

Supplementary Material

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

Harald Oey¹, Martha Zakrzewski², Kerstin Gravermann⁵, Neil D. Young⁴, Pasi K. Korhonen⁴, Geoffrey N. Gobert^{2,3}, Sujeevi Nawaratna², Shihab Hasan^{1,2}, David M. Martinez², Hong You², Martin Lavin⁹, Malcolm Jones^{2,8}, Mark A. Ragan⁶, Jens Stoye⁵, Ana Oleaga¹⁰, Aidan Emery⁷, Bonnie Webster⁷, David Rollinson⁷, Robin B. Gasser⁴, Donald P. McManus^{2,*}, Lutz Krause^{1,2,*,#}

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* orthologous 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

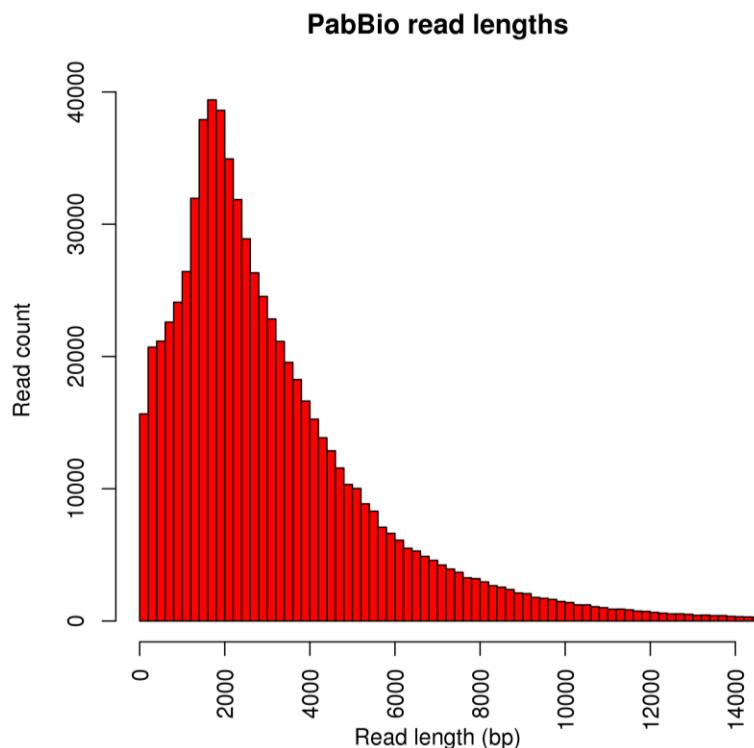


Figure A: PacBio read length distribution

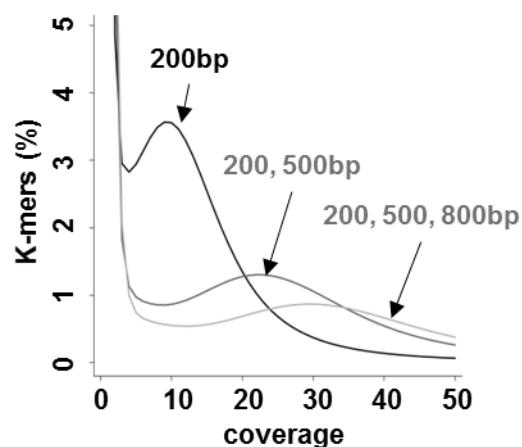


Figure B: Distribution of 17-mers in short insert Illumina libraries.

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*.

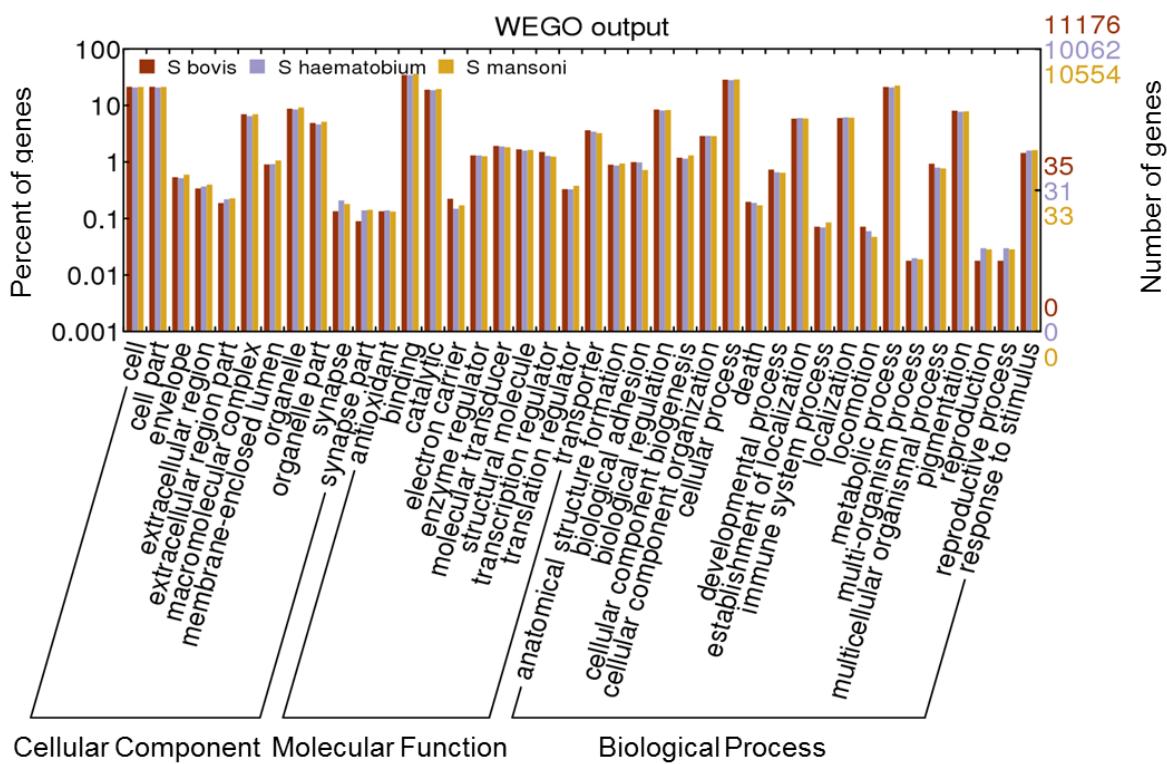


Figure C: GO categories identified across three *Schistosoma* species.

GO annotations were assigned from InterProScan results using WEGO.

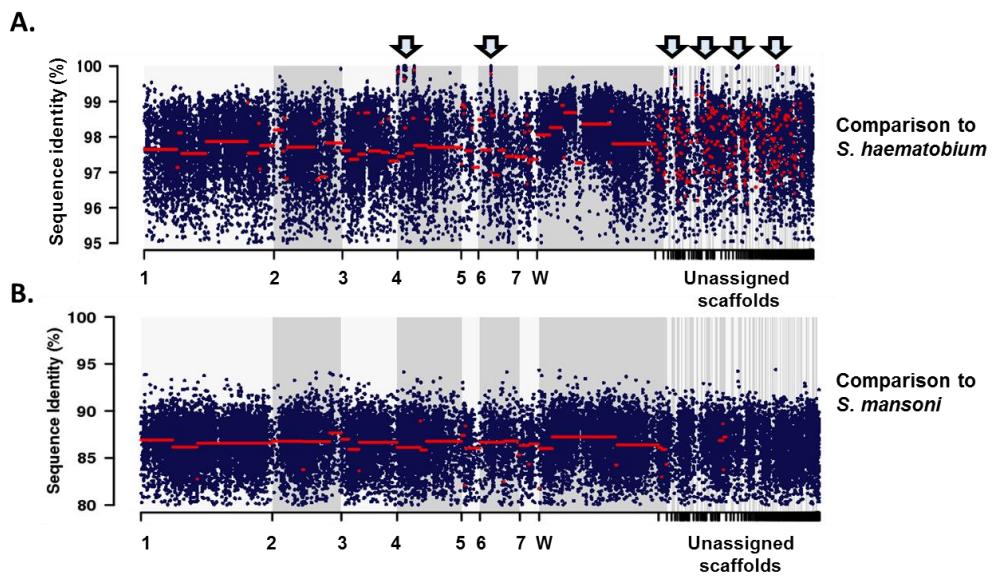
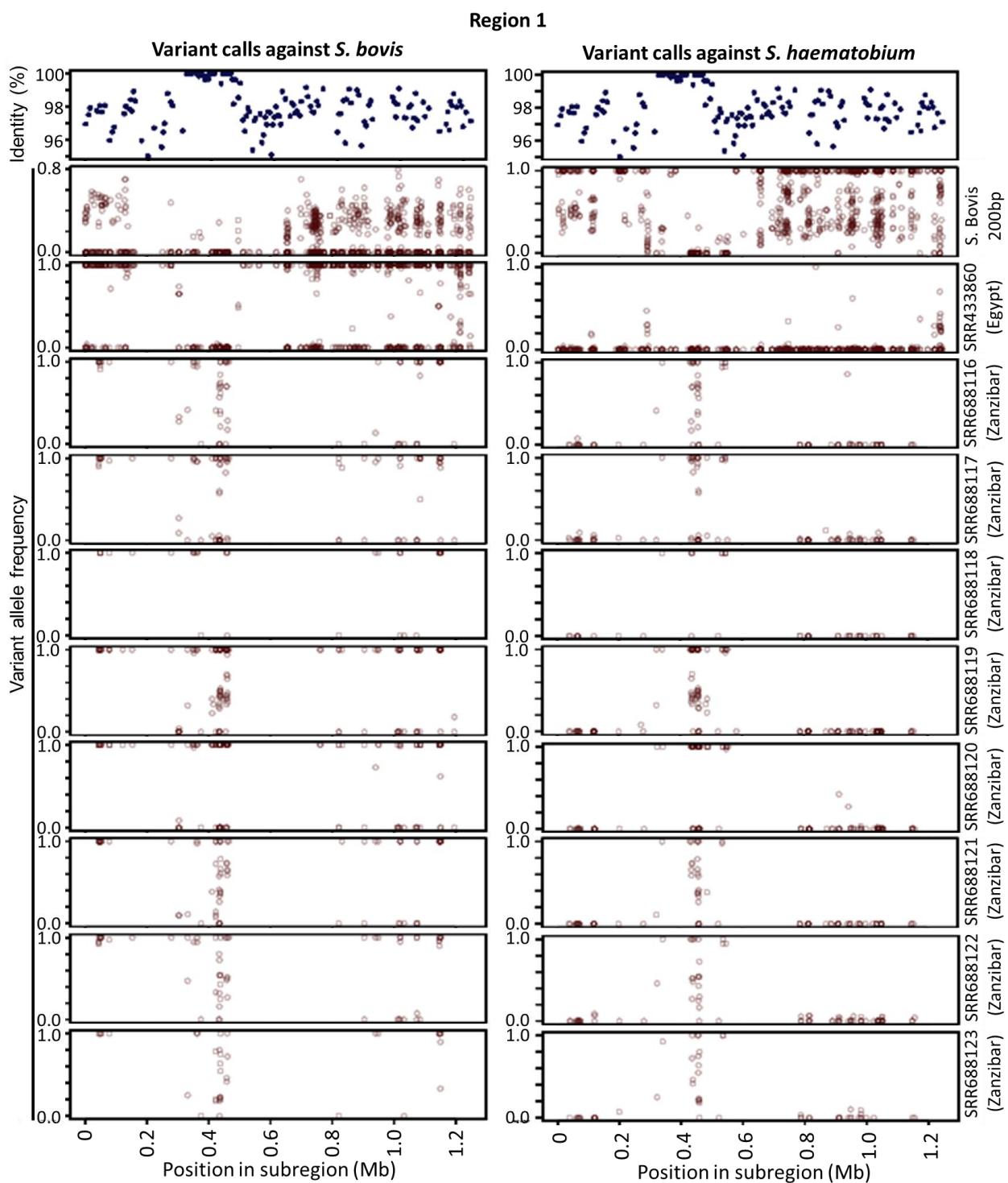


Figure D: Genome-wide comparison of *S. bovis* with *S. haematobium* and *S. mansoni*.

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)

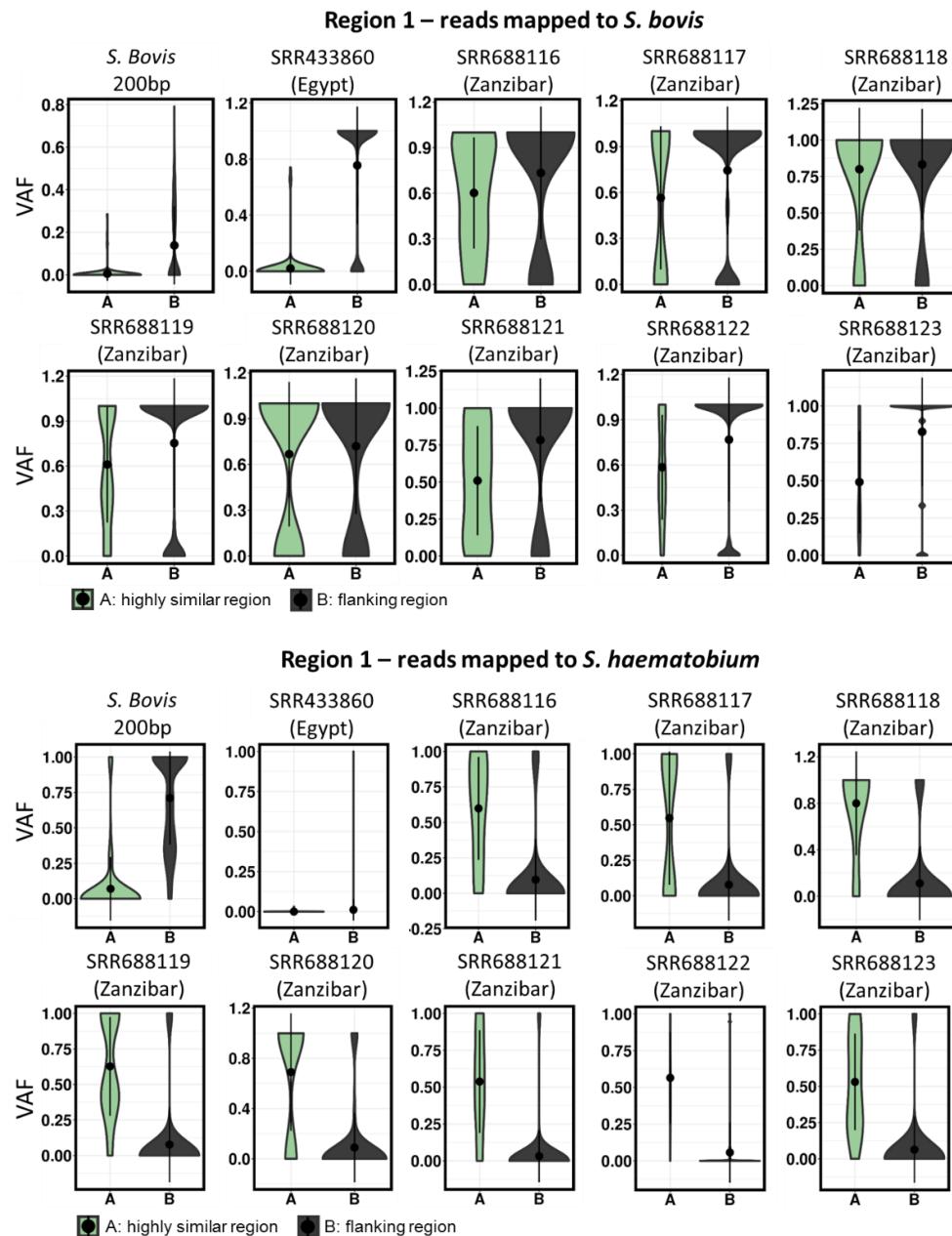
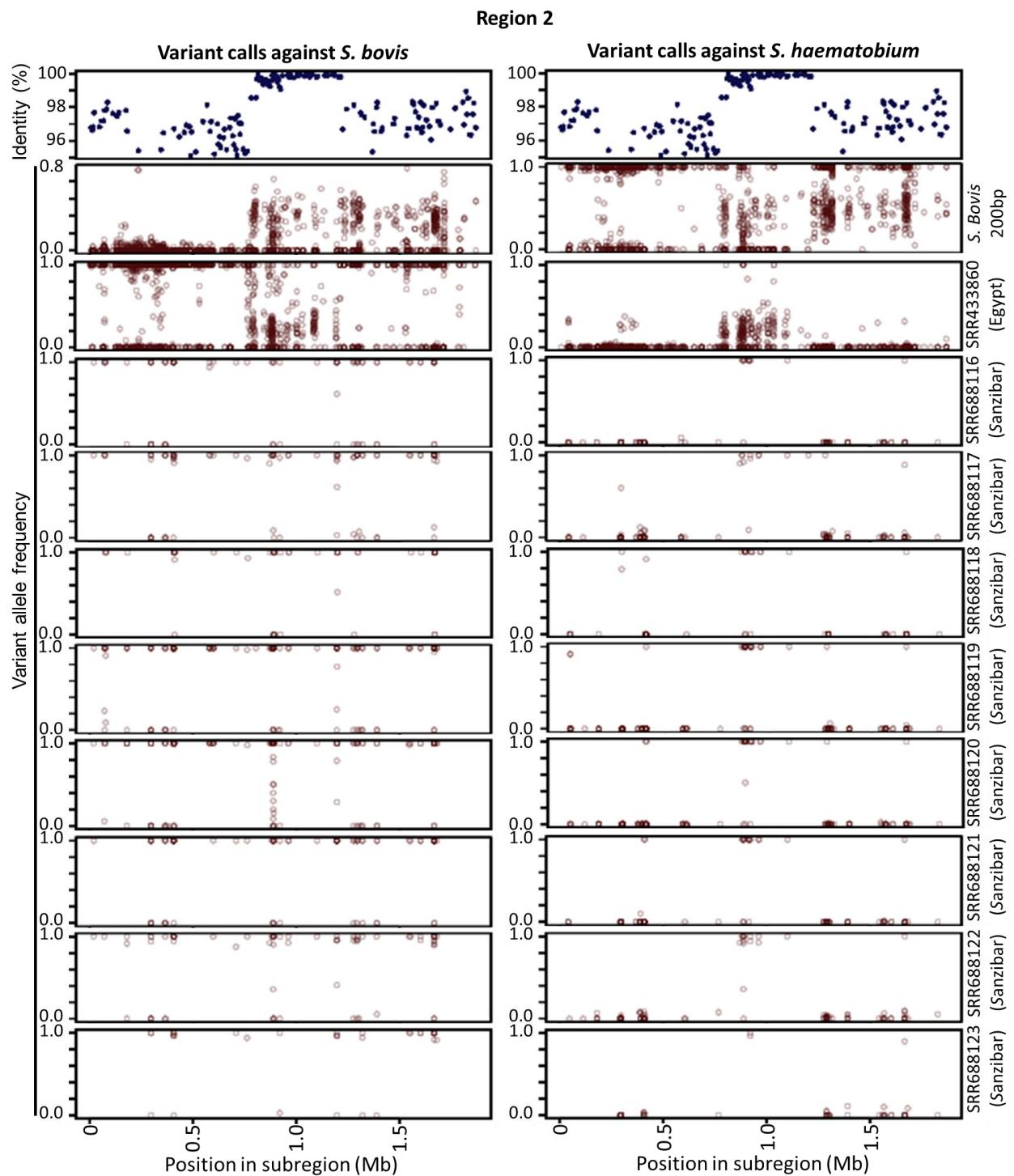


Figure E: Variant allele frequencies for *S. bovis* and *S. haematobium* polymorphisms for region 1

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)

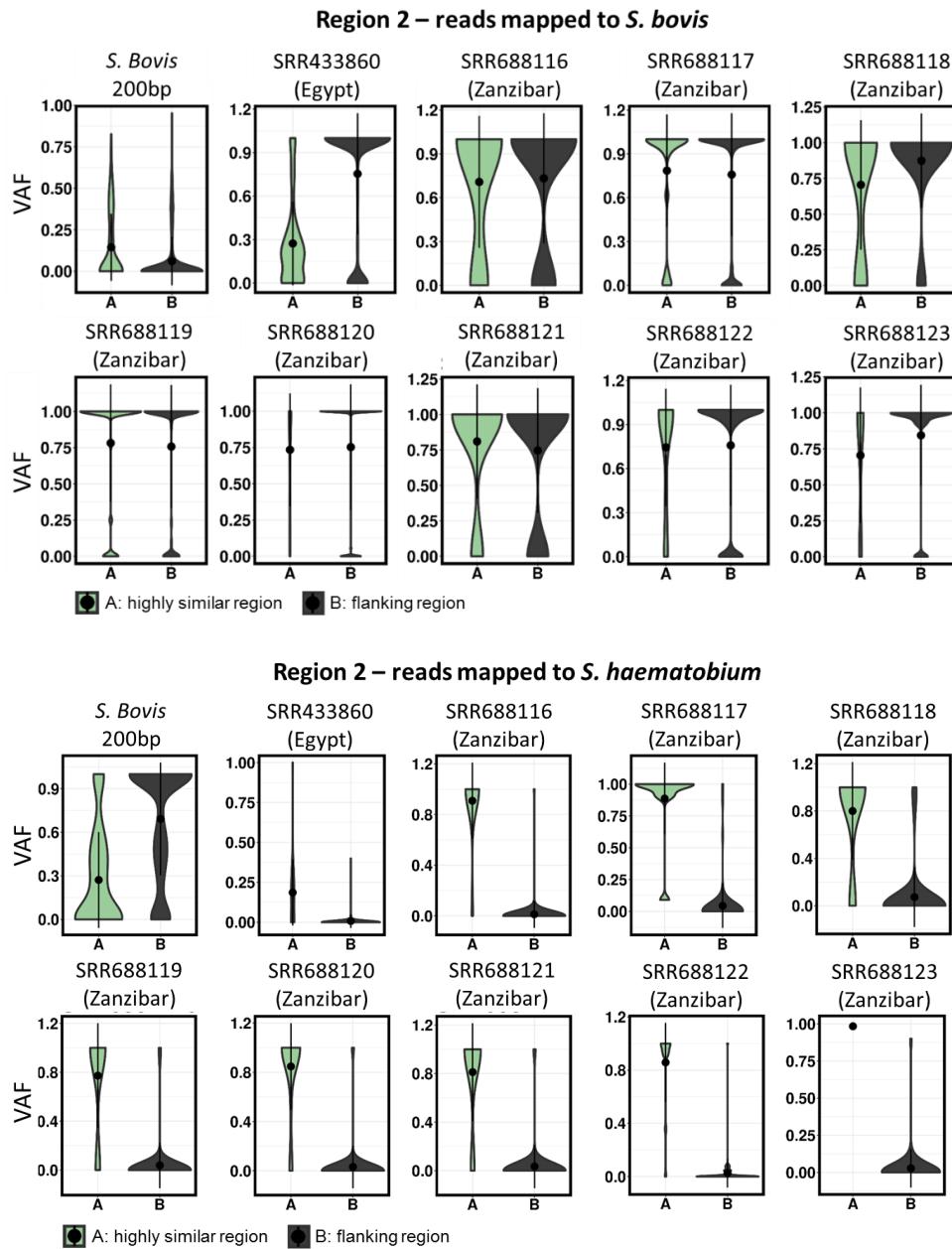


Figure F: Variant allele frequencies for *S. bovis* and *S. haematobium* polymorphisms for region 2

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	MYVDSKVIST
S.bovis	
S.haem	MRTTDYLVSIVSSPQLFYCRSGNSTAASSNQSDIYQTNQSQLSNTNQSDIRQTTESQ
S.jap	MTSSTAHLIKHNFFILTITLIITWFTFKLQLLKICLVLFTVKAWCLSRRKSLRKAGEWA
S.man	IDITSILKFPIEHWRILLTLAVSLIIYKLLOVVTICCFKFTVKKKWCFSRKTLRKAGEWA
S.bovis	-DIGKILTFIDHWRLLILFGVSLINNNQIIAVLKICFKFTLKKWLCKSRKTLRKQAGEWA
S.haem	LDIGKILTFIDHWRLLILFGVSLINNNQIIAVLKICFKFTLKKWLCKSRKTLRKQAGEWA
S.jap	MTSSTAHLIKHNFFILTITLIITWFTFKLQLLKICLVLFTVKAWCLSRRKSLRKAGEWA
S.man	IDITSILKFPIEHWRILLTLAVSLIIYKLLOVVTICCFKFTVKKKWCFSRKTLRKAGEWA
S.bovis	-DIGKILTFIDHWRLLILFGVSLINNNQIIAVLKICFKFTLKKWLCKSRKTLRKQAGEWA
S.haem	LDIGKILTFIDHWRLLILFGVSLINNNQIIAVLKICFKFTLKKWLCKSRKTLRKQAGEWA
S.jap	IVTGASSGIGEAYAAELAKEGLNILLISNDESQLRLVSERISTDYHVERTRIVVADFTQPDI
S.man	VVTGASSGIGEAYAAELAKEGLNIMLISNDEEQLSLVANRIATTYNVQTRIVVADFTKHD
S.bovis	VVTGSSGIGEAYAAELAKEGLNILLISNDEEQLSLAVSNRIANTVNQTRIVVADFTKDD
S.haem	VVTGSSGIGEAYAAELAKEGLNIMLISYDEEQLSDVSNRNIANTVNQTRIVVADFTKDD
S.jap	SYDVVIKPAIQQLSTIACLVNNGMGLPFLSFVVEDFYSPPNEOSIHDIIHCNVIVSTMTHI
S.man	VYEIIIRPAVDQLSTIACLVNNGMGLPFLSFGEINSPPNEESIRNIIHCNCILSAVTMTSI
S.bovis	VYEIIIRPAVDQLSTIACLVNNGMGLPFLFAGEVDPSPNEESIRNIIHCNCIMSTMVTNI
S.haem	VYEIIIRPAVDQLSTIACLVNNGMGLPFLFAGEVDPSPNEESIRNIIHCNCIMSTMVTNI
S.jap	VMPKMLSQGSNPQIINISSYSGLKEYPYASMYAATKAFIHQFSRC1SAEKYSRNLIVQT
S.man	ILPKMLTQKEPNPGIINIASYSGLKVFPYASLYASTKAIIQFSRCVAEEKYKKNVIIQIA
S.bovis	ILPKMLTQKGPNPGIINIIGSYAGLKVFPHATMYSSTKASIIQFSKCLAAEKYKKNVIIQIA
S.haem	ILPKMLTQKGPNPGIINIIGSYAGLKVFPHATMYSSTKASIIQFSKCLAAEKYKKNVIIQIA
S.jap	ICPLIVSTKLSSYYKSTSFFIPTAQVFAKSALDMFGVQQTTGVMQHDLQALIYNSMPNFL
S.man	ICPLFVSTTNNTNLMKTTFFIPTAKVYAKNALDMYGVQEQTSGYFRELKAYLYSSLPTSV
S.bovis	ICPLCVSTNTNQVKTTFFTPSAKFAKSALDMFGVQEQTGTYIHRDLQOTLYDLLPAII
S.haem	ICPLCVSTNTNQVKTTFFTPSAKVYAKSALDMFGVQEQTGTYIHRDLQOTLYDLLPAII
S.jap	WLRFWRRRIHRFMNESRKAK
S.man	FIRIIRLVAKSTLKK---
S.bovis	SKQLMKLKANSQRNKA---
S.haem	SKQLMKLKANSQRNKA---

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

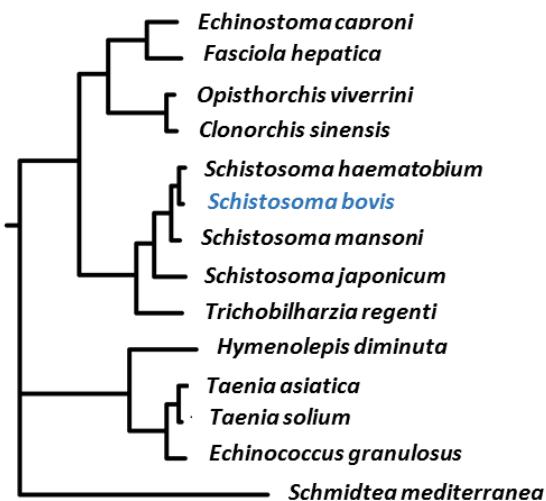


Figure H: Phylogenetic tree reconstructed from the amino acid sequences of 52 shared single-copy genes using the maximum-likelihood method.

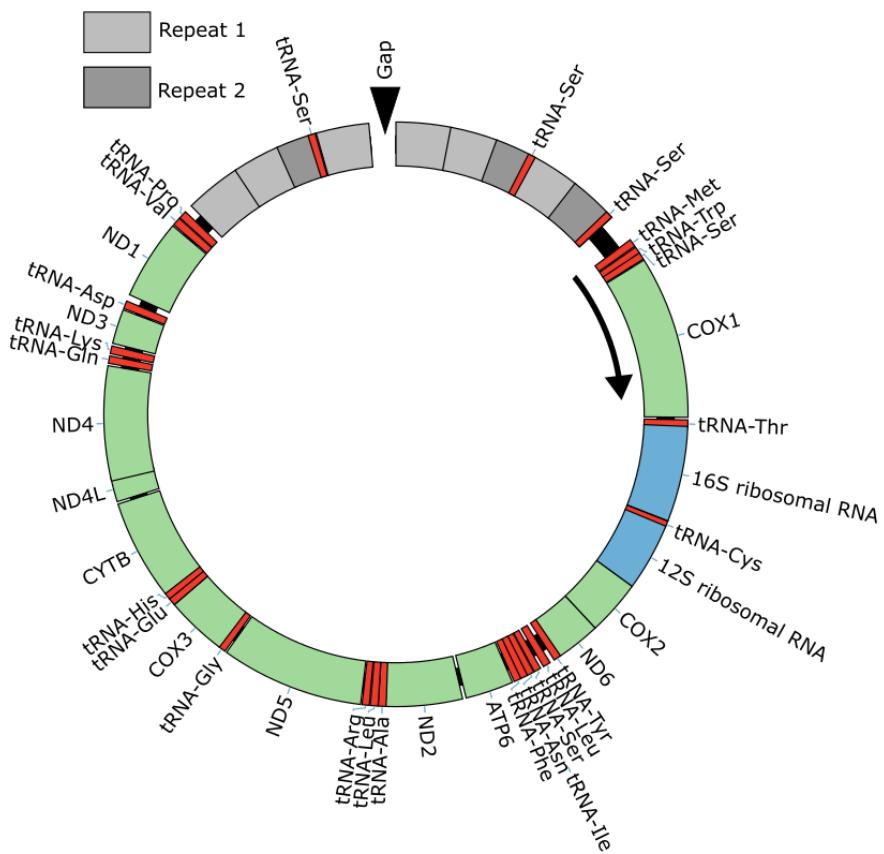


Figure I: Circular genome plot of the *S. bovis* mitochondrial genome.

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	TATTCTACGAAACTAECTCCTACTCTTATCAAAACTTATA-AGTCAGAACCTCTGTA
SB Mito rep	-----ATAAGTATGTTAGCGCTATCAT---AGTTATATAGTTA-----A ***** *
PacBio read	TAAACCCAACAATCCAACCTTTGGCCCTTACGTAATTGAATATAGGATAAGACCTTA
SB Mito rep	TAAAGCT-----TTAATTGTTGGTTTTATGAATT-----ATCTAT **** *
PacBio read	CGTTCTCCCATTATATA-----AAAGGCTACTAATAATAAAAATCATTATTAACCT----- TATTGTCATTATACATGTGCTTAAGGAGAGTTATACCCAGAGCGTTAAAGCTAG ** *
SB Mito rep	-----CTCAAACTCAAAAATAACTC-TAACATATACATAT----- GTGATATATGGGTGGGTATGTC--TTCTAGAAGTGTCCATTACTTATATGTATTTG *
PacBio read	-CCGTA--ATTGCTCAA-----AATATCTACGCCGTATAACCAAAACACTATTA- ATTGAGGGTAGTGAGGGTAGGTTAATTTAGCTTGAGATAATAAA-AGTATGAA *** *
SB_Mito_rep	-----TCAACTACTAA--AATTATAAAATAATTTTA--AATAATCCA----- TTTGAGTTAGTTATGTTAGTAAATGATTATAATGTTTGACGAAGATCCCATTGATTCC *
PacBio read	----CATAACTGACCATAAAT-----ATACTAATCACCAAAAGCACGACAAGCTCTAT GTTCATAGA-TGACTACCAATGGTCATGCTGA-AGCTAGTAGTCCTACAAGTATA-AT **** *
SB Mito rep	CTTCATAAACTAAATTAT-----TATCAACCATATGC---CCACATA ATTAAAGAGGGTATTATGAAGTTGGTTAGTCGGTGTGTCGTTATGTGGTAGATG *
PacBio read	CACATAATTACTATTATTAATCCACTTTA-----TATACCCCAACTTAAGCTCTGCCA AGTATGAGGGTAGGAATAAAATAGAGTTTCAGAGTTATTTTGTAAAGTTAGGATT ** *
SB Mito rep	TATAACCAAGAACCAAAAAAAAAAAAAAAATAACCCAAAAGGTGTAA---TATC TTTGATTGGGCTTATTATGAGGTTGATTCAAAGCGTTGACTGGCTGTGTTAAGTCTATC *
PacBio read	ATATACCGTCATAAACTATATTAATAA---TAATGATATTCTAAATCGAA----- TCATATAATGGTTAGAGTTATTATGGGTTACGATGTTAAATTATTGGC *** *
SB Mito rep	-----GACCACAAATA-----ATATTAATAGATTAAAT CTATATCTTGTCTGGTCATGGTTCTGTGTGGTATGTTTGTAAATTGATG *
PacBio read	AAG-----AAAAACTAAAAATTTCATTCAACCCATGCCCTAAACCCAC AGGCTATAATCATGTGAAATCTGGAGATGATTGATTAACTAAGTTACAGCCTTGT *
SB_Mito_rep	ATTCGAATCTGAATATCAT---TCTGACATGTTACTAAACGATTAC---CTTCGAAA TCCTGTAATTCAAGTGTATGTTA-ATGTT---TATGAGGGTTCAAGTTCCACC *
PacBio read	TGCTTACCCCAAAACACTATAGTGAACTAACCCATCCAATAAACACTGAAAATATA TACTTGC-----AATTTTTAGTGGATGATACTAGC-----ATGGTCAGAT-TG *
SB Mito rep	ACTGTTAATAACT-----AAACTCAGTTAGACACTTATTACTCGGCCA---CG GCTGTTAATAATGATATATTGACATTATTATATGATCTATTGTTAGTAGATTG ***** *
PacBio read	ATATGTTAATTATAAT-----CTAACTTAGCT---ATGCTAATAG---GTT GTAGGGTTAATGTTAGGATTAGTCTTAAATTACTGTTAGATTAGTAGATTGTT *
SB_Mito_rep	CATATCTATAACATACTAAGATTAAGAGT-CTTATGCTTTA-----ATTCTTA---A CATTACAATAATAGAAAAGTTAGGTGAAGTATAATTATAGGAATCTTAATGTTA *** *
PacBio read	GCTACTATCAGTAAAATCTATTATGCTAACTACAATCATCTTATGTTGTATGAAAAAC TTTACTTTATTGTTAGTTATGTTATAAAATTTGGTAAAGTTGTA-GTAGAAC **** *
SB Mito rep	TGGCATTAA-----CTAAACCACCT---CTTGACCAAAACAAACATGCTATA

SB Mito rep	-GATGTTAAAAACTTAGTGTATAGCATGTTTGGTCAAAGAGGTGGTTAGTTA * * *** * * * * * * * * * * * * *
PacBio read SB_Mito_rep	ACACTAAGTTTTAACATCGTACTCTACAAACAT-AGATGATTAGTGTAGCAATAA-A ATGCC-AGTTTTA-----CATACAAACATAAGTGA-TTGAGTTAGCAATAATA *
PacBio read SB Mito rep	GATTTTACTGATAGTAGCTTAAGAAT--AAGCATAAAGACTCTTAATC-TAGTATGTTAT GATTTTACTGATAGTAGCTTAAGAATTAAGCATAA-GACTCTTAATCTTAGTATGTTAT *
PacBio read SB Mito rep	AAGATATGAACCTATTAGCATAGCTAAGT--GTTAATAATT-AACATATCGTGGCCGAGT -AGATATGAACCTATTAGCATAGCTAAGTGTAGATTATAATTAAACATATCGTGGCCGAGT *
PacBio read SB_Mito_rep	AATAAA--GGTAAACTGAGTTAGTATTATAAC--TATATTTTCAGTGTATTATT AATAAAGTGTCTAAACTGAGTTAGTATTATAACAGTTATA-TTTTCAGTGTATTATT *
PacBio read SB Mito rep	GGTAGGGTTAGTCACTATAGTGTGTTGGGTAAAGCATTGCAAGGTAATCTTATAG GGTAGGGTTAGTCACTATAGTGTGTTGGGTAAAGCATTGCAAGGTAATCGTATAG *
PacBio read SB Mito rep	TAACATG-CAGAATGATATTCAAGATTG-ATGTGGGTTTAGGCATGGGTGAATGA TAACATGTCAGAATGATATTCAAGATTGCAATGTGGGTTTAGGCATGGGTGAATGA *
PacBio read SB Mito rep	AAAATTTTTAG-TTTCTTATTAAATCTATTAAATATTATTGTGGCTTCGATTAGA AAAATTTTTAGTTCTTATTAAATCTATTAAATATTATTGTGGCTTCGATTAGA *
PacBio read SB_Mito_rep	AATATCATTATTAAATATAGTTATGACGGTATATGATATTACACCTTGGT--- AATATCATTATTAAATATAGTTATGACGGTATATGATATTACACCTTGGTGTAA *
PacBio read SB Mito rep	TGTTATTTTTTTTTGGTCTTGGTT-TATGGCAGAGCTTAAGTGGGTATAT TTTTTTTTTTTTTTGGTCTTGGTTATATGGCAGAGCTTAAGTGGGTATAT *
PacBio read SB Mito rep	AAAGTGGATGTA-----AATTATGTTATGTCGCTTTGGTATTAGTATTATGGTCAGTGTAT AAAGTGGATTAATAGATAGTAATTATGTTATGTCGCTTTGGTATTAGTATTATGGTCAGTGTAT *
PacBio read SB_Mito_rep	TTTATGAAGAATAGAAGCTTGTGCTTTGGTATTAGTATTATGGTCAGTGTAT TTTATGAAGAATAGAAGCTTGTGCTTTGGTATTAGTATTATGGTCAGTGTAT *
PacBio read SB_Mito_rep	GTGGATTATTTAAAA--ATATTTAATAATTTAGTA---GATGATATAGTGGTTGGTT GTGGATTATTTAAAAATATTTAATAATTTAGTAGATTGATAATATAGTGGTTGGTT *
PacBio read SB Mito rep	TATCAGGCGTAG--ATTTGAGCGAATTACGGATATGTTATGTTAGGTTATTTGAG TATCAGGCGTAGATATTTGAGCGAATTACGGATATGTTATGTTAGGTTATTTGAG *
PacBio read SB Mito rep	TATTGAGAGTTAATAATGATTTATTAGTAGCCTTTATATAAT-GGAGAACGTA TATTGAGAGTTAATAATGATTTATTAGTAGCCTTTATATAATGGGAGAACGTA *
PacBio read SB_Mito_rep	GGCTTATCCTAT-TTCAATTACGTAAGGGGCCTAAAAAGGGTGGTATTGTTGGTTACT GGCTTATCCTATATTCAATTACGTAAGGGGCCTAAAAAGGGTGGTATTGTTGGTTACT *
PacBio read SB Mito rep	ACAGAGGTTGCTGACTTTAAAGTGTAAAGTAAGAGTAAGTTAGGAGAGTTAGGTT ACAGAGGTTGCTGACTTTAAAGTGTAAAGTAAGAGTAAGTTAGGAGAGTTAGGTT *
PacBio read SB Mito rep	TCGTAGATGATTTATTGTTAGGTTGCGTAATTTCATGCTCTTGTCAATGGT TCGTAGATGATTTATTGTTAGGTTGCGTAATTTCATGCTCTTGTCAATGGT *
PacBio read SB_Mito_rep	GGTATTATAGATTAATAATAGGAAATGTGTAGGAAATGAATATTACCTTACCTTT GGTATTATAGATTAATAATAGGAAATGTGTAGGAAATGAATATTACCTTACCTTT *
PacBio read SB_Mito_rep	AGTAGTATGTTCTGCTTAAGTTAGA---CTTATTGATAGGTTGATGTTCATGGCTAA AGTAGTATGTTCTGCTTAAGTTAGAACTTATTGATAGGTTGATGTTCATGGCTAA

PacBio read SB Mito rep	GTATGGTTAATTAGTCTATTCTGATCCCTTCTAGTGTGATGTTGAGATGATACT GTATGGTTAATTAGTCTATTCTGATCCCTTCTAGTGTGATGTTGAGATGATACT *****
PacBio read SB Mito rep	GATGTGTATAATAATTCTGGTTAATGTACGGTAGCTATGGAAACCTGAC---A GATGTGTATAATAATTCTGGTTAATGTACGGTAGCTATGGAAACCTGACATTAA ***** *
PacBio read SB_Mito_rep	TAGAA-TATGATAATTGTAGTACCTCTGGTTACATCATATGACTTATAGTGTACTAAG TAGAATTATGATAATTGTAGTACCTCTGGTTACATCATATGACTTATAGTGTACTAAG *****
PacBio read SB_Mito_rep	TGAGAGTAATCGAACACCTTGTGATTATGTAGAGTCGAGAGTGAATTAGTAAGAGGGAT TGAGAGTAATCGAACACCTTGTGATTATGTAGAGTCGAGAGTGAATTAGTAAGAGGGAT *****
PacBio read SB Mito rep	TAGAGTAGAATATAGGAGTA-TTTTTTTAATTATATTGCTTGTGAATATTTAATTAT TAGAGTAGAATATAGGAGTATTTTTTAATTATATTGCTTGTGAATATTTAATTAT *****
PacBio read SB Mito rep	GTTTATTTAGTGTAGTGAGGTTACGATATTGAGTATAAATGAGATGTTAATAGT GTTTATTTAGTGTAGTGAGGTTACGATATTGAGTATAAATGAGATGTTAATAGT *****
PacBio read SB Mito rep	AATGAATTGATGCTGTTATAATGATGCGTGGTCGTTCTCGACTTCGGTTGATGT AATGAATTGATGCTGTTATAATGATGCGTGGTCGTTCTCGACTTCGGTTGATGT *****
PacBio read SB Mito rep	ATTGTTTCAGTGATTGAGAGTATTGTTATGGAATATTGATATTTAATATGTTT ATTGTTTCAGTGATTGAGAGTATTGTTATGGAATATTGATATTTAATATGTTT *****
PacBio read SB Mito rep	ATTAGAATTATAATGAGCTATTAGTATCTGAGTAAAATTGATTAGCGTAGTAG ATTAGAATTATAATGAGCTATTAGTATCTGAGTAAAATTGATTAGCGTAGTAG *****
PacBio read SB Mito rep	TTAAAGTTAACGTTG-TTACACACAAAGATGAAGTAAGGTTGAGTATGCTCCGG TTAAAGTTAACGTTGTTACACACAAAGATGAAGTAAGGTTGAGTATGCTCCGG *****
PacBio read SB_Mito_rep	CGCCATAATTAAAAATTAGTTGATCAGTTAAGTAGTTGATAAAGAATATA CGCCATAATTAAAAATTAGTTGATCAGTTAAGTAGTTGATAAAGAATATA *****
PacBio read SB Mito rep	AGCTTGGGTGCTAACGGATCTATTAAAGTAGACTTACTGAGGTTAGGTAAAG AGCTTGGGTGCTAACGGATCTATTAAAGTAGACTTACTGAGGTTAGGTAAAG *****
PacBio read SB Mito rep	TAGAAAAAGTATAAGGCTTATAGTAAGCTCACCATCGAATGGTGAGTAAGATTAA TAGAAAAAGTATAAGGCTTATAGTAAGCTCACCATCGAATGGTGAGTAAGATTAA *****
PacBio read SB_Mito_rep	AA---TCTCTTTATAGTC----TTGTTTGTTTTTTTGGGGGGTAAGTTAG AAATCTCTATTATAGTCTGTTTGTGTTTTGGGGGGTAAGTTAG ** *** ***** * *****
PacBio read SB Mito rep	TTGTATTTAGTCCAATAAAATTTTGTAAAAAGTATGTTTAATTCAATTATAT TTGTATTTAGTCCAATAAAATTTTGTAAAAAGTATGTTTAATTCAATTATAT *****
PacBio read SB Mito rep	AATAT---TTATATAATATAATATAATGTAAAATAAGTATAGTATAGTA-GGGGGAAAA AATATATTATAATATAATATAATGTAAAATAAGTATAGTATAGTAAGGGGGAAAA *****
PacBio read SB_Mito_rep	ATAATAAGGTGAACGTATTGCATACCTATAGGCAAATGTGATTGTATTATCTGAGAATG ATAATAAGGTGAACGTATTGCATACCTATAGGCAAATGTGATTGTATTATCTGAGAATG *****
PacBio read SB_Mito_rep	GGTGTAGGCAAATATGATTACTCCTAAATTGGTTAGTCTATAGGGTTGTATTTAG GGTGTAGGCAAATATGATTACTCCTAAATTGGTTAGTCTATAGGGTTGTATTTAG *****
PacBio read SB Mito rep	AGTCGTAGTTAACGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAATTATG AGTCGTAGTTAACGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAATTATG *****

PacBio read	TTAGTCTATAGAGGGTGTATTAA-GGTTTATGTTAACGGCTAACGGATTCTAACAGCAAAT
SB Mito rep	TTAGTCTATAGAGGGTGTATTAGGGTTATGTTAACGGCTAACGGATTCTAACAGCAAAT *****
PacBio read	GTGATTGTAECTCCTAAAATTATGTTAGTCTATAGA-GGTGTATTTAAAGATAATGCAT
SB Mito rep	GTGATTGTAECTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTAAAGATAATGCAT *****
PacBio read	TTAATTCTTGTAAGGGGATGTACAACAAATTGTTAACATTGATTTATGCTTTA
SB_Mito_rep	TTAATTCTTGTAAGGGGATGTACAACAAATTGTTAACATTGATTTATGCTTTA *****
PacBio read	ATATTGCTTTCAATTGAGGTAAATTAAATGATATTTTATATTTTATGTATTCAATTGT
SB Mito rep	ATATTGCTTTCAATTGAGGTAAATTAAATGATATTTTATATTTTATGTATTCAATTGT *****
PacBio_read	GGTTATATTAAAGGTATAAAAAGTTGTTCTGTAACTTTTGTGTTTTTTGGT
SB Mito rep	GGTTATATTAAAGGTATAAAAAGTTGTTCTGTAACTTTTGTGTTTTTTGGT *****
PacBio_read	GGTGGTAAGTTAGTTGTTAGTCCAATAAAATTGGTAAAGTATGTTTA
SB Mito rep	GGTGGTAAGTTAGTTGTTAGTCCAATAAAATTGGTAAAGTATGTTTA *****
PacBio read	ATTCATTTATATAT-ATATATTATAATATAATATAATGAAAATAAGTATAGTAG
SB_Mito_rep	ATTCATTTATATATAATATATTATAATATAATGAAAATAAGTATAGTAG *****
PacBio read	GGGGAAAAATAAAGGTGAGCGTATCGCATACCTATAGGCAAATGTGATTGTTATCT
SB Mito rep	GGGGAAAAATAAAGGTGAGCGTATCGCATACCTATAGGCAAATGTGATTGTTATCT *****
PacBio read	GTAGAATGGGTGAGGCAAATATGATTACTCCTAAATTGGTTAGTCTATAGGGTT
SB Mito rep	GTAGAATGGGTGAGGCAAATATGATTACTCCTAAATTGGTTAGTCTATAGGGTT *****
PacBio read	GTATTTAGAGTCGTAGTTAACGGCTAACGGATTCTAAAGCAAATGTGATTGACTCCTA
SB_Mito_rep	GTATTTAGAGTCGTAGTTAACGGCTAACGGATTCTAAAGCAAATGTGATTGACTCCTA *****
PacBio read	AAATTATGTTAGTCTATAGAGGGTG--TTTAAAG--AAATGCATTAAATTCTGTAA
SB Mito rep	AAATTATGTTAGTCTATAGAGGGTGATTAAAGATAATGCATTAAATTCTGTAA *****
PacBio read	GGGGGATGTACAACAAATTGTTAACATTGGA-TTTATGCTTAATATTGCTTCATT
SB Mito rep	GGGGGATGTACAACAAATTGTTAACATTGGA-TTTATGCTTAATATTGCTTCATT *****
PacBio_read	GA-GTAAATTAAATGAT-TTTTTATATT-----
SB Mito rep	GAGGTAAATTAAATGATATTTTATATTGTATTCAATTGGGTTATATTAAAG ** *****
PacBio read	-----
SB_Mito_rep	GTATAAAAAGTTGTTCTGTAATCTTTTGTGTTTTGGTGGTAAGTTAGT
PacBio read	-----
SB_Mito_rep	TGTATTTAGTCCAATAAAATTGGTAAAGTATGTTTAATTCAATTATATA
PacBio read	-----
SB_Mito_rep	ATATATTATAATATAATATAATGAAAATAAGTATAGTATAGTAGGGGGAAAAATAATA
PacBio_read	-----
SB_Mito_rep	AGGTGAGCGTATCGCATCCATAGGCAAATGTGATTGATTCTGTAGAATGGGTGTA
PacBio read	-----
SB Mito rep	GGCAAATATGATTACTCCTAAATTGGTTAGTCTATAGGGTTGTATTAGAGTCGT
PacBio read	-----
SB Mito rep	AGTTAACGGCTAAGGATTCTAAAGCAAATGTGATTGACTCCTAAATTATGTTAGTCT

PacBio read	-----
SB_Mito_rep	ATAGAGGGTGTATTAGGGTTGTGTGTAGGAATGGTAAATAGTGATAGAAAGGC
PacBio read	-----
SB_Mito_rep	TTATAGTAAGCTCACCATCGAATGGTAGTAAGATTTAAAATCTCTATTATCTGT
PacBio read	-----
SB_Mito_rep	TTTGTGTTGTTTTTTTTGTGGTGGTAAGTTAGTTGTTAGTTCCAATAAAA
PacBio read	-----
SB_Mito_rep	AATTTTGTAAGTATGTTAATTCAATTATATATTATAAATAATATAATATAA
PacBio read	-----
SB_Mito_rep	TAAATAAGTATAAGTATAGTAGGGGGAAAAATAAAAGTAGGTAAACGTATTGCATACTATA
PacBio read	-----
SB_Mito_rep	GGCAAATGTGATTGTATTACTGTAGAATGGGTGAGGAAATATGATTACTCCTAA
PacBio read	-----
SB_Mito_rep	ATTGGTTAGTCATAAGGGTTGTATTAGAGTTCTAGTTAAGGCTAAGGATTCTAAA
PacBio read	-----
SB_Mito_rep	GCAAATGATTGACTCCTAAAATTATGTTAGTCTATAGGGGTGTTAGGTTATGT
PacBio read	-----
SB_Mito_rep	TTAAGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTATGTTAGTCTATA
PacBio read	-----
SB_Mito_rep	GAGGGTATTAAAGATAATGCATTTAATTCTGTAAGGGGATGTACAACAAATTG *****
PacBio read	-----
SB_Mito_rep	TTTAAATTGGATTATGCTTAATTGCTTCAATTGAGGTAAATTAAATGATATT
PacBio read	-----
SB_Mito_rep	TCAATTGTGGTTATA--TTAAGGTAT-AAAAGTTGTTG--TCTG *****
PacBio read	-----
SB_Mito_rep	TAATC-----TTTTGTTTTTTGGTGGTAGTTAGTTAGTGTAT TAATCNNNNNNNNNNNNNNCTTTTTGTTTTGGTGGTAGTTAGTTAGTGTAT *****
PacBio read	-----
SB_Mito_rep	TTTAGTTCCAATAAAATTTTGTAAGAATGTTAATTCAATTATATAAATATA TTAGTTCCAATAAAATTTTGTAAGAATGTTAATTCAATTATATAAATATA *****
PacBio read	-----
SB_Mito_rep	ATATAATAT-----ATATAATG---AATAAGTATAGTATAGTAGGGGGAAAAATAATAA TTATAATATAATATAATATAATGTAAGAATAGTATAGTATAGTAGGGGGAAAAATAATAA *****
PacBio read	-----
SB_Mito_rep	GGTAGCGTATCGCATACTATAGGCAAATGTGATTGTATTACTGTAGAATGGGTGTTAG GGTAGACGTATTGCTACCTATAGGCAAATGTGATTGTATTACTGTAGAATGGGTGTTAG *****
PacBio read	-----
SB_Mito_rep	GCAAATATGATTACTCCTAAATTGGTTAGTCTATAGGGTTGTTAGAGTTCTG GCAAATATGATTACTCCTAAATTGGTTAGTCTATAGGGTTGTTAGAGTTCTG *****
PacBio read	-----
SB_Mito_rep	A----- AGTTAAGGTCTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAAATTATGTTAGTCT *
PacBio read	-----
SB_Mito_rep	GTTTAAGGTCTAAGGATTCTAAAGCAAATGTGATTG ATAGAGGGTGTATTAGGGTTATGTTAAGGTCTAAGGATTCTAAAGCAAATGTGATTG *****
PacBio read	TACTCCTAAAATTATGTTAGTCTATAGAGGGTGTATTAAAGATAATGCATTAAATT

SB Mito rep	TACTCCTAAAATTATGTTAGTCTATAGAGGGGTGATTTAAAGATAAATGCATTTAATT *****
PacBio read SB_Mito_rep	TCTTGTAAAGGGGGATGTACAACAAATTGTTAAATTGGATTTATGCTTAATATTGC TCTTGTAAAGGGGGATGTACAACAAATTGTTAAATTGGATTTATGCTTAATATTGC *****
PacBio read SB Mito rep	TTTCAATTGAGGTAAATTAAATGATA-TTTTTATATTGTATTCAATTGTGGTTA TTTCAATTGAGGTAAATTAAATGATA-TTTTTATATTGTATTCAATTGTGGTTA *****
PacBio_read SB Mito rep	TATTTTAAGGTAT----- TATTTTAAGGTATAAAAGTTGTTCTGTAATCTTTTGTTTTTGGTGGTG *****
PacBio_read SB_Mito_rep	----- GTAAGTTAGTTGATTTAGTTCCAATAAAATTGGTAAAAAGTATGTTAAC
PacBio read SB_Mito rep	----- ATTATATAAAATATTATAATAGTATAATGTAAAAATAGTATAGTAGGGGGAAA
PacBio read SB_Mito rep	----- AATAATAAGGTGACCGTATTGCATACTATAGGCAAATGTGATTGATTCTGAGAAT
PacBio read SB_Mito rep	----- GGGTGTAGGCAAATATGATTATACTCTTAAATTGGTTAGTCTATAGGGTTGTATT
PacBio read SB_Mito rep	----- GAGTCGTAGTTAAGGTCTAAGGATTCTAAAGCAAATGTGATTGACTCTAAATTAT
PacBio read SB_Mito rep	----- GTTAGTCTATAGAGGGGTGATTTAAAGATAAATGCATTTAATTCTTGTAAAGGGGAT
PacBio read SB_Mito rep	----- GTACAACAAATTGTTAAATTGGATTTATGCTTAATATTGCTTCAATTGAGGTAA
PacBio_read SB_Mito_rep	-----AA ATTAAATGATATTTTTATATTGTATTCAATTGTGGTTATATTAAAGGTATAAA **
PacBio read SB_Mito rep	----- AAGTTGTTGTTCTGTAATT----- AAGTTGTTGTTCTGTAATCTTTTGTGTTTTGGTGGTGGTAAGTTAGTTGATT *****
PacBio read SB_Mito rep	----- TTAGTCCAATAAAATTGGTAAAGTATGTTAATTCAATTATAATATAATAT
PacBio read SB_Mito rep	----- TATAATATAATATAATGTAAAATAAGTATAGTATAGTAGGGGGAAAATAAAGGT
PacBio read SB_Mito rep	----- GAGCGTATCGCATTCTATAGGCAAATGTGATTGATTCTGTAGAATGGGTGAGGCA
PacBio read SB_Mito rep	----- AATATGATTATACTCTTAAATTGGTTAGTCTATAGGGTTGTATTAGAGTCGTAGT
PacBio_read SB_Mito rep	----- TTAAGGTCTAAGGATTCTAAAGCAAATGTGATTGACTCTAAAATTATGTTAGTCTATA
PacBio_read SB_Mito rep	----- GAGGGTGTATTAGGGTTGTGTTGTTAGGAATGGTAAATAGTTGGATAGAAAGGCTT
PacBio read SB_Mito rep	----- ATAGTAAGCTTCACCACGAAATGGTGGTAAGATTAAAAATCTCTATTAGTCTTGT

PacBio read	-----TTTTTTTTGGTGGTGGTAAGTTAGTTAGTTAGTTAGTCCAATAA
SB Mito rep	TTTTGTTTTGGTTTTGGTGGTAAGTTAGTTAGTTAGTCCAATAA *****
PacBio read	TTTTTGTAAGTATGTTAACATTATATATAATATA-TATAATAT-ATATAAT
SB Mito rep	TTTTTGTAAGTATGTTAACATTATATATAATATAATATAATATAAT *****
PacBio read	GTAAAATAAGTATAGTATAGTAGGGGGAAAAATAATAAGTGAGCGTATCGCATACCTA
SB_Mito_rep	GTAAAATAAGTATAGTATAGTAGGGGGAAAAATAATAAGTGAGCGTATCGCATACCTA *****
PacBio read	TAGGCAAATGTGATTGTATTATCTGTAGAATGGGTGAGGCAAATATGATTATACTCCTT
SB_Mito_rep	TAGGCAAATGTGATTGTATTATCTGTAGAATGGGTGAGGCAAATATGATTATACTCCTT *****
PacBio read	AAATTGGTTAGTCTATAGGGGTTGATTAGAGTTCGTAGTTAACGGTCTAAGGATTCT
SB Mito rep	AAATTGGTTAGTCTATAGGGGTTGATTAGAGTTCGTAGTTAACGGTCTAAGGATTCT *****
PacBio read	AAAGCAAATGTGATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGGTGATTAAAGA
SB Mito rep	AAAGCAAATGTGATTGTACTCCTAAAATTATGTTAGTCTATAGAGGGGTGATTAAAGA *****
PacBio read	TAAATGCATTTAATTCTGTAAAGGGGGATGTACAACAAATTGTTAAATTGGATT
SB Mito rep	TAAATGCATTTAATTCTGTAAAGGGGGATGTACAACAAATTGTTAAATTGGATT *****
PacBio read	TATGTTTAATATTGCTTCATTGAGGTAATTAAATGATATTTTATATTGT
SB Mito rep	TATGTTTAATATTGCTTCATTGAGGTAATTAAATGATATTTTATATTGT *****
PacBio read	ATTCAATTGTGGTTATATTAAAGGTATAAAAGTTGTTCTGTAACTTTTG-T
SB Mito rep	ATTCAATTGTGGTTATATTAAAGGTATAAAAGTTGTTCTGTAACTTTTGTT ***** *
PacBio read	TTTTTTGGTGGTGGTAAGTTAGTTAGTTAGTCCAATAAAATTGGTAA
SB Mito rep	TTTTTTGGTGGTGGTAAGTTAGTTAGTTAGTCCAATAAAATTGGTAA *****
PacBio read	AGTATGTTTAATTCTTATATAATATATTATAAT--TATAATGTAAAAGTA
SB_Mito_rep	AGTATGTTTAATTCTTATATAATATATAATATAATAGTATAATGTAAAAGTA *****
PacBio read	TAGTATAGTA-GGGGAAAAATAATAAGGTGAGCGTATCGCATACCTATAGGCAAATGT
SB Mito rep	TAGTATAGTAGGGGGAAAAATAATAAGGTGAGCGTATCGCATACCTATAGGCAAATGT *****
PacBio read	ATTGTATTATCTGTAGAATGGGTGAGGCAAATATGATTATACTCCTAAATTGGTTAG
SB Mito rep	ATTGTATTATCTGTAGAATGGGTGAGGCAAATATGATTATACTCCTAAATTGGTTAG *****
PacBio read	TCTATAGGGTTGTATTAGAGTTCGTAGTTAAGGTCTAAGGATTCTAAAGCAAATGT
SB_Mito_rep	TCTATAGGGTTGTATTAGAGTTCGTAGTTAAGGTCTAAGGATTCTAAAGCAAATGT *****
PacBio read	ATTGTACTCCTAAATTATGTTAGTCTATAGAGGGGTGATTAGGGTTATGTTAAGGT
SB Mito rep	ATTGTACTCCTAAATTATGTTAGTCTATAGAGGGGTGATTAGGGTTATGTTAAGGT *****
PacBio read	CTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAATTATGTTAGTCTATAGAGGGTG
SB Mito rep	CTAAGGATTCTAAAGCAAATGTGATTGTACTCCTAAATTATGTTAGTCTATAGAGGGTG *****
PacBio read	TATTTAGGGTTGTGTTGTTAGGAATGGTAAATAGTGGATAGAAAGGCTTATAGTAA
SB_Mito_rep	TATTTAGGGTTGTGTTGTTAGGAATGGTAAATAGTGGATAGAAAGGCTTATAGTAA *****
PacBio read	GCTTCACCACGGAATGGTGAAGATTTAAAATCTCTTATAGAAGATATTGATAA
SB_Mito_rep	GCTTCACCACGGAATGGTGAAGATTTAAAATCTCTTATAGAAGATATTGATAA *****
PacBio read	TAAAACTAATGGATAAGTTT-TTATATATTAGGTTATTATTTATGATT
SB Mito rep	TAAAACTAATGGATAAGTTATTATATA-TTTTAGGTTATTATTTATGATT *****

PacBio read	-TATTAGTACTCTTGTAGTATTATAGTATTAAAGTTGTG-TTTTTTAGTACTAA
SB Mito rep	ATATTAGTACTCTTGTAGTATTATAGTATTAAAGTTGTGTTTTAGTACTAA *****
PacBio read	GTTTGGTGGAGTGGTGGGGTGTGTTGTATATATTGTATATGTATTGTTAATTTGGTAA
SB Mito rep	GTTTGGTGGAGTGGTGGGGTGTGTTGTATATATTGTATATGTATTGTTAATTTGGTAA *****
PacBio read	TAGTGATGTTGTTATTATTGAATTATAGGTTGATTCTAGTATATTATGAT
SB_Mito_rep	TAGTGATGTTGTTATTATTGAATTATAGGTTGATTCTAGTATATTATGAT *****
PacBio read	TTTAG-TTAATTATGTTA---TATATTCTCGTAATAATTGATGATTAGTTAGTT
SB Mito rep	TTTAGTTAATTATGTTATTCGTAATAATTGATGATTAGTTAGTTAGTT *****
PacBio read	ATGTTATGGTATTGGTGGTTAGATATTAAGTTAAACTATGTTCATGTGC
SB Mito rep	ATGTTATGGTATTGGTGGTTAGATATTAAGTTAAACTATGTTCATGTGC *****
PacBio read	ACAAATTGCTGATTAAAGTTACG--TATCTTATTATTACTGGTTAGGCTGGTT
SB Mito rep	ACAAATTGCTGATTAAAGTTACGCACTTATTACTGGTTAGGCTGGTT *****
PacBio read	ATCTAGTTAGTAGGTTAGTTAGACTGCTGCTTCAAAGTAAGAGAGGAACTGTATT
SB_Mito_rep	ATCTAGTTAGTAGGTTAGTTAGACTGCTGCTTCAAAGTAAGAGGTGAACTGTATT *****
PacBio read	GGCCTAGGCTGAGGTGTATGGAAGTAAAATTGTTATCGCTGCTAACGATAATTAGGA
SB Mito rep	GGCCTAGGCTGAGGTGTATGGAAGTAAAATTGTTATCGCTGCTAACGATAATTAGGA *****
PacBio read	-TATGTTAGTGTAAATT-TGCTTCTTATGTTTATTAAATTATATTGTTAGCTAT-GAA
SB Mito rep	TTATGTTAGTGTAAATTCTGCTTCTTATGTTTATTAAATTATATTGTTAGCTATGGAA *****
PacBio read	GAAACAAAGAAGAAGAAGAGAGTTAAGAAGGAAGTTCT-AAGAAGAAGA--ACT
SB_Mito_rep	GAAACAAAGAAGAAGAAGAGAGTTAAGAAGGAAGTTCTACAAAGAAGAAGACT *****
PacBio read	TCAGTTGGTGGAAAGCGGTGGTCTGGGTTGGAGGTAAGTTAGATTAAATAA
SB Mito rep	TCAGTTGGTGGAAAGCGGTGGTCTGGGTTT-GAGGTAAGTTAGATTAAATAA *****
PacBio read	G--TTTAGGTGGATTAAAGAAGTCTATATGTTAC-TATTAAGTTGTAATTGGTAGTGT
SB Mito rep	GCTTTAGGTGGATTAAAGAAGTCTATATGTTACTTAAAGTTGTAATTGGTAGTGT *
PacBio read	AAGTGTGTTCAATTAAATTATATAAGTTGTTAGTATTGGTTAGATTCAATTAAAG
SB Mito rep	AAGTGTGTTCAATTAAATTATATAAGTTGTTAGTATTGGTTAGATTCAATTAAAG *****
PacBio read	TTTAGTAGTTGATTGGTGTAAATGCTGGTTAGGTTATTGGTAATTCTTCTTAA
SB_Mito_rep	TTTAGTAGTTGATTGGTGTAAATGCTGGTTAGGTTATTGGTAATTCTTCTTAA *****
PacBio read	TCATAAGCGGATAAGTTAGTGTATATTAGGTTATGAGGTGGTTATTGGTCT
SB Mito rep	TCATAAGCGGATAAGTTAGTGTATATTAGGTTATGAGGTGGTTATTGGTCT *****
PacBio read	T-GTTTAAGGCTCTAATTGGTTAAATTATGTGATCCATTATAATTGGTTCGTT
SB Mito rep	TGGTTAAGGCTCTAATTGGTTAAATTATGTGATCCATTATAATTGGTTCGTT *
PacBio read	AGAGGTTATAAATTGGTACTAATCATGGTATAGCTATGA-TTTTTTTTTGAT
SB_Mito_rep	AGAGGTTATAAATTGGTACTAATCATGGTATAGCTATGAATTGGTTTTGAT *****
PacBio read	GCCGGTATTAATAGGT-GATTGGT-AATATTCTCC-TTTTTTTATATAGATGA
SB Mito rep	GCCGGTATTAATAGGTGGATTGGTAAATATTCTCCATTGGTTATATAGATGA *****
PacBio read	TTTATTGTTACCTCGGTTGAATTCTTTAGTTAGGTTGATGATTCTTCATTGTTA
SB Mito rep	TTTATTGTTACCTCGGTTGAATTCTTTAGTTAGGTTGATGATTCTTCATTGTTA *****

PacBio read SB Mito rep	TATGGAGCTAAGTTATTATGGTTGGGGTT-GTTGAACGTTGTATCCTCCATTATC TATGGAGCTAAGTTATTATGGTTGGGGTTGGTGAACGTTGTATCCTCCATTATC ***** *****
PacBio_read SB_Mito_rep	TATATCTGAGAACTCTGGCTTGGGTAGATTATTGATGTTTCGTTGCATTAGCGGG TATATCTGAGAACTCTGGCTTGGGTAGATTATTGATGTTTCGTTGCATTAGCGGG ***** *****
PacBio read SB Mito rep	TGTGTCTAGATTGGTTGGTCAATTAAATTATTCTACAATTATGAGTTGTAGATT TGTGTCTAGATTGGTTGGTCAATTAAATTATTCTACAATTATGAGTTGTAGATT ***** *****
PacBio read SB Mito rep	TAAGACTTCGATAATACTGATCATATTATTACTTCAATTATTGTTACTGTCA TAAGACTTCGATAATACTGATCAT-ATTATTACTTCAATTATTGTTACTGTCA ***** *****
PacBio_read SB_Mito_rep	TACCCAGTTTAGCGGCTGGTATAACTATGTTATTATGATCGTAATTGGTACTG TA-CCAGTTTAGCGGCTGGTATAACTATG-TTATTAT-TGATCGTAATTGGTACTG ** *****
PacBio read SB_Mito_rep	-TTTTTTGAGCCGATGGGTGGTGGTATCCGCTGTTATTCAATTGGTT CTTTTTGAGCCGATGGGTGGTGGTATCCGCTGTTATTCAATTGGTT ***** *****
PacBio read SB Mito rep	TTGGGCATCCTGAGGTGATGTTTGATTCTGCCAGGATTC-GTATAGTTAGTCATA TTGGGCATCCTGAGGTGATGTTTGATTCTGCCAGGATTCGGTATAGTTAGTCATA ***** *****
PacBio read SB Mito rep	GTATGAGAATTAGTAATAATGATTCACTTGGATTATTGGATTAAATTGCTATGG GTATGAGAATTAGTAATAATGATTCACTTGGATTATTGGATTAAATTGCTATGG ***** *****
PacBio read SB Mito rep	CTTCAATAGTTGTTAGGTAGTGTAGTT-GGCTCACCATATGTTATGGTTGG CTTCAATAGTTGTTAGGTAGTGTAGTTGGCTCACCATATGTTATGGTTGG ***** *****
PacBio read SB Mito rep	ATTA-TTGACAGCTATATTTTAGTTCTGTGACTCATGA-TATAGGGATTCTAC-GGT ATTATTGACAGCTATATTTTAGTTCTGTGACT-ATGATTATAGGGATTCTACGGG *** *****
PacBio read SB Mito rep	ATAAAGGTTTTCTGATTATATGCTAAAAGATGTAGGTTACTCGTTGCTTGT ATAAAGGTTTTCTGATTATATGCTAAAAGATGTGG--TACTCGT--TTGCT ***** *****
PacBio_read SB_Mito_rep	GATGACTCCTGTTATTATGGTAGTTGGT-GGTTATACCTTGTCATCTAACAT GA---TCCTG---TGTATGGT-GTGGTGGGTTTATATTGGT---TTACAAAT ** *****
PacBio_read SB_Mito_rep	CGTGGTAGGGTGGTACCTGGC-TAGCTTTCTTCTGTTCTTCACTATTGGACATT -----AGGGGGTGTACTGGTAGCTTGTGGCTTCTTCACTATTGGACATT ** *****
PacBio read SB Mito rep	ATTATTCTGATATTGTTTGTGTCATGTTCTGCCATTATGTTCTCTTATTCTT TTATTCTGATACTGGTTGTGCTGCTATTTCATTATGTTCT---TTCGTTAGG * *****
PacBio read SB Mito rep	TATTGCTATTAAAGATAGTCGGTCGACTAAATTGATAGATTGGTTATTGTT TTCT-----TATAGTAGGGTATAATGTTG-----TATGATGGT * *****
PacBio read SB_Mito_rep	GGCCTTT---TATAATTGATAGAGTACATGTAATGAATTATAGTTATCCTTACAGT GGCCTTT---TATAATTGGATACAGT-----ATGA---ATAAGTATCTCTACAGGG *** *****
PacBio read SB_Mito_rep	CTATGATTATTGTCGATGGTT-GATTATATTG-TATTTCTATGCAACTAT TCATTGATTATTG-TCGATGGTTGGATTAACTTGTGTTTTCTATGCACT---TAT ** *****
PacBio read SB Mito rep	TTATGGTTATCGTATGATGGTTGGCTCGTCCGGTAGGTGCTTATGATGTTGGT TTA---GGTATTG---ATGGTTGGCTCG-TCGGTAAGT-TGCTATGATGTTGAGTAT *** *****
PacBio_read SB_Mito_rep	TAT--GGTTTATATACTGAGAAGATTATGTCGGTGTGTATATTCTGTTACT-- TATTGGGTTAAAGTACTGAGAAG---TATA---GGTGGTGTGTATCTGTTACTAG *** *****
PacBio read	TGCTTTCTGAGTATGTTTAATTCTTAGAGCAGAGTCATT-----

SB Mito rep	TGCTTTCGTGT--TGTTTTATTCTTG---AGAGTCATTGAGTGTCTGTAATCGTAT ***** * * * * * * * * * * * * * * * *
PacBio read SB_Mito_rep	-----TGT-----GTGTCATC TATAGGATTATGGGATCTGGCAGATGTATAACTAAAGTTATGGTGTGCCGGTCCAGC ***** * * * *
PacBio read SB Mito rep	TTAAATACGTTATA-----TATTA----- TCATATA-ACTTATATTGTTGGTAAGGTGTGATTATAATAAAGTAGTATTATGGCA * * *** * * * * * * * * * * * * * *
PacBio read SB Mito rep	-----TTTTTATGGGTAG---ATCGGCTGAATAGATATATCCATT----- TGCTTTATTTTTTATTGTAGTAGTATAGGTATAGCAGGTGTAGTTATTTATAAAAAT ***** * * * * * * * * * * * * * * * * * *
PacBio read SB_Mito_rep	-----TTAACTAATGTTA-----TGGTGC----- ATTAGTTTGTAACAAATGATAGTGTATGATGTCGCACTTATCTGTTTCGTGTACC * * * * * * * * * * * * * * * * * *
PacBio read SB Mito rep	-----TTGTGCCGTATAACTCAGAATC--TCAAT AGATTGATTAAAATTATTTAATGGTTATTGTACCTTTGTATCATGATCCTTGAT ***** * * * * * * * * * * * * * * * * *
PacBio read SB Mito rep	TATATTAACA-----TAGT----- TGTAATTACAGTATTGTATGCCGAAAGGTTACGATTTATTTAGTCTAACAAATAG * * * * * * * * * * * * * * * * * * *
PacBio read SB Mito rep	-----ACTTAATA CTGAATTGGCTAGAGGGGTTACAGGTATTAAAGTAGATTCTAAAGAGACTAAATA * * * * * * * * * * * * * * * * * * *
PacBio read SB_Mito_rep	-----TGTGGTTATGTGTTACACGTGTC----- AGTAG-TGATAAAATTACCCGTGATCTTATCTGGTGGTAACTGTTATCGAGTTG * * * * * * * * * * * * * * * * * * *
PacBio read SB Mito rep	-----TTGATCATATT----- TGTGGAAAGTTGATTGTGTTAGTTCTAGAGCAGAGATTAAATTGTTATGAGTAGT * * * * * * * * * * * * * * * * * * *
PacBio read SB Mito rep	-----AATCAAAT----- CTTAGATTAAGATTGTCTTGGATTAGTAAGTTATTCATAGCTTGTCACTACTGTG * * * * * * * * * * * * * * * * * * *
PacBio read SB_Mito_rep	-----ACTCAGAACGTATCCTTACA----- GTAATATCTGCTTGATTGGTAGTTACCCCTATAATTGGTTGTATGATGAAGGAAT * * * * * * * * * * * * * * * * * * *
PacBio read SB_Mito_rep	-----TCAAAT----- TAGTAGGTTATGGTAAAGTTAGTTGTCTTTTTCGGTGATGCTTCGACTGTTATT * * * * * * * * * * * * * * * * * * *
PacBio read SB Mito rep	-----CCTC AAAAACATTCTATCTGATGAGGTTAGATAGTATGACCTGCCCA * * * * * * * * * * * * * * * * * * *

Coding region:  Non-coding region:  Assembly gap: 

Figure J: Pairwise alignment of an error corrected PacBio read spanning the repetitive region of the *S. bovis* mitochondrial genome

The alignment shows a section of the *S. bovis* mitochondrial genome, centred on the mitochondrial region, aligned with a PacBio read found to span this region. The PacBio read was error corrected byCanu. The Mitochondrial sequence is rearranged such that it is centred on the repetitive region with 3kb of non-repetitive sequence included on either side. The PacBio read is the error corrected reverse complement of read m141015_175441_42138R_c100696632550000001823138502281592_s1_p0/1168 21/381_8437 RQ=0.777 id=41428_0

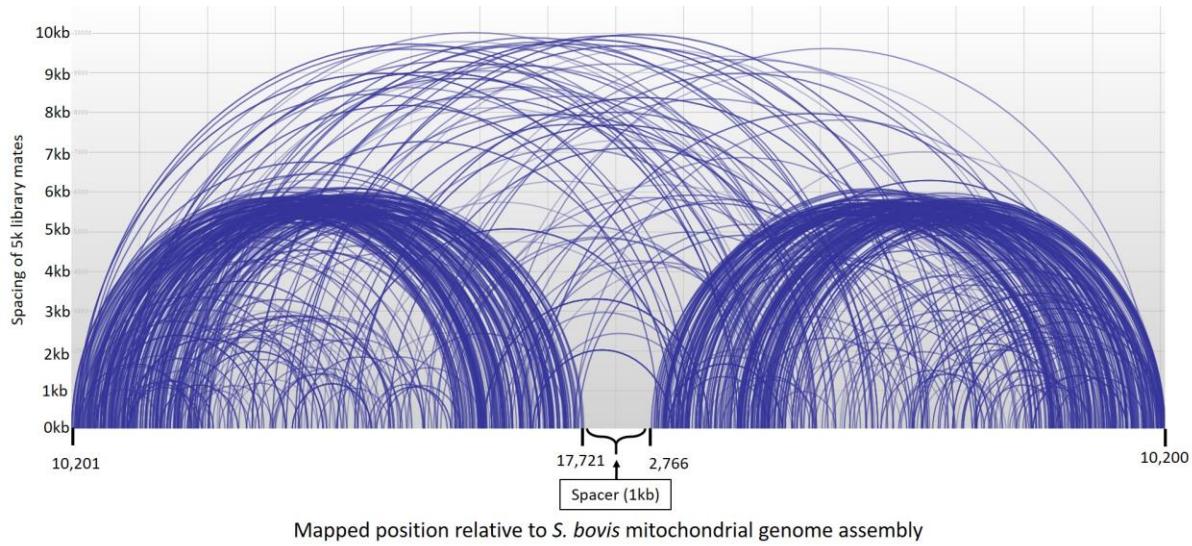


Figure K: 5kb mate pair library mapped to mitochondrial genome centered on the repetitive region

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.

Protein	Reference organism / Target organism ^a	Identity / Coverage of reference sequence	Description
Estradiol 17beta-dehydrogenase	<i>S. haematobium</i> / <i>S. bovis</i>	98% / 100%	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
Glycoprotein Omega-1	<i>S. haematobium</i> / <i>S. bovis</i>	92% / 100%	Secretion from <i>Schistosoma</i> eggs thought to trigger Th2-type immune response leading to cancerous environment

^aQuery 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 ^a	Identity ^b / 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

^aQuery and target species of the Exonerate search. ^bSequence 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.