

## Supplementary Material

### Whole-genome sequence of the bovine blood fluke *Schistosoma bovis* supports interspecific hybridization with *S. haematobium*

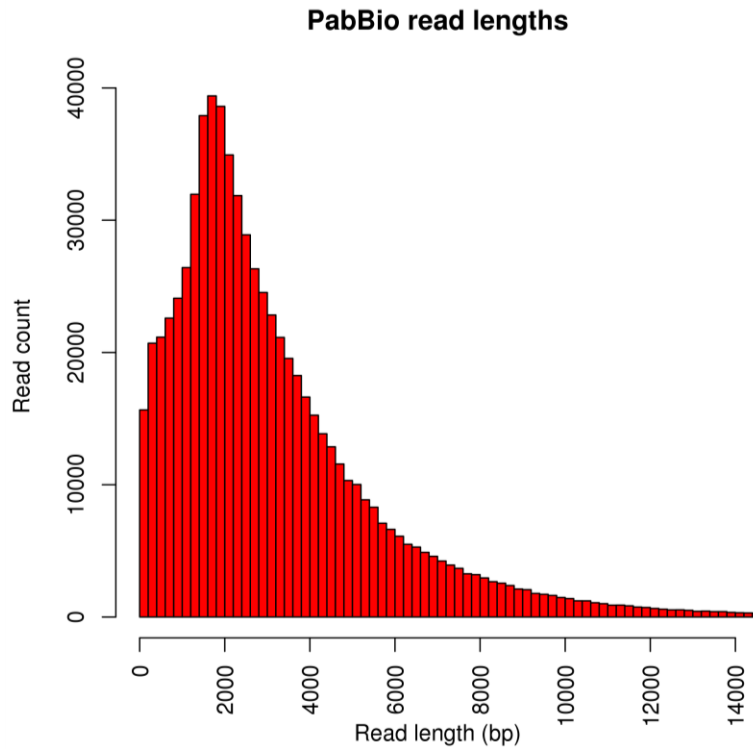
Harald Oey<sup>1</sup>, Martha Zakrzewski<sup>2</sup>, Kerstin Gravermann<sup>5</sup>, Neil D. Young<sup>4</sup>, Pasi K. Korhonen<sup>4</sup>, Geoffrey N. Gobert<sup>2,3</sup>, Sujeevi Nawaratna<sup>2</sup>, Shihab Hasan<sup>1,2</sup>, David M. Martinez<sup>2</sup>, Hong You<sup>2</sup>, Martin Lavin<sup>9</sup>, Malcolm Jones<sup>2,8</sup>, Mark A. Ragan<sup>6</sup>, Jens Stoye<sup>5</sup>, Ana Oleaga<sup>10</sup>, Aidan Emery<sup>7</sup>, Bonnie Webster<sup>7</sup>, David Rollinson<sup>7</sup>, Robin B. Gasser<sup>4</sup>, Donald P. McManus<sup>2,\*</sup>, Lutz Krause<sup>1,2,\*,#</sup>

## Supplementary Results

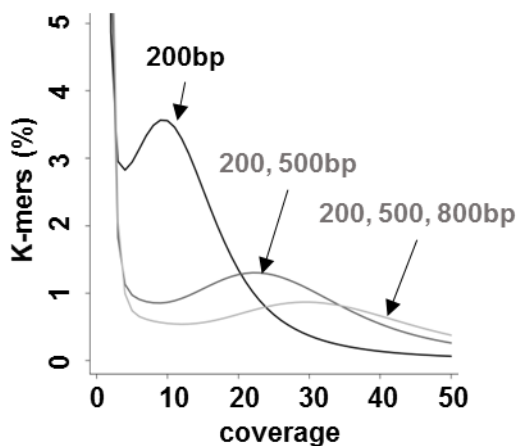
### Vaccine Targets

Schistosomes have been targeted for elimination by the World Health Organization, but despite global efforts to reduce disease infection, rates in ruminants and humans continue to be high in most endemic regions. No *Schistosoma* vaccines are currently available, and treatment relies on one drug, praziquantel [1]. Praziquantel chemotherapy is highly effective but even mass treatment does not prevent reinfection. Vaccines that induce long-term immunity represent a logical component for the future control of schistosomiasis with the final goal of complete elimination. Currently, only three schistosome vaccine antigens have entered human clinical trials. Two recombinant *S. mansoni* vaccines are in early stage clinical testing (Sm14 and Sm-TSP-2) and one *S. haematobium* specific vaccine (Sh28GST or Bilhvax) completed a Phase 3 trial [2, 3]. The *S. mansoni* molecule Smp80 (calpain) is undergoing testing in non-human primates and will likely advance to clinical development [3]. We identified highly conserved *S. bovis* orthologues of Sm14 (99% identity), Sh28GST (97.2%) and Sm-p80 (98.5%), and a moderately conserved orthologue of the tetraspanin Sm-TSP-2 (34%) (**Table S5**). Tetraspanins are composed of two domains, the short EC1 domain and longer EC2 domain. All four included schistosome species were conserved over the EC1 domain (65% identity), but showed two different variants of the EC2 domain. These results indicate that *Schistosoma* vaccine targets potentially show cross-reactivity for multiple *Schistosoma* species and represent promising vaccine target candidates against *S. bovis*.

## Supplementary Figures

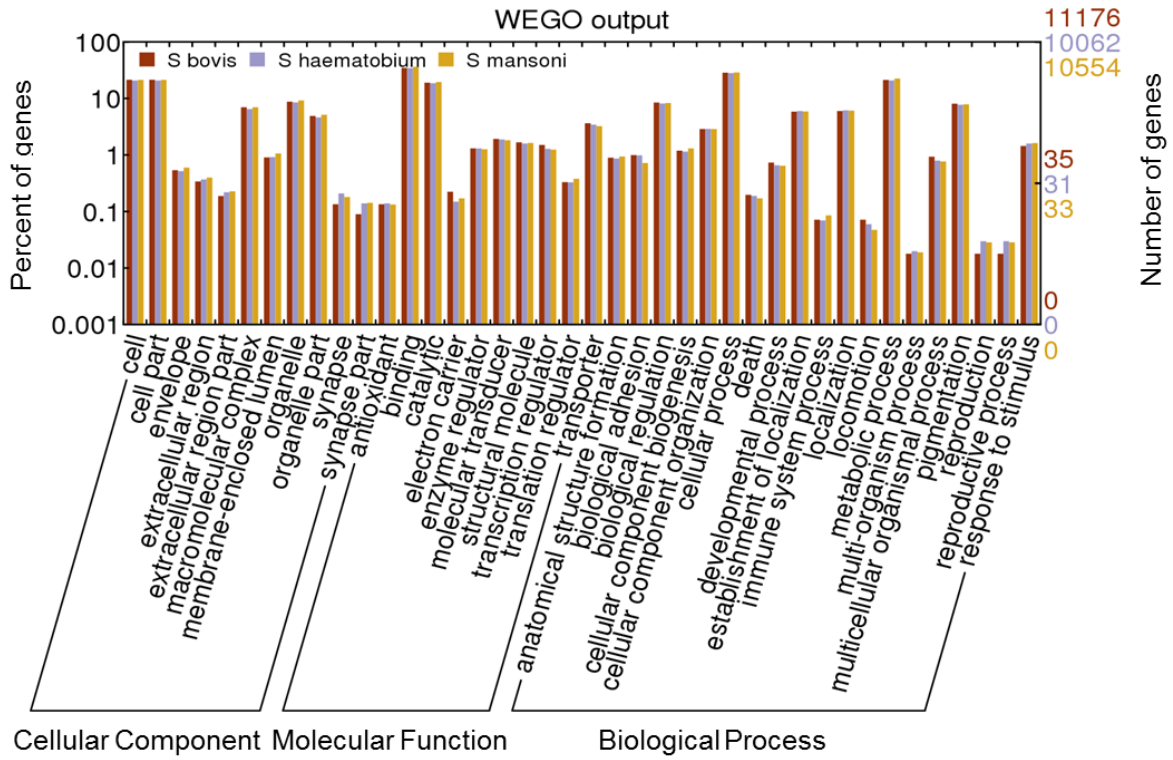


**Figure A:** PacBio read length distribution



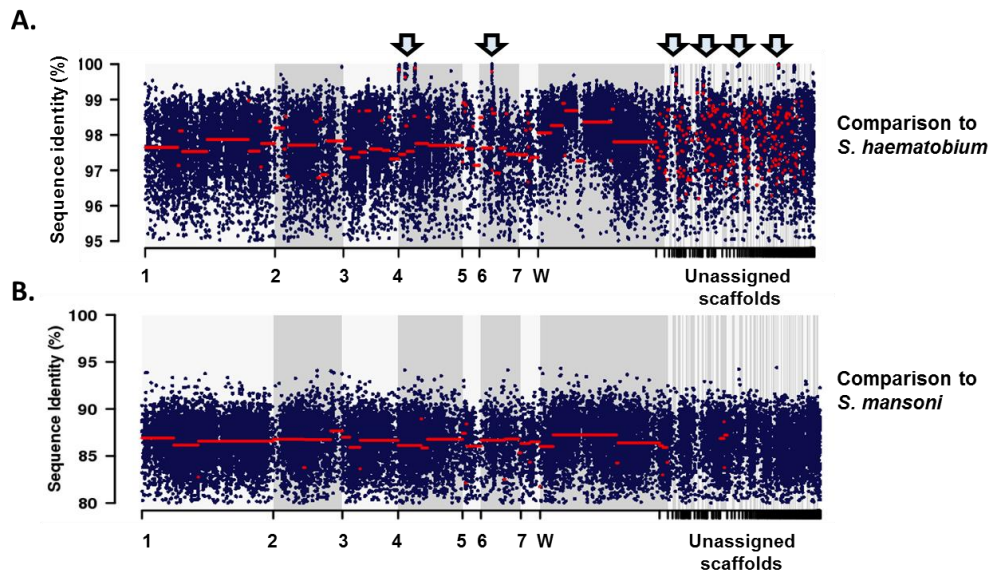
**Figure B:** Distribution of 17-mers in short insert Illumina libraries.

The x-axis represents k-mer coverage (17 base pairs), the y-axis the distribution (percentage) of k-mers with that coverage. The peak at 10 indicates that 3.5% of all detected 17-mers occurred about 10 times in the 200bp insert libraries. The frequency analysis demonstrated a low sequence heterozygosity of *S. bovis*.



**Figure C:** GO categories identified across three *Schistosoma* species.

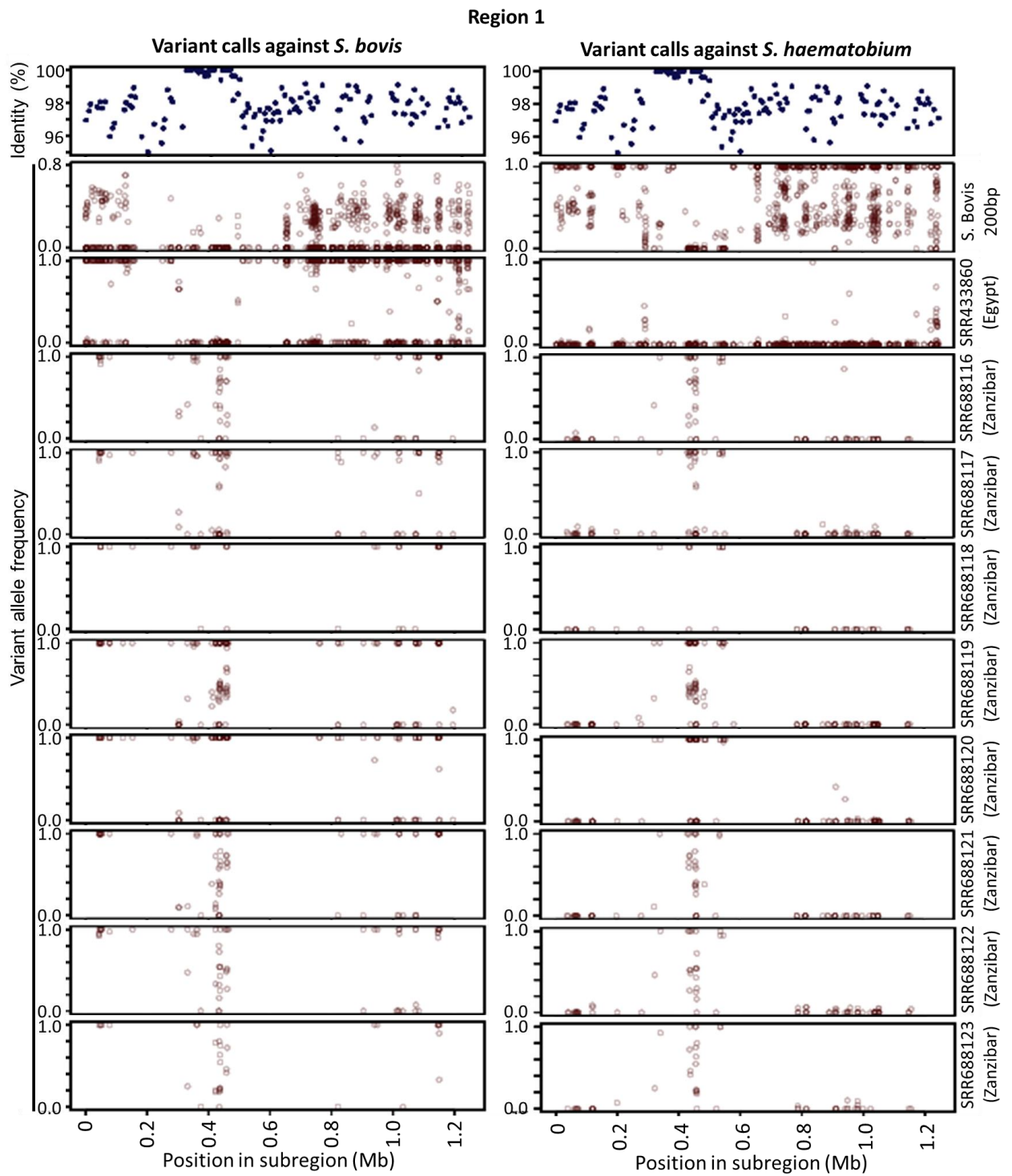
GO annotations were assigned from InterProScan results using WEGO.



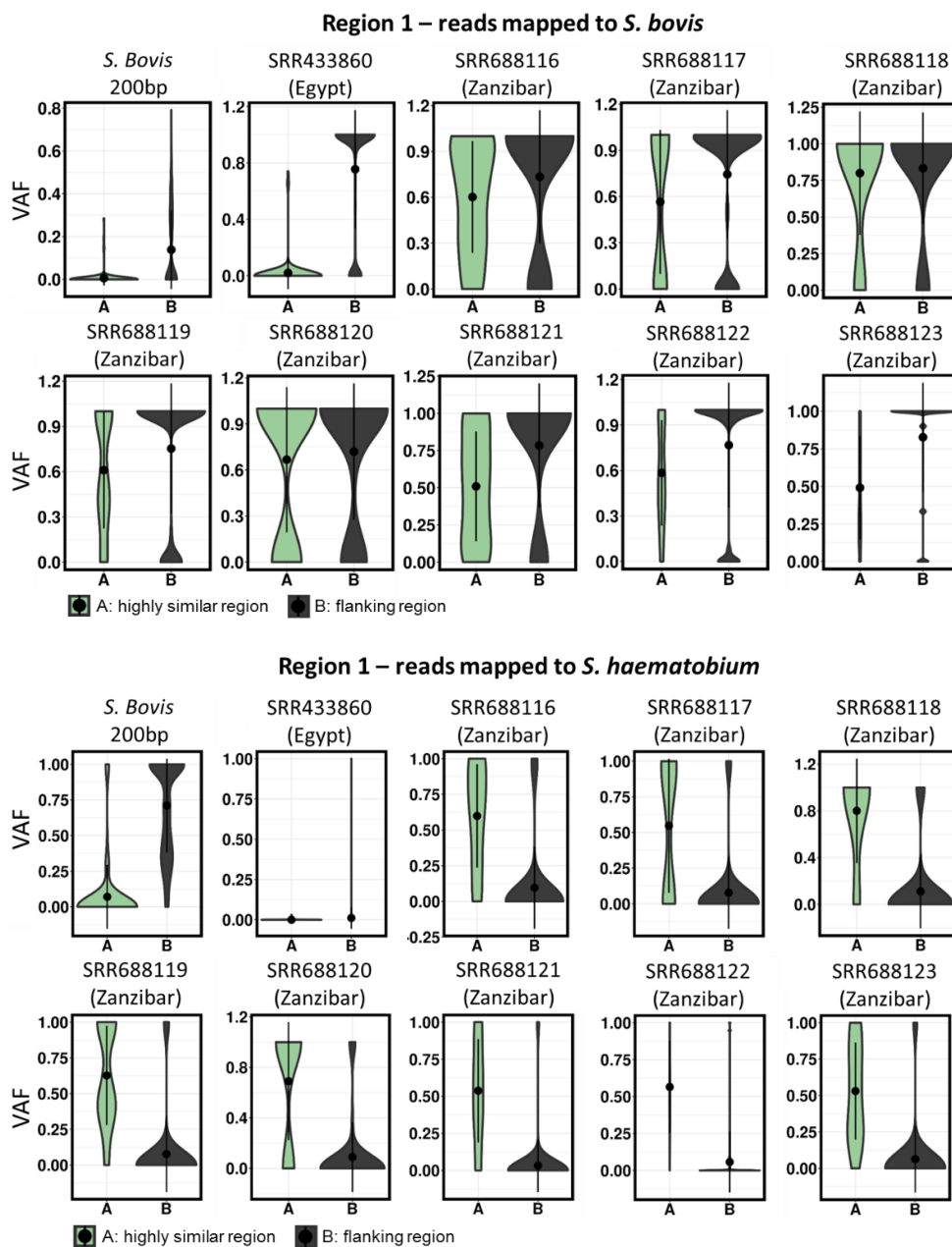
**Figure D:** Genome-wide comparison of *S. bovis* with *S. haematobium* and *S. mansoni*.

Genome-wide comparison of *S. bovis* and *S. haematobium*, panel (A), and *S. bovis* and *S. mansoni*, panel (B). Genomic similarity was measured in a sliding-window approach using a 5kb window size. Chromosome numbers were assigned by mapping *S. bovis* scaffolds to the *S. mansoni* genome assembly (autosomal chromosomes 1-7 and sex chromosome W). Several long contiguous and highly conserved regions (sequence-identity > 99%) provide evidence of recent inter-species hybridization and are marked with arrows. Red lines depict segments with similar level of similarity, which were found by fitting a piecewise constant curve using a least squares cost function.

A)



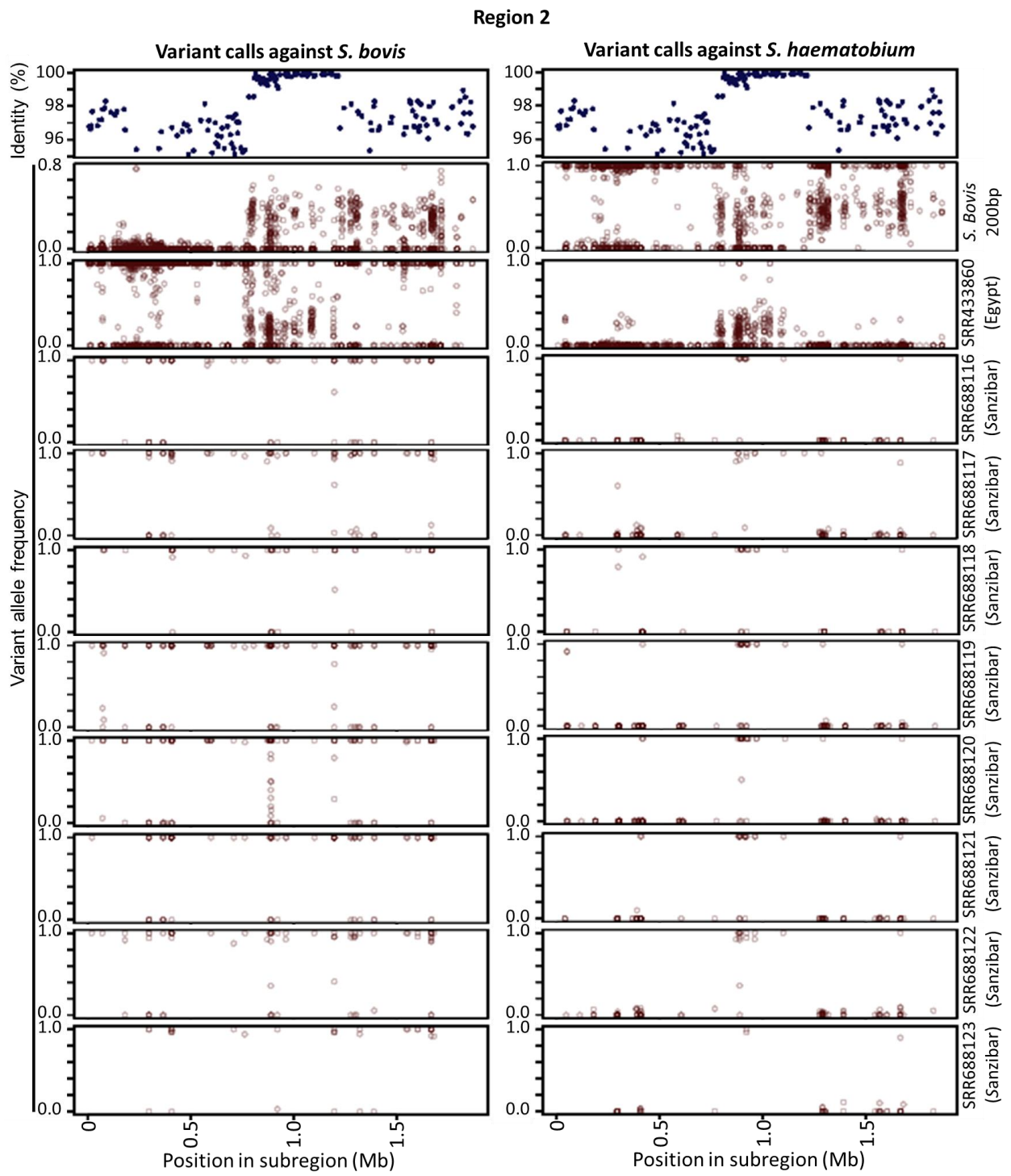
B)



**Figure E:** Variant allele frequencies for *S. bovis* and *S. haematobium* polymorphisms for region 1

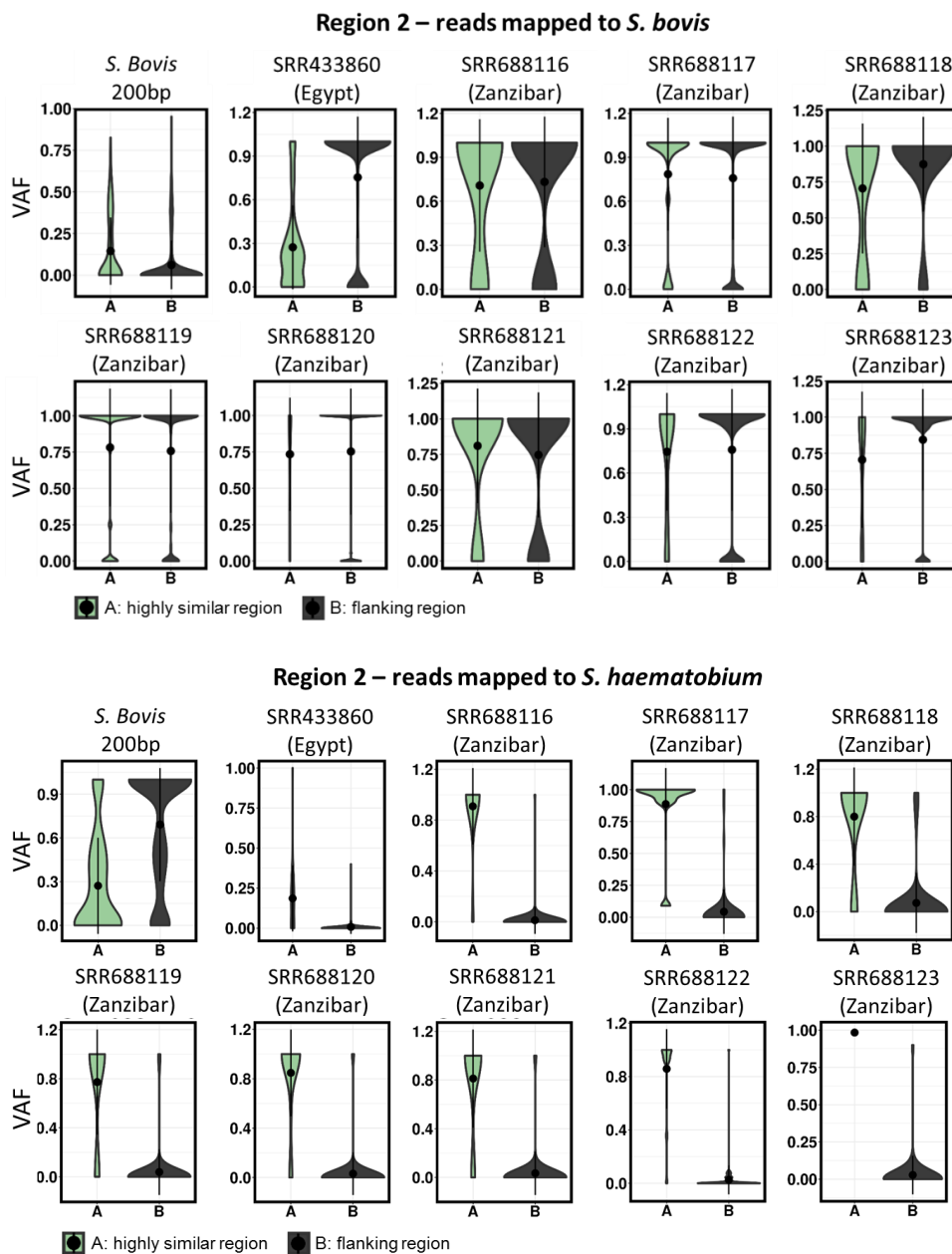
Panel (A) The figure shows the sequence identity scores for region 1, and the flanking regions (top panel) and the variant allele frequencies of variants identified by mapping whole genome sequencing reads (*S. bovis* and *S. haematobium*, Egypt isolate) and exome reads (*S. haematobium*, Zanzibar isolates) to either the *S. bovis* or the *S. haematobium* genome assemblies. Panel (B) Violin plots of the variant allele frequency distributions for variants within and flanking the highly similar region, as indicated.

A)





B)



**Figure F:** Variant allele frequencies for *S. bovis* and *S. haematobium* polymorphisms for region 2

Panel (A) The figure shows the sequence identity scores for region 2, and the flanking regions (top panel) and the variant allele frequencies of variants identified by mapping whole genome sequencing reads (*S. bovis* and *S. haematobium*, Egypt isolate) and exome reads (*S. haematobium*, Zanzibar isolates) to either the *S. bovis* or the *S. haematobium* genome assemblies. Panel (B) Violin plots of the variant allele frequency distributions for variants within and flanking the highly similar region, as indicated.



```

S.jap      -----MRTSTD-D
S.man      -----MYVDSKVIIST
S.bovis    -----
S.haem     MRTTDLVSPSVISSPQLFYCRSGNSTAASSNQSDIYQTNQSQLSNIITNQSDIRQQTESQ

S.jap      MVTSSTAHLIKHNFFILITLITWFTKLLQLLKI CLVFTVAKWCLSKRKS LRKAGWEA
S.man      IDITSILKFPPIEHWRIILLILLAVSLIYKLLQVVTICFKFTVKKWCF SKRRTL RKAGWEA
S.bovis    -DIGKILTFPIDHWRILLILFVSLINNQIIAVLKI CFKFTLKKWCLSKRRTL RQAGWEA
S.haem     LDIGKILTFPIDHWRILLILFVSLINNQIIAVLKI CFKFTLKKWCLSKRRTL RQAGWEA
          . . . * * * . . . : : : * * * * * : * * * * * : * * * * *

S.jap      IVTGASSGIGEAYAEELAKEGLNILLISNDESQRLVSE RISTDYHVETRIVVADFTQPD
S.man      VVTGASSGIGEAYAEELAKEGLNIMLISNDEEQLSLVANRIATTY NVQTRIVVADFTKHD
S.bovis    VVTGSSGIGEAYAEELAKEGLNILLISNDEEQLSAVSNRIANTY NVQTRIVVADFTKDD
S.haem     VVTGSSGIGEAYAEELAKEGLNIMLISYDEEQLSDVSNRIANTY NVQTRIVVADFTKDD
          : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

S.jap      SYDVIKPAIQLSTIACLVNNVGMGLPFELFVDFYSPNEQSIHDI IHCNIVSTMTHI
S.man      VYEIRPAVDQLSTIACLVNNVGMGLPFELFSGEINSPNEESIRNI IHCNILSAVTMTSI
S.bovis    VYEIRPAVDQLSTIACLVNNVGMALPFELFAGEVDS PNEESIRNI IHCNIMSTVTMTNI
S.haem     VYEIRPAVDQLSTIACLVNNVGMALPFELFAGEVDS PNEESIRNI IHCNIMSTVTMTNI
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

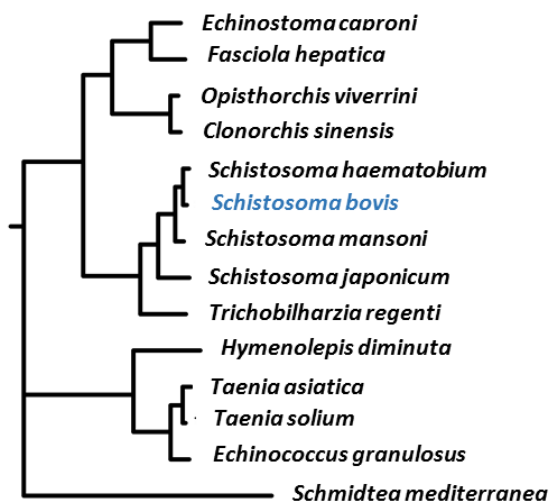
S.jap      VMPKMLSQKGSNPGIINISSYGLKEYPYASMYAATKAFIHQFS RCISAEKYSRNLIVQT
S.man      ILPKMLTQKEPNPGIINIASYGLKVFPYASLYASTKAAIIQFS RCVA AEKYKKNVIIQA
S.bovis    ILPKMLTQKGNPPIINIGSYAGLKVFPFATMYSSTKASIIQFS KCLAAEKYKKNVIIQA
S.haem     ILPKMLTQKGNPPIINIGSYAGLKVFPFATMYSSTKASIIQFS KCLAAEKYKKNVIIQA
          : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

S.jap      ICPLIVSTKLSYYKSTSFPIPTAQVFAK SALDMFGVQQTGGYMQHDLQAL IYNSMPNFL
S.man      ICPLFVSTNMTNLMKTTFFIPTAKVYAKNALDMYGV EQQTSGYFRHELKAYLSLLPTSV
S.bovis    ICPLCVSTNMTNQKTTFFTPSAKVFAK SALDMFGVEQQTGGYIRHDLQTYLYDLLPAII
S.haem     ICPLCVSTNMTNOVKTTFFTPSAKVYAK SALDMFGVEQQTGGYIRHDLQTYLYDLLPAII
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

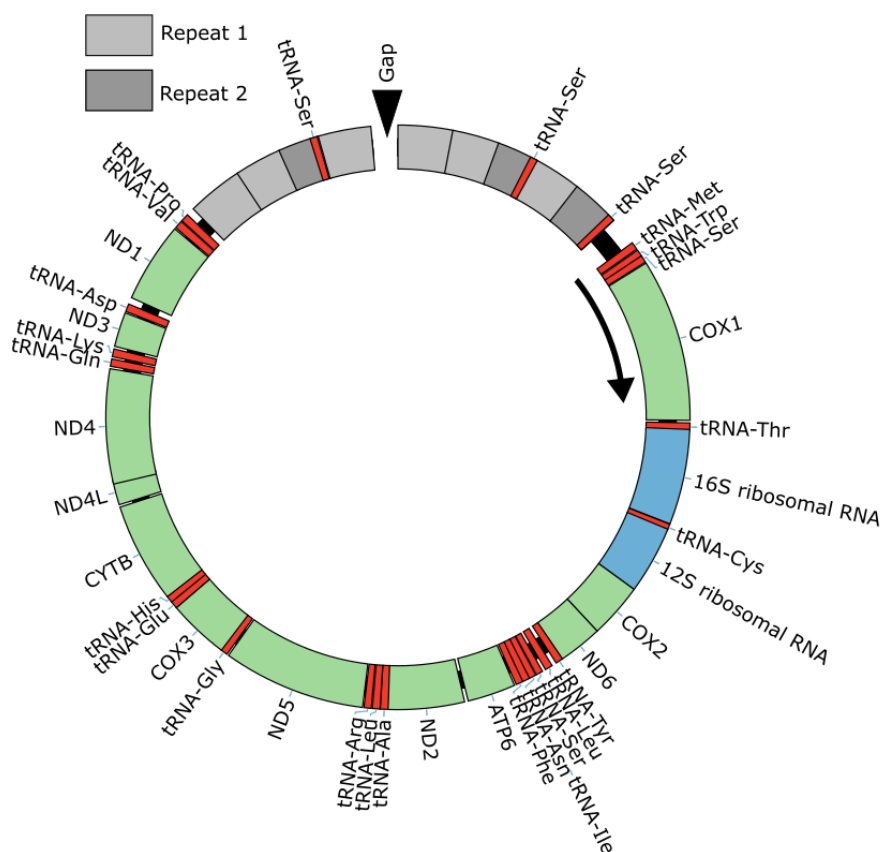
S.jap      WLRFRRRIRHRFMNESRKAK
S.man      FIRIIRLVAKSTLKKK----
S.bovis    SKQLMKLKANSQRNKA----
S.haem     SKQLMKLKANSQRNKA----
          . . . . . : : :

```

**Figure G:** Conservation of estradiol 17beta-dehydrogenase across multiple *Schistosoma* species.



**Figure H:** Phylogenetic tree reconstructed from the amino acid sequences of 52 shared single-copy genes using the maximum-likelihood method.



**Figure I:** Circular genome plot of the *S. bovis* mitochondrial genome.

The mitochondrial genome sequence was *de-novo* assembled from long PacBio sequence reads yielding a 20 kb contig. Shown are predicted proteins (green), rRNAs (blue) and tRNAs (red). All predicted genes are transcribed in the same direction, as indicated by the black arrow. A >4 kb repetitive region comprised of two distinct repeat units is depicted in grey. Repeat unit 1 (light grey) has ~520 bp; unit 2 (dark grey) has ~460 bp and contains a predicted tRNA for the amino acid serine. The repetitive region was not completely resolved in the assembly and the position of the gap is indicated at the top of the circle.





```
*****
PacBio read   GTATGGTTAATTAGTTCTATTTCGTGTATCCTTTTCTAGTGTGATGTTTGAGATGATACT
SB Mito rep   GTATGGTTAATTAGTTCTATTTCGTGTATCCTTTTCTAGTGTGATGTTTGAGATGATACT
*****

PacBio read   GATGTGTATAATAATTCTGTTTGGTTTAAATGTACGGTAGCTATGGGAAACCTGAC---A
SB Mito rep   GATGTGTATAATAATTCTGTTTGGTTTAAATGTACGGTAGCTATGGGAAACCTGACATTAA
*****

PacBio_read   TAGAA-TATGATATTTGTAGTACCTCTGGTTTACATCATATGACTTATAGTGTACTAAG
SB_Mito_rep   TAGAATATGATATTTGTAGTACCTCTGGTTTACATCATATGACTTATAGTGTACTAAG
*****

PacBio_read   TGAGAGTAATCGAACACCTTGTGATTATGTAGAGTCCGAGAGTGAATTAGTAAGAGGGAT
SB_Mito_rep   TGAGAGTAATCGAACACCTTGTGATTATGTAGAGTCCGAGAGTGAATTAGTAAGAGGGAT
*****

PacBio_read   TAGAGTAGAATATAGGAGTA-TTTTTTTTAAATATATTTGCTTGTGAATATTTAATTAT
SB Mito rep   TAGAGTAGAATATAGGAGTATTTTTTTTTTAAATATATTTGCTTGTGAATATTTAATTAT
*****

PacBio_read   GTTTATATTTAGTTGAGTGAGGTTTACGATATTTTGAAGTATAAATGAGATGTTAATAGT
SB Mito rep   GTTTATATTTAGTTGAGTGAGGTTTACGATATTTTGAAGTATAAATGAGATGTTAATAGT
*****

PacBio_read   AATGAATTTGATGCTGTTTATAATGATGCGTGGTTCGTTTCCCTCGACTTCGGTTTGATGT
SB Mito rep   AATGAATTTGATGCTGTTTATAATGATGCGTGGTTCGTTTCCCTCGACTTCGGTTTGATGT
*****

PacBio_read   ATTTGTTTCAGTGATTTGAGAGTATTGTGTTATGGTAATATTGATATATTTAATATGTTT
SB Mito rep   ATTTGTTTCAGTGATTTGAGAGTATTGTGTTATGGTAATATTGATATATTTAATATGTTT
*****

PacBio_read   ATTTAGAATATTATAATGAGTCTATTTTAGTATCTGTAGTAAAATTGATTAGCGTAGTAG
SB Mito rep   ATTTAGAATATTATAATGAGTCTATTTTAGTATCTGTAGTAAAATTGATTAGCGTAGTAG
*****

PacBio_read   TTAAGTTTAACGTTTG-TTTACACACAAAAGATGAAGTAAGGTTTGGATGCTTCCGG
SB Mito rep   TTAAGTTTAACGTTTGTTTTACACACAAAAGATGAAGTAAGGTTTGGATGCTTCCGG
*****

PacBio_read   CGCCATAATTATTTTTTAAATTTAGTTTGATCAGTTAAGTAGTTTGTATAAAGAATATA
SB_Mito_rep   CGCCATAATTATTTTTTAAATTTAGTTTGATCAGTTAAGTAGTTTGTATAAAGAATATA
*****

PacBio_read   AGCTTTGGGTGCTTAAGGATCTATTTTAAAGTAGACTTACTGAGGTTATAGGTTAAAG
SB Mito rep   AGCTTTGGGTGCTTAAGGATCTATTTTAAAGTAGACTTACTGAGGTTATAGGTTAAAG
*****

PacBio_read   TAGTAAAAAGTATATAAGGCTTATAGTAAGCTTCACCATCGAATGGTGAGTAAGATTTAA
SB Mito rep   TAGTAAAAAGTATATAAGGCTTATAGTAAGCTTCACCATCGAATGGTGAGTAAGATTTAA
*****

PacBio_read   AA---TCTCTTATAGTC-----TTGTTTTGTTTTTTTTTTGGTGGTGGTAAAGTTTAA
SB_Mito_rep   AAATCTCTATTTATAGTCTTGTTTTCTTTTTGTTTTTTTTTTGGTGGTGGTAAAGTTTAA
**   **   *****   **

PacBio_read   TTGTATTTAGTTCCAATAAAAAATTTTTGTAAAAAGTATGTTTAAATTCATTATATAT
SB Mito rep   TTGTATTTAGTTCCAATAAAAAATTTTTGTAAAAAGTATGTTTAAATTCATTATATAT
*****

PacBio_read   AATAT---TTATATAATATAATATAATGTAAAATAAGTATAGTATAGTA-GGGGAAAA
SB Mito rep   AATATATTATAATATAATATAATATAATGTAAAATAAGTATAGTATAGTAGGGGGAAAA
*****

PacBio_read   ATAATAAGGTGAACGTATTGCATACCTATAGGCAAATGTGATTGTATTATCTGTAGAATG
SB_Mito_rep   ATAATAAGGTGAACGTATTGCATACCTATAGGCAAATGTGATTGTATTATCTGTAGAATG
*****

PacBio_read   GGTGTAGGCAAATATGATTATACTCCTTAAATTTGGTTTAGTCTATAGGGTTGTATTTAG
SB_Mito_rep   GGTGTAGGCAAATATGATTATACTCCTTAAATTTGGTTTAGTCTATAGGGTTGTATTTAG
*****

PacBio_read   AGTTCGTAGTTTAAAGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTATG
SB Mito rep   AGTTCGTAGTTTAAAGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTATG
*****
```

PacBio read TTAGTCTATAGAGGGTGTATTTA-GGTTTATGTTAAGGTCTAAGGATTCTAAAGCAAAT  
SB Mito rep TTAGTCTATAGAGGGTGTATTTAGGGTTTATGTTTAAAGGTCTAAGGATTCTAAAGCAAAT  
\*\*\*\*\*

PacBio read GTGATTGTACTCCTAAAATTATGTTAGTCTATAGA-GGTGATTTTAAAGATAAATGCAT  
SB Mito rep GTGATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTTAAAGATAAATGCAT  
\*\*\*\*\*

PacBio read TTTAATTTCTTGTAAAGGGGATGTACAACAAATTTGTTTAAATTTGGATTTTATGCTTTA  
SB Mito rep TTTAATTTCTTGTAAAGGGGATGTACAACAAATTTGTTTAAATTTGGATTTTATGCTTTA  
\*\*\*\*\*

PacBio read ATATTGCTTTC AATTGAGGTA AATTTAATGATATTTTTATATTTTTATGTATTCAATTGT  
SB Mito rep ATATTGCTTTC AATTGAGGTA AATTTAATGATATTTTTATATTTTTATGTATTCAATTGT  
\*\*\*\*\*

PacBio read GGTATATTTTTAAGGTATAAAAAGTTGTTGTTTCTGTAATCTTTTTGTTTTTTTTGGT  
SB Mito rep GGTATATTTTTAAGGTATAAAAAGTTGTTGTTTCTGTAATCTTTTTGTTTTTTTTGGT  
\*\*\*\*\*

PacBio read GGTGGTAAGTTTAGTTGATTTTTAGTTCCAATAAAAATTTTTTGTA AAAAGTATGTTT  
SB Mito rep GGTGGTAAGTTTAGTTGATTTTTAGTTCCAATAAAAATTTTTTGTA AAAAGTATGTTT  
\*\*\*\*\*

PacBio read ATTCATTTATATAT-ATATATTATAATATAATATAATGTAAAATAAGTATAGTATAGTAG  
SB Mito rep ATTCATTTATATATAATATAATATAATATAATGTAAAATAAGTATAGTATAGTAG  
\*\*\*\*\*

PacBio read GGGGAAAAATAAATAAGGTGAGCGTATCGCATACCTATAGGCAAATGTGATTGTATTATCT  
SB Mito rep GGGGAAAAATAAATAAGGTGAGCGTATCGCATACCTATAGGCAAATGTGATTGTATTATCT  
\*\*\*\*\*

PacBio read GTAGAATGGGTG TAGGCAAATATGATTATACTCCTTAAATTTAGTCTATAGGGGTT  
SB Mito rep GTAGAATGGGTG TAGGCAAATATGATTATACTCCTTAAATTTAGTCTATAGGGGTT  
\*\*\*\*\*

PacBio read GTATTTAGAGTTCG TAGTTTAAAGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTA  
SB Mito rep GTATTTAGAGTTCG TAGTTTAAAGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTA  
\*\*\*\*\*

PacBio read AAATTATGTTAGTCTATAGAGGGTG--TTTAAAG--AAATGCATTTTAAATTTCTTGTA  
SB Mito rep AAATTATGTTAGTCTATAGAGGGTGTATTTAAAGATAAATGCATTTTAAATTTCTTGTA  
\*\*\*\*\*

PacBio read GGGGGATGTACAACAAATTTGTTTAAATTTGGA-TTTATGCTTTAATATTGCTTTCAATT  
SB Mito rep GGGGGATGTACAACAAATTTGTTTAAATTTGGATTTTATGCTTTAATATTGCTTTCAATT  
\*\*\*\*\*

PacBio read GA-GTAAATTTAATGAT-TTTTTTATATTTT-----  
SB Mito rep GAGGTAATTTAATGATATTTTTTATATTTTTATGATTCAATTGGTTATATTTTAAAG  
\*\* \*\*\*\*\*

PacBio read -----  
SB Mito rep GTATAAAAAGTTGTTGTTTCTGTAATCTTTTTGTTTTTTTTGGTGGTGGTAAGTTTAGT

PacBio read -----  
SB Mito rep TGTATTTTAGTTCCAATAAAAATTTTTTGTA AAAAGTATGTTTAAATTCATTTATATATA

PacBio read -----  
SB Mito rep ATATATTATAATATAATATAATGTAAAATAAGTATAGTATAGTAGGGGGAAAAATAATA

PacBio read -----  
SB Mito rep AGGTGAGCGTATCGCATTCCTATAGGCAAATGTGATTGATTATCTGTAGAATGGGTGTA

PacBio read -----  
SB Mito rep GGCAAATATGATTATACTCCTTAAATTTAGTCTATAGGGTTGATTTAGAGTTCGT

PacBio read -----  
SB Mito rep AGTTTAAAGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTTATGTTAGTCT

```
PacBio_read -----
SB_Mito_rep ATAGAGGGTGTATTAGGGTTTGTGTTTGTAGGAATGGTTAAATAGTTGGATAGAAAGGC

PacBio_read -----
SB_Mito_rep TTATAGTAAGCTTCACCATCGAATGGTGAGTAAGATTAAAAATCTCTATTTATATCTGT

PacBio_read -----
SB_Mito_rep TTTGTTTTGTFTTTTTTTTTTTTTTGTGGTGGTAAGTTTAGTTGTATTTTAGTTCCAATAAA

PacBio_read -----
SB_Mito_rep AATTTTTGTAAAAGTATGTTTTAATTCATTTATATATATATTATAAATAATATAATATAA

PacBio_read -----
SB_Mito_rep TAAATAAGTATAGTATAGTAGGGGGGAAAAATAATAAGGTGAACGTATTGCATACCTATA

PacBio_read -----
SB_Mito_rep GGCAAATGTGATTGTATTATCTGTAGAATGGGTGTAGGCAAATATGATTATACTCCTTAA

PacBio_read -----
SB_Mito_rep ATTGGTTTAGTCTATAGGGTTGATTTAGAGTTCGTAGTTAAGGTCTAAGGATTCTAAA

PacBio_read -----
SB_Mito_rep GCAAATGATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTTAGGGTTTATGT

PacBio_read -----
SB_Mito_rep TTAAGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTATGTTAGTCTATA

PacBio_read -----ATGTA-----
SB_Mito_rep GAGGGTATTTAAAGATAAATGCATTTAATTCTGTAAGGGGGATGTACAACAAATTTG
*****

PacBio_read -----
SB_Mito_rep TTTAAATTTGGATTTTATGCTTTAATATTGCTTCAATTGAGGTAAATTTAATGATATTT

PacBio_read -----TCAATTGTGGTTATA--TTAAGGTAT-AAAAGTTGTTG--TCTG
SB_Mito_rep TTTTATATTTTGTATTCAATTGTGGTTATATTTTAAAGGTATAAAAAGTTGTTGTTTCTG
*****

PacBio_read TAATC-----TTTGTGTTTTTTTTTGGTGGTGGTAAGTTTAGTTGTAT
SB_Mito_rep TAATCNNNNNNNNNNNNNNCTTTTTGTTTTTTTTTGGTGGTGGTAAGTTTAGTTGTAT
*****

PacBio_read TTTAGTTCCAATAAAAAATTTTTGTAAAAGTATGTTTTAATTCATTTATATATAATATA
SB_Mito_rep TTTAGTTCCAATAAAAAATTTTTGTAAAAGTATGTTTTAATTCATTTATATATAATATA
*****

PacBio_read ATATAATAT-----ATATAATG---AATAAGTATAGTATAGTAGGGGGAAAAATAATAA
SB_Mito_rep TTATAATATAATATAATATAATGTAAAATAAGTATAGTATAGTAGGGGGAAAAATAATAA
*****

PacBio_read GGTGAGCGTATCGCATACCTATAGGCAAATGTGATTGTATTATCTGTAGAATGGGTGTAG
SB_Mito_rep GGTGAACGTATTGCATACCTATAGGCAAATGTGATTGTATTATCTGTAGAATGGGTGTAG
*****

PacBio_read GCAAATATGATTATACTCCTTAAATTGGTTTAGTCTATAGGGGTGTATTTAGAGTTCGT
SB_Mito_rep GCAAATATGATTATACTCCTTAAATTGGTTTAGTCTATAGGGGTGTATTTAGAGTTCGT
*****

PacBio_read A-----
SB_Mito_rep AGTTTAAAGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTATGTTAGTCT
*

PacBio_read -----GTTTAAAGGTCTAAGGATTCTAAAGCAAATGTGATTG
SB_Mito_rep ATAGAGGGTGTATTTAGGGTTTATGTTTAAAGGTCTAAGGATTCTAAAGCAAATGTGATTG
*****

PacBio_read TACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTTTAAAGATAAATGCATTTAATT
```



```

SB Mito_rep      TACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTTTAAAGATAAATGCATTTAATT
*****

PacBio_read     TCTTGTAAGGGGGATGTACAACAAATTTGTTTAAATTTGGATTTTATGCTTTAATATTGC
SB_Mito_rep     TCTTGTAAGGGGGATGTACAACAAATTTGTTTAAATTTGGATTTTATGCTTTAATATTGC
*****

PacBio_read     TTTCAATTGAGGTAAATTTAATGATA-TTTTTATATTTTATGTATTCAATTGTGGTTA
SB_Mito_rep     TTTCAATTGAGGTAAATTTAATGATATTTTTTATATTTTATGTATTCAATTGTGGTTA
*****

PacBio_read     TATTTTAAGGTAT-----
SB_Mito_rep     TATTTTAAGGTATAAAAAGTTGTTGTTCTGTAATCTTTTTTGTTTTTTTTTTGGTGGTG
*****

PacBio_read     -----
SB_Mito_rep     GTAAGTTTAGTTGTATTTTAGTCCAATAAAAATTTTTGTAAAAGTATGTTTAAATC

PacBio_read     -----
SB_Mito_rep     ATTTATATAAATATATTATAATATAGTATAATGTAAAATATAGTATAGTAGGGGGAAA

PacBio_read     -----
SB_Mito_rep     AATAATAAGGTGAGCGTATTGCATACCTATAGGCAATGTGATTGTATTATCTGTAGAA

PacBio_read     -----
SB_Mito_rep     GGGTGTAGGCAATATGATTATACTCCTTAAATTGGTTTAGTCTATAGGGTTGTATTTA

PacBio_read     -----
SB_Mito_rep     GAGTTCGTAGTTAAGGTCTAAGGATTCTAAGCAATGTGATTGTACTCCTAAAATTAT

PacBio_read     -----
SB_Mito_rep     GTTAGTCTATAGAGGGTGTATTTTAAAGATAAATGCATTTTAAATTTCTGTAAAGGGGAT

PacBio_read     -----
SB_Mito_rep     GTACAACAAATTTGTTTAAATTTGGATTTTATGCTTTAATATTGCTTTCAATTGAGGTAA

PacBio_read     -----AA
SB_Mito_rep     ATTTAATGATATTTTTTATATTTTATGTATTCAATTGTGGTTATATTTAAGGTATAAA
**

PacBio_read     AAGTTGTTGTTTCTGTAATTTTTTT-----
SB_Mito_rep     AAGTTGTTGTTTCTGTAATCTTTTTTGTTTTTTTTGGTGGTAAAGTTTAGTTGTATT
*****

PacBio_read     -----
SB_Mito_rep     TTAGTCCAATAAAAATTTTTGTAAAAGTATGTTTAAATTCATTTATATATAATATAT

PacBio_read     -----
SB_Mito_rep     TATAATATAATATATAATGTAAAATAAGTATAGTATAGTAGGGGGAAAAATAATAAGGT

PacBio_read     -----
SB_Mito_rep     GAGCGTATCGCATTCCTATAGGCAATGTGATTGTATTATCTGTAGAAATGGGTGTAGGCA

PacBio_read     -----
SB_Mito_rep     AATATGATTATACTCCTTAAATTGGTTTAGTCTATAGGGTTGTATTTAGAGTTCGTAGT

PacBio_read     -----
SB_Mito_rep     TTAAGGTCTAAGGATTCTAAGCAATGTGATTGTACTCCTAAAATTATGTTAGTCTATA

PacBio_read     -----
SB_Mito_rep     GAGGGTGTATTTAGGGTTTGTGTTTGTAGGAATGGTTAAATAGTTGGATAGAAAGGCTT

PacBio_read     -----
SB_Mito_rep     ATAGTAAGCTTCACCATCGAATGGTGAGTAAGATTTAAAATCTCTATTTATAGTCTTGT

```

PacBio\_read -----TTTTTTTTTGGTGGTGGTAAGTTAGTTGATTTTAGTTCCAATAAAAA  
SB\_Mito\_rep TTTTGTPTTTGGTTTTTTTTTGGTGGTGGTAAGTTAGTTGATTTTAGTTCCAATAAAAA  
\*\*\*\*\*

PacBio\_read TTTTTGTAAAAAGTATGTTTTAATTCATTTATATATAATATA-TATAATAT-ATATAAT  
SB\_Mito\_rep TTTTTGTAAAAAGTATGTTTTAATTCATTTATATATAATATAATATAATATAATATAAT  
\*\*\*\*\*

PacBio\_read GTAAAAAAGTATAGTATAGTAGGGGGAAAAATAATAAGGTGAGCGTATCGCATACCTA  
SB\_Mito\_rep GTAAAAAAGTATAGTATAGTAGGGGGAAAAATAATAAGGTGAGCGTATCGCATACCTA  
\*\*\*\*\*

PacBio\_read TAGGCAATGTGATTGTATTATCTGTAGAATGGGTGTAGGCAATATGATTATACTCCTT  
SB\_Mito\_rep TAGGCAATGTGATTGTATTATCTGTAGAATGGGTGTAGGCAATATGATTATACTCCTT  
\*\*\*\*\*

PacBio\_read AAATTGGTTTAGTCTATAGGGGTTGATTTAGAGTTCGTAGTTAAGGTCTAAGGATTCT  
SB\_Mito\_rep AAATTGGTTTAGTCTATAGGGGTTGATTTAGAGTTCGTAGTTAAGGTCTAAGGATTCT  
\*\*\*\*\*

PacBio\_read AAAGCAATGTGATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTTAAAGA  
SB\_Mito\_rep AAAGCAATGTGATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTTAAAGA  
\*\*\*\*\*

PacBio\_read TAAATGCATTTTAATTTCTTGAAGGGGGATGTACAACAAATTTGTTAAATTTGGATT  
SB\_Mito\_rep TAAATGCATTTTAATTTCTTGAAGGGGGATGTACAACAAATTTGTTAAATTTGGATT  
\*\*\*\*\*

PacBio\_read TATGCTTTAATATTGCTTTCAATTGAGGTAATTTAATGATATTTTTTATATTTTATGT  
SB\_Mito\_rep TATGCTTTAATATTGCTTTCAATTGAGGTAATTTAATGATATTTTTTATATTTTATGT  
\*\*\*\*\*

PacBio\_read ATTCAATGTGGTTATATTTTAAAGGTATAAAAAAGTTGTTGTTCTGTAATCTTTTTTG-I  
SB\_Mito\_rep ATTCAATGTGGTTATATTTTAAAGGTATAAAAAAGTTGTTGTTCTGTAATCTTTTTTGT  
\*\*\*\*\*

PacBio\_read TTTTTTTGGTGGTGGTAAGTTTAGTTGATTTTAGTTCCAATAAAAAATTTTTGTAAAA  
SB\_Mito\_rep TTTTTTTGGTGGTGGTAAGTTTAGTTGATTTTAGTTCCAATAAAAAATTTTTGTAAAA  
\*\*\*\*\*

PacBio\_read AGTATGTTTAAATTCATTTATATATAATATAATTATAATAT--TATAATGTAAAAAAGTA  
SB\_Mito\_rep AGTATGTTTAAATTCATTTATATATAATATAATTATAATATAAGTATAAAGTAAAAAAGTA  
\*\*\*\*\*

PacBio\_read TAGTATAGTA-GGGGAAAAATAATAAGGTGAGCGTATCGCATACCTATAGGCAAATGTG  
SB\_Mito\_rep TAGTATAGTAGGGGGAAAAATAATAAGGTGAGCGTATCGCATACCTATAGGCAAATGTG  
\*\*\*\*\*

PacBio\_read ATTGTATTATCTGTAGAATGGGTGTAGGCAAAATATGATTATACTCCTTAAATTTGGTTAG  
SB\_Mito\_rep ATTGTATTATCTGTAGAATGGGTGTAGGCAAAATATGATTATACTCCTTAAATTTGGTTAG  
\*\*\*\*\*

PacBio\_read TCTATAGGGGTTGTATTTAGAGTTCGTAGTTAAGGTCTAAGGATTCTAAAGCAAATGTG  
SB\_Mito\_rep TCTATAGGGGTTGTATTTAGAGTTCGTAGTTAAGGTCTAAGGATTCTAAAGCAAATGTG  
\*\*\*\*\*

PacBio\_read ATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTTAGGGTTATGTTAAGGT  
SB\_Mito\_rep ATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTTAGGGTTATGTTAAGGT  
\*\*\*\*\*

PacBio\_read CTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGTG  
SB\_Mito\_rep CTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGTG  
\*\*\*\*\*

PacBio\_read TATTTAGGGTTTGTGTTTGTAGGAATGGTTAAATAGTTGGATAGAAAGGCTTATAGTAA  
SB\_Mito\_rep TATTTAGGGTTTGTGTTTGTAGGAATGGTTAAATAGTTGGATAGAAAGGCTTATAGTAA  
\*\*\*\*\*

PacBio\_read GCTTCACCATCGAATGGTGAGTAAGATTTAAAAATCTCTATTTATAGAAGATATTGATA  
SB\_Mito\_rep GCTTCACCATCGAATGGTGAGTAAGATTTAAAAATCTCTATTTATAGAAGATATTGATA  
\*\*\*\*\*

PacBio\_read TAAAACTAATGGATAAGTTT-TTATATATTTTLAGGTTATTTATCTTTTATGATT  
SB\_Mito\_rep TAAAACTAATGGATAAGTTTATTATATA-TTTTLAGGTTATTTATCTTTTATGATT  
\*\*\*\*\*

PacBio read -TATTAGTACTCCTTTGTTAGTATTTATAGTATTTAAGTTTGTG-TTTTTTACTACTAA  
SB Mito rep ATATTAGTACTCCTTTGTTAGTATTTATAGTATTTAAGTTTGTGTTTTTTTACTACTAA  
\*\*\*\*\*

PacBio read GTTTGTTGAGTGGTGGGGTGTGTTGATATATTGTTATATGTTATTGTTTTTATTGTTAA  
SB Mito rep GTTTGTTGAGTGGTGGGGTGTGTTGATATATTGTTATATGTTATTGTTTTTATTGTTAA  
\*\*\*\*\*

PacBio read TAGTGATGTTTGTATTATTGTAATTATAGGTTGATTATTCTAGTATATTATTATGAT  
SB Mito rep TAGTGATGTTTGTATTATTGTAATTATAGGTTGATTATTCTAGTATATTATTATGAT  
\*\*\*\*\*

PacBio read TTTAG-TTAATTTATGTTA---TATATTCTGCGTAATAATTTGATGTATTAGTTAGTTT  
SB Mito rep TTTAGTTTAATTTATGTTATATTCTGCGTAATAATTTGATGTATTAGTTAGTTT  
\*\*\*\*\*

PacBio read ATGTTATGTTATTTTGGTGGTTAGATATTAAGTTAATAAATACTATGTGTTTCATGTGC  
SB Mito rep ATGTTATGTTATTTTGGTGGTTAGATATTAAGTTAATAAATACTATGTGTTTCATGTGC  
\*\*\*\*\*

PacBio read ACAAAATGCTGTATTTAAGTTACG--TATCTTTATTTATTATACTGGTTAGGCTGGTT  
SB Mito rep ACAAAATGCTGTATTTAAGTTACGCATATCTTTATTTATTATACTGGTTAGGCTGGTT  
\*\*\*\*\*

PacBio read ATCTAGTTTAGTAGGTTAGTTAGACTGCTTGCTTCAAAGTAAGAGGATACTGTATTT  
SB Mito rep ATCTAGTTTAGTAGGTTAGTTAGACTGCTTGCTTCAAAGTAAGAGGATACTGTATTT  
\*\*\*\*\*

PacBio read GGTCCTAGGCTGAGGTGTATGGAAGTAAAATTTGTTATCGCTGCTAACGATAATTTAGGA  
SB Mito rep GGTCCTAGGCTGAGGTGTATGGAAGTAAAATTTGTTATCGCTGCTAACGATAATTTAGGA  
\*\*\*\*\*

PacBio read -TATGTAGTGAATT-TGCTTCTTTATGTTTTATTAATTATATTGTTATAGCTAT-GAA  
SB Mito rep TTATGTAGTGAATTCGCTTCTTTATGTTTTATTAATTATATTGTTATAGCTATGGAA  
\*\*\*\*\*

PacBio read GAAACAAGAAGAAGAAGAAGAGAGTTAAGAAGGAAGTTTCTA-AAAGAAGAAG--ACT  
SB Mito rep GAAACAAGAAGAAGAAGAAGAGAGTTAAGAAGGAAGTTTCTAACAAGAAGAAGAAGACT  
\*\*\*\*\*

PacBio read TCAGTTGGTGAAGCGGTGGTTCTGGGTTTTGGAGGTAAGTTATTAGATTTATTAATAA  
SB Mito rep TCAGTTGGTGAAGCGGTGGTTCTGGGTTTT-GAGGTAAGTTATTAGATTTATTAATAA  
\*\*\*\*\*

PacBio read G---TTTAGGTGGATTTAAGAAGTCTATATGTTTAC-TATTAAGTTTGTAAATGGTAGTGT  
SB Mito rep GCTTTTAGGTGGATTTAAGAAGTCTATATGTTTACTTATTAAGTTTGTAAATGGTAGTGT  
\*\*\*\*\*

PacBio read AAGTGTGTTTCATTTTAAATTTATATAAGTTTGTAGTATTGGGTTAGATTCATTAAATAG  
SB Mito rep AAGTGTGTTTCATTTTAAATTTATATAAGTTTGTAGTATTGGGTTAGATTCATTAAATAG  
\*\*\*\*\*

PacBio read TTTAGTAGTTGATTTTGTGTAATGCTGTGGTTATGGTTATGGTTAATTTCTTTAGA  
SB Mito rep TTTAGTAGTTGATTTTGTGTAATGCTGTGGTTATGGTTATGGTTAATTTCTTTAGA  
\*\*\*\*\*

PacBio read TCATAAGCGGATAAGTTTAGTGTATATATTTTAGGTTTATGAGGTGGTTTTATTGGTCT  
SB Mito rep TCATAAGCGGATAAGTTTAGTGTATATATTTTAGGTTTATGAGGTGGTTTTATTGGTCT  
\*\*\*\*\*

PacBio read T-GTTTAAAGGCTTCTAATTCGGTTAAATTTATGTGATCCATATTATAAATTTGGTTTCGTT  
SB Mito rep TGGTTTAAAGGCTTCTAATTCGGTTAAATTTATGTGATCCATATTATAAATTTGGTTTCGTT  
\*\*\*\*\*

PacBio read AGAGGTTTATAAATTTTGTACTAATCATGGTATAGCTATGA-TTTTTTTTTTTGAT  
SB Mito rep AGAGGTTTATAAATTTTGTACTAATCATGGTATAGCTATGATTTTTTTTTTTTGGAT  
\*\*\*\*\*

PacBio read GCCGGTATTAATAGGT-GATTTGGT-AATATTTCTTCC-TTTTTTTTATATATAGATGA  
SB Mito rep GCCGGTATTAATAGGTGGATTTGGTAAATTTCTTCCTTTTTTTTTATATATAGATGA  
\*\*\*\*\*

PacBio read TTTATTGTTACCTCGGTTGAATTCCTTTAGTTTATGGTTGATGATTCCTTCATTTTTTTA  
SB Mito rep TTTATTGTTACCTCGGTTGAATTCCTTTAGTTTATGGTTGATGATTCCTTCATTTTTTTA  
\*\*\*\*\*

PacBio\_read TATGGAGCTAAGTTTATATTATGGTTGTGGGGTT-GTTGAACGTTGTATCCTCCATTATC  
SB\_Mito\_rep TATGGAGCTAAGTTTATATTATGGTTGTGGGGTTGGTTGAACGTTGTATCCTCCATTATC  
\*\*\*\*\*

PacBio\_read TATATCTGAGAATTCTGGCTTGGGTGTAGATTATTTGATGTTTTTCGTTGCATTAGCGGG  
SB\_Mito\_rep TATATCTGAGAATTCTGGCTTGGGTGTAGATTATTTGATGTTTTTCGTTGCATTAGCGGG  
\*\*\*\*\*

PacBio\_read TGTGTCTAGATTGGTTGGTTC AATTAATTTTATTTCTACAATTATGAGTTGTGTAGATT  
SB\_Mito\_rep TGTGTCTAGATTGGTTGGTTC AATTAATTTTATTTCTACAATTATGAGTTGTGTAGATT  
\*\*\*\*\*

PacBio\_read TAAGACTTCGATAATAATCTGATCATATTTTATTTACTTCAATTTTATGTTACTGTGCAT  
SB\_Mito\_rep TAAGACTTCGATAATAATCTGATCATA-TTTATTTACTTCAATTTTATGTTACTGTGCAT  
\*\*\*\*\*

PacBio\_read TACCCAGTTTTAGCGGCTGGTATAACTATGTTTATTATATTGATCGTAAATTTGGTACTG  
SB\_Mito\_rep TA-CCAGTTTTAGCGGCTGGTATAACTATG-TTATTAT-TTGATCGTAAATTTGGTACTG  
\*\* \*\*\*\*\*

PacBio\_read -TTTTTTTGTAGCCGATGGGTGGTGGTATCCGCTGTTATTTTCAGCATTATTTTGGTTTT  
SB\_Mito\_rep CTTTTTTTGTAGCCGATGGGTGGTGGTATCCGCTGTTATTTTCAGCATTATTTTGGTTTT  
\*\*\*\*\*

PacBio\_read TTGGGCATCCTGAGGTGTATGTTTTGATTCTGCCAGGATTC-GTATAGTTAGTCATATAT  
SB\_Mito\_rep TTGGGCATCCTGAGGTGTATGTTTTGATTCTGCCAGGATTCGTTATAGTTAGTCATATAT  
\*\*\*\*\*

PacBio\_read GTATGAGAATTAGTAATAATGATTCATCATTTGGATATTATGGATTAATTTGTGCTATGG  
SB\_Mito\_rep GTATGAGAATTAGTAATAATGATTCATCATTTGGATATTATGGATTAATTTGTGCTATGG  
\*\*\*\*\*

PacBio\_read CTCAATAGTTTGTTTAGGTAGTGTAGTTT-GGCTCACCATATGTTTATGGTTGGTTTGG  
SB\_Mito\_rep CTCAATAGTTTGTTTAGGTAGTGTAGTTTGGCTCACCATATGTTTATGGTTGGTTTGG  
\*\*\*\*\*

PacBio\_read ATTA-TTGACAGCTATATTTTTTAGTCTGTGACTCATGA-TATAGGGATTCCTAC-GGT  
SB\_Mito\_rep ATTATTTGACAGCTATATTTTTTAGTCTGTGACT-ATGATTATAGGGATTCCTACGGGT  
\*\*\* \*\*\*\*\*

PacBio\_read ATAAAGGTTTTTCTTGATTATATATGCTTAAAAGATGAGGCTACTCGTTGCTTGTCT  
SB\_Mito\_rep ATAAAGGTTTTTCTTGATTATATATGCTTAAAAGATGTGG---TACTCGT---TTGCTC  
\*\*\*\*\* \* \*\*\*\*\*

PacBio\_read GATGACTCCTGTGTTATTTATGGTGAGTTGGT-GGTTATATACCTTTGTCATCTAACAAT  
SB\_Mito\_rep GA----TCCTG----TGTTATGGTG-GTTGGTGGGTTTTATATTTTTGT---TTACAAT  
\*\* \*\*\*\*\* \* \*\*\*\*\*

PacBio\_read CGTGGTAGGGTGGTGTACCTGGC-TAGCTTTTTTCTTGTCTTTTCTACTATTGGACATT  
SB\_Mito\_rep -----AGGGGTGTTACTGTGTAGCTTTGTCGGCTTCCT-----CATTAGACATA  
\*\* \*\*\*\*\* \* \*\*\*\*\*

PacBio\_read ATTATTCATGATATTTGTTTTGTTGTGCATGTTCTGCCATTATGTTCTCTTATCTTTAGA  
SB\_Mito\_rep TATATTCATGATACTTGGTTTTGTCGTTGCTCATTTTCATTATGTTCT---TTCGTTAGG  
\* \*\*\*\*\* \* \*\*\*\*\*

PacBio\_read TATTGCTATTTAATAGATAGTCGGTCGACTAAATTGATAGATTGGTTTATTTGTATTTGT  
SB\_Mito\_rep TTCT-----TATAGTAGGGTGATAAATGTTGG-----TATGATGGT  
\* \* \*\*\*\*\* \* \* \* \* \*

PacBio\_read GGCCTTTTATAAAAATTTGATAGAGTACATGTAATGAATTATAGTTATCTCCTTACAGT  
SB\_Mito\_rep GGCCTTT---TATAATGGATACAGT-----ATGA--ATAAGTATCTCCTACAGGG  
\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

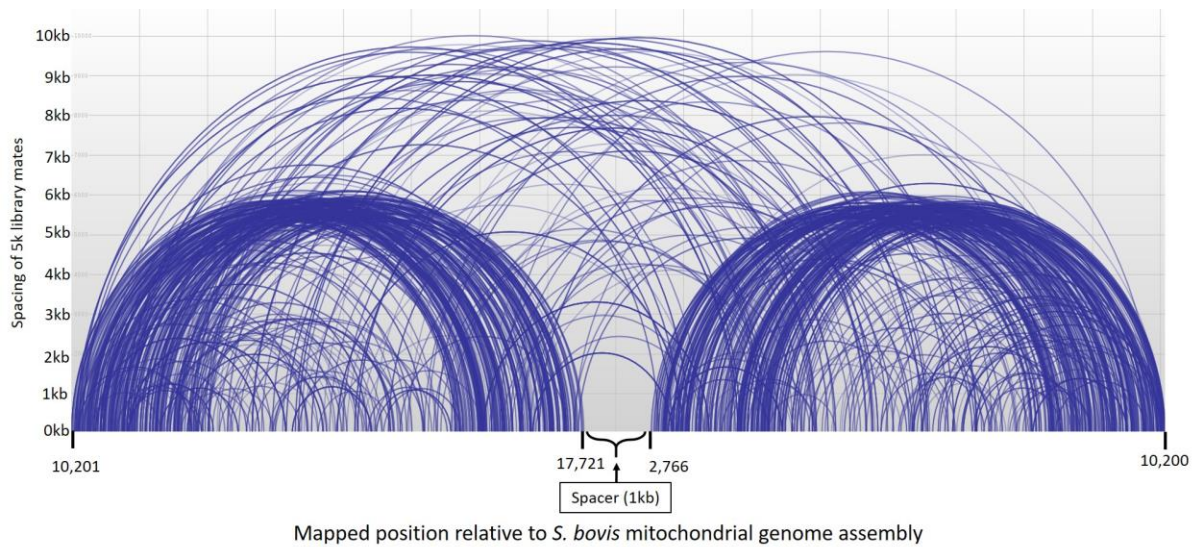
PacBio\_read CTATGATTTATTGTTTCGATGGTT-GATTTTATATTG-TATTTTCTATGCATGAACAT  
SB\_Mito\_rep TCATGATTTATTG-TCGATGGTTGGATTTAATCTTTGTTTTTTTCTATGCAT---TAT  
\*\* \*\*\*\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

PacBio\_read TTATGGTTATTCGATGATGGTTTGCCTCGTTCGGGTAGGTGTCTTATGATGTGTTGTT  
SB\_Mito\_rep TTA--GGTATTC----ATGGTTGCCTCG-TCGGGTAAGT-TGCTATGATTTGAGTAT  
\*\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

PacBio\_read TAT--GGTTATATACTGAGAAGATTATATTGTGTCGGTGTGTATATTTCTGTACT--  
SB\_Mito\_rep TATTGGGTTAAAGTACTGAGAAG--TATA---GGTGGTGTGTTA---TCTGTACTAG  
\*\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

PacBio\_read TGCTTTCGTAGTATGTTTTAATCTTTAGAGCAGAGTCATT-----





**Figure K:** 5kb mate pair library mapped to mitochondrial genome centered on the repetitive region

The figure shows the spacing of read-pairs mapped to the *S. bovis* mitochondrial genome in the reverse-forward orientation in the form of arcs with the height of the arcs representing the distance between the mapped reads in a pair (visualized in the Savant Genome Browser [4]). Read pairs spaced by more than 10kb or less than 1kb were omitted. The mitochondrial sequence was centered on the repetitive region with the region itself substituted for by a spacer comprised of 1,000 Ns.

## Supplementary Tables

**Table A:** Overview of generated *S. bovis* sequence data.

Sequencing Platform	Average read length	Number of reads (M)	Bases (Gb)	GC content (%)
Illumina	90bp	623	56.1	35.91
PacBio	3.2kb	0.714	2.3	37.05

**Table B:** Summary of sequenced Illumina libraries.

Library	Insert size (bp)	Reads (M)	Sequenced bases (Gb)	High-Quality Reads (M)
Small insert / paired-end	200	158.8	14.3	137.0
Small insert / paired-end	500	174.4	15.7	133.5
Small insert / paired-end	800	91.6	8.2	58.2
Large insert / paired-end	2000	130.6	11.8	88.3
Large insert / paired-end	5000	68.0	6.1	45.4



**Table C:** Characterization of repetitive elements found in four *Schistosoma* genomes.

Repeat Type	<i>S. bovis</i>	<i>S. mansoni</i>	<i>S. haematobium</i>	<i>S. japonicum</i>
Total interspersed repeats (%)	35.83	39.66	34.44	36.85
SINEs (%)	2.57	3.25	2.19	2.26
LINEs (%)	17.38	19.4	19.14	18.69
LTR elements (%)	4.18	5.86	3.96	3.65
DNA elements (%)	0.56	0.76	0.56	0.94
Unclassified (%)	11.12	10.39	8.6	11.31

**Table D:** Characteristics of protein coding genes (CDSs) across four *Schistosoma* species.

Species	Total genes predicted	Average coding domain length (kb)	GC of coding sequences (%)	Average exon number per gene	Average exon length (bp)	Average intron length (kb)
<i>S. bovis</i>	11,631	1.2	36	4.8	256	1.95
<i>S. haematobium</i>	13,073	1.3	35	5.2	247	2.51
<i>S. mansoni</i>	11,774	1.4	36	6.4	225	2.43
<i>S. japonicum</i>	12,657	1.2	36	5.3	220	2.08

**Table E:** Conservation of parasite proteins previously linked to *S. haematobium*-induced cancer across *S. bovis* and *S. haematobium*. Proteins were mapped to the genome sequences using Exonerate.

<b>Protein</b>	<b>Reference organism / Target organism<sup>a</sup></b>	<b>Identity / Coverage of reference sequence</b>	<b>Description</b>
<b>Estradiol 17beta-dehydrogenase</b>	<i>S. haematobium</i> / <i>S. bovis</i>	<b>98% / 100%</b>	Involved in the synthesis of estradiol. Proposed to be conducive of tumorigenesis in <i>S. haematobium</i> , potentially via estrogen receptor mediated cell proliferation leading to an increase of DNA mutations during replication
<b>Glycoprotein Omega-1</b>	<i>S. haematobium</i> / <i>S. bovis</i>	<b>92% / 100%</b>	Secretion from <i>Schistosoma</i> eggs thought to trigger Th2-type immune response leading to cancerous environment

<sup>a</sup>Query and target species of the Exonerate search.

**Table F:** Conservation of *Schistosoma* vaccine targets in *S. bovis*. Known vaccine targets were mapped to the *S. bovis* genome sequence using Exonerate.

Vaccine / Major antigen	Species targeted	Reference organism / Target organism <sup>a</sup>	Identity <sup>b</sup> / Coverage of reference sequence	Clinical development stage	Description
Bilhvax / 28-kDa recombinant glutathione-S-transferase Sh28GST	<i>S. haematobium</i>	<i>S. haematobium</i> / <i>S. bovis</i>	97% / 100%	Completed Phase 3 trial	Important for fatty acid metabolism and prostaglandin synthesis. Localised in tegument
Sm14 / 14-kDa recombinant fatty acid binding protein Sm14	<i>S. mansoni</i>	<i>S. mansoni</i> / <i>S. bovis</i>	99% / 100%	Successfully completed Phase 1 trial	Schistosomes are unable to synthesize fatty acids. Sm14 is required to import fatty acids from diet
Sm-TSP-2 / 9-kDa recombinant tetraspanin Sm-TSP-2	<i>S. mansoni</i>	<i>S. mansoni</i> / <i>S. bovis</i>	34% / 89%	Successfully completed Phase 1 trial	Expressed in tegument providing high levels of protection as recombinant vaccine in mouse models
Sm-p80 (calpain)	<i>S. mansoni</i> and <i>S. haematobium</i>	<i>S. mansoni</i> / <i>S. bovis</i>	99% / 91%	Pre-clinical development	Located in tegument and may play role in immune evasion by renewing surface membrane. Vaccines against large subunit Sm-p80 protected from <i>S. mansoni</i> and <i>S. haematobium</i> in model organisms

<sup>a</sup>Query and target species of the Exonerate search. <sup>b</sup>Sequence identity on amino acid level.

## References

1. Doenhoff MJ, Cioli D, Utzinger J: **Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis.** *Curr Opin Infect Dis* 2008, **21**:659-667.
2. Tebeje BM, Harvie M, You H, Loukas A, McManus DP: **Schistosomiasis vaccines: where do we stand?** *Parasit Vectors* 2016, **9**:528.
3. Merrifield M, Hotez PJ, Beaumier CM, Gillespie P, Strych U, Hayward T, Bottazzi ME: **Advancing a vaccine to prevent human schistosomiasis.** *Vaccine* 2016, **34**:2988-2991.
4. Fiume M, Williams V, Brook A, Brudno M: **Savant: genome browser for high-throughput sequencing data.** *Bioinformatics* 2010, **26**:1938-1944.