

Database

Open Access

VSGdb: a database for trypanosome variant surface glycoproteins, a large and diverse family of coiled coil proteins

Lucio Marcello¹, Suraj Menon^{1,2}, Pauline Ward¹, Jonathan M Wilkes¹, Nicola G Jones³, Mark Carrington³ and J David Barry*¹

Address: ¹Wellcome Centre for Molecular Parasitology, University of Glasgow, Glasgow Biomedical Research Centre, 120 University Place, Glasgow G12 8TA, UK, ²Department of Pathology, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK and ³Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK

Email: Lucio Marcello - l.marcello.1@research.gla.ac.uk; Suraj Menon - menons1@Cardiff.ac.uk; Pauline Ward - j.d.barry@bio.gla.ac.uk; Jonathan M Wilkes - jw149j@udcf.gla.ac.uk; Nicola G Jones - ngj20@cam.ac.uk; Mark Carrington - mc115@cam.ac.uk; J David Barry* - j.d.barry@bio.gla.ac.uk

* Corresponding author

Published: 2 May 2007

Received: 11 September 2006

BMC Bioinformatics 2007, 8:143 doi:10.1186/1471-2105-8-143

Accepted: 2 May 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/143>

© 2007 Marcello et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Trypanosomes are coated with a variant surface glycoprotein (VSG) that is so densely packed that it physically protects underlying proteins from effectors of the host immune system. Periodically cells expressing a distinct VSG arise in a population and thereby evade immunity. The main structural feature of VSGs are two long α -helices that form a coiled coil, and sets of relatively unstructured loops that are distal to the plasma membrane and contain most or all of the protective epitopes. The primary structure of different VSGs is highly variable, typically displaying only ~20% identity with each other. The genome has nearly 2000 VSG genes, which are located in subtelomeres. Only one VSG gene is expressed at a time, and switching between VSGs primarily involves gene conversion events. The archive of silent VSGs undergoes diversifying evolution rapidly, also involving gene conversion. The VSG family is a paradigm for α helical coiled coil structures, epitope variation and GPI-anchor signals. At the DNA level, the genes are a paradigm for diversifying evolutionary processes and for the role of subtelomeres and recombination mechanisms in generation of diversity in multigene families. To enable ready availability of VSG sequences for addressing these general questions, and trypanosome-specific questions, we have created VSGdb, a database of all known sequences.

Description: VSGdb contains fully annotated VSG sequences from the genome sequencing project, with which it shares all identifiers and annotation, and other available sequences. The database can be queried in various ways. Sequence retrieval, in FASTA format, can deliver protein or nucleotide sequence filtered by chromosomes or contigs, gene type (functional, pseudogene, etc.), domain and domain sequence family. Retrieved sequences can be stored as a temporary database for BLAST querying, reports from which include hyperlinks to the genome project database (GeneDB) CDS Info and to individual VSGdb pages for each VSG, containing annotation and sequence data. Queries (text search) with specific annotation terms yield a list of relevant VSGs, displayed as identifiers leading again to individual VSG web pages.

Conclusion: VSGdb <http://www.vsgdb.org/> is a freely available, web-based platform enabling easy retrieval, via various filters, of sets of VSGs that will enable detailed analysis of a number of general and trypanosome-specific questions, regarding protein structure potential, epitope variability, sequence evolution and recombination events.

Background

The variant surface glycoprotein (VSG) is essential for the survival of *Trypanosoma brucei* in mammalian hosts. There are $\sim 5.5 \times 10^6$ VSG homodimers per cell and the cell surface monolayer that the VSG forms is considered to provide general protection from innate immune mechanisms [1,2]. The coat nevertheless elicits a specific, trypanocidal immune response. This is countered by antigenic variation, in which trypanosomes switch to expression of a distinct VSG which, if antigenically novel, allows clonal proliferation of the switched cells, generating a new parasitaemia peak. Each trypanosome expresses only one VSG gene but has the potential to switch to any of probably hundreds of others [3,4].

The VSG is a structural paradigm for α helical coiled coil proteins and for B cell epitope variation [5,6]. This is because its hundreds or thousands of isoforms have limited similarity in peptide sequence and antigenicity, but strong conservation in higher level structure. In *T. brucei*, mature VSGs contain 400 – 500 amino acids, most having between 420 and 460 residues. Most of the protein is an N-terminal domain of ~ 350 residues, which is followed by a C-terminal domain, containing one or two smaller subdomains of 40–80 residues each [5,7]. N-terminal domains usually have only $\sim 20\%$ identity between different VSGs, although some are more closely related. The most conserved primary structure feature is the cysteine pattern, of which there are three, resulting in the classification of this domain into types A, B and C [5]. In contrast to the extensive sequence diversity, secondary structure potential is conserved, with a consequent overall similarity between the N-terminal domains of distinct VSGs. The backbone comprises two long, antiparallel α helices that form a coiled coil. It is not known exactly which elements of this domain contain the exposed epitopes that are the basis of antigenic variation, but it has been inferred that they are conformational rather than linear and are located in the loops exposed at the most extracellular end of the N-terminal domain, outwith the helices [8-10]. One important question that is yet to be definitively answered is why N-terminal domains vary over their entire sequence, rather than in just the region encoding the exposed protective epitopes.

The C-terminal domain is hidden from antibodies (NJ & MC, unpublished), presumably due to its membrane-proximal location, so is thought not to contribute to antigenic variation. It has $\sim 40\%$ identity between different VSGs. The genome project has revealed that, rather than the four C-terminal domain types (1 – 4) previously recognized, based on their cysteine patterns, there are six types. Types 2, 4 and 5 are single domains, each containing four cysteine residues, whereas types 1, 3 and 6 contain eight cysteines and appear to be composed of two

subdomains, each containing four cysteines. Individual VSGs can have any combination of N- and C-terminal domains (A1, A2, A3, A4, B1, B2 etc.), and, as judged from the VSGs analysed so far, there appears to be no restriction on combinations. At its C-terminal end, this domain contains a signal sequence for the addition of a glycosylphosphatidylinositol- (GPI-) anchor [2]. Although sequence features specifying GPI signal sequences have been identified [11], their full diversity across VSGs is not known, and study of as many as possible potential signal sequences could enable deeper understanding.

Despite about 1000 VSG sequences being currently available, mainly through genome and cDNA sequencing, it is not facile to retrieve a complete set from general databases. We have therefore created a database allowing retrieval of criterion-based subsets. This should facilitate a more detailed analysis of VSG structure and more general questions about protein structure including:

- what are the sequence requirements for coiled coil structures?
- can epitope diversity be correlated with diversity in primary and higher-order structure?
- How diverse are GPI-anchor signal sequences?
- How does evolutionary selection for diversification fit within a conserved protein structure?

It is worth noting that relatively few of the silent VSGs sequenced in the genome project are considered to be fully functional, and verification of function of any element, for example GPI anchor signal sequences, requires demonstration of expression. In contrast, the non-genome VSG sequences are based mainly on expression, most having been derived from cDNA sequences, and query returns in FASTA format report their derivation from cDNA or genomic DNA.

At the genetic level, trypanosomes use a strategy common to antigenic variation in a diverse range of microbial pathogens: accessing an archive of silent genes. In *T. brucei*, there is a large archive of silent, distinct VSG genes, effectively all of which are telomeric and subtelomeric. In the genome strain, about 1600 VSGs are arranged as tandem arrays in subtelomeres of a range of chromosomes, and it is likely that different strains contain substantially larger archives [12]. Only 4.5% of this set are annotated as intact genes, the rest consisting of atypical genes (do not convincingly encode maturable VSGs), pseudogenes (include frameshifts and/or stop codons), and VSG fragments. Another set, of up to ~ 200 genes, are located telomere-proximally in the ~ 100 minichromosomes; so far, based

on three genes [13,14], this set appear to be intact VSGs. Despite the enormous size of the silent archive, each trypanosome expresses only one VSG. Expression occurs only from specialized, telomere-proximal transcription units termed bloodstream expression sites (BESs). For archival genes, activation therefore involves duplication into a BES at the expense of the previously transcribed VSG, which is lost. Even pseudogenes can be activated this way, as part of their sequence can contribute to the formation of mosaic genes. Although it is known that homologous recombination participates actively in VSG switching, it is thought that limited sequence homology within the coding sequence is involved in the formation of mosaic genes, and an important function of a database could be enabling identification of such sequence homologies [15]. Duplication of intact VSGs apparently can utilize homologous, imperfect repeats upstream of most VSGs and can end at the other flank in conserved sequences towards the 3' end of the coding region or further downstream, where the 3' untranslated region is encoded. Sometimes the incoming gene duplicate inherits part of the C-terminal domain encoding sequence from the VSG already in the BES.

The VSG archive is very diverse, to the extent that different trypanosome strains have widely different gene sets. How the archive evolves is unknown, but it is evident from the dispersed nature of the VSG gene arrays [4], and from analysis of duplication events within the archive (LM, JDB, unpublished), that homologous recombination, involving primarily gene conversion, plays a major role. It is now becoming clear that subtelomeres of various organisms, including humans [16], are preferential sites for the rapid evolution of multigene families [17], possibly due to the preferred use of particular recombination mechanisms [18]. Due to the availability of the sequence of most VSGs in the silent gene arrays, the trypanosome has now become an experimentally tractable paradigm for the role of subtelomere recombination in multigene family diversification. Thus, ready accessibility to the individual gene sequences in a dedicated database can help address a number of questions about chromosomes and recombination, such as:

- How do sequences spread and diversify within and between subtelomeres?
- What is the contribution of partial gene conversion, (micro)homologous recombination and point mutation to diversification of the gene family?
- What is the rate of evolution of coding sequences and of pseudogenes that can donate partial coding information?

Construction and content

VSGdb has been constructed as a specialised database to store a definitive, annotated set of VSG sequences that can be retrieved for analysis at the nucleic acid or peptide level. The source sequence data include the genome sequence and other cDNA and genomic sequences in public databases. For the genome project sequences, final annotation was achieved through Artemis [19] and VSGdb shares all identifiers and annotation with the genome project database at geneDB [20]. Of necessity, due to the limits of empirical knowledge, annotation in the genome project is parsimonious, with features being scored as negative if there is any doubt. This applies in particular to GPI anchor signals, where manual annotation based on known VSG sequences was undertaken, allied with the parameters of the bigPI [21] and DGPI [22] prediction programmes. The GPI signals appear to be much more varied amongst the silent array genes than in the expressed set of genes available to date, so a conservative annotation approach was taken, envisaging stringent requirements for expression. As biochemical knowledge improves, it will be possible to evaluate more directly the stringency of this crucial surface anchoring process, and indeed VSGdb should be a catalyst for such biochemical study. A parsimonious approach has been taken also for the potential structure of the C-terminal domain. It has been assumed that a full set of four or eight cysteines would be a requirement for a correctly folded C-terminal domain to contribute to an expressed VSG: there are only few examples of C-terminal domains completely devoid of cysteines, and as yet no instance of domains with one to three cysteines has been recorded. Therefore, intact domains lacking these conserved features have been described as "atypical", in a grey area that is intermediate between putative functional and pseudogene domains.

VSGdb is freely accessible [23]. The user interface consists of web pages allowing users to query the database in several different ways. Queries are handled by CGI scripts that extract information from the source files and return them to the user in dynamically created web pages. Figure 1 shows the flow of information to and from the database. The source files are essentially of two types. The first type constitutes the majority of files currently present in the database and are EMBL-format sequence files [24] of chromosomes 1–11 and contigs of *Trypanosoma brucei* stock TREU 927, the genome of which has been sequenced [20]. Primary annotation of VSGs was carried out via the Artemis annotation software, and secondary annotation of the sequence files was then carried out manually to enable extraction of sequences of different parts of the VSGs and to make the source files parsable by BioPerl [25] modules. Details of the annotation can be seen in Figure 2. The second type of source files comprises EMBL-format sequence files of VSG cDNAs and genomic DNA

sequences, some of which have been annotated, independently, from various trypanosomal species. These files were obtained from public databases.

The CGI scripts that process queries and return results were written in Perl and utilize BioPerl modules that parse EMBL-format files, and return sequence information quickly and accurately. It is due to the availability of these modules that it was decided initially to run the scripts on sequence files, but it is planned to move across to a standard database management system like MySQL. We shall update the database annually.

Utility and discussion

The following features are available in the VSGdb for both the annotated genome project VSGs and the non-genome project VSGs:

- **VSG sequence retrieval** – The user is allowed to choose from several parameters and retrieve sequence data from the source files in the popular FASTA format [26]. For the annotated VSGs, these parameters include the type of sequence (DNA or protein), one or all chromosomes or contigs, the type of VSG (functional, pseudogene, etc.), the part of the VSG (full-length VSG, N-terminal domain, etc.), and the types of N- and C-terminal domains. Selection for N-terminal domains allows retrieval of the whole set of putative α -helical coiled coil sequences, or subsets thereof. For the VSGs not from the genome project, only the full sequence is returned. The user can select sequences from this set on the basis of species and strain/VSG repertoire, and can choose linkage to other databases. Figure 3 shows a screenshot of the web form where the user can select parameters.

- **VSG BLAST** – Using the same list of parameters as above, the user has the option of using the sequences

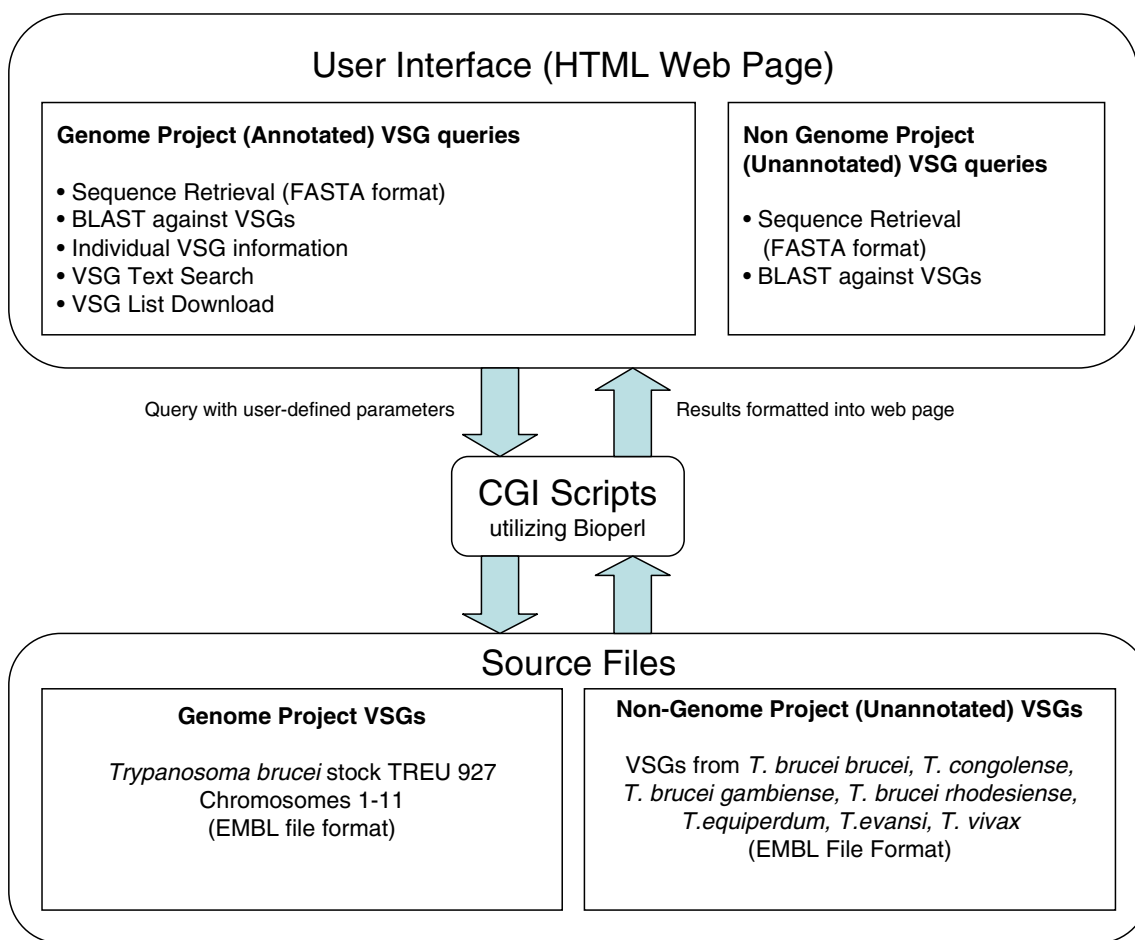


Figure 1 Information flow in a VSGdb query. The user inputs a query with defined parameters through a web interface, which is processed by CGI scripts to extract information from the source files and present the results as a web page.

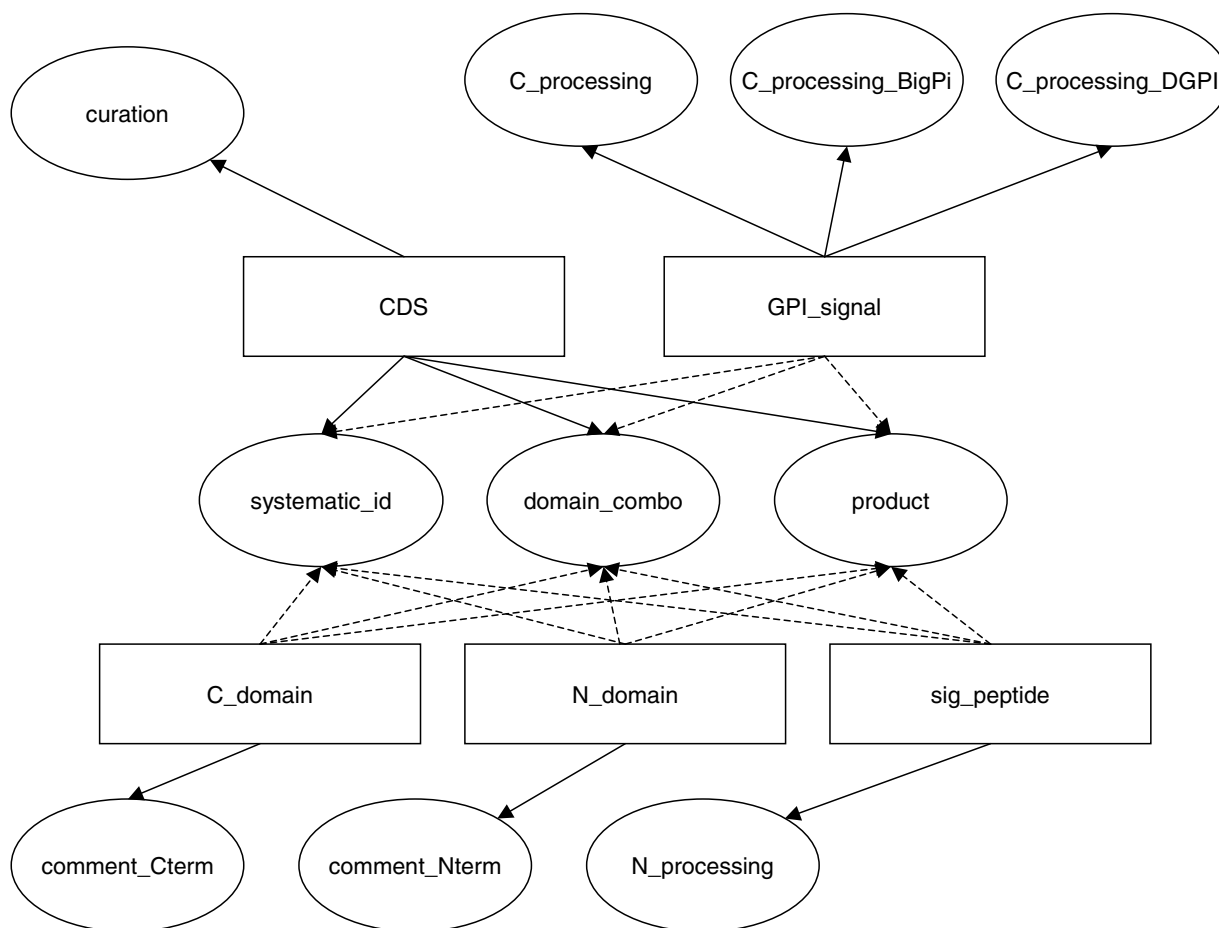


Figure 2

Annotation components utilized by VSGdb. The rectangles represent the primary tags which define parts of a VSG (CDS defines an entire VSG). The ovals represent features of each primary tag. Secondary annotation was carried out to link the systematic_ID, domain_combo and product features of each VSG to all of its components. These features are used in the sequence retrieval (as FASTA or as a database for BLAST). All features are used in the construction of individual VSG pages. Features containing curation and comments are used in the text search.

retrieved as a temporary database against which to BLAST [27] their own query sequence. For genome project VSGs, the BLAST reports also include hyperlinks to GeneDB CDS Info [20,28], and VSGdb pages for individual VSGs.

VSGdb also has the following features:

- **VSG Text Search** – This utility searches through annotation terms containing comments, curation and other information for search terms input by the user. Currently it treats multiple terms as one, analogous to putting quotes around search terms in a Google search. It outputs a list of identifiers of VSGs where there are hits, each of which is a hyperlink to the individual VSG web page.
- **VSG List Download** – This allows a user to input a list of VSG identifiers and retrieve various types of data

regarding those VSGs. This is available only for genome project VSGs, since they are annotated.

- **Individual VSG Web Pages** – These contain all the annotation and sequence data regarding any one VSG and are accessible either by user-input of the VSG identifier or through result pages of the VSG BLAST and text search utilities described above. Again, these are available only for genome project VSGs, since they are annotated. The pages also have links to the corresponding GeneDB CDS info pages. Figure 4 shows a screenshot of an individual VSG page.

It is our aim that this facility will be of general and specific use. Its quality and development depend on feedback from users, and all contributions and suggestions will be most welcome.

The screenshot shows the VSGdb website in a Windows Internet Explorer browser. The page title is "VSGdb: A database of trypanosomal variant surface glycoproteins". It features logos for WCMP (The Wellcome Centre for Molecular Parasitology) and the University of Glasgow. A search bar at the top contains the URL "http://www.vsgdb.org/". Below the search bar, there is a "Go to VSG" field with a placeholder "(type in id, eg Tb927.1.520)" and a "VSG Text Search" button. The main content area is titled "Genome project VSGs - Sequence Retrieval and BLAST" and includes a note: "**Currently only chromosomes 1-11 of *Trypanosoma brucei* stock TREU 927.**". The instructions state: "Choose search parameters and retrieve VSG sequences in FASTA format, or create a database of VSGs to BLAST your query sequence against." The search parameters are: "Type of Sequence: Protein", "Chromosome No.: 6", "Type of VSG: Only Pseudogene VSGs", "Part of VSG: N Domains- only full length pseudo", "Type of N Domain: A", and "Type of C Domain: 2". There are two buttons: "Get FASTA sequences" and "Create database for BLAST". On the left, there is a "Navigation" menu with links to Home, News, What are VSGs?, Search the VSGdb, How to use this site, and Links. Below the menu is a 3D protein structure visualization. At the bottom left, it says "Page author: Suraj Menon" and "Maintained by: Jon Wilkes".

Figure 3

Screenshot of the web form to define parameters for retrieval of sequences from genome project (annotated) VSGs in FASTA format or as a temporary database against which to BLAST a query sequence. For non-genome project (unannotated) VSGs, only the type of sequence and the organism need to be specified.

Conclusion

VSGs and their genes are important for trypanosome survival and growth, but also display features of general biological interest. Because the family has expanded and diversified very extensively, it is a unique biological resource for addressing questions about protein structure, evolution and genetic mechanisms. VSGdb allows all VSG sequence and annotation data to be accessed via a user-friendly, web-based interface. The database can be queried using various criteria, and retrieval, at either protein or nucleotide level, includes specific information on each VSG, especially the large set fully annotated in the genome project. Retrieval of subsets as temporary databases allows further detailed analyses. Besides contributing to general areas of biology, VSGdb should help enhance our understanding of trypanosome biology.

Availability and requirements

Project name: VSGdb: a database of trypanosomal variant surface glycoproteins

Project home page: <http://www.vsgdb.org/>

Operating systems: HTML 4.x-compliant browsers

Programming language: server side processing via perl; server Apache 2.0.40

Licence: none

Authors' contributions

LM provided community-end annotation of all genome project VSG sequences and interacted with programmers to develop the database and user interface. SM, PW and JMW programmed the database and interface and installed the sequence files. NGJ and MC provided specialist knowledge on VSG sequences. JDB oversaw the project and helped define user requirements. All authors read and approved the final manuscript.

WVMP
the WILLCOMBE CENTRE FOR
MOLECULAR PARASITOLOGY

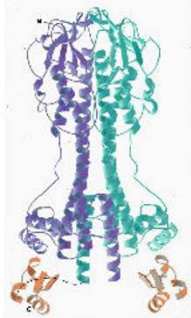
VSGdb
A database of trypanosomal variant surface glycoproteins

UNIVERSITY of GLASGOW

Go to VSG (type in id, eg Tb927.1.520)

Navigation

Home
News
What are VSGs?
Search the VSGdb
How to use this site
Links



VSGdb Page for VSG Tb927.1.520

PRODUCT = variant surface glycoprotein (VSG, pseudogene), putative, fragment
DOMAIN_COMBO = ?1
CURATION = truncated by INGI

DNA SEQUENCE:
GGGACCTTACGTGTCTGCTCTGCTCAAACCCAGAACGGCAACCTCTGCGGAGGCACAACG
GGGGCCGGAACGTATGGCAAAGGCGGCACAGACACGGCAGCGCAGGCACTAAAAGTGTGG
GAAACAATCAAACCACTCTGTGAAGCGGCAGCGGTGGAACCTGAAACCAGGAAACCGGCG
ATAACAACATGTATTCCCGGTTTTAAGGCCGCCCTCGGTGCCGGAATAACCGCAGCCGGC
GACAGCGCGCTAAAAGCTAGGCAACGGCGCAGCAACAAACATGGTGAATGACGGCGGATTC
GCTTCAGCATCGCTAGAATACATCCAAAAGGCGAGCCGAAAGGCGCGGTGGGGAAGATACT
CGCGTGAACAATGCAACTAGCCGCGGCTGCGCTTGACAAAGGAAAGTTGAGAGCAGCGCA
ACTAACGGCTGCACGAAACGCAAAATGCAGAAAAGCAAGCTCAATAACATAGCGAAATTTGA
GCTTCCCATTTTCTGTGGCGGCTCCTGCCGCCAGCAGCGCAACAGCAAAACTTGACACA
GCAACCGAAGGAGGCTGCAACAACACACAAAAAATGAATGCAATGACCCCTGCAAAATGG
GAAGAGGATGCGGCTGACAAAGACAAAACTGCTCATTAGACACCAAAAAGGCAGCAGAA
CAACAAGCAACCCAAAGCAGGAAAAGGAGGAAAGACTTCAAAGTGTCCAAATAAGAAAGCT
GAGGCCGAGTGTAAATCTCCGGGTTGTAAGTGGGATGGAAACACTTGCAAGAAATTCAGT
TTCTTGTAAAGAAATTTGCGCTGAGTATGGCTGCTGCCTTTATCAACTTGGTAGTATT
TAA

PROTEIN SEQUENCE:
GTLRVCSAQTQNGNLGGTTGAGTYGKGGTDAAQALKVWETIKPLCEAAAVELETTEPA
ITTCIPGFKAALGAGITAAAGDSALKLGNAAATWCNDADSASACVEYIQKAAEGAVGKIL
RVKHATSRGCACTREVESSATNGCTNANAESKLNNIAKFGASHFLAAPAASSATAKLD
ATEGGCNKHKNECNDPCKWEEDAADKDKNCSLDTKKAEEQQTQAGKGGKTSKCSNKK
EAECKSPGCKWDGNTCKNFSFLVKNFALSMAAAF INLVVF

N DOMAIN:
GTLRVCSAQTQNGNLGGTTGAGTYGKGGTDAAQALKVWETIKPLCEAAAVELETTEPA
ITTCIPGFKAALGAGITAAAGDSALKLGNAAATWCNDADSASACVEYIQKAAEGAVGKIL
RVKHATSRGCACTRE

COMMENT_NTERM = N?, incomplete, truncated by INGI upstream

C DOMAIN:
VESSATNGCTNANAESKLNNIAKFGASHFLAAPAASSATAKLDTEGGCNKHKNECN
DPCKWEEDAADKDKNCSLDTKKAEEQQTQAGKGGKTSKCSNKKAEAECKSPGCKWDGNT
CKNFSFLVKNFALSMAAAF INLVVF

COMMENT_CTERM = C1, degenerate

GPI SIGNAL:
FSFLVKNFALSMAAAF INLVVF

Page author:
Suraj Menon
Maintained by
Jon Wilkes

Figure 4
Screenshot of an individual VSG page (here Tb927.1.520). It contains all the sequence and annotation information available within the VSGdb for this particular VSG. Individual VSG pages can be accessed by typing in the VSG identifier in the searchbar or through the results of the text search and VSG BLAST functions. These are available only for annotated VSGs.

Acknowledgements

We acknowledge extensive input from our colleagues in the Sanger Institute and The Institute for Genome Research, and in particular Matt Berriman, Christiane Hertz-Fowler, Hubert Renauld, and Najib El-Sayed and Gaelle Blandin. We acknowledge also the input about the website from an anonymous reviewer. We thank the Wellcome Trust for funding: JDB is a Wellcome Trust Principal Research Fellow.

References

- Cross GAM: **Antigenic variation in trypanosomes - secretes surface slowly.** *Bioessays* 1996, **18**:283-291.
- Ferguson MAJ: **The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research.** *J Cell Sci* 1999, **112**:2799-2809.
- Barry JD, McCulloch R: **Antigenic variation in trypanosomes: Enhanced phenotypic variation in a eukaryotic parasite.** *Adv Parasitol* 2001, **49**:1-70.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, MacLeod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpborg AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El Sayed NM: **The genome of the African trypanosome *Trypanosoma brucei*.** *Science* 2005, **309**(5733):416-422.
- Carrington M, Miller N, Blum M, Roditi I, Wiley D, Turner M: **Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues.** *J Mol Biol* 1991, **221**(3):823-835.
- Blum ML, Down JA, Gurnett AM, Carrington M, Turner MJ, Wiley DC: **A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*.** *Nature* 1993, **362**(6421):603-609.
- Allen G, Gurnett LP: **Locations of the six disulfide bonds in a variant surface glycoprotein (VSG 117) from *Trypanosoma brucei*.** *Biochem J* 1983, **209**:481-487.
- Miller EN, Allan LM, Turner MJ: **Topological analysis of antigenic determinants on a variant surface glycoprotein of *Trypanosoma brucei*.** *Mol Biochem Parasitol* 1984, **13**:67-81.
- Miller EN, Allan LM, Turner MJ: **Mapping of antigenic determinants within peptides of a variant surface glycoprotein of *Trypanosoma brucei*.** *Mol Biochem Parasitol* 1984, **13**(3):309-322.
- Masterson WJ, Taylor D, Turner MJ: **Topologic analysis of the epitopes of a variant surface glycoprotein of *Trypanosoma brucei*.** *J Immunol* 1988, **140**(9):3194-3199.
- Bohme U, Cross GAM: **Mutational analysis of the variant surface glycoprotein GPI- anchor signal sequence in *Trypanosoma brucei*.** *J Cell Sci* 2002, **115**(4):805-816.
- Callejas S, Leech V, Reitter C, Melville S: **Hemizygous subtelomeres of an African trypanosome chromosome may account for over 75% of chromosome length.** *Genome Res* 2006, **16**(9):1109-1118.
- Williams RO, Young JR, Majiwa PAO: **Genomic environment of *T. brucei* VSG genes - presence of a minichromosome.** *Nature* 1982, **299**:417-421.
- Alsford S, Wickstead B, Ersfeld K, Gull K: **Diversity and dynamics of the minichromosomal karyotype in *Trypanosoma brucei*.** *Mol Biochem Parasitol* 2001, **113**(1):79-88.
- Barbet AF, Kamper SM: **The importance of mosaic genes to trypanosome survival.** *Parasitol Today* 1993, **9**(2):63-66.
- Linaridopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ: **Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication.** *Nature* 2005, **437**(7055):94-100.
- Barry JD, Ginger ML, Burton P, McCulloch R: **Why are parasite contingency genes often associated with telomeres?** *Int J Parasitol* 2003, **33**(1):29-45.
- Ricchetti M, Dujon B, Fairhead C: **Distance from the chromosome end determines the efficiency of double strand break repair in subtelomeres of haploid yeast.** *J Mol Biol* 2003, **328**(4):847-862.
- Artemis: a DNA sequence viewer and annotation tool** [<http://www.sanger.ac.uk/Software/Artemis/>]
- Trypanosoma brucei GeneDB** [<http://www.genedb.org/genedb/tryp/>]
- big-PI Predictor** [http://mendel.imp.ac.at/gpi/gpi_server.html]
- DGPI** [http://129.194.185.165/dgpi/index_en.html]
- VSGdb A database of trypanosomal variant surface glycoproteins** [<http://www.vsgdb.org>]
- The EMBL Nucleotide Sequence Database** [<http://www.ebi.ac.uk/emb/Documentation/>]
- BioPerl** [http://www.bioperl.org/wiki/Main_Page]
- Fasta format description** [<http://www.ncbi.nlm.nih.gov/blast/fasta.shtml>]
- Altschul SF, Gish W, Miller W, Myers EV, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
- Hertz-Fowler C, Peacock CS, Wood V, Aslett M, Kerhornou A, Mooney P, Tivey A, Berriman M, Hall N, Rutherford K, Parkhill J, Ivens AC, Rajandream MA, Barrell B: **GeneDB: a resource for prokaryotic and eukaryotic organisms.** *Nucl Acids Res* 2004, **32**(Database issue):D339-D343.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

