

1 Preferential suppression of *Anopheles gambiae* host sequences allows detection
2 of the mosquito eukaryotic microbiome

3

4 Eugeni Belda^{1,2}, Boubacar Coulibaly³, Abdrahamane Fofana³,
5 Abdoul H. Beavogui⁴, Sekou F. Traore³, Daryl M. Gohl⁵, Kenneth D. Vernick^{1,2,*},
6 Michelle M. Riehle⁶

7

8 ¹Department of Parasites and Insect Vectors, Unit of Genetics and Genomics of
9 Insect Vectors, Institut Pasteur, Paris, France.

10 ²CNRS Unit of Hosts, Vectors and Pathogens (URA3012), Paris, France

11 ³Malaria Research and Training Centre (MRTC), Faculty of Medicine and
12 Dentistry, University of Mali, Bamako, Mali.

13 ⁴Centre de Formation et de Recherche en Santé Rurale de Mafèrinyah, Conakry,
14 Republic of Guinea.

15 ⁵University of Minnesota Genomics Center, Minneapolis, Minnesota, USA

16 ⁶Department of Microbiology and Immunology, University of Minnesota,
17 Minneapolis, MN, USA.

18

19

20 * Correspondence:

21 email: kvernick@pasteur.fr

22 address: Institut Pasteur, GGIV, 28 rue du Dr Roux, 75724 Paris Cedex 15,

23 FRANCE

894 **Supplementary Figure Legends**

895

896 **Figure S1:** Rarefaction plots based on phylogenetic diversity. Samples were
897 grouped according to their amplification with and without PNA (for PNA
898 samples, data from all 4 PNA concentrations are pooled in this analysis).
899 Phylogenetic diversity was computed from rarified OTU tables at sampling
900 depths from 10 to 1000 sequences. At each sampling depth, the average and
901 standard error of 10 rarified OTU tables is represented.

902

903 **Figure S2:** Taxonomic profile of larval sample from 18s rRNA gene amplicons of
904 V4 and V9 hypervariable regions amplified at different concentrations of PNA
905 blocker after exclusion of mosquito and mammalian sequences. Chi-square tests
906 were carried out to test whether the presence or concentration of PNA blocker
907 affects the relative fractions of the eukaryotic lineages detected. There was no
908 significant difference in the composition of eukaryotic microbiota (chi-square
909 $p=0.92$ for V4 and $p=1$ for V9) across all samples, with and without PNA blocker,
910 showing that the presence of PNA blockers does not influence differential
911 amplification of any particular eukaryotic microbe, and instead functions solely
912 to block the amplification of the mosquito host template. Note that this Figure, as
913 compared to Figures 3 and 4, provides the full taxonomic list of the larval sample
914 only, with the Other category uncollapsed to show all lineages. The larval sample
915 yielded most of the microbial diversity, and for simplicity the other two samples
916 represented in Figures 3 and 4, Animal Bloodmeal and Human Bloodmeal, are
917 not included here, so some classes in Figures 3 and 4 are not represented in
918 Figure S4.

919

920 **Figure S3:** Multiple sequence alignment of representative 18s rRNA gene
921 sequences from mosquito and mammal (Silva 119 database) with *Mammalia* and
922 *Anopheles* anneal-inhibiting blocking primers targeting the V4 hypervariable
923 region used in the present study. The alignment region corresponds to the 5' end
924 of the V4 hypervariable region.

925

926 **Figure S4:** Multiple sequence alignment of representative 18s rRNA gene
927 sequences from mosquito and mammal (Silva 119 database) with *Mammalia* and
928 *Anopheles* anneal-inhibiting blocking primers targeting the V9 hypervariable
929 region used in the present study. The alignment region corresponds to the 3' end
930 of the V9 hypervariable region.

931

932 **Figure S5:** Multiple sequence alignment of the 18s rRNA gene V9 hypervariable
933 region of wild *A. coluzzii* (Acol) and *A. gambiae* (Agam) from Burkina Faso.
934 Alignments represent the reversed and complemented genome region between
935 nucleotide coordinates X:23669218-23669319 of the Vectorbase AgamP4
936 annotation. Sequence of the V9 PNA blocker (as shown in Table S1) is
937 highlighted in purple.

938

939 **Figure S6:** Multiple sequence alignments as in Figure S5, but for the V4
940 hypervariable region. Alignments represent the genome region between
941 nucleotide coordinates X:24339450-24339510 of the Vectorbase AgamP4
942 annotation. Sequence of the V4 PNA blocker (as shown in Table S1) is
943 highlighted in purple.

Figure S1

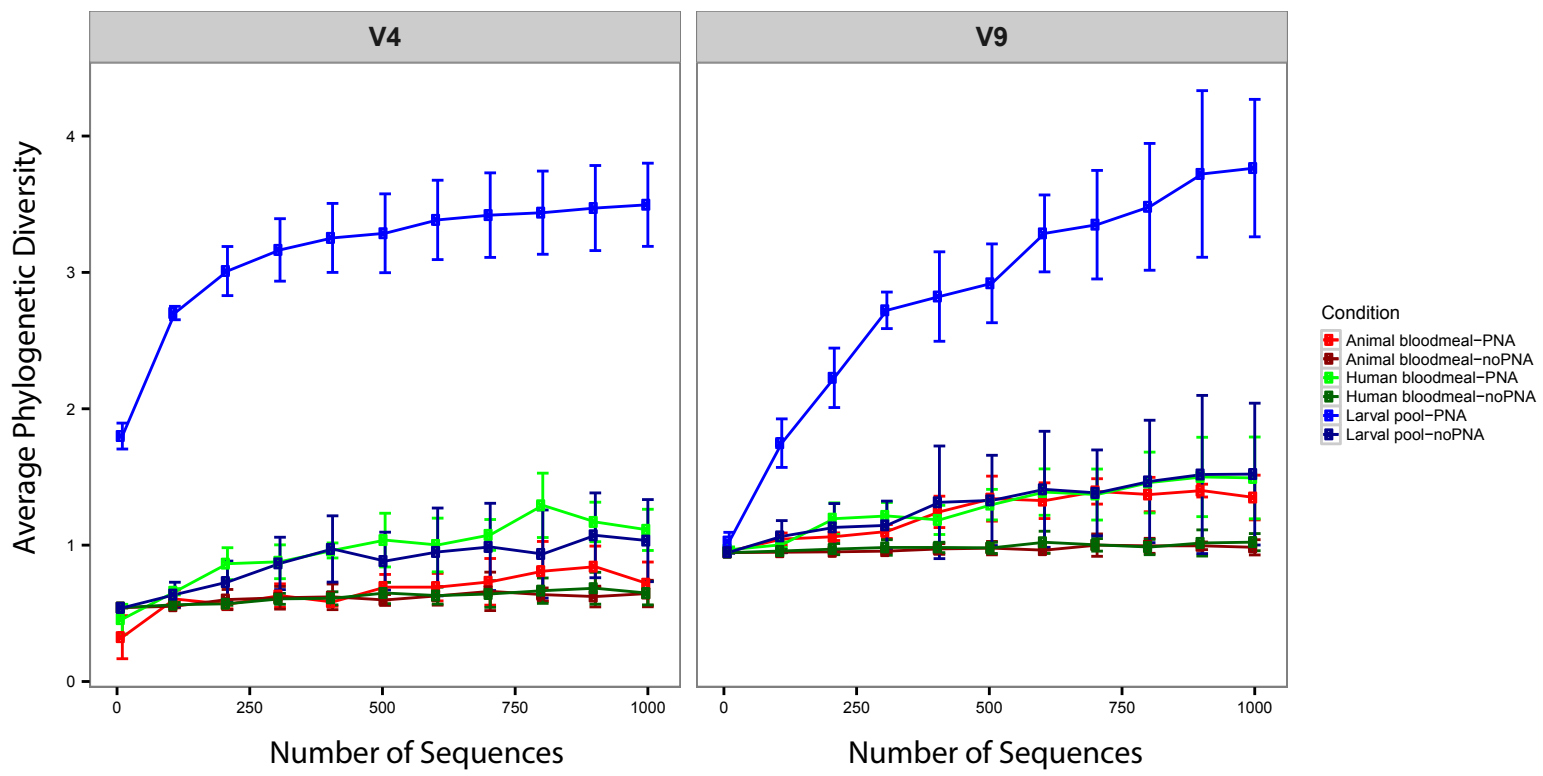
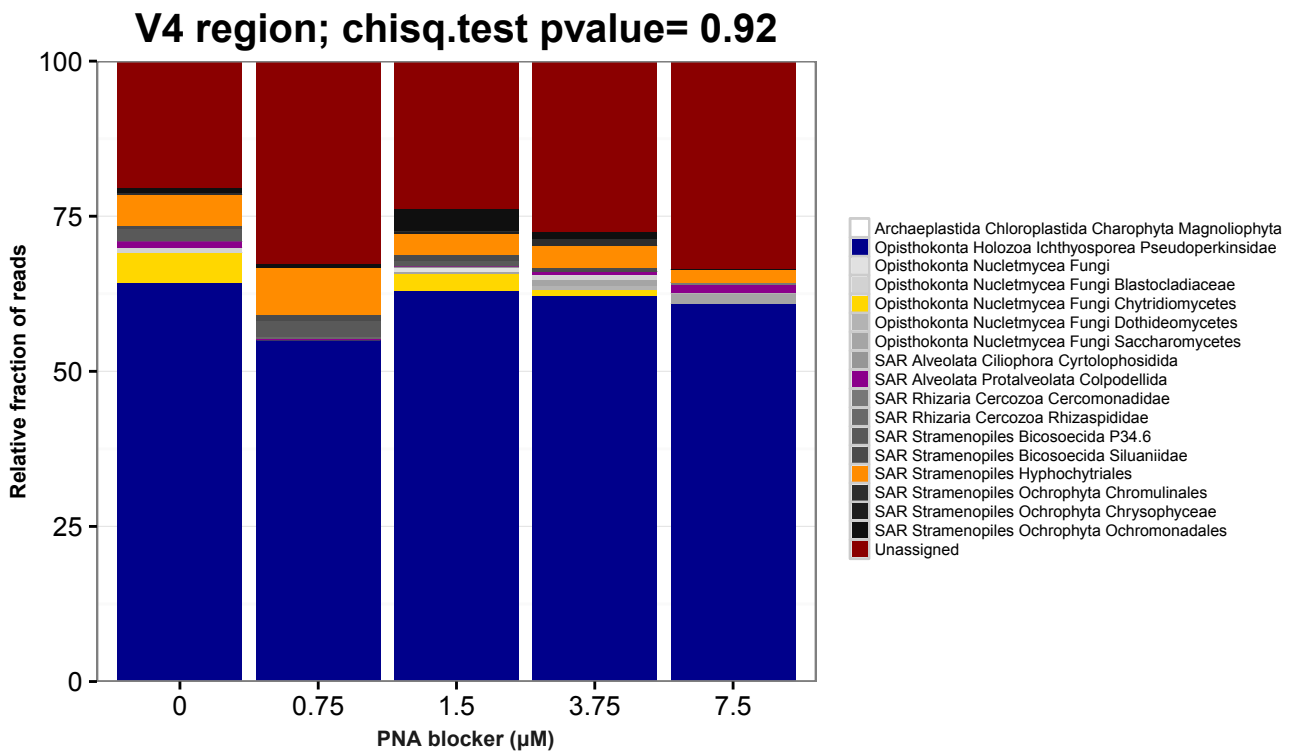
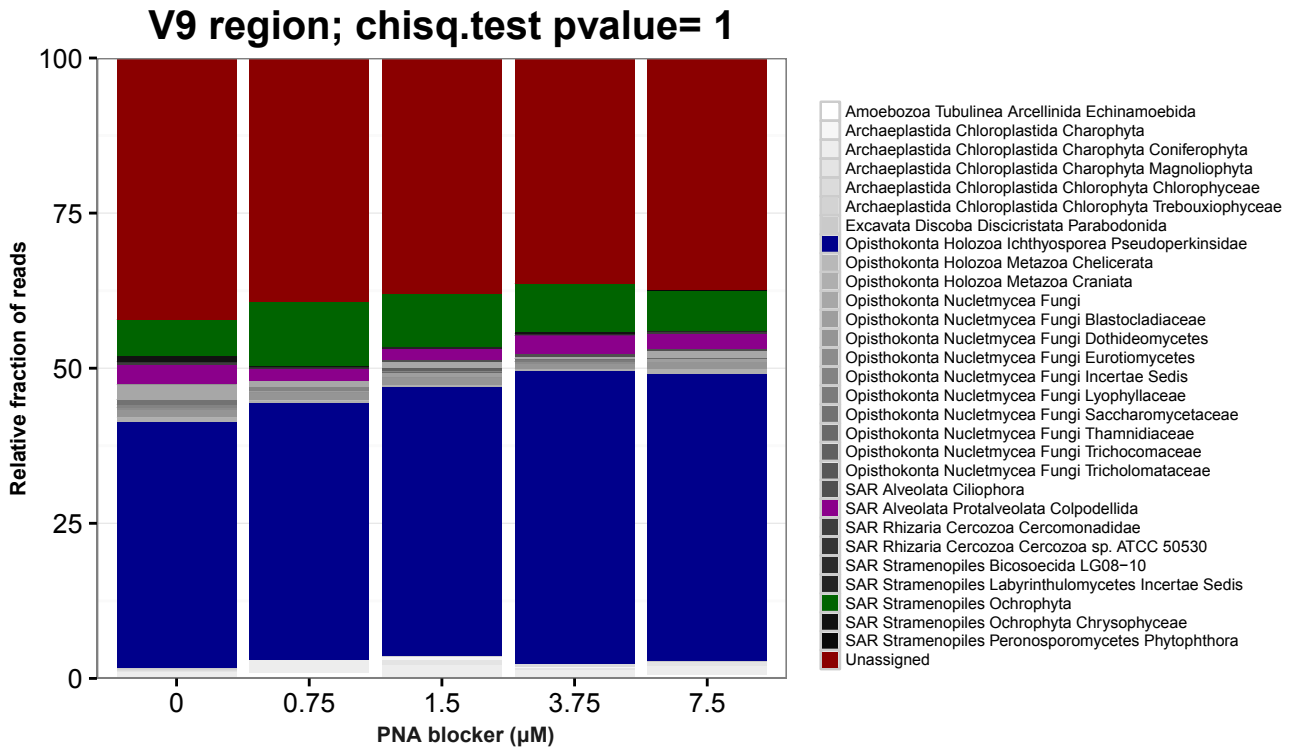


Figure S2



Mammalian Sequences

Mosquito Sequences

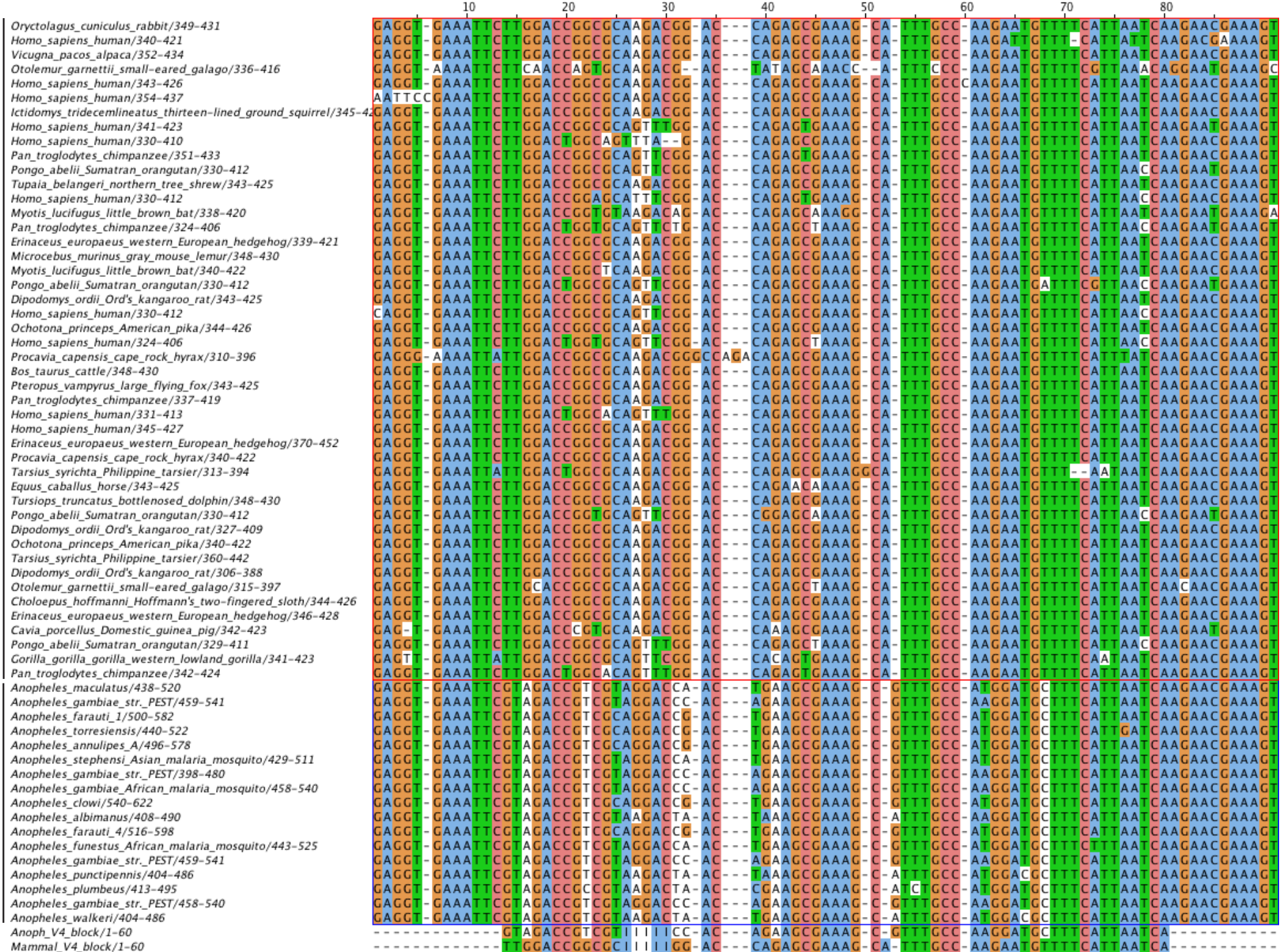


Figure S3

Consensus

GAGGTCGAAATTCCTTGGACCGGGCGAAGACGGGACAGACAGAGCGGAAAGGCCAGTTTGCCCAAGAATGTTTTCACTAATCAAGAACGAAAGT

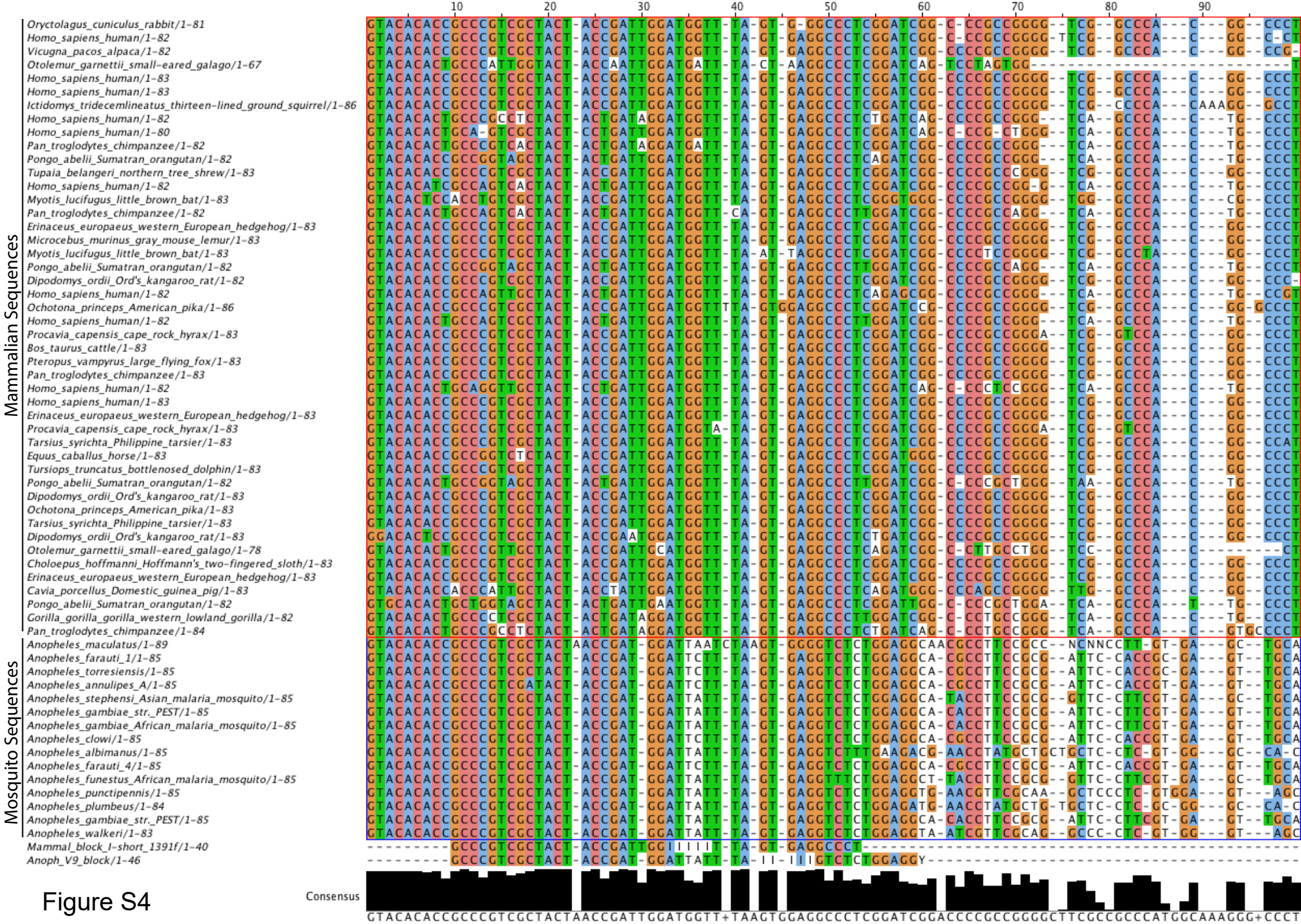


Figure S4

Figure S5



Table S1. Primers used for Anneal Blocking and PNA Blocking

Anneal Blockers

Hypervariable Region	Name	Sequence (5'->3')
V9	Mammal_block_l-short_1391f	GCCCGTCGCTACTACCGATTGG/ideoxyl//ideoxyl//ideoxyl//ideoxyl//ideoxyl/TTAGTGAGGCCCT/3SpC3/
	Anoph_V9_block	GCCCGTCGCTACTACCGATTGGATTATTA/ideoxyl//ideoxyl//ideoxyl//ideoxyl//ideoxyl/GTCTCTGGAGGY/3SpC3/
	18S_V9_1391_F_Nextera*	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GTACACACCGCCCGTC
	18S_V9_EukBr_R_Nextera*	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG TGATCCTTCTGCAGGTTACCTAC
V4	Mammal_V4_block	TGATTAATGAAAACATTCTTGGCAAATGCTTTCGCTCTGGTCC/ideoxyl//ideoxyl//ideoxyl//ideoxyl//ideoxyl/GCGCCGGTCCAA/3SpC3/
	Anoph_V4_block	TGATTAATGAAAGCATCCTTGGCAAACGCTTTCGCTCTGGTGG/ideoxyl//ideoxyl//ideoxyl//ideoxyl//ideoxyl/ACGACGGTCTAC/3SpC3/
	18S_V4F_Nextera*	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CCAGCASCYCGGGTAATTCC
	18S_V4R_Nextera*	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG ACTTTCGTTCTTGATYRA

*The 18S targeting portions of the primer are in bold, adaptor tails for sequencing are in plain text

PNA Blockers

Hypervariable Region	Name	Sequence (5'->3')
V9	AgV9-PNA	CCGTGCCAACTGCAAC
V4	AgV4-PNA	ACGCCAGGTACACC
Hypervariable region 1 Forward primer*		Reverse primer*
V9	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GTACACACCGCCCGTC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG TGATCCTTCTGCAGGTTACCTAC
V4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CCAGCASCYCGGGTAATTCC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG ACTTTCGTTCTTGATYRA

*The 18S targeting portions of the primer are in bold, adaptor tails for sequencing are in plain text