Evolution of Antigenic Variation in African Trypanosomes: Variant Surface Glycoprotein Expression, Structure, and Function

James D. Bangs

The process of antigenic variation in parasitic African trypanosomes is a remarkable mechanism for outwitting the immune system of the mammalian host, but it requires a delicate balancing act for the monoallelic expression, folding and transport of a single variant surface glycoprotein (VSG). Only one of hundreds of VSG genes is expressed at a time, and this from just one of \( \approx 15 \) dedicated expression sites. By switching expression of VSGs the parasite presents a continuously shifting antigenic facade leading to prolonged chronic infections lasting months to years. The basics of VSG structure and switching have been known for several decades, but recent studies have brought higher resolution to many aspects this process. New VSG structures, in silico modeling of infections, studies of VSG codon usage, and experimental ablation of VSG expression provide insights that inform how this remarkable system may have evolved.

1. Introduction

Of all microbes inhabiting the total planetary biome, only very few have successfully evolved to lead a pathogenic life style. Whether bacterial, viral, fungal, or parasitic, these pathogens have all been shaped by unique selection pressures in adapting to host defense mechanisms. Perhaps one of the most remarkable outcomes of this process is found in the African trypanosomes. These kinetoplastid protozoan parasites (Trypanosoma brucei ssp and related species) cause fatal sleeping sickness in humans, and related diseases in domestic livestock. Trypanosomes have a complex life cycle that alternates between the insect vector, the tsetse fly, and the blood and tissues of the mammalian host. There is no intracellular stage, unlike the related South American trypanosome T. cruzi, and infection can last for months to years. Consequently bloodstream parasites must bear the full brunt of host acquired immunity throughout this extended period. To achieve this end trypanosomes have evolved a remarkable process of antigenic variation based on a unique protein, variant surface glycoprotein (VSG) [ref. [1] and references therein].

The T. brucei genome has \( \approx 2000 \) VSG genes,\(^{[2]} \) \( \approx 80\% \) of which are pseudogenes, and only one of which can be expressed at a time. VSGs are homodimers that are anchored in membranes by glycosylphosphatidylinositol anchors (GPI, two per dimer). Remarkably abundant, \( \approx 10\% \) of total cell protein, some \( 5 \times 10^6 \) dimers form a monolayer covering the entire cell body and flagellum. This \( \approx 15 \) nm thick surface coat shields underlying invariant proteins from host-derived antibodies that inevitably arise during prolonged infection. VSG itself is highly immunogenic and ultimately, as parasitemia rises, parasites are recognized and killed by primary humoral responses. However, VSG expression switches stochastically, such that new serotypes continuously arise in the background. This dynamic leads to waves of parasitemia in the bloodstream, each wave containing a mixed population of cells expressing up to 100 distinct VSGs.\(^{[3,4]} \) Thus antigenic variation ensures the survival of the population, and consequently transmission to new hosts, at the expense of the individual.

2. Trypanosome Gene Expression and Antigenic Variation

Kinetoplastid protozoa are atypical eukaryotes in that housekeeping gene expression is polycistronic across large segments of the 11 diploid megabase chromosomes.\(^{[5]} \) Individual mRNAs are excised from the poly-transcript by combined 5' trans-splicing/capping and 3' polyadenylation. This unique situation liberates gene expression from transcriptional initiation, thus modulation of expression levels, at least within a given life cycle stage, is by post-transcriptional mechanisms. In contrast, VSG genes are transcribed from short (\( \approx 30-60 \) Kb) telomeric expression sites (ES) of which there are \( \approx 15-20. \)\(^{[6]} \) These have similar conserved structures, an RNA polymerase I promoter at the 5' end, a VSG
gene at the telomeric end, and a variable number of expression site associated genes (ESAG1-13) in between (Figure 1A). All other VSG genes are located in silent haploid subtelomeric arrays and on minichromosomes. Only one ES, residing in a unique extranuclear structure called the expression site body, is active at any given time thereby ensuring strict monoallelic VSG expression. There are two main ways in which antigenic variation can occur (Figure 1B and C). First is in situ activation, in which the active ES residing in the expression site body is replaced by another inactive ES. This mechanism effectively changes expression of all genes in the ES. Second is gene conversion, in which all or segments of a gene, either from the sub-telomeric arrays or another ES, over-writes the VSG in the active site. This process is driven by conserved upstream 70 bp repeats and sequence homology throughout the open reading frame and 3’ UTR. As only \( \approx 20\% \) of the VSG repertoire are complete genes, and the rate of switching is up to \( 10^{-3} \), segmental gene conversion plays a critical role in generating antigenic diversity late in chronic infection.

3. Chicken or Egg: How VSG Function Impacts VSG Structure

Our understanding of how VSG performs its two essential functions – presentation of an ever-changing antigenic profile and shielding of the underlying cell surface – is informed by our understanding of VSG structure. Early DNA sequencing revealed a remarkable degree of heterogeneity weighted toward the N-terminus, while concurrent biochemical studies indicated a bipartite structure with a large outer N-terminal domain (NTD) and a smaller C-terminal domain (CTD) proximal to the plasma membrane. Based on size and disulfide patterns these domains can be sorted into subgroups (NTD, A–C; CTD, 1–4) that have been interchangeably mixed during evolution of the VSG repertoire. The two domains are connected by a flexible linker, and consequently when the first crystal structures were determined only the protoxylotyzed NTDs were visualized. These were all A-type NTDs and, despite low sequence identity, they have strikingly similar structures (Figure 2A). The central feature is an \( \approx 10\text{ nm} \) N-terminal coiled-coil that comes together homotypically to form the dimer interface. Each dimer has a “bulb” at the exterior end formed by internal sequences, which not coincidentally are the most diverse between individual VSG genes, and another at the membrane proximal end formed by C-terminal sequences. Thus VSGs have evolved to present maximum antigenic diversity to the outside world while conserving the ability to assemble a structurally coherent surface coat. Natively perhaps, given so few structures all of the same subtype and without any knowledge of the CTD, the common depiction was of elongate VSGs packed shoulder-to-shoulder such that each dimer could diffuse laterally, but so that potentially lytic antibodies could not penetrate. This packing would be particularly advantageous during antigenic variation, since it is critical to maintain an intact coat as the new VSG replaces the old by dilution during cell division.

Recent developments paint a more nuanced picture. First, the solution structure of several CTDs have been solved by NMR and wedded to corresponding NTDs by small-angle X-ray scattering to reveal VSG structure in toto. The results yield a two state model of VSG within the surface coat (Figure 2B). In the low-density relaxed state the independent CTDs, each grounded by a GPI anchor, are spread apart drawing the dimeric NTD down toward the membrane and allowing for broader coverage of the cell surface. In the high-density compact state the CTDs are drawn toward each other thereby lifting the dimeric NTD away from the cell surface allowing for clearance of imbedded polytopic proteins. This model also allows for lateral “breathing” to accommodate varying VSG densities that might occur during cell growth and division, or during switching to a new VSG.

All of this may inform two related peculiarities of African trypanosomes. First, a priori there is no obvious need for a dimeric coat protein. For instance, T. cruzi has a complex glycocalyx composed of multiple monomeric GPI-anchored proteins and free GPI lipids, perhaps reflecting the fact that T. cruzi is a mostly intracellular parasite that is only transiently exposed to host immunity. However, this may be overly simplistic as T. theileri, which is more closely related to T. cruzi, is modeled to have a similarly complex glycocalyx, yet is exclusively an extracellular parasite in the bovine host. That T. theileri also establishes long-term infections indicates there are strategic alternatives to the paradigm of antigenic variation seen in African trypanosomes. Nevertheless, the requirement for a monotype dimeric coat in T. brucei is not obvious. Second, and uniquely among eukaryotes, bloodstream form T. brucei GPI anchors exclusively contain the short chain (14-carbon) fatty acid myristate, which alone is not sufficient to maintain membrane association. In contrast, T. cruzi bases its GPI structures on longer chain (C18-24) phosphorylglycerol and phosphorylceramide lipids. Given the abundance of VSG, the use of longer chain acyl groups might disrupt normal membrane function, driving the need for myristate and consequently dimerization to maintain a stable surface coat. An additional driving force for dimerization is that two GPIs are critical for trafficking to the cell surface – monomeric GPI anchored proteins are primarily endosomal and are preferentially degraded in the lysosome, while VSG is long lived being constantly endocytosed and efficiently recycled to the cell surface. We can now add the need for a dynamic coat vis-à-vis the two-state model as contributing to the interrelated evolution of VSG dimers and short chain GPI anchors.

The second development is the crystal structure of the NTD of a class B subtype, VSG3 (aka MITat1.3, VSG224), which yielded two surprises. First is the previously unknown presence of a short variable O-linked glycan on the outer face of the VSG. This modification, also found on two other NTD-B VSGs and likely more, increases virulence by shielding the surface coat itself from immune recognition thereby prolonging infection in a mouse model. The second surprise is that the VSG3 NTD crystallizes as a monomer. Whether the full length VSG is a monomer in vivo is not known, but the answer to this question will have consequences for the role of dimerization in VSG functions as discussed above.

4. Just Right: How Do Trypanosomes Precisely Regulate VSG Synthesis?

Ten percent of total protein per cell cycle (\( \approx 30\text{,000 VSGs per min} \)) is a huge metabolic commitment and trypanosomes walk a
A. VSG Genomic Organization

Sub-Telomeric Arrays

Telomeric Expression Sites (~15-20)

RNA Pol I

B. In Situ Activation

ES2

ES1

C. Gene Conversion

Sub-Telomeric Array

Gene Conversion

Limits of Gene Conversion

Figure 1. VSG Genomic Organization and Expression. A) Antigenically distinct VSG genes are found in silent haploid telomeric arrays (top), each preceded by 70 bp repeats (black boxes) that facilitate recombination events. The single expressed VSG gene is found in the active expression site, modeled here on ES1 of Lister Strain 427 of T. b. brucei (bottom). An RNA polymerase I promoter (flag) is at the 5' end of the transcription unit and the VSG gene (V) is at the extreme 3' end just upstream of the telomeric repeats (black circles). The VSG gene is preceded by 70 bp repeats that facilitate recombination events. Variable complements of Expression Site Associated Genes (numbered ESAGs) are found in between. ESAG6 and ESAG7, which encode the heterodimeric transferrin receptor, are always found proximal to the promoter. B) In situ switching involves displacement of the active ES from the expression site body by one of the other 15-20 silent ESs. This results in a change in expression for the entire transcription unit. C) Gene conversion over-writes the VSG gene in the active ES with a new VSG gene from the silent arrays, or from another ES. Gene conversion is facilitated by the 70 bp repeats and by sequence homology throughout the open reading frame and the 3' UTR. Only expression of the VSG is affected, all other ESAGs remain the same.

5. When Good VSG Goes Bad: Coping with Failed Antigenic Variation

That trypanosomes carefully monitor VSG production is evident from studies blocking VSG synthesis. Conditional RNAi silencing of VSG expression results in a rapid (within one cell...
cycle) pre-cytokinesis growth arrest that can be rescued by introducing a second VSG gene into the active expression site. When the second gene is introduced the levels (mRNA and protein) of the first VSG drop by half, and subsequently when the first gene is silenced levels of the second VSG rise by two. The implication is that trypanosomes carefully monitor and regulate total VSG production, and when this is insufficient growth and division cease in order to maintain an intact surface coat. Coinciding with arrest is a global shutdown of protein synthesis. Cells can survive in this state in vitro for several days, but in infected mice arrested cells are cleared within 12 h, suggesting that even a slight chink in the armor can be fatal in vivo. This result is actually the first formal proof that VSG is an essential virulence factor. How VSG levels are monitored, and how this is tied to cell cycle regulation, is not known, but it must be post-translational since selectively blocking VSG synthesis with antisense morpholino oligonucleotides also induces precise pre-cytokinesis arrest without affecting VSG mRNA. Presumably then, it is total VSG protein, or some aspect of its synthesis and/or transport, that is monitored [To Do #4].

Newly synthesized VSG is the overwhelmingly major secretory cargo, and the trypanosome secretory pathway is designed for its efficient transport to the cell surface. Not surprisingly then, sudden loss of VSG has pronounced effects on this pathway. All secretory cargo departs the endoplasmic reticulum from specific ER exit sites (ERES) for delivery to the Golgi apparatus, and subsequent post-Golgi trafficking involves selective sorting to either endolysosomal compartments or the cell surface. In interphase trypanosomes there are just two ERES, each closely juxtaposed with a single Golgi, and this number increases to four in pre-cytokinetic cells. Loss of VSG results in distortion of the ER and a significant reduction of ERES/Golgi across all stages of the cell cycle, indicating that maintenance of these organelles is dependent on cargo flux. However, distal Golgi cisternae are actually grossly distended. Why shutting down cargo flux would result in Golgi expansion is not intuitive, but the thought is that ongoing anterograde membrane flow into the Golgi without compensatory out flow results in swollen trans-cisternae. Consistent with this model, forward trafficking of endogenous cathepsin L, a lysosomal

---

**A. VSG NTD Structures**

- MITat1.2
- ILTat1.24
- MITat1.1

---

**B. VSG Two-State Model**

Figure 2. VSG Structure. A) The solved structures of the NTDs of three type-A VSGs are presented: MITat1.2, pdb 1VSG; ILTat1.24, pdb 2VSG; and MITat1.1, pdb 5LY9. The N-terminal membrane distal end is up; the C-terminal membrane proximal end is down. The view is front on looking through the dimer interface. The central coiled-coil is blue. B) The two-state model of VSG packing on the cell surface. Side-view illustration of membranes occupied by VSGs at two different densities. The conformation of tightly packed VSGs (left) can elevate the VSG above the transmembrane proteins (dark blue, red), whereas the relaxed conformation (right) could allow the maintenance of a protective coat on the cell surface even at a reduced protein density. Reproduced with permission. Copyright 2017, The Authors.
protease, from the ER to the Golgi is unimpaired, but so too is subsequent transport from the Golgi to the lysosome. One would have to suppose then that it is the distinct secretory arm of post-Golgi trafficking, the route VSG takes to the cell surface, which is affected by loss of cargo [To Do #5].

Whatever the mechanism, is seems unlikely that this arrest phenomenon evolved to cope with sudden experimental loss of VSG. More likely it is designed to tweak growth rates in response to subtle fluctuations of VSG synthesis. [44] Or maybe not! There is one natural scenario in which trypanosomes could be confronted with the catastrophic loss of functional VSG – failed antigenic variation. The primary mode of antigenic variation is gene conversion, which late in infection – after the repertoire of intact VSG genes is exhausted – relies on segmental events in which parts of related VSG genes are assembled in the active expression site. This has the salubrious effect of progressively expanding antigenic diversity, but inevitably given the high natural rate of switching, it will result in fusion of incompatible protein segments, or creation of a premature stop codon. The upshot is that the cell will have gone in an instant from synthesizing a perfectly functional VSG to one that is incapable of folding, assembly, and ER exit; a situation that might be perceived in a similar manner as VSG silencing. One can envision three things that must happen for the unfortunate cell to survive. First, it must shut down growth, just as occurs with VSG silencing, to maintain an intact surface coat. Second, it must cope with the massive accumulation of misfolded VSG in the ER. Misfolded secretory proteins are typically retained in the ER and disposed of by ER-associated degradation (ERAD). [45] The misfolded protein is recognized and targeted for retrotranslocation to the cytosol where it is degraded by the proteasome. Such a process would seem to be the most efficient mechanism for disposal of misfolded VSG. However, in yeast and mammalian cells misfolded GPI-anchored proteins are actually exported to the lysosome for degradation. [46, 47] This is because the GPI anchor is a forward trafficking signal for ER exit. This sets up a dynamic tension between retention by ER quality control versus packaging into secretory cargo transport vesicles – in these model organisms ER exit is the winner. GPI anchors are also ER exit signals in trypanosomes, [48, 49] but in this case it is retention and disposal by ERAD that dominates. [29, 50] Whether trypanosomes have the capacity to cope with VSG levels of misfolded protein remains to be tested [To Do #6]. Finally, the afflicted cell must switch to another functional VSG. It is unlikely that further rounds of segmental gene conversion could repair the damage, but fortunately there are other expression sites that can be activated to restore functional VSG production. One weakness of this model is that, as discussed above, arrested VSG deficient parasites are rapidly cleared from circulation in infected mice, a process that likely involves complement opsinization. [38] However, T. brucei is well known to invade extravascular tissues; indeed invasion of the central nervous system is the hallmark event of late stage sleeping sickness. Trypanosomes also have a particular tropism for adipose tissue and it is possible that these locales provide a safe haven from circulating innate immune factors such as complement. [51, 52]

6. If Two Are Better Than One . . . Why Not 15?

The idea that trypanosomes will inevitably have to cope with the catastrophic formation of non-functional VSGs may offer insight into a longstanding mystery about antigenic variation. Why are there as many as 20 VSG expression sites (ES) when a priori there is no need for so many? Gene conversion alone is an efficient mechanism for switching, and most certainly at some point in the evolution of T. brucei there was an ancestor with a single expression site. What then was the driving selection pressure for expansion of this system in T. brucei? It is likely to have come in response to multifactorial pressures. One proposal centers on another ES encoded protein – transferrin receptor (TfR). [53, 54] TfR, which is structurally related to VSG, [55] is a heterodimer of ESAG7 and ESAG6, the genes for which are found proximal to the promoter in every ES (Figure 1A). Both genes have polymorphisms spread across all the ESs that result in different affinities for transferrin from different mammalian hosts, which can vary up to 35% in sequence. As the transmissive metacyclic form in the salivary glands of a tsetse fly could end up in any mammal, it would be advantageous to have an adaptive repertoire of optimal TfRs that could be accessed by ES switching. This model has been challenged, [56] and it would be difficult (but not impossible) to assess in natural infections, but in vitro it has been demonstrated that switching growth media from optimal (bovine) to suboptimal (canine) sera results in multiple forms of gene rearrangements at the ESAG6/ESAG7 loci, including switching to a different ES. [53, 54] Critically, all of these effects are blocked by inclusion of bovine transferrin, indicating that the biological stimulus is the change of transferrin, not other serum factors. Another scenario where an expanded repertoire of ESs could be beneficial is in different tissue locations within the mammalian host. As mentioned above T. brucei has a tropism for adipose tissue, and parasites in this location have a distinct metabolic profile from parasites in the circulation (β-oxidation vs. glycolysis). [51] With this in mind it has been suggested that switching to an ES with a different complement of ESAGs may be an adaptive response to tissue specific nutritional requirements. [57]

We can now also envision the inevitability of misfolded VSGs as contributing to the expansion of ESs. Telomeres are fluid structures and the random duplication of the first ES would give that cell a leg up in the event of failed gene conversion. And if two is good, perhaps more are better, but at some point there has to be a diminishing return in terms of selection pressure. For discussion’s sake, if ES duplication reaches the point of five, is there really enough fitness advantage to drive that number to six, seven, and so on? Perhaps early in the evolution of this system the process of gene conversion was not as efficient as it is now, and there was a real benefit to having many ESs in order for antigenic variation to occur in a timely manner in the infected host. It is even possible that the immediate ancestor of T. brucei had yet more ESs, and that these were reduced to what we see today. In either case, it is likely that the needs for nutritional responsiveness, coping with misfolded VSGs, and whatever additional benefits derive from the other poorly understood ESAG gene products, all combined to create the system we see today in T. brucei. That said, it is important to note that in the related African species, T. congolense and T. vivax, much less is
known about the mechanisms of VSG expression and switching, but it is clear that there are significant differences compared with *T. brucei*. For instance, they are both much less dependent on gene conversion for switching. TFR orthologues are not associated with telomeres in *T. congolense*, and no obvious orthologues of TFR are found in *T. vivax*. Without question then, our understanding of the evolution of multiple ESs in *T. brucei* will change as more knowledge is gained in these related species [To Do #7].

7. Whatever To Do?: Conclusions and Outlook

We have known of the process of antigenic variation as a survival strategy for trypanosomes since the beginning of the 20th century, but we only began to understand it at a mechanistic level with the discovery of VSG, the early molecular studies into the mechanisms of antigen switching, and the first VSG structures. The last decade has provided a much finer focus on how and why the process of antigenic variation is so successful, which in turn allows thoughtful consideration of how it might have evolved in the first place. But as is always the case the answer to one question leads to many more. There is much still to do:

1) The crystal structure of more VSGs of the other non-A subtypes are needed. Just the one new structure of VSG3 has raised interesting issues about O-glycosylation and dimer vs. monomer structure, e.g., is full length VSG3 really a monomer in vivo. It is reasonable that more surprises await new structures. The one prediction that can be made with a degree of confidence is that an elongated structure will be conserved, even if the fine details vary, to allow packing into a functional barrier as one coat is replaced by another.

2) The model for controlling the hierarchy of VSG switching in chronic infection is based on growth rate being inversely proportional to VSG size. It will be critical to confirm this relationship experimentally. This can be accomplished by low dose infection of immunosuppressed mice with clonal lines expressing different VSGs and monitoring infection during the initial rise of circulating parasitemia.

3) Recoding an expressed VSG gene to maximize the use of optimal codons. Does VSG synthesis go up? If not, do VSG mRNA levels drop to compensate for higher translation efficiency? If so, does growth rate drop in response to the greater metabolic strain of elevated VSG synthesis?

4) How is the loss of VSG monitored? What is the signaling pathway to global shut down of translation? Can any mechanistic insights be gleaned from known stress response pathways in other eukaryotic systems?

5) What leads to the counter intuitive expansion of the distal Golgi in the absence of VSG cargo? If the hypothesis that outward flow of membrane is reduced is correct, then it cannot be the endolysosomal arm of post-Golgi sorting since TbCatL trafficking to the lysosome is unaffected. This focuses attention on the secretory arm by which VSG goes to the cell surface. What effect does VSG silencing have on trafficking of other cargo using this pathway, such as transferrin receptor?

6) Trypanosomes are able to cope with significant levels of accumulated misfolded GPI-anchored proteins in the ER, but can they really cope with accumulation of 10% of cell protein in a misfolded state? Development of a system to conditionally induce a switch from a functional VSG to a deliberately misfolded VSG from within the active ES would allow many questions to be asked. Do cells arrest as seen with VSG loss by RNAi? Can they cope with the massive overload in the ER? Is there global translation arrest? Can they escape by switching to a new ES?

7) *T. vivax* and *T. congolense* are a much bigger cause of veterinary trypanosomiasis than *T. brucei*, but studies in these species have lagged behind, in part due to technical issues, e.g., *T. vivax* cannot be cultured, but also due to the facile nature of working with *T. brucei*. Going forward it is critical that more effort be made on the basic cellular and molecular biology of these species. Already from the available genomic and transcriptomic data it is clear that both species differ significantly in their approach to antigenic variation. *T. vivax* diverged closer to the evolutionary root of the African trypanosomes, while the split of *T. brucei* and *T. congolense* is a more recent event. Illuminating this issue in these species will without doubt significantly impact our understanding of the evolution of antigenic variation in *T. brucei*.

The answers to these questions and more are certain to keep "trypologists" gainfully employed for many years to come.

Abbreviations

CTD, C-terminal domain; ERAD, ER associated degradation; ERES, ER exit site; ES, expression site; ESAG, expression site associated gene; GPI, glycosylphosphatidylinositol; NTD, N-terminal domain; TFR, transferrin receptor; VSG, variant surface glycoprotein.

Acknowledgements

This work was supported by United States Public Health Service Grant R01 AI035739, and funds from the Jacobs School of Medicine and Biomedical Sciences to JDB. The author would like to thank the members of the Bangs laboratory, past and present, who have made contributions to the evolution of the ideas presented, and where relevant for suggestions for the current manuscript. Thanks also to Professor Andrew Jackson, University of Liverpool, for critical reading and helpful suggestions.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

African trypanosome, antigenic variation, gene conversion, protein folding, secretion, variant surface glycoprotein

Received: September 11, 2018
Revised: October 4, 2018
Published online: October 29, 2018
[58] A. P. Jackson, A. Berrya, M. Asletta, A. C. Allison, P. Burton, J. Vavrova-Anderson, R. Brown, H. Browne, N. Cortona, H. Hausera, J. Gamble, R. Gilderthorp, L. Marcelloc, J. McQuillan, T. D. Ottoa, M. Quail, M. J. Sanders, A. van Tondera, M. L. Gingerd, FMC, J. D. Barry, C. Hertz-Fowler, M. Berriman, Proc. Natl. Acad. Sci. USA 2012, 109, 3416.

[59] A. P. Jackson, H. Allison, J. D. Barry, M. C. Field, C. Hertz-Fowler, M. Berriman, PLoS Neg. Dis. 2013, 7, e2121.

[60] A. H. Abbas, S. S. Pereira, S. D’Archivio, B. Wickstead, L. J. Morrison, N. Hall, C. Hertz-Fowler, A. C. Darby, A. P. Jackson, Gene. Biol. Evol. 2018, 10, 2458.