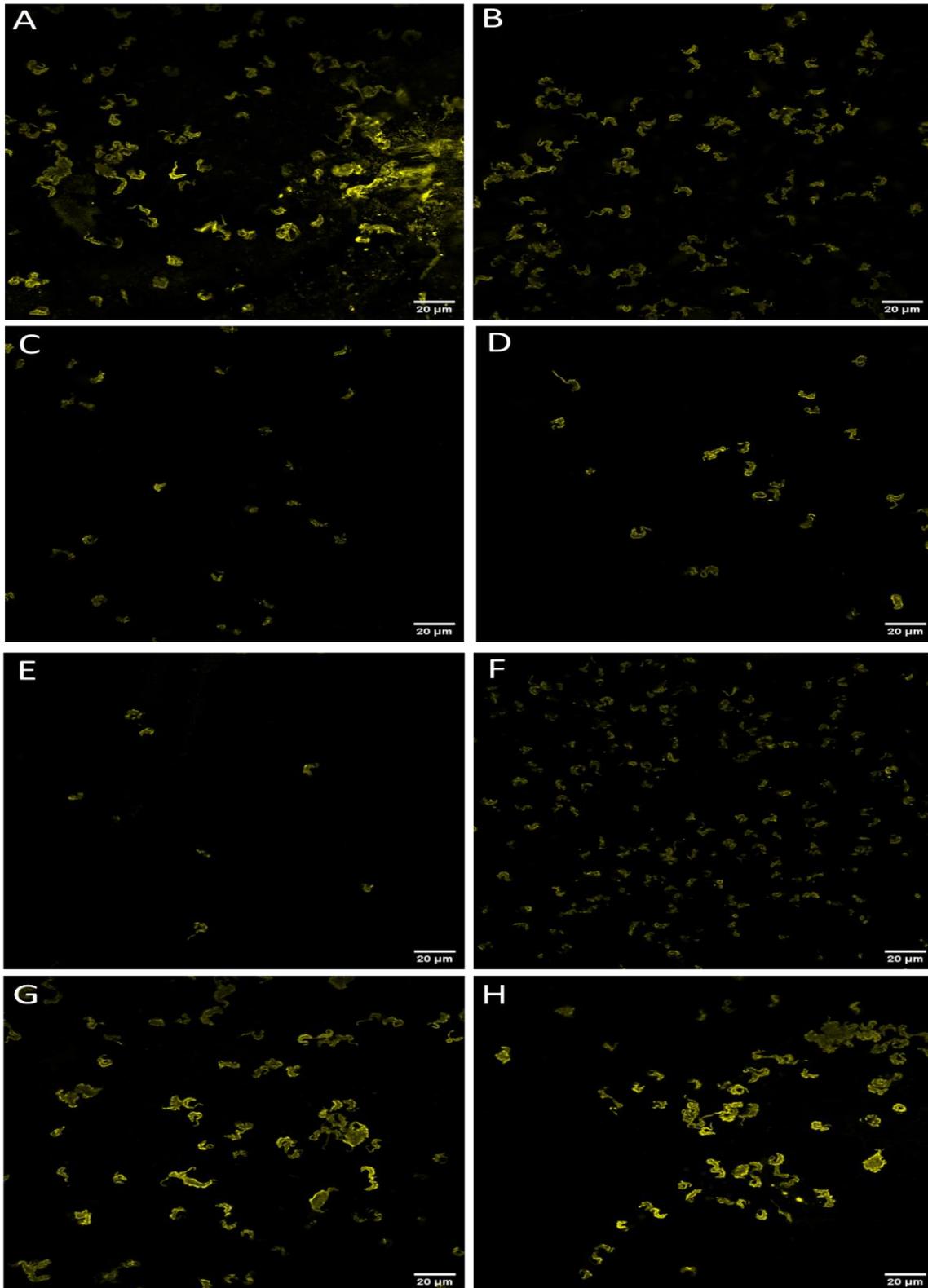
Figure 3-23. VSG immunofluorescence assay for expression of VSG121. Images obtained using Leica DiM8 microscope and analysed using FIJI. Scale bar of 20 µm. A) Promoter clone 1 B) Promoter clone 2 C) Promoter clone 3 D) Promoter clone 4 E) Promoter clone 5 F) WT G) 70 bp region 2 clone 3 H) 70 bp region 2 clone 4.

3.4.2 Oxford Nanopore sequencing

To attempt to provide a definitive assessment of whether or not any of the genetic manipulations of BES1 had been successful, Oxford Nanopore sequencing was completed. Oxford Nanopore sequencing identifies the sequence of nucleotides by measuring the electric current as single-stranded DNA passes through the nanopores in a linear sequence. Oxford Nanopore sequencing reveals single nucleotide polymorphisms and requires different Nanopore reads to be aligned together to reveal a consensus of the real base at the given location. DNA was extracted from a total of $2x10^8$ cells for each generated *T. brucei* cell line, 70 bp repeat deletion 1, 70 bp repeat deletion 2 clone 2 and clone 7, VSG121 insertion into promoter clone 1 and 4, VSG121 insertion into 70 bp repeat deletion 2 clone 3 and clone 4. Cells were also extracted from TbCas9 WT and a *T. brucei* RAD51 knockout cell line generated by Girasol et al. (2023). Each of these cell lines were subjected to Nanopore sequencing, and subsequent analysis was completed by Marija Krasilnikova. To look at the read depth coverage, sequencing data was mapped to the TriTrypDB v.64 *T. brucei brucei* Lister 427 genome (http://tritrypdb.org), sequencing output files were sorted into bamfiles, and bamCoverage (of deepTools) used to assess read depth coverage for each sample. A minimum mapping quality (MAPQ) score was applied at 1 to filter out reads with a mapping quality below this. This cutoff removed the lowest quality reads which map in many locations but would not remove all reads that map to more than one place in the genome.

The read depth coverage was first mapped to the active BES, BES1, containing the endogenous *VSG*, *VSG221* (Figure 3-24). Three regions of interest (ROI) in BES1 were defined: ROI1 ~21,050 – 21,900 across the promoter region, ROI2 ~ 67,850 – 68,500 across the small region of 70 bp repeats, region 2, and ROI3 ~ 70,960 – 74,550 across the large section of 70 bp repeats, region 1. This mapping gave a first indication that the genetic manipulations to BES1 were correct. In the RO1, *VSG121* insertion into the promoter region in clones 1 and 4 both have sequencing which does not match the reference genome in the promoter region, indicating that this sequence may have been removed (Figure 3-24; Read depth coverage in purple and annotated as Prom C1 and Prom C4). In the ROI2 which covers the small section of 70 bp repeats, region 2, four clones have a read depth coverage which do not match the reference genome which would be expected if the genetic manipulations to this region were successful. Two of these clones had VSG121 attempted to be inserted into this region so this mapping gives an indication that this may have been successful (Figure 3-24; Read depth coverage in blue and annotated as Del.2 C3 and Del.2 C4). Two other clones, clone 2 and 7 for 70 bp repeat

deletion 2, do not have reads mapped to this region which would match successful knockout (Figure 3-24; Read depth coverage in orange and annotated as 70bp Del.2 C2 and 70bp Del.2 C7). The ROI3 covers the large section of 70 bp repeats, region, 1, and 70 bp deletion 1 does not map fully to the reference genome in this region indicating part of the repeats may have been successfully removed (Figure 3-24; Read depth coverage in green and annotated as 70bp del. 1). However, there is some read mapping to the main stretch of 70 bp repeats in 70 bp deletion 1, although this may be mismapping of short reads to BES1, which are from other BESs, as the MAPQ of 1 would not have removed all reads, which map to more than one place in the genome.
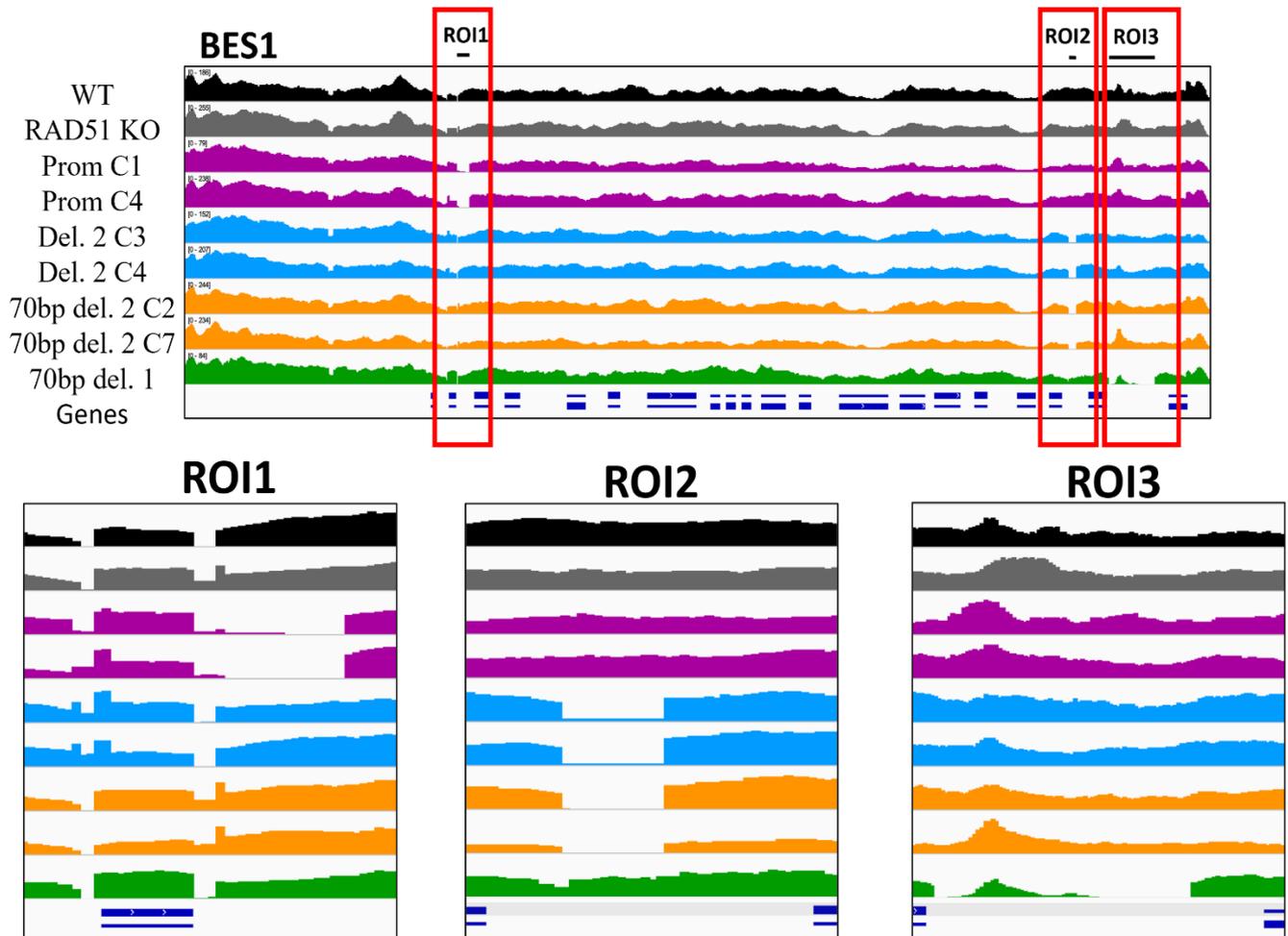
Figure 3-24. Mapping of read depth coverage in Nanopore sequencing reads to BES1.
Regions of interest to the generated cell lines are defined. Region of interest 1 (ROI1)
~21,050 – 21,900 is across the promoter region. ROI2 ~ 67,850 – 68,500 is across the small
region of 70 bp repeats, region 2. ROI3 ~ 70,960 – 74,550 is across the large section of 70 bp
repeats, region 1. Cell lines analysed include the wild-type TbCas9 (WT), knockout of
RAD51 (RAD51 KO), VSG121 inserted into the promoter region clones 1 and 4 (Prom C1;
Prom C4), VSG121 inserted into the small section of 70 bp repeats clones 3 and 4 (Del. 2 C3
and Del. 2 C4), the small section of 70 bp repeats removed from the BES clones 2 and 7
(70bp del. 2 C2 and 70bp del.2 C7), and deletion of the larger section of 70 bp repeats from
the BES (70bp del. 1).

The nanopore sequencing reads were also mapped to BES3, a normally inactive BES, as this is where *VSG121* is usually found (Figure 3-25). As mapping is relative to the reference genome where *VSG121* resides in BES3, all reads with sequencing matching *VSG121* will be placed in BES3. The read depth coverage across BES3 was mapped to try to detect an increase in the read depth coverage in BES3 across *VSG121* in cell lines where insertion of *VSG121* was attempted. The ROI4 is ~ 62,890 – 64,550 and covers the *VSG* in BES3. In generated cell lines with a copy of *VSG121* inserted, in purple and blue, there is a corresponding increase in read depth coverage over the region of *VSG121*, suggesting *VSG121* had been inserted.
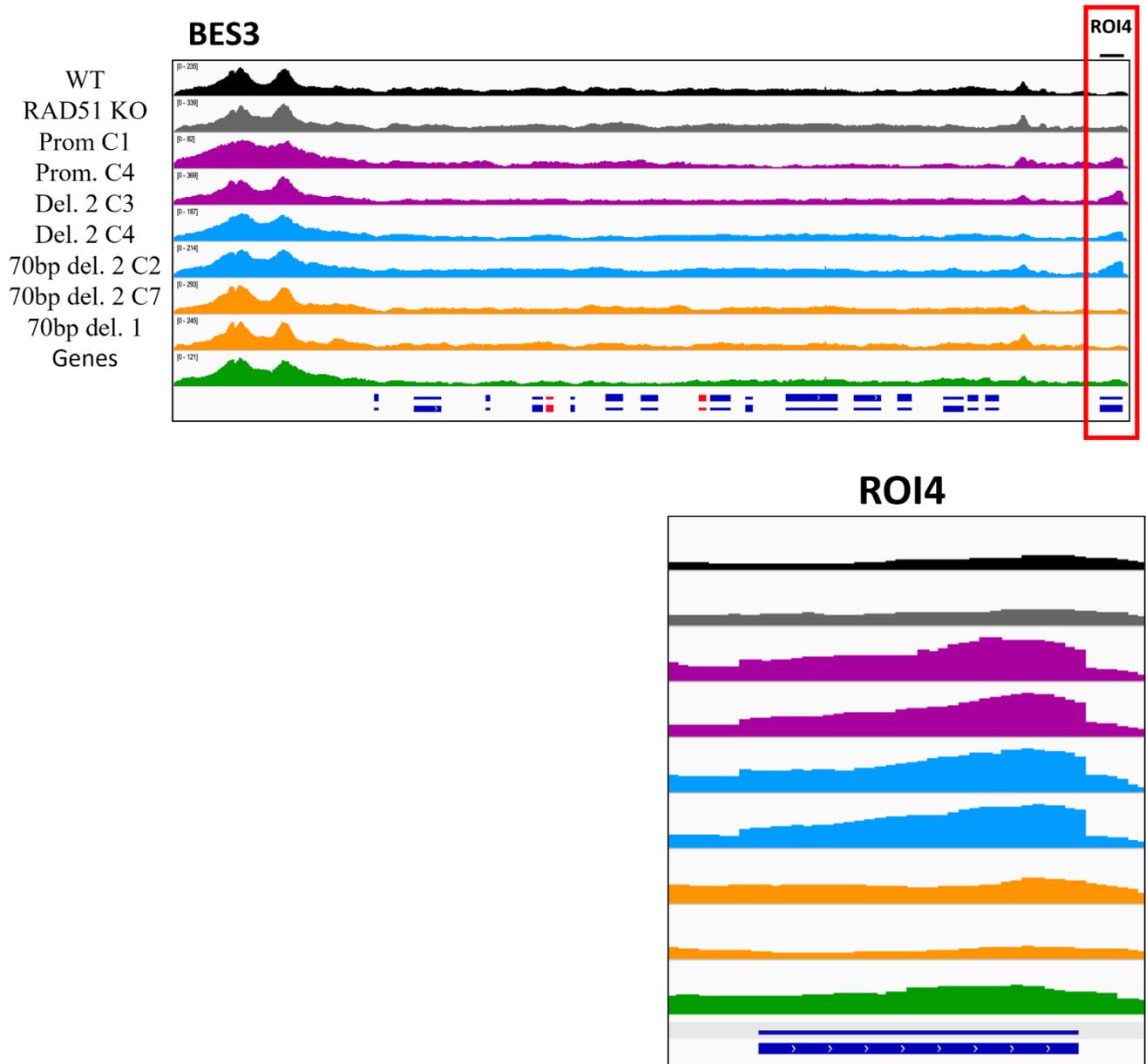
Figure 3-25. Mapping of read depth coverage in Nanopore sequencing reads to BES3.
Region of interest 4 (ROI4) is depicted from ~62,890 – 64,550 and covers *VSG121*. Cell lines
analysed include the wild-type TbCas9 (WT), knockout of RAD51 (RAD51 KO), VSG121
inserted into the promoter region clones 1 and 4 (Prom C1; Prom C4), VSG121 inserted into
the small section of 70 bp repeats clones 3 and 4 (Del. 2 C3 and Del. 2 C4), the small section
of 70 bp repeats removed from the BES clones 2 and 7 (70bp del. 2 C2 and 70bp del.2 C7),
and deletion of the larger section of 70 bp repeats from the BES (70bp del. 1).

Individual reads that cover the regions of interest, which were suggested by the read depth coverage mappings were then analysed. Between 3-6 reads per line were analysed and viewed on Benchling (www.benchling.com), comparing the reads to annotated sequence maps of what should have been present in the generated cell lines if gene knockout or insertion was successful.

3.4.2.1 Successful removal of 70 bp repeat region 2

Nanopore sequencing confirmed successful insertion of BSD resistance cassette in both clone 2 (Figure 3-26) and clone 7 (Figure 3-28) by comparing individual sequence reads with the annotated sequence map of BES1. Sequencing also confirmed that insertion of BSD had successfully removed 70 bp repeat region 2 in both clone 2 (Figure 3-27) and clone 7 (Figure 3-29) by comparing individual sequence reads with the annotated sequence map.

Figure 3-26. Oxford Nanopore sequencing for confirmation of BSD resistance cassette in 70 bp repeat deletion 2 clone 2. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, actin 5' indicated in purple, BSD in blue, aldolase UTR 3' in green, aldolase 5' in orange. Primers used for PCR screening are indicated in pink. The sequences used as donor DNA template in CRISPR/Cas9 are indicated in lime green. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.

Figure 3-27. Oxford Nanopore sequencing for confirmation of knockout of the 70 bp repeat region 2 in clone 2. Identifies the location of the inserted plasmid in BES1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, actin 5' indicated in purple, BSD in blue, aldolase UTR 3' in green, aldolase 5' in orange, ESAG8 in lime green, and pseudogene in grey.The sequences used as donor DNA template in CRISPR/Cas9 are indicated in lime green. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.

Figure 3-28. Oxford Nanopore sequencing for confirmation of BSD resistance cassette in 70 bp repeat region 2 clone 7. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, actin 5' indicated in purple, BSD in blue. aldolase UTR 3' in green, aldolase 5' in orange. Primers used for PCR screening are indicated in pink. The sequences used as donor DNA template in CRISPR/Cas9 are indicated in lime green. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.
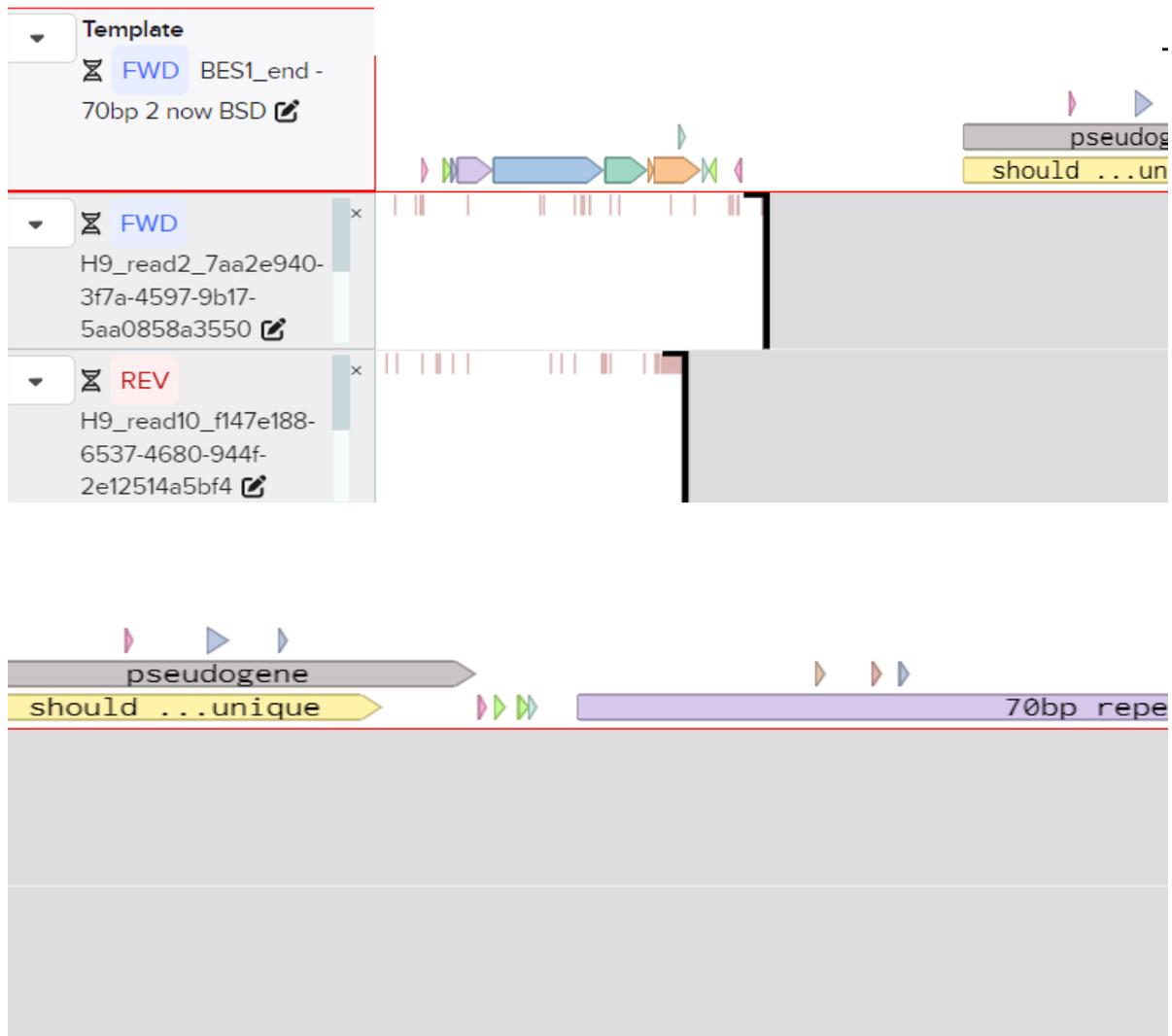
Figure 3-29. Oxford Nanopore sequencing for confirmation of knockout of the 70 bp repeat region 2 in clone 7. Identifies the location of the inserted plasmid in BES1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, actin 5' indicated in purple, BSD in blue, aldolase UTR 3' in green, aldolase 5' in orange, pseudogene in grey, and the large section of 70 bp repeats in purple. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.

3.4.2.2 Successful removal of 70 bp repeat region 1

Successful removal of 70 bp repeat region 1 and replacement with the BSD resistance cassette was confirmed by Nanopore sequence, by comparing individual sequence reads with the annotated sequence map of BES1 (Figure 3-30).

Nanopore sequencing further upstream of the *VSG* revealed a difference in the sequence of the 70 bp repeats region 2 compared to the endogenous BES1. The previously small region of 70

bp repeats had expanded from ~ 500 bp to ~3500 bp (Figure 3-31). The sequence of the expanded 70 bp repeats was compared to that of 70 bp repeats region 1 in the endogenous BES1 to determine if this sequence had relocated rather than being successfully removed. The sequences did not match indicating that 70 bp region 1 had been removed and this was a separate expansion of region 2.
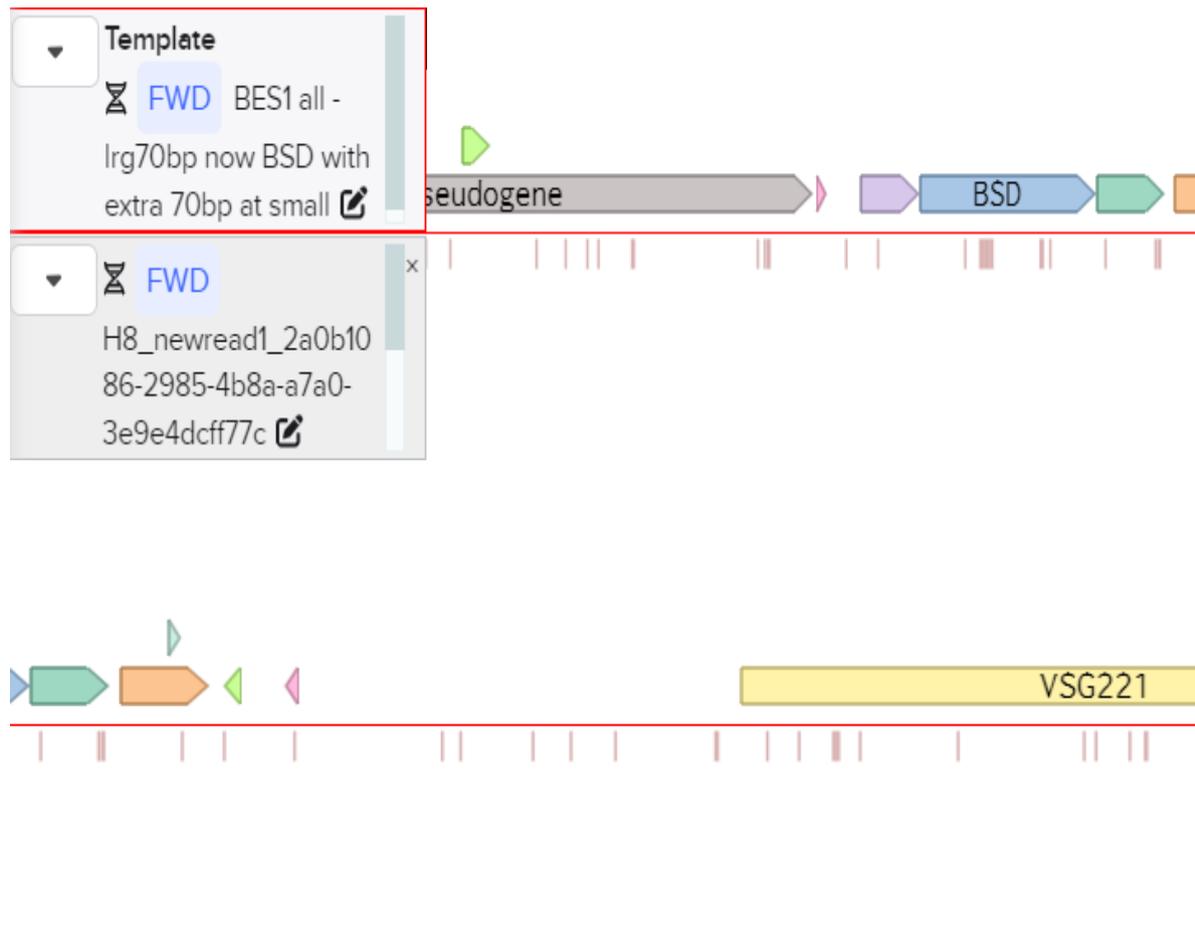
Figure 3-30. Oxford Nanopore sequencing for confirmation of knockout of the 70 bp repeat region 1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, actin 5' indicated in purple, BSD in blue, aldolase UTR 3' in green, aldolase 5' in orange, pseudogene in grey, and *VSG221* in yellow. Primers used for PCR screening are indicated in pink. The sequences used as donor DNA template in CRISPR/Cas9 are indicated in lime green. Pink lines in the sequence indicate mismatches in the sequence. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.
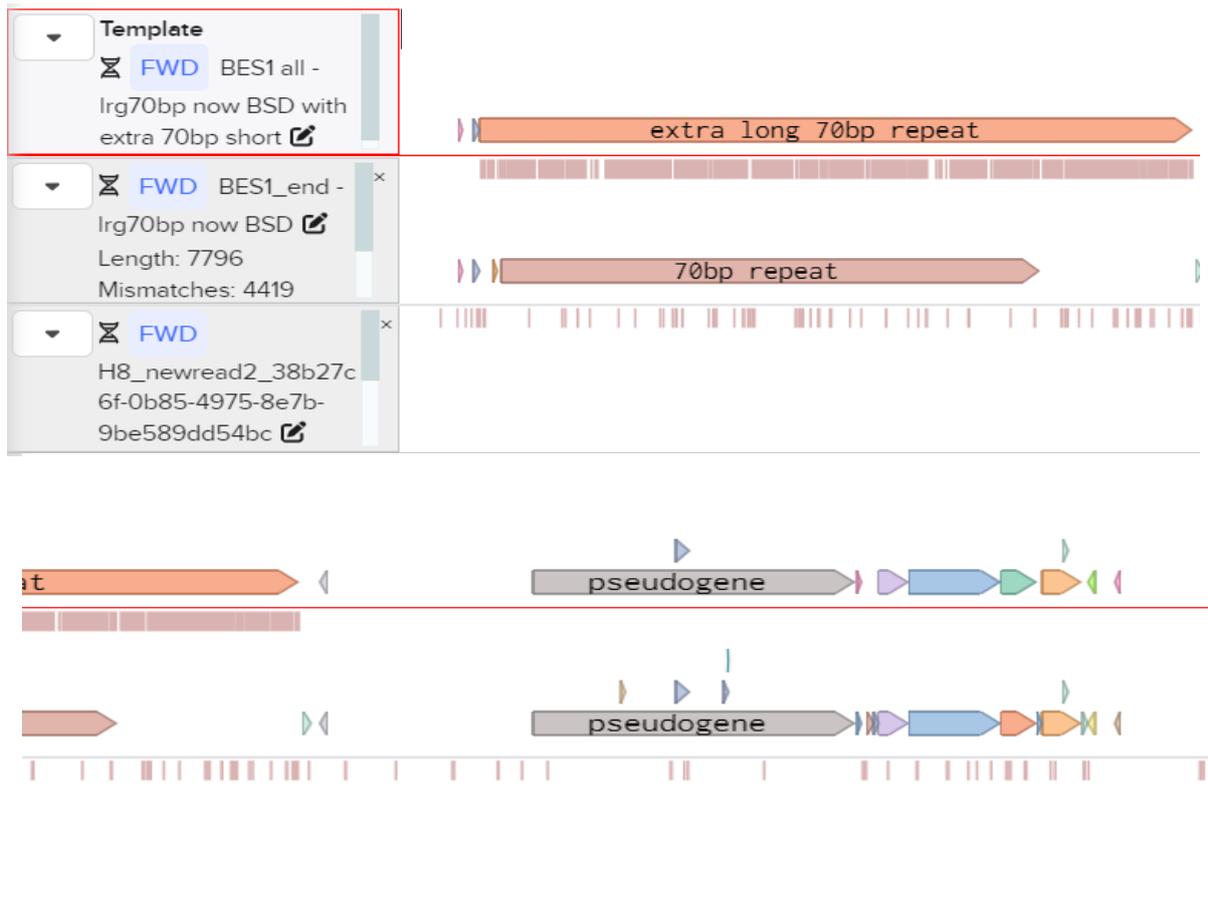
Figure 3-31. Oxford Nanopore sequencing upstream of the *VSG* in 70 bp repeat deletion 1 cell line. The first line is the organisation of BES1 with the small section of 70 bp repeats enlarged (in orange), the pseudogene in grey and the predicted insertion from pPOTv7 BSD upstream. The second line is the original organisation of BES1 with a small section of 70bp repeats (also orange); the pseudogene (grey) and the same predicted organisation of the pPOTv7 BSD insertion. The third line shows a nanopore read. The pink lines in the sequence indicate mismatches - as can be seen, the 70bp repeat region in the second line is mostly mismatched; while the nanopore read has only few pink lines indicating SNPs as expected for long read sequences. The upstream pPOTv7 BSD region matches in all lines.

3.4.2.3 Successful insertion of *VSG121* into 70 bp repeat region 2

Nanopore sequencing confirmed successful insertion of *VSG121* and NEO resistance cassette in both clone 3 (Figure 3-32) and clone 4 (Figure 3-34) by comparing individual sequence reads with the annotated sequence map of BES1. Sequencing also confirmed that this insertion successfully removed and replaced 70 bp repeat region 2 in clone 3 (Figure 3-33) and clone 4 (Figure 3-35) by comparing individual sequence reads with the annotated sequence map of BES1.
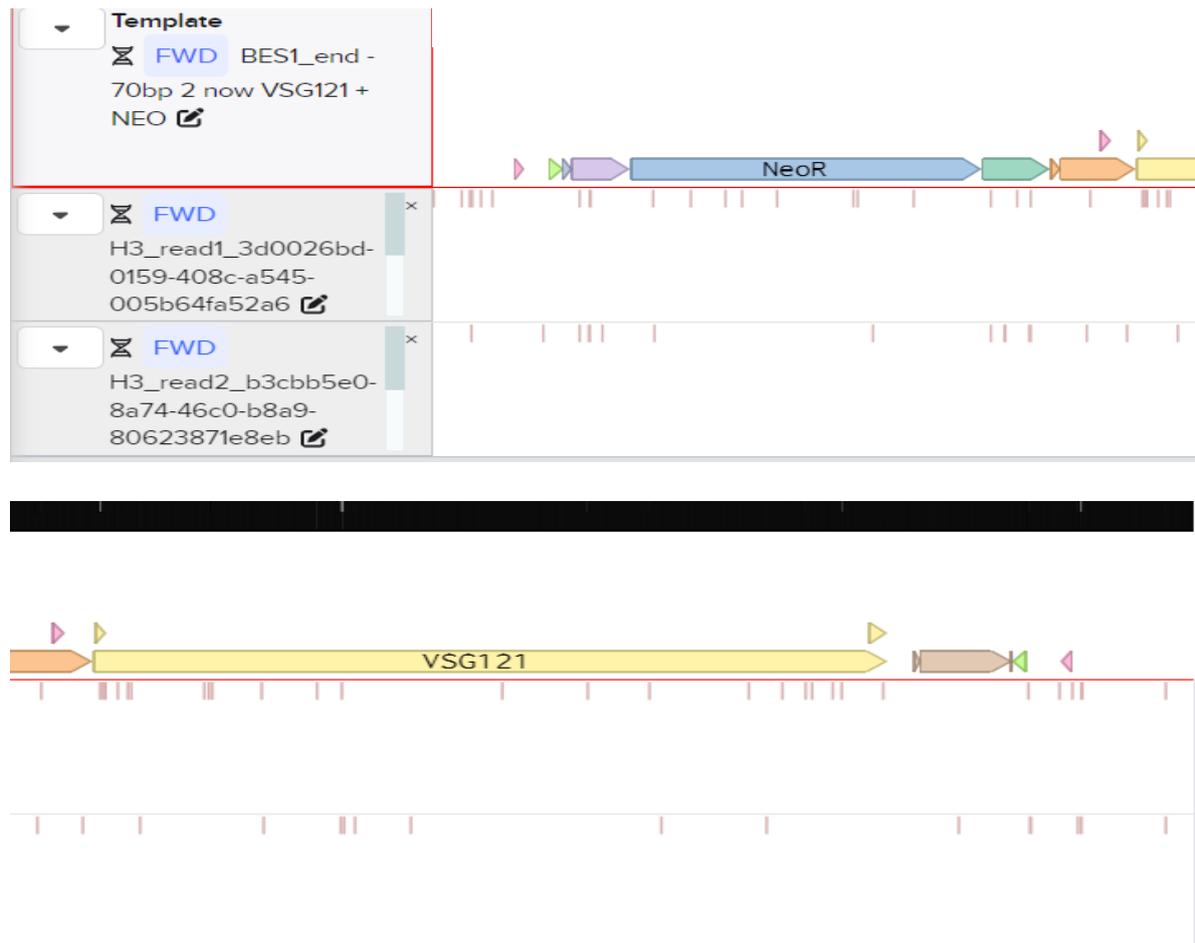
Figure 3-32. Oxford Nanopore sequencing for confirmation of insertion of *VSG121* into 70 bp repeat region 2 in clone 3. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, actin 5' indicated in purple, NEO in blue, aldolase UTR 3' in green, aldolase 5' in orange, and *VSG121* in yellow, PFR2 IGS in brown. Primers used for PCR screening are indicated in pink. The sequences used as donor DNA template in CRISPR/Cas9 are indicated in lime green. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.
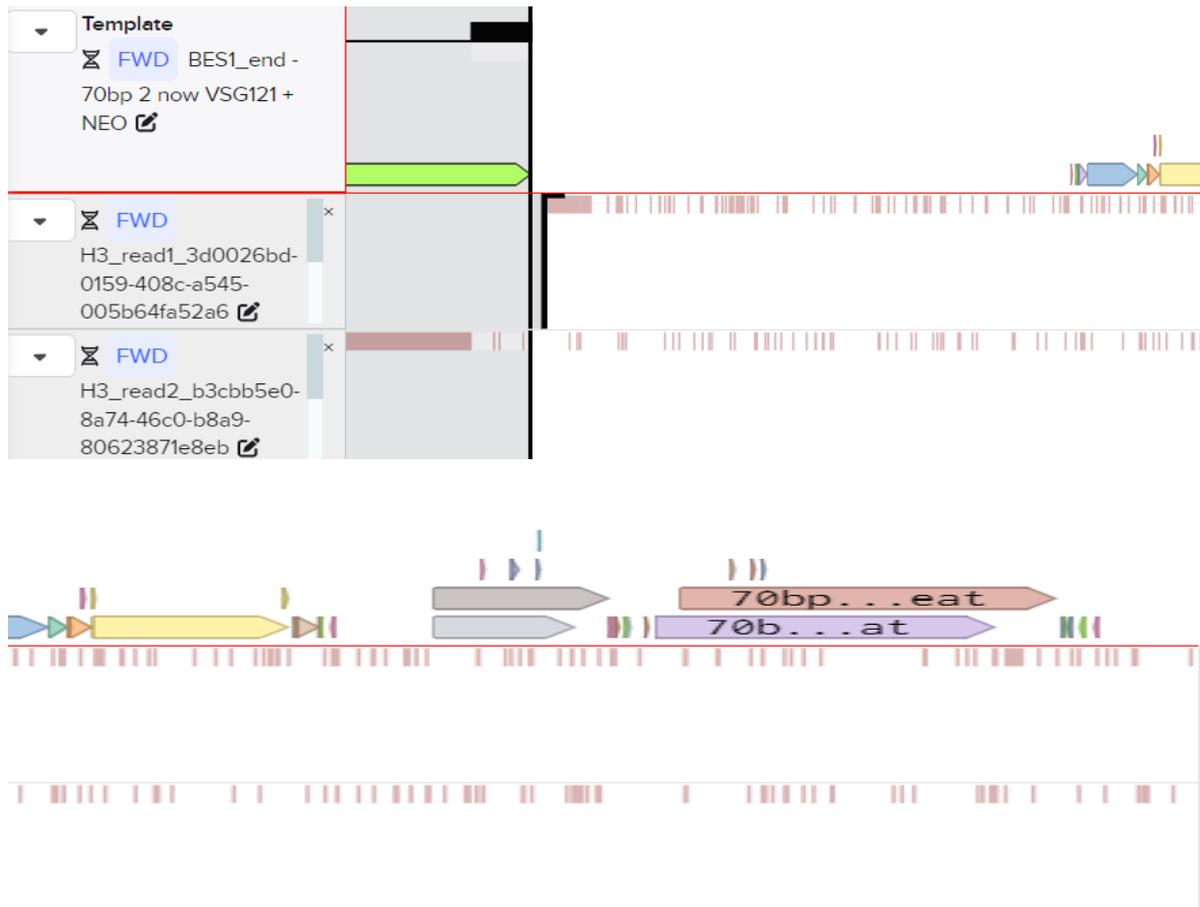
Figure 3-33. Confirmation that *VSG121* replaced the small region of 70 bp repeats in clone 3. Identifies the location of the inserted plasmid in BES1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, ESAG8 is in lime green, NEO in blue, *VSG121* in yellow, *VSG* pseudogene in grey, and the large section of 70 bp repeats in purple and brown. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.
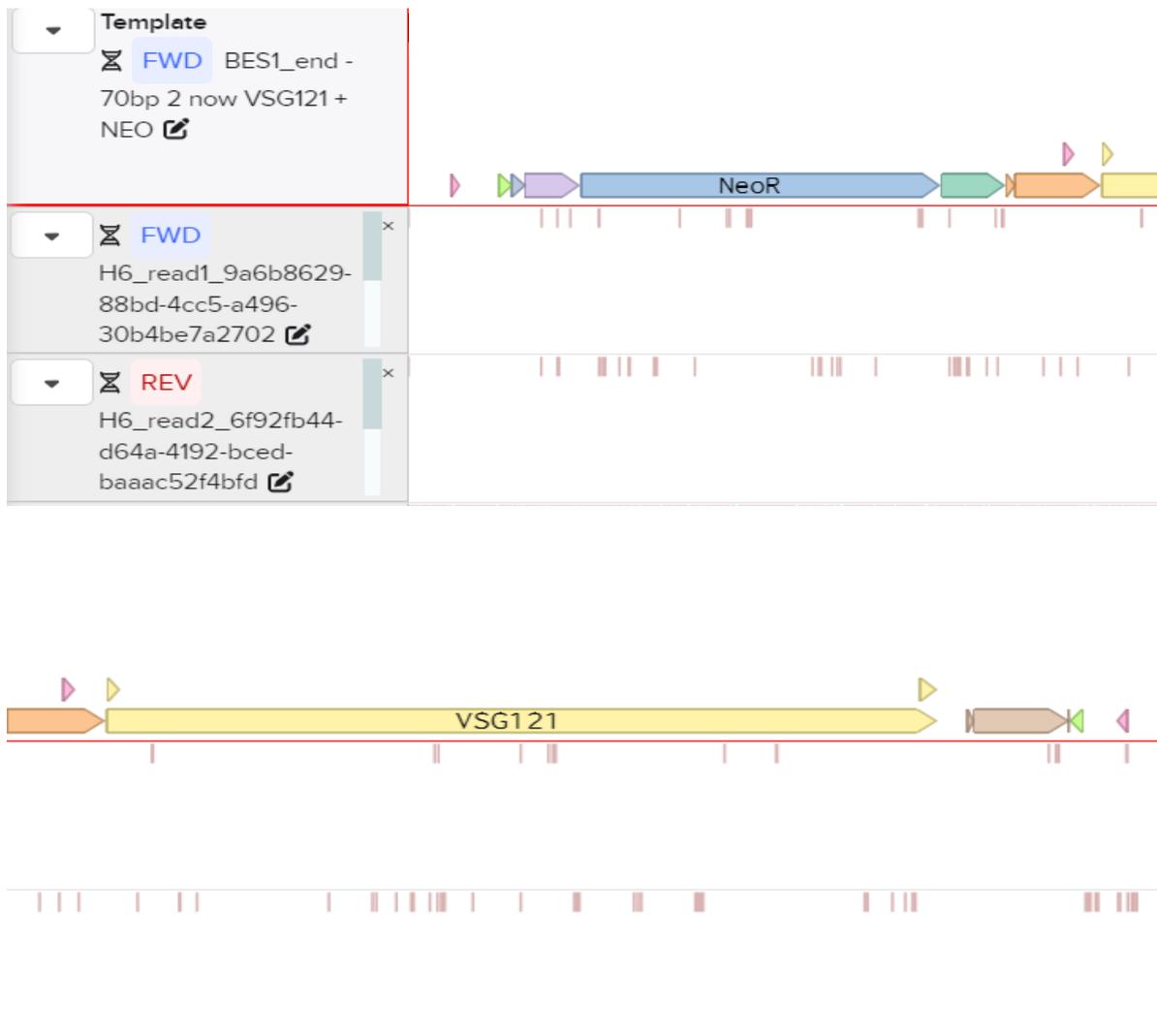
Figure 3-34. Oxford Nanopore sequencing for confirmation of insertion of *VSG121* into 70 bp repeat region 2 in clone 4. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, actin 5' indicated in purple, NEO in blue, aldolase UTR 3' in green, aldolase 5' in orange, *VSG121* in yellow, and PFR2 IGS in brown. Primers used for PCR screening are indicated in pink. The sequences used as donor DNA template in CRISPR/Cas9 are indicated in lime green. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.
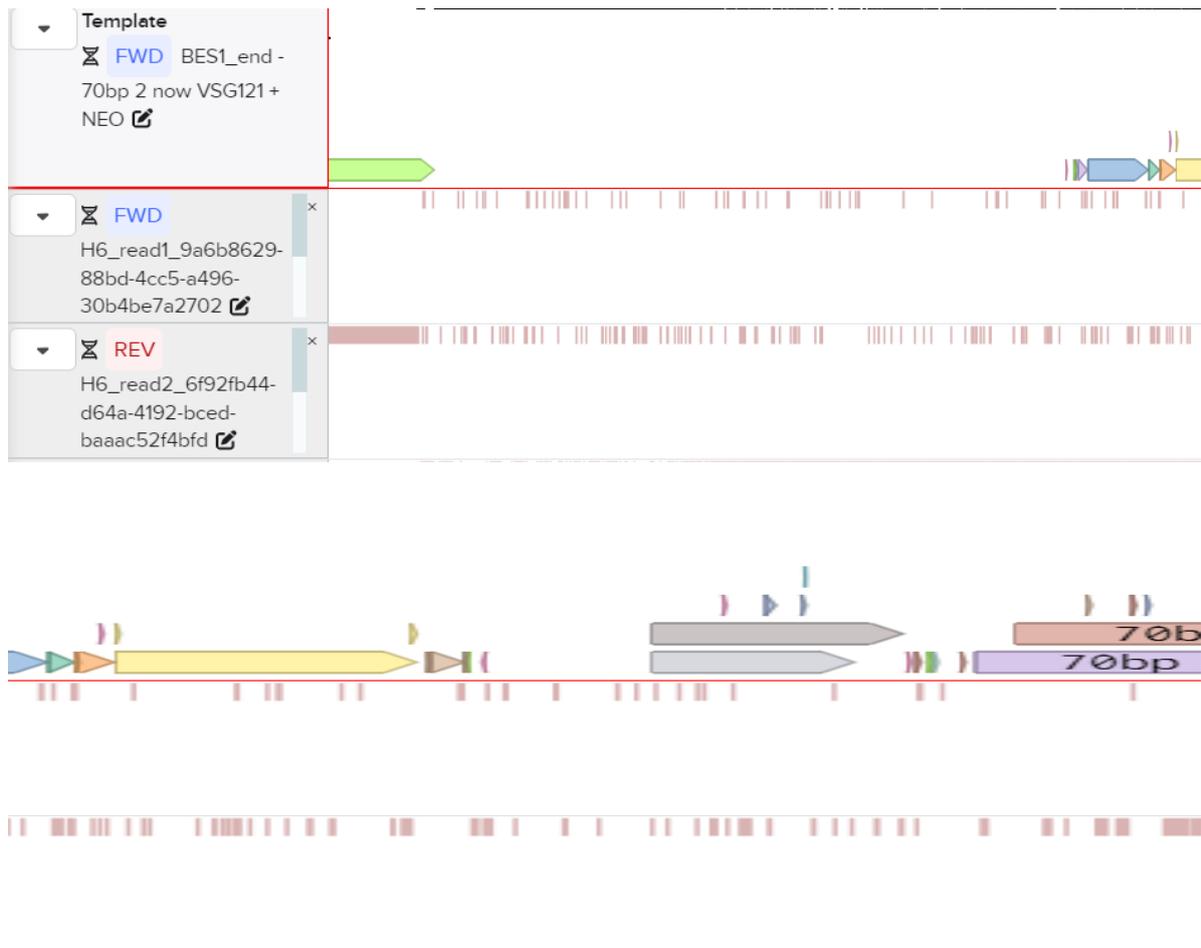
Figure 3-35. Confirmation that *VSG121* replaced the small region of 70 bp repeats in clone 4. Identifies the location of the inserted plasmid in BES1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, ESAG8 is in lime green, NEO in blue, *VSG121* in yellow, *VSG* pseudogene in grey, and the large section of 70 bp repeats in purple and brown. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.

3.4.2.4 Successful insertion of *VSG121* into the promoter region

Nanopore sequencing confirmed successful insertion of *VSG121* and NEO resistance cassette in both clone 1 (Figure 3-36) and clone 4 (Figure 3-38) by comparing individual sequence reads with the annotated sequence map of BES1. Sequencing also confirmed that this insertion successfully removed and replaced the promoter region in clone 1 (Figure 3-37) and clone 4 (Figure 3-39) by comparing individual sequence reads with the annotated sequence map.
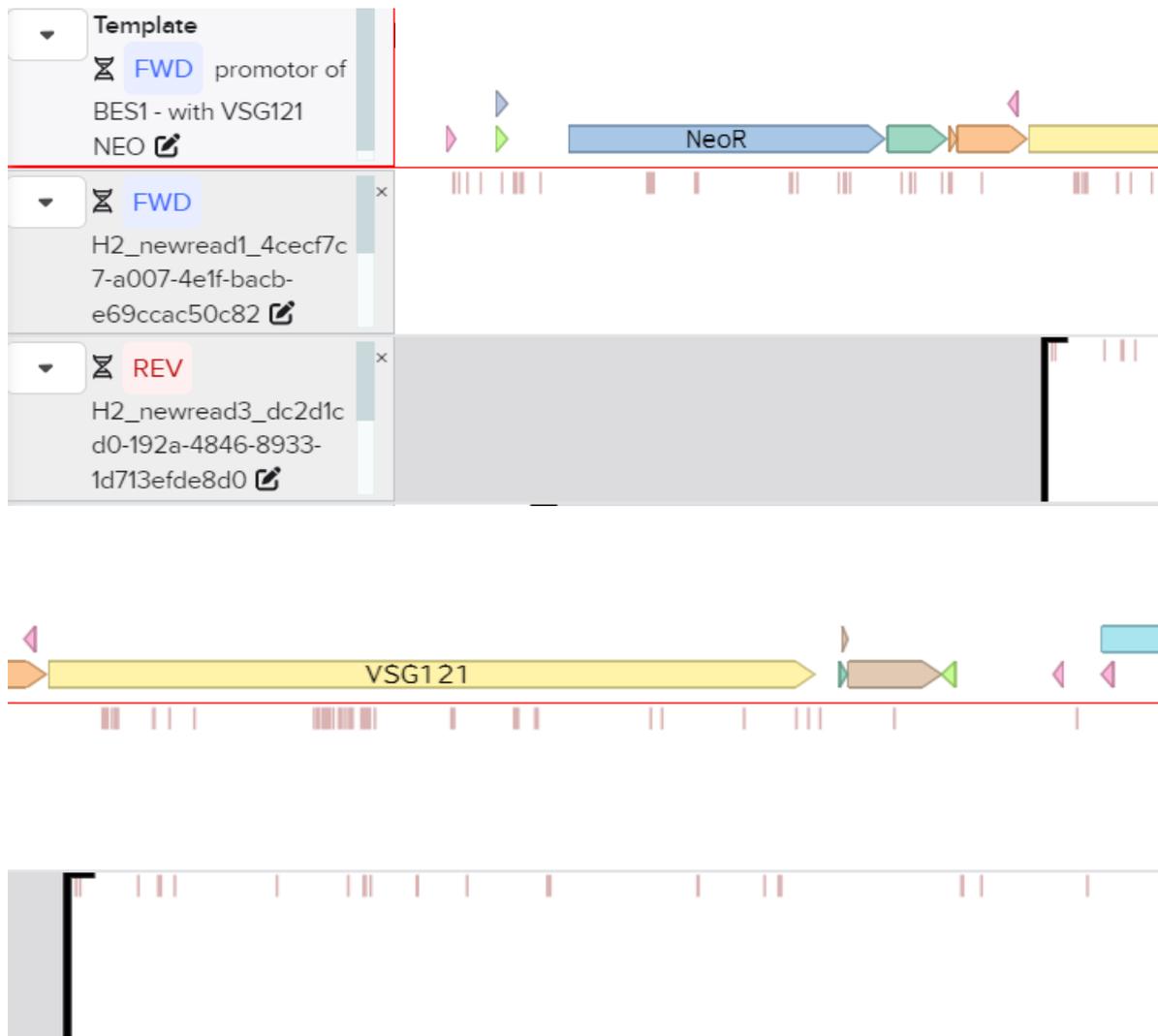
Figure 3-36. Oxford Nanopore sequencing for confirmation of insertion of *VSG121* in clone 1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, NEO in blue, aldolase UTR 3' in green, aldolase 5' in orange, *VSG121* in yellow, PFR2 IGS in brown and the start of ESAG7 in blue. Primers used for PCR screening are indicated in pink. The sequences used as donor DNA template in CRISPR/Cas9 are indicated in lime green. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.
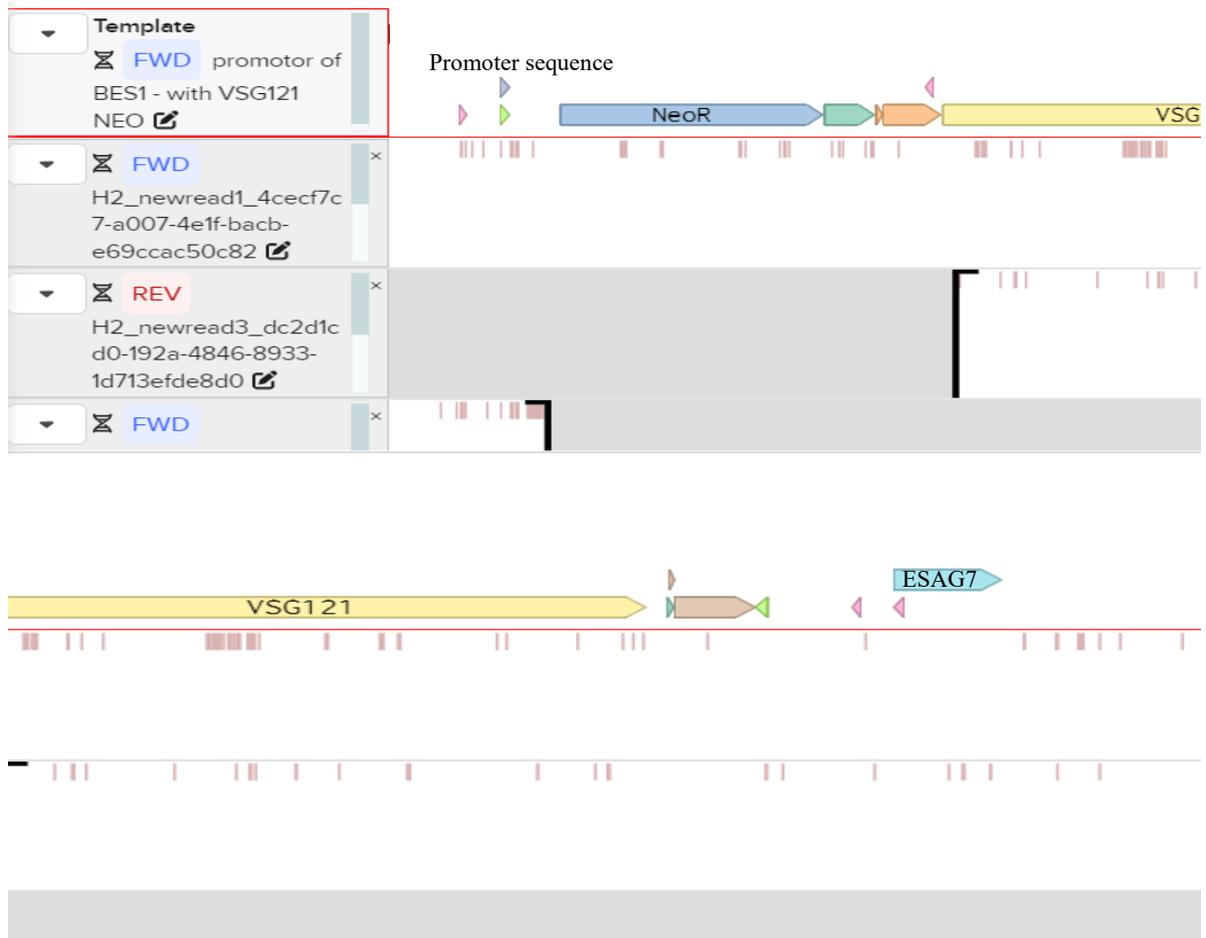
Figure 3-37. Oxford Nanopore sequencing for confirmation of insertion of *VSG121* into the promoter region in clone 1. Identifies the location of the inserted plasmid in BES1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, the sequence of the promoter is indicated in blue, the inserted plasmid with NEO in blue, aldolase UTR 3' in green, aldolase 5' in orange, *VSG121* in yellow, and PFR2 IGS in brown followed by ESAG7 in blue. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.
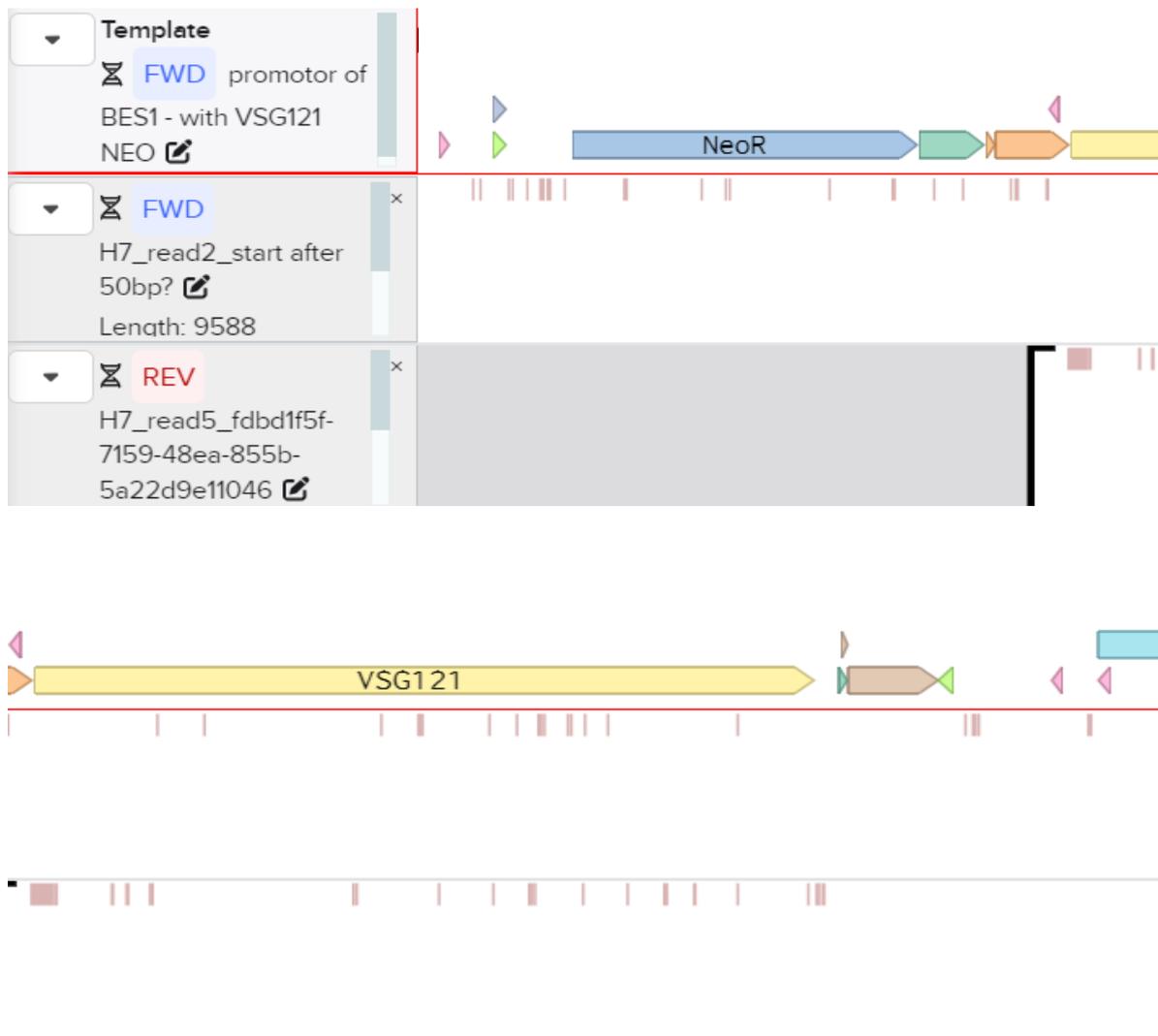
Figure 3-38. Oxford Nanopore sequencing for confirmation of insertion of *VSG121* into the promoter region in clone 4. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, NEO in blue, aldolase UTR 3' in green, aldolase 5' in orange, *VSG121* in yellow, PFR2 IGS in brown and the start of ESAG7 in blue. Primers used for PCR screening are indicated in pink. The sequences used as donor DNA template in CRISPR/Cas9 are indicated in lime green. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.
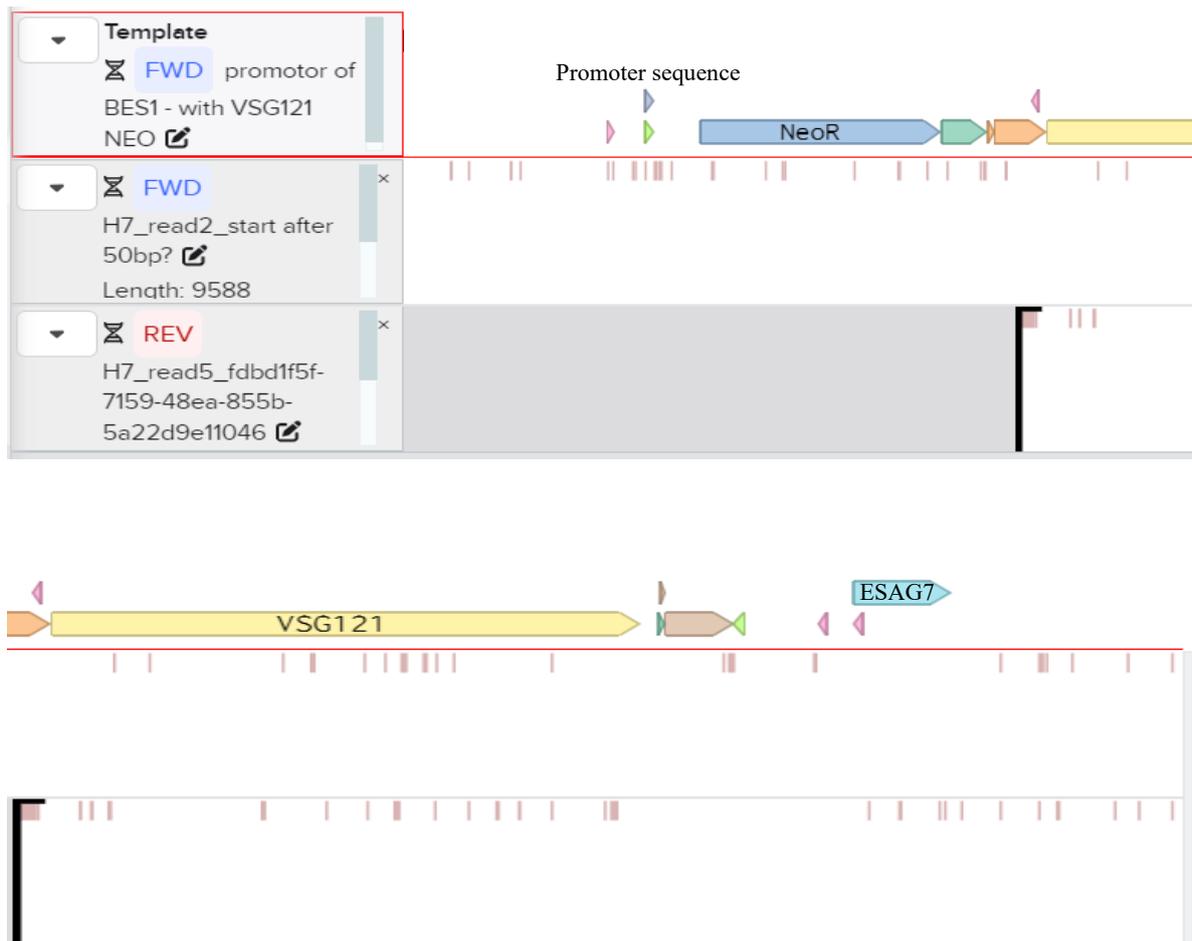
Figure 3-39. Oxford Nanopore sequencing for confirmation of insertion of *VSG121* into the promoter region in clone 4. Identifies the location of the inserted plasmid in BES1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, the sequence of the promoter is indicated in blue, the inserted plasmid with NEO in blue, aldolase UTR 3' in green, aldolase 5' in orange, *VSG121* in yellow, and PFR2 IGS in brown followed by ESAG7 in blue. Pink lines in the sequence indicate mismatches in the sequence, which are individual SNPs in the Nanopore reads, as is expected due to the nature of the long read sequence.

3.4.3 Western blot of γH2A expression

Having used Nanopore sequencing to determine that a number of the BES1 modifications have occurred as expected, despite problems with PCR validation, we next began to ask if the changes affected cell function. First, to determine if any of the cells showed altered levels of nuclear DNA damage, Thr130-phoshorylated histone H2A (γH2A) expression levels were assessed by western blot (Figure 3-40A). γH2A is a marker for nuclear DNA damage and expression of γH2A increases in response to DNA damage as it accumulates at DNA repair foci (Glover and Horn, 2012). EF1α, a transcription elongation factor, was used as a loading control at 49 kDa. The γH2A expression in each cell line was normalised relative to EF1α, and expression in each clone plotted relative to the WT (Figure 3-40). All cells appeared to have an increase in expression of γH2A when compared to the parental TbCas9 (WT) cells (Figure 3-40B). Deletion of the 70 bp repeats deletion 2 resulted in the largest increase in expression of γH2A relative to the WT. Clone 2 had the highest expression at an increase of 4.68, and more than double the expression of clone 7 with an increase of 1.11 in comparison to the WT. Insertion of *VSG121* into the BES1 promoter region in clone 1 was associated with a slight increase in γH2A (0.16) while clone 4 had a higher increase (0.7) in comparison to the WT. *VSG121* insertion into the 70 bp repeats region 2 had an increase in γH2A expression of 1.21 in clone 4 but, due to cell contamination issues, the expression level in clone 3 could not be assessed. A *T. brucei* RAD51 knockout cell line, generated by Girasol et al. (2023a), was used for comparison as RAD51 KO is known to increase DNA damage and, as expected, the RAD51 KO cells had an increase in γH2A expression compared to WT cells (0.35). These data are from a single experiment, and each cell line shows variation, but suggest that modification of the active BES may lead to DNA damage.
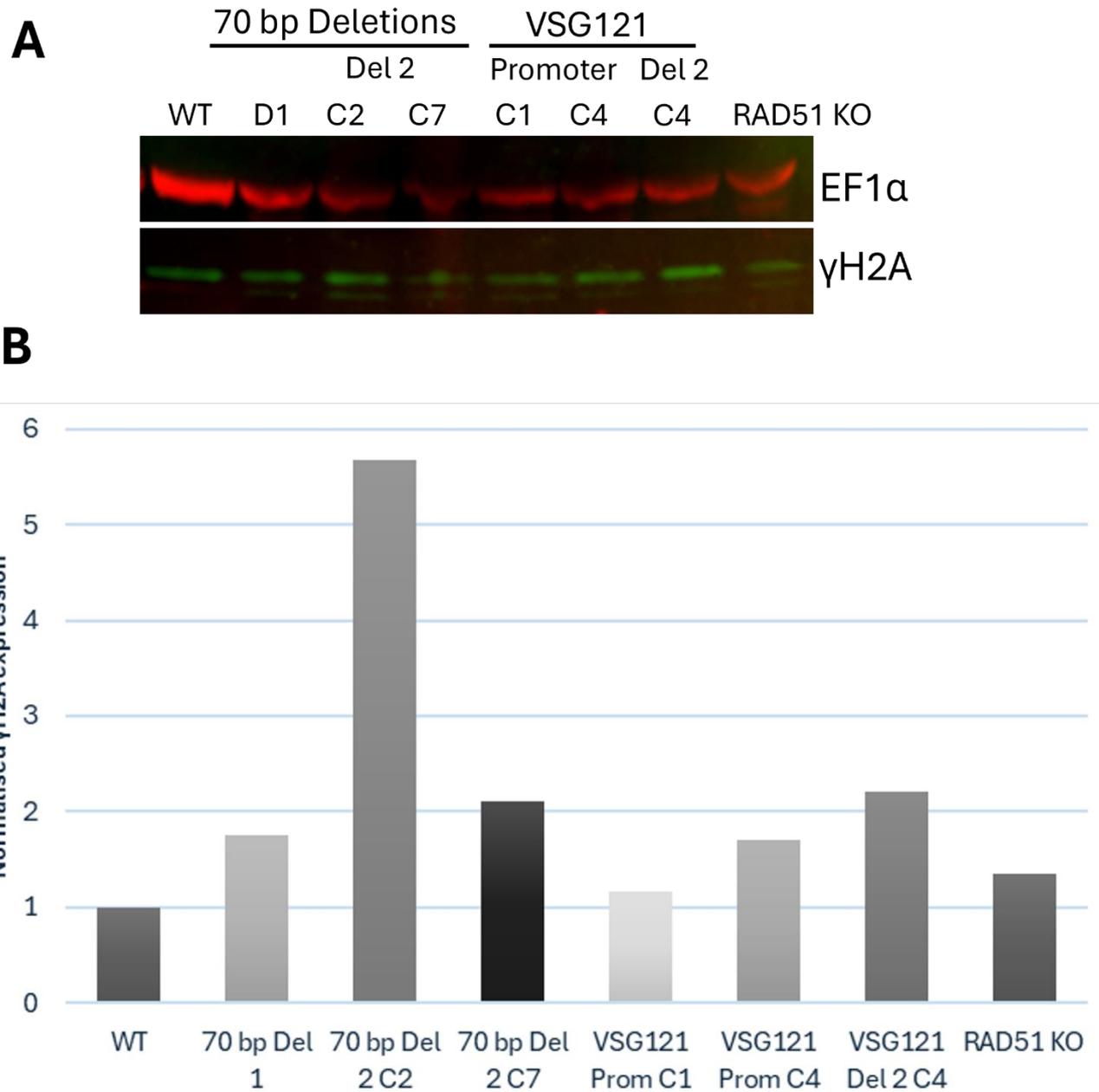
Figure 3-40. Relative γH2A expression levels in generated *T. brucei* cell lines.
A. Western blot of γH2A expression. TbCas9/T7 cells are denoted as WT. EF1α used as a loading control. C = Clone.
B. Normalised γH2A expression level in cell lines compared to the WT.

3.4.4 INDUCE-seq

A total of $2x10^8$ cells from each generated cell line, as well as from TbCas9 WT and a *T. brucei* RAD51 knockout cell line generated by Girasol et al. (2023), were used to attempt to map DNA DSBs using a process termed INDUCE-seq (Dobbs *et al*., 2022). INDUCE-seq is used for mapping DNA DSBs and is considered to be a simpler methodology than BLISS, which was used by Girasol et al (2023) and, hence, was preferred in the short timeframe of this project. The RAD51 KO cell line has had DNA DSBs mapped previously by BLISS and was used as a control to compare DNA DSBs mapping via INDUCE-seq (Girasol *et al.,* 2023).

Only a single attempt at this mapping was possible, given the technical issues in establishing and validating the cells described above. Unfortunately, no sequence was recovered after library preparation.

3.4.5 DRIP-seq

DNA/RNA hybrid immunoprecipitation (DRIP) using the S9.6 antibody followed by high throughput sequencing (DRIP-seq) allows for R-loop mapping in *T. brucei* (Briggs *et al.,* 2018a). A total of $2x10^8$ cells from each generated cell line, and from TbCas9 WT, and a *T. brucei* RAD51 knockout cell line generated by Girasol et al. (2023), were processed for DRIP-seq by disrupting cells and purifying DNA through phenol/chloroform extraction and ethanol precipitation. The RAD51 KO cell line has had R-loops mapped previously using DRIP-seq and was used as a control to compare R-loop mapping (Girasol *et al.,* 2023). After DNA extraction, each sample was digested with BamHI, NcoII, ApaLI, and PvuII followed by a phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation and resuspending in 100 μl ultrapure TE buffer. Unfortunately, low levels of DNA were recovered after chromatin preparation.

Only a single attempt at this mapping was possible, given the technical issues in establishing and validating the cells described above, and the time constraints of this project. Due to the low levels of input DNA, libraries failed to be generated, and sequencing was not possible.

4. Discussion and Future Work

This study focused on the development of *T. brucei* cell lines with alterations in the active bloodstream expression site (BES) to allow investigation of the role of differing structures within the BES on the formation of double strand breaks and R-loops, which we hypothesis are needed to initiate VSG switching. Two structures were investigated, based on the results of Girasol et al. (2023b), the 70 bp repeats regions and the expressed *VSG*. Two regions of 70 bp repeats were successfully removed and a *VSG* sequence relocated to two other locations within the active BES, BES1. Transformants with the 70 bp repeats removed from BES1 were successfully generated in two deletions with one deletion consisting of removal of a large section of repeats immediately upstream of the native *VSG* and the second deletion, a smaller section of repeats further upstream from the *VSG*. A distinct *VSG*, *VSG121*, was successfully inserted into the promoter region of BES1 and was also used to replace the smaller region of 70 bp repeats in BES1.

The project suffered from a number of unexpected technical challenges. Cloning of VSGs proved highly problematic, perhaps because they cause problems for bacterial replication. Using CRISPR-Cas9 to manipulate the BES was not as efficient as expected, with many transformation failures and antibiotic-resistant cells recovered that proved incorrect; the reasons for this are unknown, but it is possible that Cas9-induced DSBs are a problem for cell survival. DSB repair in the BES may not be as effective as hoped, and targeting only the active BES could be an issue due to there being ~15 BESs in *T. brucei* Lister 427 (Hertz-Fowler *et al.,* 2008). Finally, PCR to validate recovered clones was a consistent problem, which may also stem from the similarity between BESs, since Nanopore sequencing showed many clones that could not be confirmed by PCR were correct. Despite these problems, *T. brucei* cell lines with modifications to the active BES1 were generated.

Two individual sections of the 70 bp repeats in BES1 could be removed, but not both together. The reason for being unable to remove all sections of the 70 bp repeats is unclear, and it may simply be due to VSG expression being impaired as the changes to the BES are made. This may have been a problem in removing the larger section of repeats, which are in close proximity of *VSG221* as this region immediately upstream of the *VSG*, known as the co-transposed region, has been found to be essential for *T. brucei* and so manipulations here may have caused issues (Davies *et al.,* 1997). However, this same region was able to be modified to remove the larger, downstream section of repeats only, and in the clone recovered, the upstream smaller region of

repeats had expanded. The expansion of this section of 70 bp repeats may be the reason the successful transfection for this deletion took 14 days for cells to be visible under light microscopy, as there were more changes occurring in the BES than had been directed, with perhaps the expansion of the repeats taking more time. Analysis of the expanded 70 bp repeats identified that it was not the large section of repeats relocated, and not several copies of the smaller section of 70 bp repeats together. Due to only slight variations in the repetitive sequence these variations were unable to be mapped elsewhere in the genome using sequence comparison tools. The slow recovery of transformants may mirror a study by Hover-Minor et al (2016) that found growth problems before and after DSB induction in cells with deletion of 70 bp repeats. The results in this study, reveal it is possible that the 70 bp repeats play so far unexplored and essential roles in *T. brucei* biology. If so, previous studies in which the 70 bp repeats were reported to be removed (McCulloch *et al*, 1997; Boothroyd *et al*, 2009; Hovel-Minor *et al*, 2016) may need to be reassessed as in each study, deletion of the 70 bp repeats was tested by PCR and Southern blot, and not by Nanopore sequencing.

Multiple attempts to repeat the knockout of the large section of the 70 bp repeats with the same methodology have failed, with no antibiotic-resistant cells surviving (Jane Munday – personal communication). Current work is ongoing to generate an inducible deletion line, using a cre-lox system (Duncan *et al.,* 2021) and use this to induce deletion of the large section of 70 bp repeats in both a WT cell line and another modified *T. brucei* cell line, with the smaller section of 70 bp repeats artificially expanded, to identify what cells can survive. This would provide further insight into the length requirements for 70 bp repeats in an active BES.

A distinct *VSG*, *VSG121*, was successfully added to two locations in the BES: downstream of the promoter and replacing the small region of 70 bp repeats. It was not possible to insert a *VSG* copy, of either *VSG121* or *VSG221,* into the middle of the BES by replacing the third copy of the expression site-associated gene 8 (ESAG8) in the active BES. The active BES, BES1, contains three copies of ESAG8 interspersed with ESAG4 and ESAG3 (Hoek and Cross, 2001). The genome of *T. brucei* Lister 427 from TriTrypDB (http://tritrypdb.org) was used to design primers targeting the third copy of ESAG8. There is minimal sequence variation amongst copies of ESAG8 genes, and after comparing the sequence of the primers used against the annotated Oxford Nanopore sequence, it was identified that the primers were targeting all ESAG8 copies (Hoek *et al.,* 2000; Jane Munday – personal communication). The forward primer was not generating specific sgRNA and was targeting the first copy of ESAG8 and so, was trying to replace a large section of the BES, all ESAG8 copies and all genes in between

them, which is essential for *T. brucei* survival. Copies of ESAG8 have been successfully deleted from the actively transcribed ES by Hoek and Cross (2001), targeting the upstream copy and the tandem duplication of ESAG8. The majority of ESAG8 is nucleolar and has been suggested to be involved in regulating RNA polymerase I and hence, regulating transcription of the active ES (Hoek *et al.,* 2002). ESAG8 also interacts with PUF1, an RNA regulatory protein, in the cytoplasm and analysis of structural features from the amino acid sequence of ESAG8 suggest it may be a cell cycle regulator (Hoek *et al.,* 2002; Pays *et al.,* 2001). However, the ES is functional without ESAG8 suggesting it may not be essential, and that with specific primers, this is still a valid target for future work (Hoek and Cross, 2001). Primers to generate sgRNA and donor DNA have been redesigned to only target the third copy of ESAG8 as this appears to have specific, different UTRs, which can be targeted and in future work, insertion of the VSG copy will be attempted using these (Jane Munday – personal communication).

This study failed to relocate the endogenous *VSG*, *VSG221,* into other locations within the active BES but vectors, pPOTv7 BSD and pPOTv7 HYG, with *VSG221* inserted were successfully generated. Transfections using the vectors failed to relocate *VSG221* to any of the three selected regions within the active BES, within the time available. There were cell growth and cell contamination issues in the laboratory during the time transfections to relocate the *VSG* were carried out, and this may have impacted on the success of transfections for insertion of *VSG221*. However, the plasmid generated was functional, as it has subsequently been possible to insert *VSG221* using these vectors into both the promoter region of BES1 and the smaller region of 70 bp repeats, region 2 (Jane Munday – personal communication). Generated cell lines will be used in future work, to compare to cell lines with the relocation of a distinct *VSG*, *VSG121*, and to map DNA breaks and R-loops.

All generated cell lines had an increase in γH2A expression suggesting these alterations may result in an accumulation of nuclear DNA damage compared to the WT. However, due to variations in γH2A expression between clones of the same alterations, this data needs to be repeated for confirmation of results. Further tests, including growth rate and cell cycle measurements, would have been beneficial to this study but were could notbe completed due to time constraints.

Unfortunately, INDUCE-seq failed to map DNA breaks and no sequencing results were received. Due to time constraints and the large number of cells required, INDUCE-seq could not be repeated for this project. For future work, the results of INDUCE-seq will be discussed

with the sequencing team at the Glasgow Shared Research Facility to investigate the failure in INDUCE-seq. Illumina sequencing was used which would sequence DNA with P5 adaptors attached to indicate where the DNA breaks were, and one possibility for the failure is that the P5 adaptors failed to anneal to the DNA hence no sequencing results. INDUCE-seq was the chosen method for mapping of DNA breaks as it involves less processing of the native break sites than the alternative, breaks labelling *in situ* and sequencing (BLISS), which has previously been used in the McCulloch laboratory (Girasol *et al.,* 2023b). If INDUCE-seq fails to work when repeated, BLISS can be used to map the DNA breaks of the generated cell lines.

DRIP-seq failed to progress to the mapping of R-loops due to the low levels of DNA recovered from each cell line after chromatin preparation. Due to time constraints and the large number of cells required, DRIP-seq could not be repeated for this project; but will be attempted in future.

In conclusion, multiple *T. brucei* cell lines were generated with modifications to the active BES – removal of two sections of 70 bp repeats from the active BES and relocation of a distinct *VSG* (*VSG121*) into other locations in the active BES (Figure 4-1). In future work, these cell lines will allow for comparison of DRIP-seq and INDUCE-seq/BLISS signal with the natural organisation of the active BES and so, give an insight into whether the 70 bp repeats and/or VSG sequence drive R-loop and DNA DSB formation and if necessary for break formation linked to VSG switching.
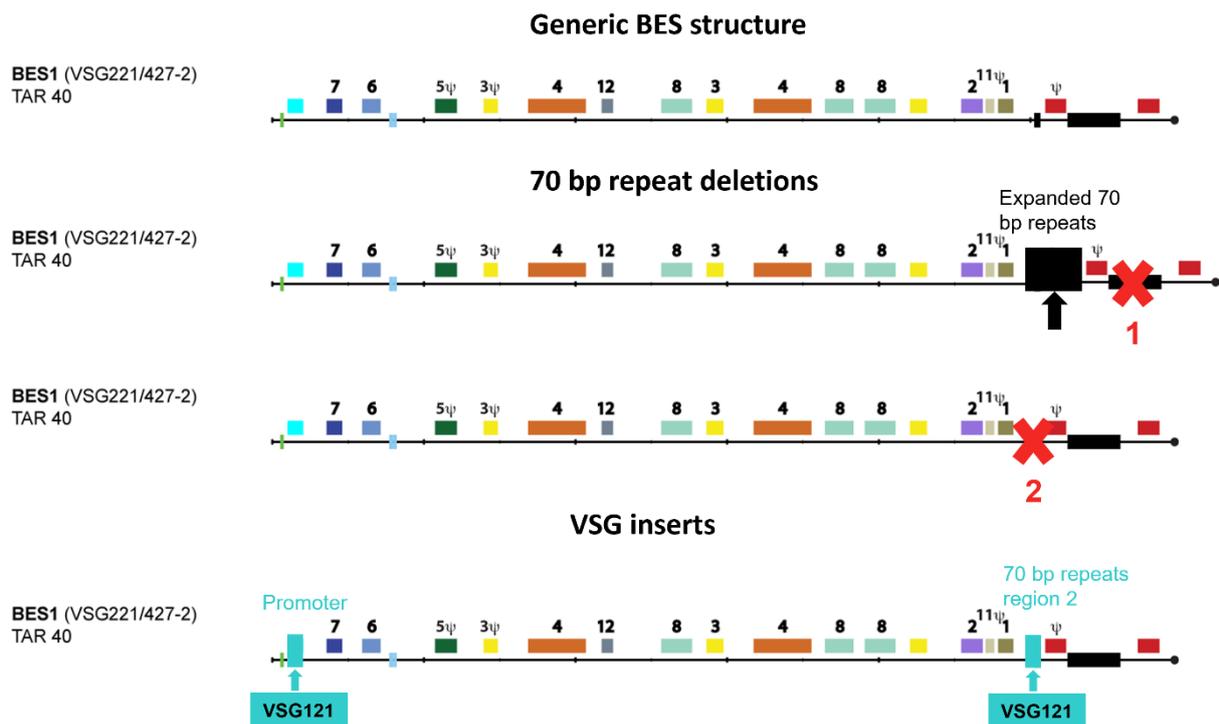
Figure 4-1. Schematic of *T. brucei* cell lines with alterations to BES1 generated in this study. (Adapted from Hertz-Fowler *et al.,* 2008).

References

Ackerson, S. M., Romney, C., Schuck, P. L. and Stewart, J. A. (2021) 'To Join or Not to Join: Decision Points Along the Pathway to Double-Strand Break Repair vs. Chromosome End Protection', Front Cell Dev Biol, 9, pp. 708763.

Aitcheson, N., Talbot, S., Shapiro, J., Hughes, K., Adkin, C., Butt, T., Sheader, K. and Rudenko, G. (2005) 'VSG switching in *Trypanosoma brucei*: antigenic variation analysed using RNAi in the absence of immune selection', Molecular Microbiology, 57(6), pp. 1608-1622.

Álvarez-Rodríguez, A., Jin, B. K., Radwanska, M. and Magez, S. (2022) 'Recent progress in diagnosis and treatment of Human African Trypanosomiasis has made the elimination of this disease a realistic target by 2030', Front Med (Lausanne), 9, pp. 1037094.

Amin, D. N., Ngoyi, D. M., Nhkwachi, G. M., Palomba, M., Rottenberg, M., Büscher, P., Kristensson, K. and Masocha, W. (2010) 'Identification of stage biomarkers for human African trypanosomiasis', Am J Trop Med Hyg, 82(6), pp. 983-90.Barnes, R. L. and McCulloch, R. (2007) '*Trypanosoma brucei* homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases', Nucleic Acids Res, 35(10), pp. 3478-93.

Barry, J. D. (1997) 'The relative significance of mechanisms of antigenic variation in African trypanosomes', Parasitology Today, 13(6), pp. 212-218.

Barry, J. D., Hall, J. P. J. and Plenderleith, L. (2012) 'Genome hyperevolution and the success of a parasite', Annals of the New York Academy of Sciences, 1267(1), pp. 11-17.

Barry, J. D. and McCulloch, R. (2001) 'Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite', Adv Parasitol, 49, pp. 1-70.

Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J. and Gluenz, E. (2017) 'A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids', R Soc Open Sci, 4(5), pp. 170095.

Berriman, M. and Ghedin, E. and Hertz-Fowler, C. and Blandin, G. and Renauld, H. and Bartholomeu, D. C. and Lennard, N. J. and Caler, E. and Hamlin, N. E. and Haas, B. and Böhme, U. and Hannick, L. and Aslett, M. A. and Shallom, J. and Marcello, L. and Hou, L. and Wickstead, B. and Alsmark, U. C. and Arrowsmith, C. and Atkin, R. J. and Barron, A. J. and Bringaud, F. and Brooks, K. and Carrington, M. and Cherevach, I. and Chillingworth, T. J. and Churcher, C. and Clark, L. N. and Corton, C. H. and Cronin, A. and Davies, R. M. and Doggett,

J. and Djikeng, A. and Feldblyum, T. and Field, M. C. and Fraser, A. and Goodhead, I. and Hance, Z. and Harper, D. and Harris, B. R. and Hauser, H. and Hostetler, J. and Ivens, A. and Jagels, K. and Johnson, D. and Johnson, J. and Jones, K. and Kerhornou, A. X. and Koo, H. and Larke, N. and Landfear, S. and Larkin, C. and Leech, V. and Line, A. and Lord, A. and Macleod, A. and Mooney, P. J. and Moule, S. and Martin, D. M. and Morgan, G. W. and Mungall, K. and Norbertczak, H. and Ormond, D. and Pai, G. and Peacock, C. S. and Peterson, J. and Quail, M. A. and Rabbinowitsch, E. and Rajandream, M. A. and Reitter, C. and Salzberg, S. L. and Sanders, M. and Schobel, S. and Sharp, S. and Simmonds, M. and Simpson, A. J. and Tallon, L. and Turner, C. M. and Tait, A. and Tivey, A. R. and Van Aken, S. and Walker, D. and Wanless, D. and Wang, S. and White, B. and White, O. and Whitehead, S. and Woodward, J. and Wortman, J. and Adams, M. D. and Embley, T. M. and Gull, K. and Ullu, E. and Barry, J. D. and Fairlamb, A. H. and Opperdoes, F. and Barrell, B. G. and Donelson, J. E. and Hall, N. and Fraser, C. M. and Melville, S. E. and El-Sayed, N. M. (2005) 'The genome of the African trypanosome *Trypanosoma brucei*', Science, 309(5733), pp. 416-22.

Bentivoglio, M. and Bertini, G. (2018) 'Alive and Ticking: *Trypanosoma brucei* Assaults the Circadian Clocks', Trends in Parasitology, 34(4), pp. 265-267.

Bitter, W., Gerrits, H., Kieft, R. and Borst, P. (1998) 'The role of transferrin-receptor variation in the host range of *Trypanosoma brucei*', Nature, 391(6666), pp. 499-502.

Bonnet, J., Boudot, C. and Courtioux, B. (2015) 'Overview of the Diagnostic Methods Used in the Field for Human African Trypanosomiasis: What Could Change in the Next Years?', Biomed Res Int, 2015, pp. 583262.

Boothroyd, C. E., Dreesen, O., Leonova, T., Ly, K. I., Figueiredo, L. M., Cross, G. A. and Papavasiliou, F. N. (2009) 'A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*', Nature, 459(7244), pp. 278-81.

Borst, P., Bitter, W., Blundell, P., Chaves, I., Cross, M., Gerrits, H., Fred, McCulloch, R., Taylor, M., and Rudenko, G. (1998) 'Control of VSG gene expression sites in *Trypanosoma brucei*', Molecular and Biochemical Parasitology, 91(1), pp. 67-76.

Brambati, A., Barry, R. M. and Sfeir, A. (2020) 'DNA polymerase theta (Polθ) - an error-prone polymerase necessary for genome stability', Curr Opin Genet Dev, 60, pp. 119-126.

Brickman, M. J. and Balber, A. E. (1994) '*Trypanosoma brucei* brucei and T. b. gambiense: stumpy bloodstream forms express more CB1 epitope in endosomes and lysosomes than slender forms', J Eukaryot Microbiol, 41(6), pp. 533-6.

Briggs, E., Crouch, K., Lemgruber, L., Hamilton, G., Lapsley, C. and McCulloch, R. (2019) '*Trypanosoma brucei* ribonuclease H2A is an essential R-loop processing enzyme whose loss causes DNA damage during transcription initiation and antigenic variation', Nucleic Acids Research, 47(17), pp. 9180-9197.

Briggs, E., Crouch, K., Lemgruber, L., Lapsley, C. and McCulloch, R. (2018b) 'Ribonuclease H1-targeted R-loops in surface antigen gene expression sites can direct trypanosome immune evasion', PLoS Genet, 14(12), pp. e1007729.

Briggs, E., Hamilton, G., Crouch, K., Lapsley, C. and McCulloch, R. (2018a) 'Genome-wide mapping reveals conserved and diverged R-loop activities in the unusual genetic landscape of the African trypanosome genome', Nucleic Acids Research, 46(22), pp. 11789-11805.

Briggs, E. M., Rojas, F., McCulloch, R., Matthews, K. R. and Otto, T. D. (2021) 'Single-cell transcriptomic analysis of bloodstream *Trypanosoma brucei* reconstructs cell cycle progression and developmental quorum sensing', Nature Communications, 12(1).

Burri, C. F. C., Reto Brun (2014) '45 - Human African Trypanosomiasis', pp. 606-621.e4.

Burton, P., McBride, D. J., Wilkes, J. M., Barry, J. D. and McCulloch, R. (2007) 'Ku Heterodimer-Independent End Joining in *Trypanosoma brucei* Cell Extracts Relies upon Sequence Microhomology', Eukaryotic Cell, 6(10), pp. 1773-1781.

Capewell, P., Cren-Travaillé, C., Marchesi, F., Johnston, P., Clucas, C., Benson, R. A., Gorman, T. A., Calvo-Alvarez, E., Crouzols, A., Jouvion, G., Jamonneau, V., Weir, W., Stevenson, M. L., O'Neill, K., Cooper, A., Swar, N. K., Bucheton, B., Ngoyi, D. M., Garside, P., Rotureau, B. and MacLeod, A. (2016) 'The skin is a significant but overlooked anatomical reservoir for vector-borne African trypanosomes', Elife, 5.

CDC. (2019) African Trypanosomiasis. Available at: https://www.cdc.gov/dpdx/trypanosomiasisafrican/index.html (Accessed: 03/08/2024 2024).

Cestari, I. and Stuart, K. (2018) 'Transcriptional Regulation of Telomeric Expression Sites and Antigenic Variation in Trypanosomes', Curr Genomics, 19(2), pp. 119-132.

Chattopadhyay, A., Jones, N. G., Nietlispach, D., Nielsen, P. R., Voorheis, H. P., Mott, H. R. and Carrington, M. (2005) 'Structure of the C-terminal Domain from *Trypanosoma brucei* Variant Surface Glycoprotein MITat1.2 *', Journal of Biological Chemistry, 280(8), pp. 7228-7235.

Claes, F., Vodnala, S. K., van Reet, N., Boucher, N., Lunden-Miguel, H., Baltz, T., Goddeeris, B. M., Büscher, P. and Rottenberg, M. E. (2009) 'Bioluminescent imaging of *Trypanosoma brucei* shows preferential testis dissemination which may hamper drug efficacy in sleeping sickness', PLoS Negl Trop Dis, 3(7), pp. e486.

Conway, C., Proudfoot, C., Burton, P., Barry, J. D. and McCulloch, R. (2002) 'Two pathways of homologous recombination in *Trypanosoma brucei*', Mol Microbiol, 45(6), pp. 1687-700.

Cosentino, R. O., Brink, B. G. and Siegel, T. N. (2021) 'Allele-specific assembly of a eukaryotic genome corrects apparent frameshifts and reveals a lack of nonsense-mediated mRNA decay', NAR Genom Bioinform, 3(3), pp. lqab082.

Cross, G. A., Kim, H. S. and Wickstead, B. (2014) 'Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427', Mol Biochem Parasitol, 195(1), pp. 59-73.

da Silva, M. (2021) 'DNA Double-Strand Breaks: A Double-Edged Sword for Trypanosomatids', Front Cell Dev Biol, 9, pp. 669041.

Davies, K. P., Carruthers, V. B. and Cross, G. A. (1997) 'Manipulation of the vsg co-transposed region increases expression-site switching in *Trypanosoma brucei*', Mol Biochem Parasitol, 86(2), pp. 163-77.

Dean, S., Sunter, J., Wheeler, R. J., Hodkinson, I., Gluenz, E. and Gull, K. (2015) 'A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids', Open Biol, 5(1), pp. 140197.

Devlin, R., Marques, C. A., Paape, D., Prorocic, M., Zurita-Leal, A. C., Campbell, S. J., Lapsley, C., Dickens, N. and McCulloch, R. (2016) 'Mapping replication dynamics in \textit *Trypanosoma brucei* reveals a link with telomere transcription and antigenic variation', eLife, 5, pp. e12765.

Desquesnes, M., Gonzatti, M., Sazmand, A., Thévenon, S., Bossard, G., Boulangé, A., Gimonneau, G., Truc, P., Herder, S., Ravel, S., Sereno, D., Jamonneau, V., Jittapalapong, S.,

Jacquiet, P., Solano, P. and Berthier, D. (2022) 'A review on the diagnosis of animal trypanosomoses', Parasit Vectors, 15(1), pp. 64.

Dewar, C. E., MacGregor, P., Cooper, S., Gould, M. K., Matthews, K. R., Savill, N. J. and Schnaufer, A. (2018) 'Mitochondrial DNA is critical for longevity and metabolism of transmission stage *Trypanosoma brucei*', PLoS Pathog, 14(7), pp. e1007195.

Dobbs, F. M., van Eijk, P., Fellows, M. D., Loiacono, L. and Nitsch, R., Reed, Simon H. (2022) 'Precision digital mapping of endogenous and induced genomic DNA breaks by INDUCE-seq', Nature Communications, 13(1).

Duncan, S. M., Nagar, R., Damerow, M., Yashunsky, D. V., Buzzi, B., Nikolaev, A. V. and Ferguson, M. A. J. (2021) 'A *Trypanosoma brucei* β3 glycosyltransferase superfamily gene encodes a β1-6 GlcNAc-transferase mediating N-glycan and GPI anchor modification', J Biol Chem, 297(4), pp. 101153.

Faria, J., Glover, L., Hutchinson, S., Boehm, C., Field, M. C. and Horn, D. (2019) 'Monoallelic expression and epigenetic inheritance sustained by a *Trypanosoma brucei* variant surface glycoprotein exclusion complex', Nat Commun, 10(1), pp. 3023.

Faria, J., Briggs, E., Black, J., and McCulloch, R. (2022) 'Emergence and adaptation of the cellular machinery directing antigenic variation in the African trypanosome', Current Opinion in Microbiology, 70, pp. 102209.

Ginger, M. L., Blundell, P. A., Lewis, A. M., Browitt, A., Günzl, A. and Barry, J. D. (2002) 'Ex vivo and in vitro identification of a consensus promoter for VSG genes expressed by metacyclic-stage trypanosomes in the tsetse fly', Eukaryot Cell, 1(6), pp. 1000-9.

Girasol, M. J., Briggs, E. M., Marques, C. A., Batista, J. M., Beraldi, D., Burchmore, R., Lemgruber, L. and McCulloch, R. (2023a) 'Immunoprecipitation of RNA–DNA hybrid interacting proteins in *Trypanosoma brucei* reveals conserved and novel activities, including in the control of surface antigen expression needed for immune evasion by antigenic variation', Nucleic Acids Research, 51(20), pp. 11123-11141.

Girasol, M. J., Krasilnikova, M., Marques, C. A., Damasceno, J. D., Lapsley, C., Lemgruber, L., Burchmore, R., Beraldi, D., Carruthers, R., Briggs, E. M. and McCulloch, R. (2023b) 'RAD51-mediated R-loop formation acts to repair transcription-associated DNA breaks driving

antigenic variation in *Trypanosoma brucei*', Proc Natl Acad Sci U S A, 120(48), pp. e2309306120.

Glover, L., Alsford, S. and Horn, D. (2013) 'DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes', PLoS Pathog, 9(3), pp. e1003260.

Glover, L., Hutchinson, S., Alsford, S. and Horn, D. (2016) 'VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes', Proc Natl Acad Sci U S A, 113(26), pp. 7225-30.

Glover, L., and Horn, D. (2012) 'Trypanosomal histone γH2A and the DNA damage response', Molecular and Biochemical Parasitology, 183(1), pp. 78-83.

Glover, L., Jun, J. and Horn, D. (2011) 'Microhomology-mediated deletion and gene conversion in African trypanosomes', Nucleic Acids Research, 39(4).

Glover, L., McCulloch, R. and Horn, D. (2008) 'Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes', Nucleic Acids Res, 36(8), pp. 2608-18.

Graham, S. V., Matthews, K. R., Shiels, P. G. and Barry, J. D. (1990) 'Distinct, developmental stage-specific activation mechanisms of trypanosome VSG genes', Parasitology, 101 Pt 3, pp. 361-7.

Graham, S. V., Terry, S. and Barry, J. D. (1999) 'A structural and transcription pattern for variant surface glycoprotein gene expression sites used in metacyclic stage *Trypanosoma brucei*', Mol Biochem Parasitol, 103(2), pp. 141-54.

Hall, J. P., Wang, H. and Barry, J. D. (2013) 'Mosaic VSGs and the scale of *Trypanosoma brucei* antigenic variation', PLoS Pathog, 9(7), pp. e1003502.

Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I., Heath, P., Kartvelishvili, M., Mungall, K., Harris, D., Hauser, H., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Taylor, J. E., Walker, D., White, B., Young, R., Cross, G. A., Rudenko, G., Barry, J. D., Louis, E. J. and Berriman, M. (2008) 'Telomeric expression sites are highly conserved in *Trypanosoma brucei*', PLoS One, 3(10), pp. e3527.

Hoek, M. and Cross, G. A. (2001) 'Expression-site-associated-gene-8 (ESAG8) is not required for regulation of the VSG expression site in *Trypanosoma brucei*', Mol Biochem Parasitol, 117(2), pp. 211-5.

Hoek, M., Engstler, M. and Cross, G. A. M. (2000) 'Expression-site-associated gene 8 (ESAG8) of *Trypanosoma brucei* is apparently essential and accumulates in the nucleolus', Journal of Cell Science, 113(22), pp. 3959-3968.

Hoek, M., Zanders, T., and Cross, G. A. M (2002) '*Trypanosoma brucei* expression-site-associated-gene-8 protein interacts with a Pumilio family protein', Molecular and Biochemical Parasitology, 120(2), pp. 269-283.

Hollingshead, CM., and Bermudez., R. (2024) 'Human African Trypanosomiasis (Sleeping Sickness)', StatPearls [Internet].

Horn, D. (2014) 'Antigenic variation in African trypanosomes', Mol Biochem Parasitol, 195(2), pp. 123-9.

Hovel-Miner, G., Mugnier, M. R., Goldwater, B., Cross, G. A. and Papavasiliou, F. N. (2016) 'A Conserved DNA Repeat Promotes Selection of a Diverse Repertoire of *Trypanosoma brucei* Surface Antigens from the Genomic Archive', PLoS Genet, 12(5), pp. e1005994.

Hughes, K., Wand, M., Foulston, L., Young, R., Harley, K., Terry, S., Ersfeld, K. and Rudenko, G. (2007) 'A novel ISWI is involved in VSG expression site downregulation in African trypanosomes', Embo j, 26(9), pp. 2400-10.

Hutchinson, S., Foulon, S., Crouzols, A., Menafra, R., Rotureau, B., Griffiths, A. D. and Bastin, P. (2021) 'The establishment of variant surface glycoprotein monoallelic expression revealed by single-cell RNA-seq of *Trypanosoma brucei* in the tsetse fly salivary glands', PLoS Pathog, 17(9), pp. e1009904.

Kelso AA, W. S., Luthman AJ and Sehorn MG (2017) 'Homologous Recombination in Protozoan Parasites and Recombinase Inhibitors', Frontiers in Microbiology, 8, pp. 1716.

Kim, H. S. and Cross, G. A. (2010) 'TOPO3alpha influences antigenic variation by monitoring expression-site-associated VSG switching in *Trypanosoma brucei*', PLoS Pathog, 6(7), pp. e1000992.

Kim, H. S. and Cross, G. A. (2011) 'Identification of *Trypanosoma brucei* RMI1/BLAP75 homologue and its roles in antigenic variation', PLoS One, 6(9), pp. e25313.

Kuepfer, I., Hhary, E. P., Allan, M., Edielu, A., Burri, C. and Blum, J. A. (2011) 'Clinical presentation of T.b. rhodesiense sleeping sickness in second stage patients from Tanzania and Uganda', PLoS Negl Trop Dis, 5(3), pp. e968.

Kuo, L. J. and Yang, L. X. (2008) 'Gamma-H2AX - a novel biomarker for DNA double-strand breaks', In Vivo, 22(3), pp. 305-9.

Leal, A. Z., Schwebs, M., Briggs, E., Weisert, N., Reis, H., Lemgruber, L., Luko, K., Wilkes, J., Butter, F., McCulloch, R. and Janzen, C. J. (2020) 'Genome maintenance functions of a putative *Trypanosoma brucei* translesion DNA polymerase include telomere association and a role in antigenic variation', Nucleic Acids Research, 48(17), pp. 9660-9680.

Li, B. (2015) 'DNA double-strand breaks and telomeres play important roles in *Trypanosoma brucei* antigenic variation', Eukaryot Cell, 14(3), pp. 196-205.

Li, B. and Zhao, Y. (2021) 'Regulation of Antigenic Variation by *Trypanosoma brucei* Telomere Proteins Depends on Their Unique DNA Binding Activities', Pathogens, 10(8), pp. 967.

Lieber, M. R. (2010) 'The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway', Annu Rev Biochem, 79, pp. 181-211.

López-Escobar, L., Hänisch, B., Halliday, C., Ishii, M., Akiyoshi, B., Dean, S., Sunter, J. D., Wheeler, R. J. and Gull, K. (2022) 'Stage-specific transcription activator ESB1 regulates monoallelic antigen expression in *Trypanosoma brucei*', Nature Microbiology, 7(8), pp. 1280-1290.

Marcello, L. and Barry, J. D. (2007) 'Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure', Genome Res, 17(9), pp. 1344-52.

Melville, S. E., Leech, V., Gerrard, C. S., Tait, A. and Blackwell, J. M. (1998) 'The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers', Mol Biochem Parasitol, 94(2), pp. 155-73.

Oethinger, M. and Campbell, S. M. (2010) 'Chapter 3 - Infection and Host Response', pp. 25-43.

Marin, P. A., da Silva, M. S., Pavani, R. S., Machado, C. R. and Elias, M. C. (2018) 'Recruitment kinetics of the homologous recombination pathway in procyclic forms of *Trypanosoma brucei* after ionizing radiation treatment', Sci Rep, 8(1), pp. 5405.

Matthews, K. R. (2005) 'The developmental cell biology of *Trypanosoma brucei*', J Cell Sci, 118(Pt 2), pp. 283-90.

McCulloch, R. and Barry, J. D. (1999) 'A role for RAD51 and homologous recombination in *Trypanosoma brucei* antigenic variation', Genes Dev, 13(21), pp. 2875-88.

McCulloch, R. and Field, M. C. (2015) 'Quantitative sequencing confirms VSG diversity as central to immune evasion by *Trypanosoma brucei*', Trends Parasitol, 31(8), pp. 346-9.

McCulloch, R., Rudenko, G. and Borst, P. (1997) 'Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences', Mol Cell Biol, 17(2), pp. 833-43.

McCulloch, R., G. R. and Piet, B. (1997) 'Gene Conversions Mediating Antigenic Variation in *Trypanosoma brucei* Can Occur in Variant Surface Glycoprotein Expression Sites Lacking 70-Base-Pair Repeat Sequences', Molecular and Cellular Biology, 17(2), pp. 833--843.

McVey, M. and Lee, S. E. (2008) 'MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings', Trends Genet, 24(11), pp. 529-38.

Mehnert, A.-K., Prorocic, M., Dujeancourt-Henry, A., Hutchinson, S., McCulloch, R. and Glover, L. (2020) 'RAD50 promotes DNA repair by homologous recombination and restrains antigenic variation in African trypanosomes', bioRxiv, pp. 2020.03.17.994905.

Mehnert, A.-K., Prorocic, M., Dujeancourt-Henry, A., Hutchinson, S., McCulloch, R. and Glover, L. (2021) 'The MRN complex promotes DNA repair by homologous recombination and restrains antigenic variation in African trypanosomes', Nucleic Acids Research, 49(3), pp. 1436-1454.

Moreno, C. J. G., Temporão, A., Torres, T. and Sousa Silva, M. (2019) '*Trypanosoma brucei* Interaction with Host: Mechanism of VSG Release as Target for Drug Discovery for African Trypanosomiasis', Int J Mol Sci, 20(6).

Morrison, L. J., Marcello, L. and McCulloch, R. (2009) 'Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity', Cellular Microbiology, 11(12), pp. 1724-1734.

Mugnier, M. R., Cross, G. A. and Papavasiliou, F. N. (2015) 'The in vivo dynamics of antigenic variation in *Trypanosoma brucei*', Science, 347(6229), pp. 1470-3.

Mugnier, M. R., Stebbins, C. E. and Papavasiliou, F. N. (2016) 'Masters of Disguise: Antigenic Variation and the VSG Coat in *Trypanosoma brucei*', PLoS Pathog, 12(9), pp. e1005784.

Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T. D., Saliba, A. E., Sebra, R. P. and Siegel, T. N. (2018) 'Genome organization and DNA accessibility control antigenic variation in trypanosomes', Nature, 563(7729), pp. 121-125.

Nanavaty, V., Sandhu, R., Jehi, S. E., Pandya, U. M. and Li, B. (2017) '*Trypanosoma brucei* RAP1 maintains telomere and subtelomere integrity by suppressing TERRA and telomeric RNA:DNA hybrids', Nucleic Acids Research, 45(10), pp. 5785-5796.

Navarro, M., Cross, G. A. M. and Wirtz, E. (1999) '*Trypanosoma brucei* variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites', The EMBO Journal, 18(8), pp. 2265-2272.

Nolan, D. P., Rolin, S., Rodriguez, J. R., Van Den Abbeele, J. and Pays, E. (2000) 'Slender and stumpy bloodstream forms of *Trypanosoma brucei* display a differential response to extracellular acidic and proteolytic stress', Eur J Biochem, 267(1), pp. 18-27.

Ooi, C. P. and Bastin, P. (2013) 'More than meets the eye: understanding *Trypanosoma brucei* morphology in the tsetse', Front Cell Infect Microbiol, 3, pp. 71.

Ooi, C. P., Smith, T. K., Gluenz, E., Wand, N. V., Vaughan, S. and Rudenko, G. (2018) 'Blocking variant surface glycoprotein synthesis alters endoplasmic reticulum exit sites/Golgi homeostasis in *Trypanosoma brucei*', Traffic, 19(6), pp. 391-405.

Passos-Silva, D. G., Rajão, M. A., Nascimento de Aguiar, P. H., Vieira-da-Rocha, J. P., Machado, C. R. and Furtado, C. (2010) 'Overview of DNA Repair in Trypanosoma cruzi, *Trypanosoma brucei*, and Leishmania major', J Nucleic Acids, 2010, pp. 840768.

Pays, E., Lips, S., Nolan, D., Vanhamme, L., and Pérez-Morga, D. (2001) 'The VSG expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts', Molecular and Biochemical Parasitology, 114(1), pp. 1-16.

Ponte-Sucre, A. (2016) 'An Overview of *Trypanosoma brucei* Infections: An Intense Host-Parasite Interaction', Front Microbiol, 7, pp. 2126.

Proudfoot, C. and McCulloch, R. (2005) 'Distinct roles for two RAD51-related genes in *Trypanosoma brucei* antigenic variation', Nucleic Acids Res, 33(21), pp. 6906-19.

Ramey-Butler, K., Ullu, E., Kolev, N. G. and Tschudi, C. (2015) 'Synchronous expression of individual metacyclic variant surface glycoprotein genes in *Trypanosoma brucei*', Mol Biochem Parasitol, 200(1-2), pp. 1-4.

Reuter, C., Hauf, L., Imdahl, F., Sen, R., Vafadarnejad, E., Fey, P., Finger, T., Jones, N. G., Walles, H., Barquist, L., Saliba, A.-E., Groeber-Becker, F. and Engstler, M. (2023) 'Vector-borne *Trypanosoma brucei* parasites develop in artificial human skin and persist as skin tissue forms', Nature Communications, 14(1).

Rijo-Ferreira, F., Carvalho, T., Afonso, C., Sanches-Vaz, M., Costa, R. M., Figueiredo, L. M. and Takahashi, J. S. (2018) 'Sleeping sickness is a circadian disorder', Nat Commun, 9(1), pp. 62.

Robinson, N., McCulloch, R., Conway, C., Browitt, A., and Barry, J. (2002). 'Inactivation of Mre11 Does Not Affect VSG Gene Duplication Mediated by Homologous Recombination in *Trypanosoma brucei*' Journal of Biological Chemistry, 277(29), pp. 26185-93.

San Filippo, J., Sung, P. and Klein, H. (2008) 'Mechanism of eukaryotic homologous recombination', Annu Rev Biochem, 77, pp. 229-57.

Schuster, S., Lisack, J., Subota, I., Zimmermann, H., Reuter, C., Mueller, T., Morriswood, B. and Engstler, M. (2021) 'Unexpected plasticity in the life cycle of *Trypanosoma brucei*', Elife, 10.

Schwede, A., Jones, N., Engstler, M. and Carrington, M. (2011) 'The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes', Mol Biochem Parasitol, 175(2), pp. 201-4.

Sfeir, A., Tijsterman, M. and McVey, M. (2024) 'Microhomology-Mediated End Joining Chronicles: Tracing the Evolutionary Footprints of Genome Protection', Annu Rev Cell Dev Biol.

Shapiro, S. Z., Naessens, J., Liesegang, B., Moloo, S. K. and Magondu, J. (1984) 'Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes', Acta Trop, 41(4), pp. 313-23.

Sheader, K., Vaughan, S., Minchin, J., Hughes, K., Gull, K. and Rudenko, G. (2005) 'Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes', Proc Natl Acad Sci U S A, 102(24), pp. 8716-21.

Silva Pereira, S., Jackson, A. and Figueiredo, L (2022) 'Evolution of the variant surface glycoprotein family in African trypanosomes', Trends in Parasitology, 38(1), pp. 23-36.

Silvester, E., McWilliam, K. R. and Matthews, K. R. (2017) 'The Cytological Events and Molecular Control of Life Cycle Development of *Trypanosoma brucei* in the Mammalian Bloodstream', Pathogens, 6(3).

Sima, N., Dujeancourt-Henry, A., Perlaza, B. L., Ungeheuer, M. N., Rotureau, B. and Glover, L. (2022) 'SHERLOCK4HAT: A CRISPR-based tool kit for diagnosis of Human African Trypanosomiasis', EBioMedicine, 85, pp. 104308.

Steverding, D. (2008) 'The history of African trypanosomiasis', Parasites & Vectors, 1(1).

Stockdale, C., Swiderski, M. R., Barry, J. D. and McCulloch, R. (2008) 'Antigenic variation in *Trypanosoma brucei*: joining the DOTs', PLoS Biol, 6(7), pp. e185.

Tihon, E., Rubio-Peña, K., Dujeancourt-Henry, A., Crouzols, A., Rotureau, B. and Glover, L. (2022) 'VEX1 Influences mVSG Expression During the Transition to Mammalian Infectivity in *Trypanosoma brucei*', Front Cell Dev Biol, 10, pp. 851475.

Trenaman, A., Hartley, C., Prorocic, M., Passos-Silva, D. G., van den Hoek, M., Nechyporuk-Zloy, V., Machado, C. R. and McCulloch, R. (2013) '*Trypanosoma brucei* BRCA2 acts in a life cycle-specific genome stability process and dictates BRC repeat number-dependent RAD51 subnuclear dynamics', Nucleic Acids Res, 41(2), pp. 943-60.

Trindade, S., Rijo-Ferreira, F., Carvalho, T., Pinto-Neves, D., Guegan, F., Aresta-Branco, F., Bento, F., Young, S. A., Pinto, A., Van Den Abbeele, J., Ribeiro, R. M., Dias, S., Smith, T. K. and Figueiredo, L. M. (2016) '*Trypanosoma brucei* Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice', Cell Host Microbe, 19(6), pp. 837-48.

van der Heijden, T., Modesti, M., Hage, S., Kanaar, R., Wyman, C. and Dekker, C. (2008) 'Homologous recombination in real time: DNA strand exchange by RecA', Mol Cell, 30(4), pp. 530-8.

Vanhollebeke, B., Uzureau, P., Monteyne, D., Pérez-Morga, D. and Pays, E. (2010) 'Cellular and molecular remodeling of the endocytic pathway during differentiation of *Trypanosoma brucei* bloodstream forms', Eukaryot Cell, 9(8), pp. 1272-82.

Vickerman, K. (1965) 'Polymorphism and mitochondrial activity in sleeping sickness trypanosomes', Nature, 208(5012), pp. 762-6.

Vink, C., Rudenko, G. and Seifert, H. S. (2012) 'Microbial antigenic variation mediated by homologous DNA recombination', FEMS Microbiology Reviews, 36(5), pp. 917-948.

Wang, Q.-P., Kawahara, T. and Horn, D. (2010) 'Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes', Molecular Microbiology, 77(5), pp. 1237-1245.

WHO (2023). Trypanosomiasis, human African (sleeping sickness). Available at: https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness) [Accessed 01/09/2024]

Wold, M. S. (1997) 'Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism', Annu Rev Biochem, 66, pp. 61-92.

Yang, H., Li, Q., Fan, J., Holloman, W. K. and Pavletich, N. P. (2005) 'The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction', Nature, 433(7026), pp. 653-7.

Yang, X., Figueiredo, L. M., Espinal, A., Okubo, E. and Li, B. (2009) 'RAP1 is essential for silencing telomeric variant surface glycoprotein genes in *Trypanosoma brucei*', Cell, 137(1), pp. 99-109.

Appendix

Appendix 1. Oligonucleotide sequences used for CRISPR/Cas9 gene editing. Primers used to amplify donor DNA and sgRNA.

#	Target	Sense	Sequence (5' -> 3')	Purpose
HS1	70 bp repeat deletion 1	R	GAAATTAATACGACTCACTATAGGGAGAGTAATCGTGAGTATAAGTTTTAGAGCTAGAAATAGC	3' sgRNA amplification
HS2		F	GAAATTAATACGACTCACTATAGGTTTGTATGATAGTGGTCTGTGTTTTAGAGCTAGAAATAGC	5' sgRNA amplification
HS9		F	TACAAAAATTCGCAATTCTTGAATATTTTGGTATAATGCAGACCTGCTGC	5' donor amplification
HS10		R	CACAACCATAACGCTAGTGTTTGAATCTGCCCGGAACCACTACCAGAACC	3' donor amplification
HS3	70 bp repeat deletion 2	R	GAAATTAATACGACTCACTATAGGTAAAGGCGGAGCCCAAATAGGTTTTAGAGCTAGAAATAGC	3' sgRNA amplification
HS4		F	GAAATTAATACGACTCACTATAGGCGTTGATGAAAAGACTGTAAGTTTTAGAGCTAGAAATAGC	5' sgRNA amplification
HS7		F	GACTGTTGCTTGTATTGCCTCTCCCCTCTCGTATAATGCAGACCTGCTGC	5' donor amplification
HS8		R	AGCTGCGCATATACAATATCACTTTCTATACCGGAACCACTACCAGAACC	3' donor amplification
HS5	70 bp repeat deletion 3	R	GAAATTAATACGACTCACTATAGGGAGAGTAATCGTGAGTATAAGTTTTAGAGCTAGAAATAGC	3' sgRNA amplification
HS6		F	GAAATTAATACGACTCACTATAGGCGTTGATGAAAAGACTGTAAGTTTTAGAGCTAGAAATAGC	5' sgRNA amplification
HS11		F	GACTGTTGCTTGTATTGCCTCTCCCCTCTCGTATAATGCAGACCTGCTGC	5' donor amplification
HS12		R	CACAACCATAACGCTAGTGTTTGAATCTGCCCGGAACCACTACCAGAACC	3' donor amplification
HS30	ESAG8	F	GAAATTAATACGACTCACTATAGGTAGCACGTATGGGATGTGCGGTTTTAGAGCTAGAAATAGC	5' sgRNA amplification
HS31		R	GAAATTAATACGACTCACTATAGGCGTCGCAACACGCTTCGGTTGTTTTAGAGCTAGAAATAGC	3' sgRNA amplification
HS32		F	TGTTGTCCCTTTGGTGCGATGTGTGCTTCGGTATAATGCAGACCTGCTGC	5' donor amplification
HS33		R	ATCTCTACTACCATTATCCAAATATCTGCAATATCAAGTTGCACAACAGT	3' donor amplification
HS34	Promoter	F	GAAATTAATACGACTCACTATAGGGGGTTAGGAAAAACGTGAAGGTTTTAGAGCTAGAAATAGC	5' sgRNA amplification
HS35		R	GAAATTAATACGACTCACTATAGGTTCTTGTGGCATAGTCATGTGTTTTAGAGCTAGAAATAGC	3' sgRNA amplification
HS36		F	CGAAGAGCAGGGGTGCAACGGAAGAGTGAAGTATAATGCAGACCTGCTGC	5' donor amplification
HS37		R	TAAAGCTCTAATATTTGTAGAAAATATGAGATATCAAGTTGCACAACAGT	3' donor amplification
HS38	70bp repeat deletion 2	R	AGCTGCGCATATACAATATCACTTTCTATAATATCAAGTTGCACAACAGT	3' donor amplification

Appendix 2. Oligonucleotides used for confirmation of CRISPR/Cas9 clones. Primers used for PCR confirmation and Sanger sequencing.

#	Target	Sense	Sequence (5' -> 3')	Purpose
HS18	70 bp repeat deletion 1	F	GCTGCATTTGTGAGTCTGTC	Confirmation of 70 bp deletion 1 knockout and Sanger sequencing
HS19		R	TCCAATTCAGATCCTTCTGTGC	
HS20	70 bp repeat deletion 2	F	TCTTGTGAGGTTCTGGCAA	Confirmation of 70 bp deletion 2 knockout
HS21		R	TCGCATCCCAACTCGACTAC	
GG27	Blasticidin	F	GCATGATATCATGGCCAAGCCTTTGTCT	Confirmation of blasticidin resistance
GG28		R	GCATCCATGGTTAGCCCTCCCACACATA	
M106	Neomycin	F	CGACGAGATCTATGATTGAACAAGATGGATT	Confirmation of neomycin resistance
M107		R	CGACGGTCGACTCAGAAGAACTCGTCAAGAA	
HS45	ESAG8	F	GTCGTTACTGGAGTGTTGTATCT	Confirmation of VSG insertion at ESAG8
HS46		R	CGAGGTTAGAATTATCCACATGT	
HS43	Promoter	F	CGTTGTCGGTTCATGCTTGACT	Confirmation of VSG insertion at promoter
HS44		R	CAGCACAAACCAAAACTTCATCAT	

Appendix 3. Oligonucleotide sequences used for plasmid generation. Primers used to amplify *VSG* fragments, confirm plasmid constructs, and for Sanger sequencing.

#	Target	Sense	Sequence (5' -> 3')	Purpose
HS24	*VSG221*	F	GGTACCATGCCTTCCAATCAGGAGGCCCG	5' amplification of *VSG221* fragment
HS25		R	GAGCTCTTAAAAAAGCAAAACTGCAAGCCAAAGAGGGGTCTTGC	3' amplification of *VSG221* fragment
HS26	*VSG121*	F	GGTACCATGGCCGTGCACAGAGCCCTAGC	5' amplification of *VSG121* fragment
HS27		R	GAGCTCTTAAAAAAGCAAGGCCACAAATGCAGCAGAAAC	3' amplification of *VSG121* fragment
M13F	pGEM®-T Easy vector	F	TGTAAAACGACGGCCAGT	Confirmation of correct plasmid construct and Sanger sequencing
M13R		R	GAGCGGATAACAATTTCACACAGG	
HS28	pPOTv7 plasmid	F	GCGATAGCTTAGCTATCAGCCA	Confirmation of pPOTv7 insert and Sanger sequencing
HS29		R	GCACTAATGAACCAGATGTCA	
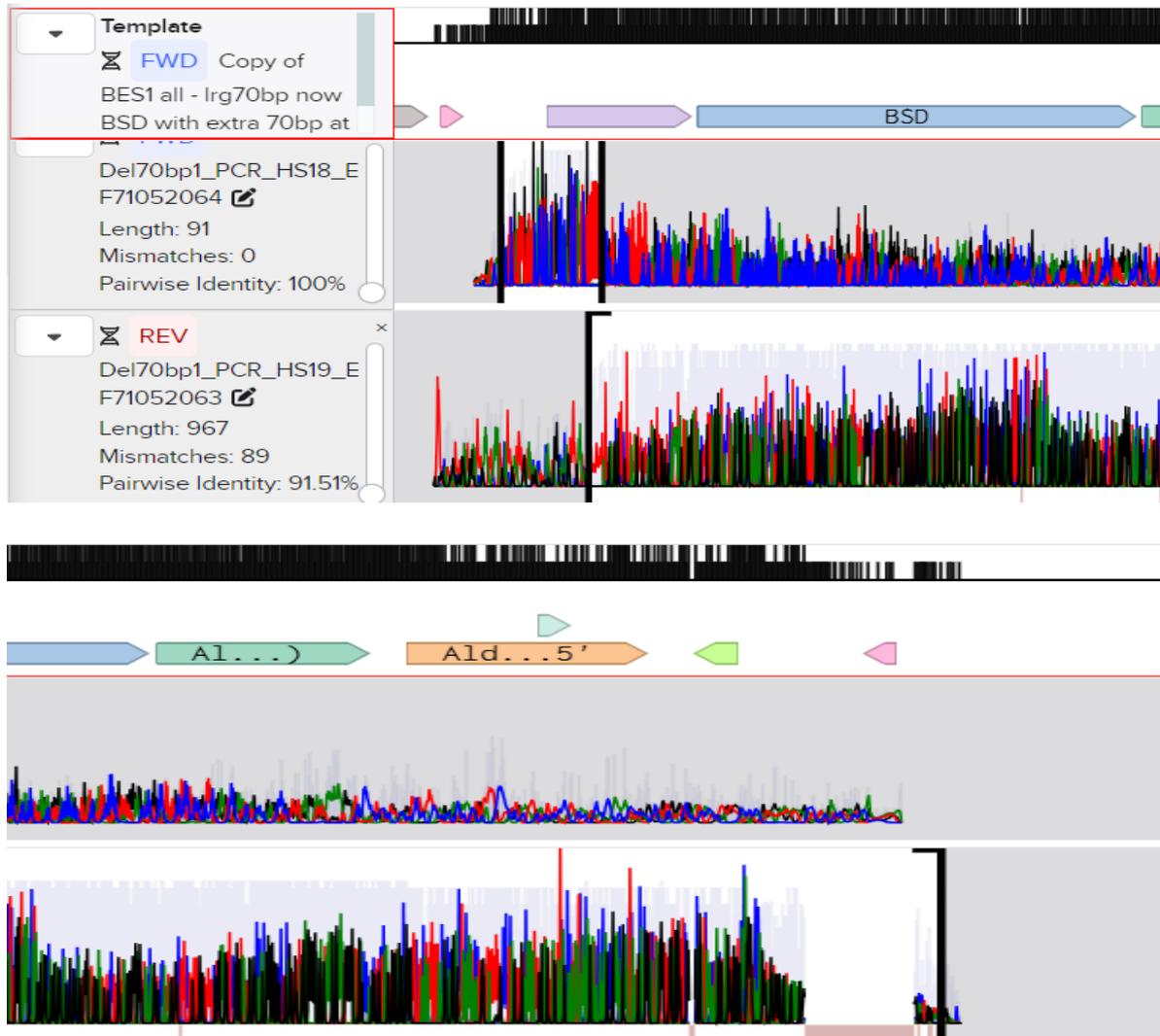
Appendix 4. Sanger sequencing confirmation of 70 bp repeat deletion 1. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, BSD is indicated in blue, actin in purple, aldolase UTR 3' in green, and aldolase 5' in orange. Primers used for sequencing are indicated in pink with donor DNA templates in lime green.
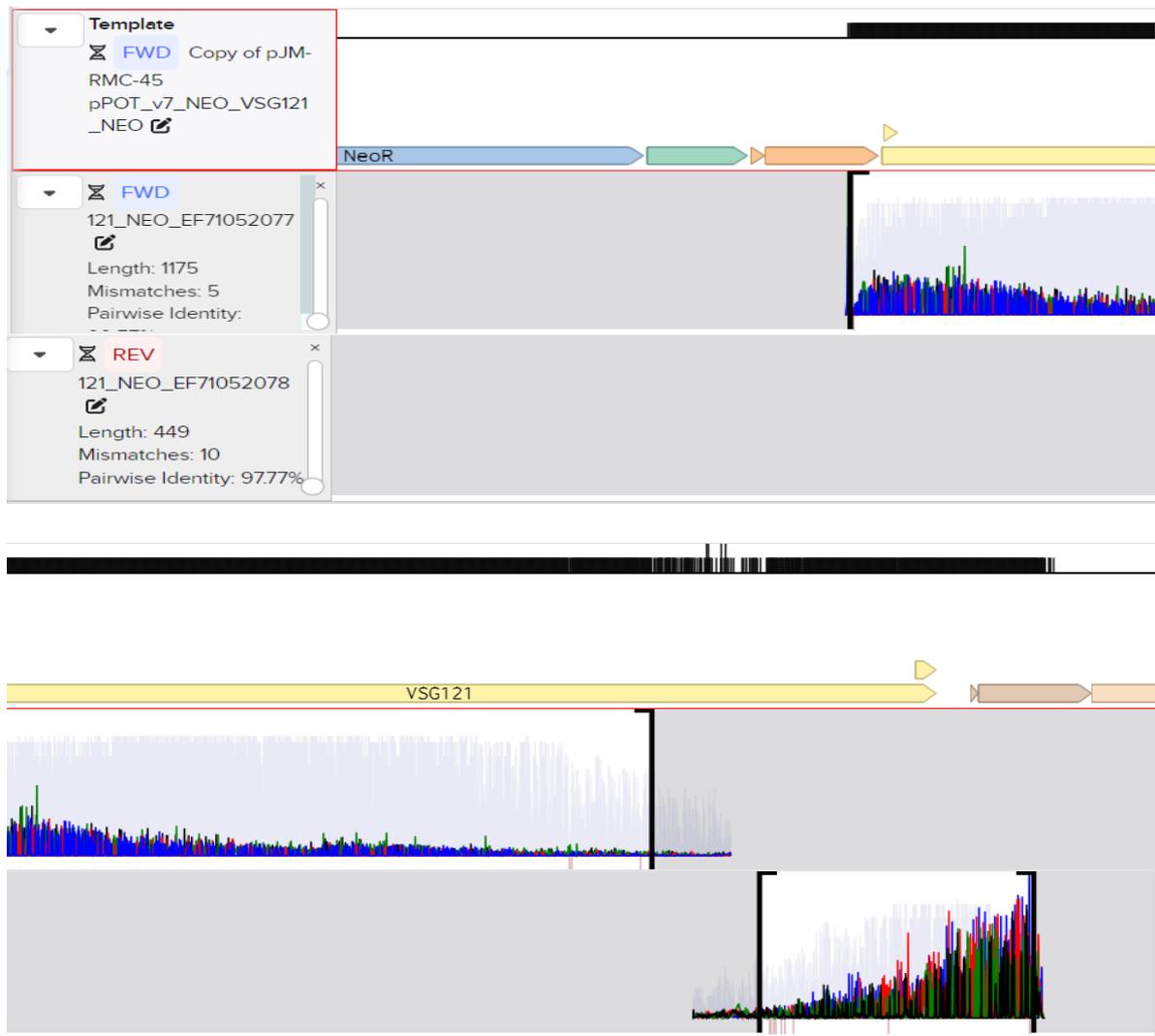
Appendix 5. Sanger sequencing confirmation of *VSG121* insertion into pPOTv7 NEO. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, NEO is indicated in blue, aldolase UTR 3' in green, aldolase 5' in orange, *VSG221* in yellow, and PFR2 IGS in brown.
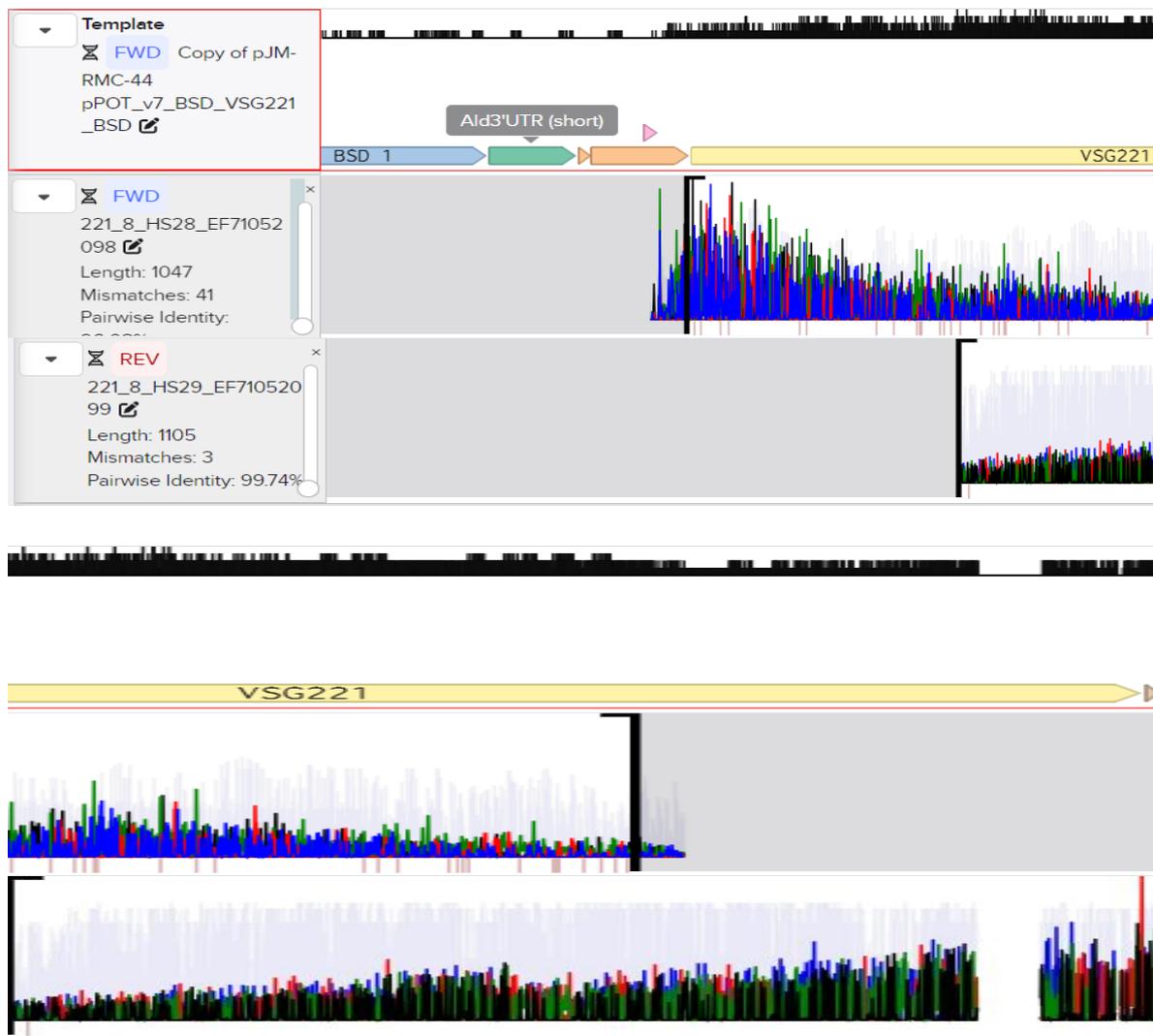
Appendix 6. Sanger sequencing confirmation of *VSG221* insertion into pPOTv7 BSD. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence, BSD is indicated in blue, aldolase UTR 3' in green, aldolase 5' in orange, *VSG121* in yellow, and PFR2 IGS in brown. Primers used for sequencing are indicated in pink.
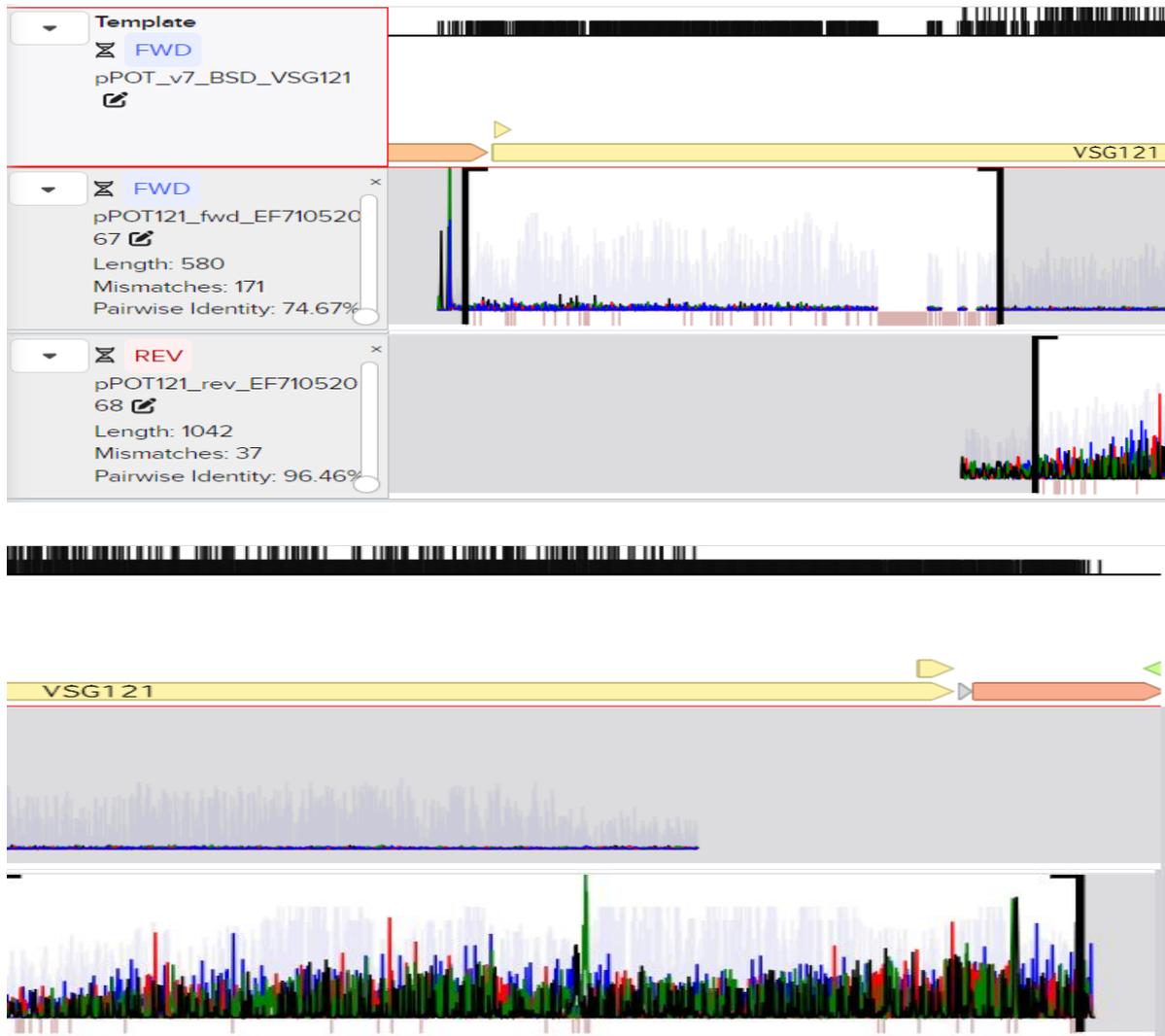
Appendix 7. Sanger sequencing confirmation of *VSG121* insertion into pPOTv7 BSD. In the template, made in Benchling by adding the inserted vector sequence into the sequence of BES1 from the TritrypDB genome sequence aldolase 5' in orange and *VSG121* in yello**w.**

Appendix 8. Sequence of amplified VSG fragments.

	Sequence (5' – 3')
VSG121	GCCGTGCACAGAGCCCTAGCGGCGTACGCGATTAGTCTTTACGTTTTACTACCCAGAA AATCGGGAGCAACAGACAAAGGCGCGATCAAGTTTGAGACGTGGGAGCCGCTCTGTTT ACTGACACAAGACTTCGGTAACCTTTACAACAGAGCGCACAAACTTAATCTCGACATC GACACCTACGTAACCGCAGCCCAGGCCGACCAACTGCGGCTCCAGGTTCTACTAAGTC GCGCTAGCAGCAAGATAGAAGCTGCTGCAGCCGCGGCAGCAACAGCAGCAATAGCAG CCGATCTAGCCGGCAAAGCTAAGCACGTCGCCAGCTGCAAGCTGGCAGCAACAACACT GACAGCCACAACAGGGTACTTGCACGGCAGAATAGCTGAATTCCTAGAGGTCATGACA GCTCACAGGGGAACAACAAACAAATACGGCTGCCTTTCAAAAAGCAGATCAGACAAC TCAAATTCGATAACGGACAGCGTCGCAAATCTAAAAGACAAGTGCAAGCTGACAGTG CAACAAATAAGTCCGGACAACAAAGCAACCGAGCAGATAACCAAAGCAGGCTTCACG AAACTAACGGCGACCGGCGGGCTAACAAGCAGCGACCTGGGAGGCAGCGGCCAAGCA GTCTGCATGATTCTGAGCACCACAGCCAGCGAAGTGGTAAACAACGGAAACCTAGATG AACCGGTCCCTTACGCCGGCGGCTACCTAAGACGTAAACACGACCTGACATCGGACGG TAACGACAACTTAGCGACCATCACAGACTCAGCTTCAGCGCCAGCAACCCGGGCCAAG ACAGACCCCTATCTCCAGATCTGGCGGGCCTTCAAAAACCTAGAAGACTGCGAATCCA CGTTCACAAGCGGCTACAGCCGACCATCGCCAGAAACATTAAAAGCAGCTGACGAAA CAAAAACAGCAATTAAAAACTATGTAGTACAGAAGGAAGGCAAATACGACCAGGCTA CAGACAAAGAAGACGACTACAAGAACCTCGACAAGATATTCAAGGATGGCAAGGACT TTTACCCGCAAAAACTCTGGGATGCCATGGATAAGAAAGATCTGCTAAAAGATGCGAC GCAAACCAACGAGATCAAAAAACTTGCCGACATAACCGACAGAAGCGAACTAAACAA GGTTCTGTTATATTACACAAGACAGAAGGAACAAACCTTAACTAAAGAACTAAAAGA AGCACAAGAAAAAGCAACACAGGCGAACCAGAATGACGCTGCAGCCAAAGCAGCAG AAGACTCTTGTAATAAGCTAGTTGGCGGAGAAAAATGCAACGCTGATAAAAAATGCA GCTATGAGACAGAAACTGATGGCACCAAAAAGTGCAAATTTAATGCTACCAAAGCTG AAAAAAGTGGTGCCCCTGTAACACAAGCTCAAACGGTAGGAGAAACCGAAGCAACGC CAGAAAAGTGCAAAGGCAAAGACGCCAAGACATGCGGCACCACTCAGGGTTGCAAAT GGGAGGGTGAAACTTGCAAAGATTCCTCTATTCTAGTAACCAAGAAATTCGCCCTCAC CGTGGTTTCTGCTGCATTTGTGGCCTTGCTTTTTTAA
VSG221	CCTTCCAATCAGGAGGCCCGGCTTTTCCTCGCCGTCTTGGTCCTAGCCCAAGTTCTTCC AATTCTTGTCGATTCGGCGGCTGAAAAAGGTTTCAAACAAGCTTTTTGGCAACCTCTTT GCCAGGTCTCCGAGGAGCTAGACGACCAACCGAAGGGTGCGTTGTTTACGCTGCAAGC AGCGGCGAGCAAAATCCAGAAAATGAGGGACGCGGCACTGCGAGCAAGTATATACGC TGAAATAAATCACGGCACCAACAGGGCCAAGGCAGCCGTTATAGTCGCCAACCACTAT GCCATGAAAGCTGATAGCGGCCTAGAGGCCCTAAAACAAACGTTAAGCAGCCAAGAG GTAACAGCTACTGCAACAGCGAGCTACCTAAAAGGAAGAATAGACGAATACTTAAAT CTCCTTCTACAAACAAAGGAGAGCGGCACCAGCGGCTGCATGATGGACACCAGCGGA ACAAACACGGTAACGAAGGCCGGCGGCACCATCGGAGGCGTTCCTTGCAAGCTGCAG TTGTCGCCGATACAGCCGAAGCGACCCGCAGCGACCTACCTAGGTAAAGCGGGCTACG TAGGCCTAACACGACAAGCAGATGCAGCCAACAATTTCCACGATAACGACGCCGAAT GCAGGCTAGCCAGTGGGCACAACACCAACGGCCTCGGCAAAAGCGGCCAGCTTTCTGC AGCGGTCACTATGGCGGCCGGCTATGTCACAGTAGCGAACAGCCAAACAGCCGTCACG GTCCAGGCGCTCGATGCATTACAGGAAGCGAGCGGAGCAGCGCACCAACCGTGGATC GACGCCTGGAAGGCCAAGAAAGCGCTAACAGGAGCAGAAACCGCTGAGTTCAGAAAC GAAACAGCCGGAATAGCTGGCAAAACAGGCGTTACCAAGCTTGTTGAAGAAGCTTTAC TAAAGAAAAAAGACTCAGAGGCCTCAGAAATACAAACAGAATTAAAAAAATACTTTA GCGGCCACGAAAATGAACAGTGGACAGCAATAGAAAAGCTCATATCCGAGCAGCCAG TGGCGCAAAACCTGGTAGGCGACAACCAGCCAACCAAGCTAGGGGAACTGGAGGGCA ATGCCAAGTTAACGACTATACTTGCCTATTACCGAATGGAAACAGCAGGGAAATTTGA AGTTTTAACCCAGAAGCACAAGCCCGCTGAAAGCCAACAACAAGCAGCAGAAACAGA AGGCAGCTGCAACAAGAAGGACCAAAATGAGTGCAAATCCCCATGCAAATGGCATAA CGATGCGGAAAACAAAAAGTGCACATTGGATAAGGAGGAGGCAAAAAAGGTAGCAG ATGAGACTGCAAAAGATGGGAAAAACTGGAAACACAAACACCACAGGAAGCAGCAATT CTTTTGTCATTAGCAAGACCCCTCTTTGGCTTGCAGTTTTGCTTTTTTAA
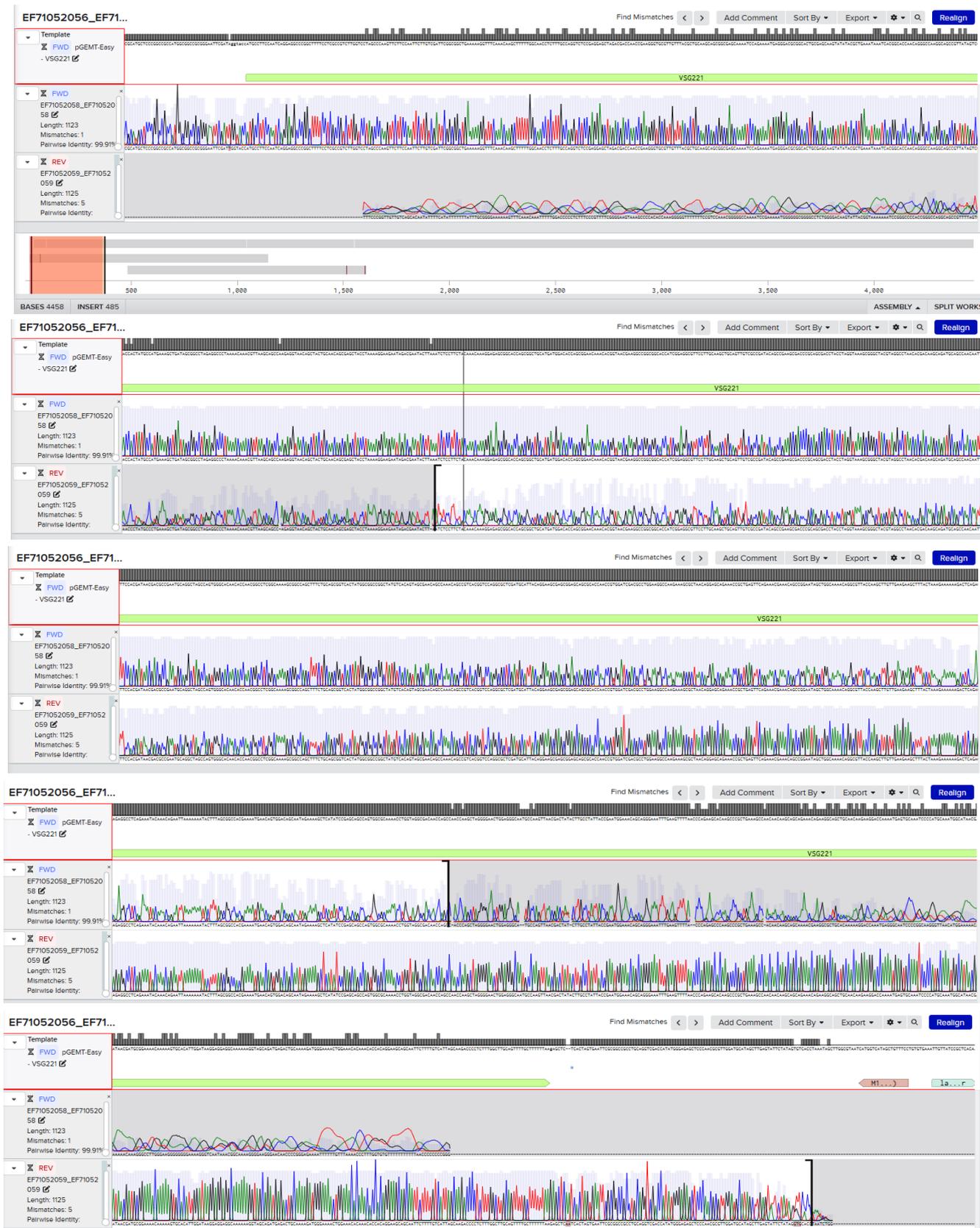
Appendix 9. Sanger sequencing of VSG221 in pGEM®-T Easy
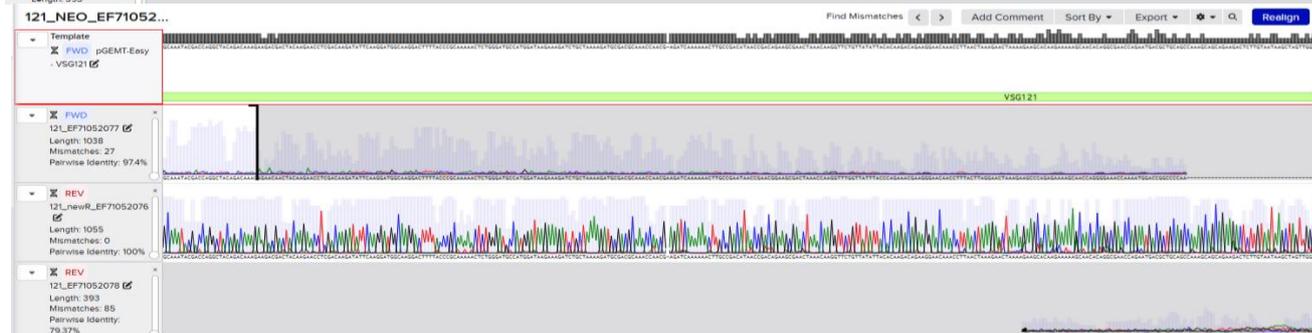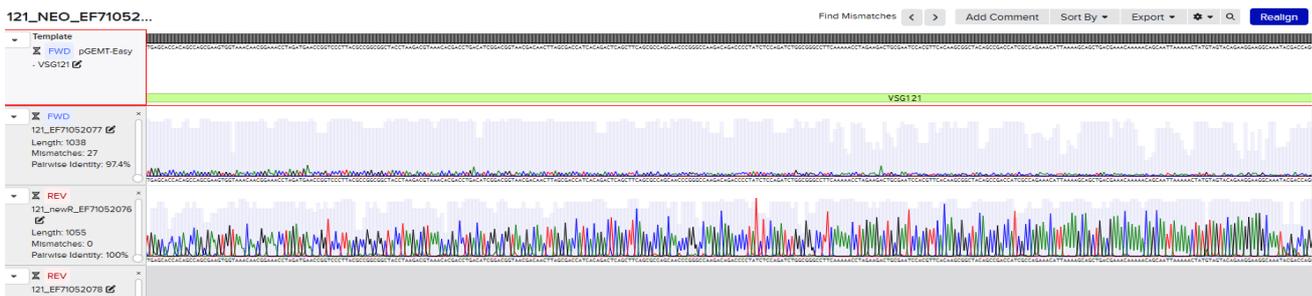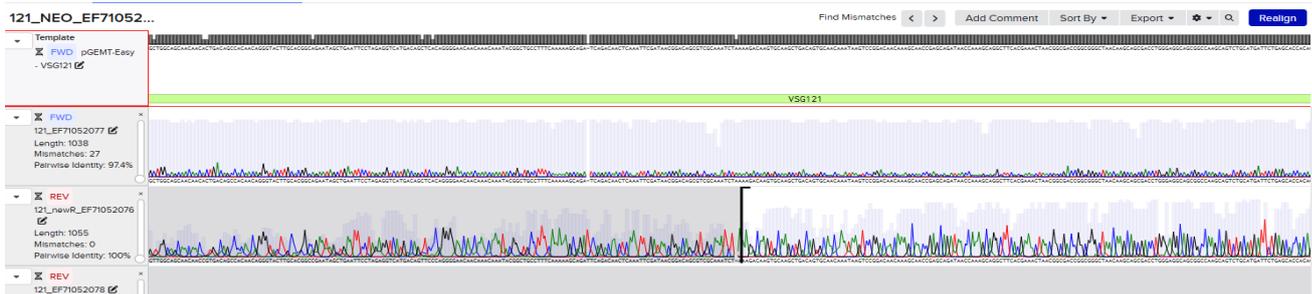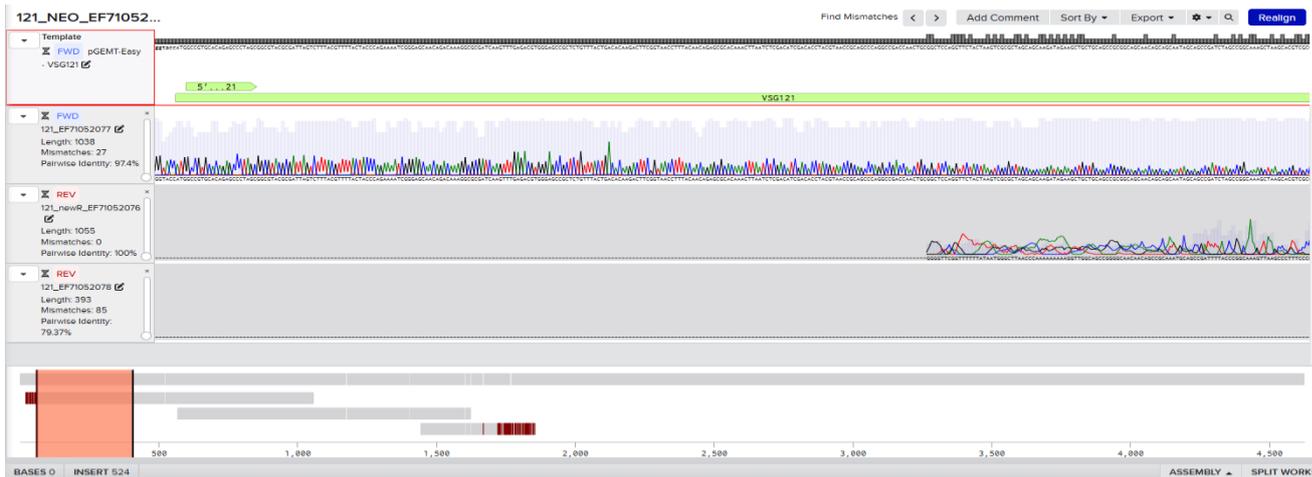
Appendix 10. Sanger sequencing of VSG121 in pGEM®-T Easy.
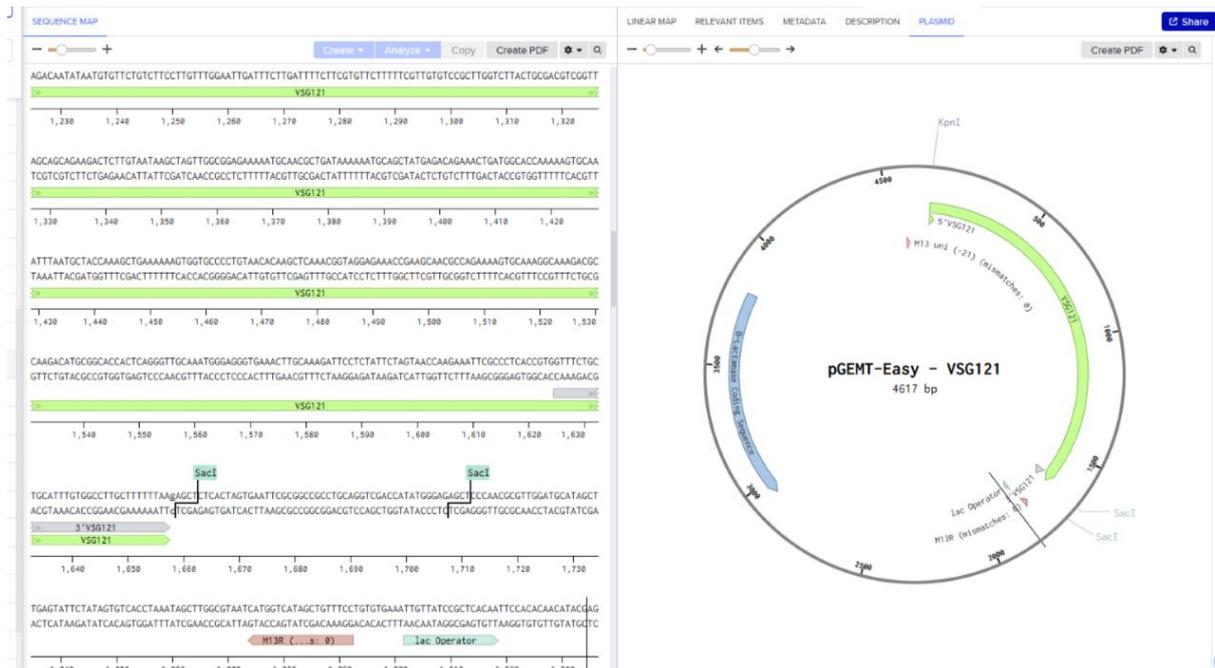
Appendix 11. Plasmid map of pGEM®-T Easy containing VSG121. Relevant *Kpn*I and *Sac*I restriction sites are indicated.