

Evolutionary insight into the Trypanosomatidae using alignment-free phylogenomics of the kinetoplast

Alexa Kaufer ^{1,*}, Damien Stark ² and John Ellis ¹

Supplementary Materials S1 – S5

S1 – Table of all trypanosomatid species used in this study

*Sequences assembled in the study

S2 -Table of the overall nucleotide frequency and skew throughout the maxicircle genome of various trypanosomatid species.

S3 - Self dottup plot comparative analysis of the entire maxicircle genome (right panel) and divergent region (left panel) of various trypanosomatid species

S4 - Analysis of repeated sequences in the maxicircle divergent region

S5 - Determination of optimal feature/k-mer length (k) for 46 trypanosomatid species.

(A) Graph shows the determination of the minimum feature/k-mer length for DNA. The optimal lower-feature length (k) is considered the point where the most features can be found in most genomes (i.e. peak of curve). (B) Graph shows the determination of the maximum feature/k-mer length for DNA. The optimal upper-feature length (k) is considered the point where the genome curves start having zero REF or begins falling to <10% of their REF maximum values [24].

S1 – Table of all trypanosomatid species used in this study.

Species	Genbank/TriTryp	Species	Genbank/TriTryp
<i>A. deanei</i>	KJ778684	<i>L. pyrrhocoris</i>	BK010873
<i>B. ayalai</i>	Baya_253	<i>L. seymouri</i>	Lsey_0394
<i>C. fasciculata</i>	CfaCI_maxi	<i>L. shawi</i>	BK010883
<i>E. herreri</i>	MK514112	<i>L. tarentolae</i>	MK514114
<i>E. monterogeii</i>	EMOLV88	<i>L. tropica</i>	MK514115*
<i>E. schaudinni</i>	BK010874	<i>L. turanica</i>	BK010887
<i>H. megaseliae</i>	ERX922862	<i>P. confusum</i>	SRX3339697
<i>L. aethiopica</i>	BK010882*	<i>P. deanei</i>	BK010886
<i>L. amazonensis</i>	SRX5187732*	<i>P. hertigi</i>	BK010885
<i>L. arabica</i>	BK010878	<i>T. brucei_brucei</i>	M94286.1
<i>L. braziliensis</i>	MK514111*	<i>T. brucei_rhodesiense</i>	SRX3199071*
<i>L. donovani</i>	CP022652.1	<i>T. copemani</i>	MG948557
<i>L. enriettii</i>	BK010880	<i>T. cruzi_CL</i>	DQ343645.1
<i>L. guyanensis</i>	BK010876*	<i>T. cruzi_Esmeraldo</i>	DQ343646
<i>L. infantum</i>	BK010877*	<i>T. cruzi_Marinkellei</i>	KC427240
<i>L. lainsoni</i>	BK010879	<i>T. cruzi_Silvio</i>	FJ203996
<i>L. macropodum</i>	SRX5006815*	<i>T. grayi</i>	SRX620256*
<i>L. major</i>	MK514113	<i>T. lewisi</i>	KR072974
<i>L. martiniquensis</i>	SRX5006816*	<i>T. rangeli</i>	KJ803830
<i>L. mexicana</i>	SRX5187730*	<i>T. vivax_Liem</i>	KM386509
<i>L. panamensis</i>	BK010875	<i>T. vivax_MTI</i>	KM386508
<i>L. peruviana</i>	BK010881	<i>Z. australiensis</i>	MK514117
<i>L. pifanoi</i>	BK010884	<i>P. francai</i>	SRX2165265*

S2. The overall nucleotide frequency and skew throughout the maxicircle genome of various trypanosomatid species.

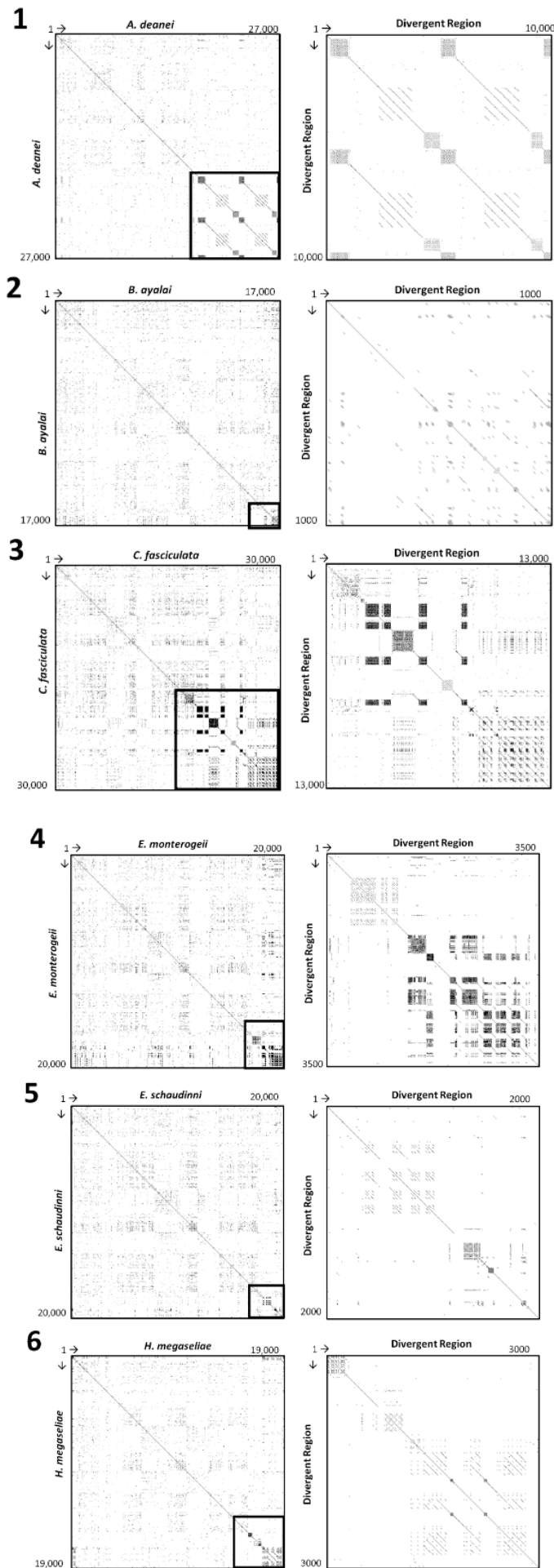
	Ltur	Lara	Lmaj	Ltro	Laet	Linf	Ldon	Lmex	Lpif	Lama	Ltar	Lguy	Lpan	Lsha	Lbra
% A	33	37	33	27	27	35	44	24	35	39	38	35	37	36	40
% C	10	7	10	36	8	9	12	10	10	9	9	8	8	8	9
% G	14	11	14	6	17	13	9	20	12	11	12	12	12	12	11
% T	43	45	43	31	48	43	35	46	43	41	41	45	43	44	40
% A + T	76	82	76	58	75	78	19	70	78	80	79	80	80	80	80
% G + C	24	18	24	42	25	22	21	30	22	20	21	20	20	20	20
AT skew	-0.13	-0.09	-0.13	-0.07	-0.28	-0.10	0.11	-0.31	-0.10	-0.03	-0.04	-0.13	-0.08	-0.1	0
GC skew	0.17	0.22	0.17	-0.71	0.36	0.18	-0.14	0.33	0.09	0.10	0.14	0.2	0.2	0.2	0.1

	Lper	Llai	Lmar	Lmac	Lenr	Pher	Pdea	Esch	Emon	Eher	Zaus	Lsey	Lpyr	Cfas	Hmeg
% A	36	36	34	34	35	38	39	37	38	36	33	34	34	35	38
% C	8	9	9	10	10	10	11	10	9	10	12	11	12	12	13
% G	12	13	14	14	13	13	13	12	12	13	15	13	14	13	14
% T	44	42	43	42	42	39	37	41	41	41	40	42	40	40	35
% A + T	80	78	77	76	77	77	76	78	79	77	73	76	74	75	73
% G + C	20	22	23	24	23	23	24	22	21	23	27	24	26	25	27
AT skew	-0.1	-0.08	-0.11	-0.10	-0.09	-0.01	0.02	-0.05	-0.04	-0.06	-0.09	-0.11	-0.08	-0.06	0.04
GC skew	0.2	0.18	0.21	0.17	0.13	0.13	0.08	0.09	0.14	0.13	0.11	0.08	0.08	0.04	0.04

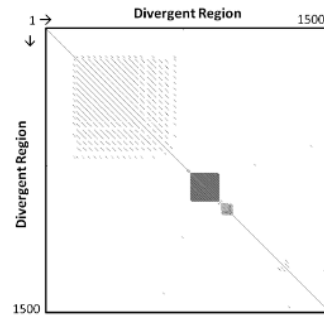
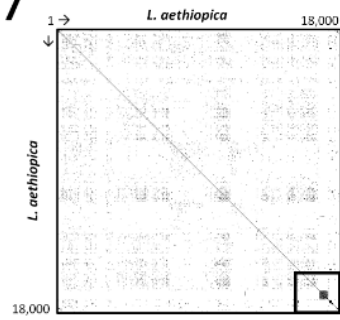
	Adea	Baya	Tvmt	Tvli	Tbru	Tgra	Tcop	Tlew	Tran	Tcma	Tces	Tcsi	Tccl	Pcon
% A	34	41	38	38	42	39	37	41	45	39	40	27	39	36
% C	16	9	10	10	9	12	13	12	13	10	11	9	11	13
% G	15	13	12	13	14	13	15	12	11	13	13	14	14	15
% T	35	37	40	39	35	36	35	35	31	38	36	50	36	36
% A + T	69	78	78	77	77	75	72	76	76	77	76	77	75	72
% G + C	31	22	22	23	23	25	28	24	24	23	24	23	25	28
AT skew	-0.01	0.05	-0.02	-0.01	0.09	0.04	0.03	0.08	0.18	0.01	0.05	0.30	0.04	0
GC skew	-0.03	0.18	0.09	0.13	0.22	0.04	0.07	0	-0.08	0.13	0.08	0.22	75	0.07

Abbreviated species: Ltur (*L. turanica*); Lara (*L. arabica*); Lmaj (*L. major*); Ltro (*L. tropica*); Laet (*L. aethiopica*); Linf (*L. infantum*); Ldon (*L. donovani*); Lmex (*L. mexicana*); Lpif (*L. pifanoi*); Lama (*L. amazonensis*); Ltar (*L. tarentolae*); Lguy (*L. guyanensis*); Lpan (*L. panamensis*); Lsha (*L. shawi*); Lbra (*L. braziliensis*); Lper (*L. peruviana*); Llai (*L. lainsoni*); Lmar (*L. martiniquensis*); Lmac (*L. macropodum*); Lenr (*L. enriettii*); Pher (*P. hertigi*); Pdea (*P. deanei*); Esch (*E. schaudinni*); Emon (*E. monterogeii*); Eher (*E. herreri*); Zaus (*Z. australiensis*); Lsey (*L. seymouri*); Lpyr (*L. pyrrhocoris*); Cfas (*C. fasciculata*); Hmeg (*H. megaseliae*); Adea (*A. deanei*); Baya (*B. ayalai*); Tvmt (*T. vivax MT1*); Tvli (*T. vivax Liem*); Tbru (*T. brucei brucei*); Tgra (*T. grayi*); Tcop (*T. copemani*); Tlew (*T. lewisi*); Tran (*T. rangeli*); Tcma (*T. cruzi marinkellei*); Tces (*T. cruzi esmeraldo*); Tcsi (*T. cruzi Silvio*); Tccl (*T. cruzi CL*); Pcon (*P. confusum*)

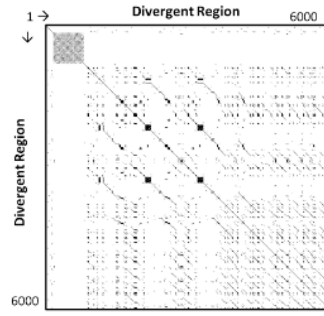
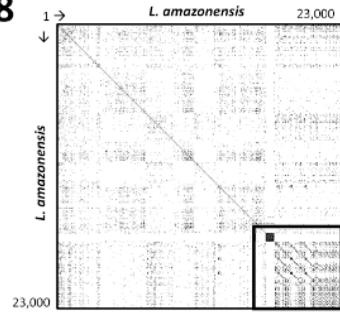
S3 - Self dottup plot comparative analysis of the entire maxicircle genome (right panel) and divergent region (left panel) of various trypanosomatid species



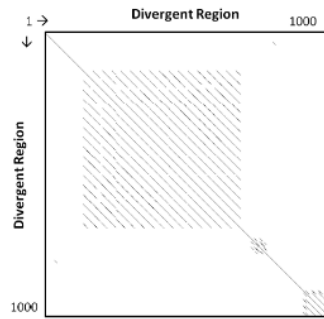
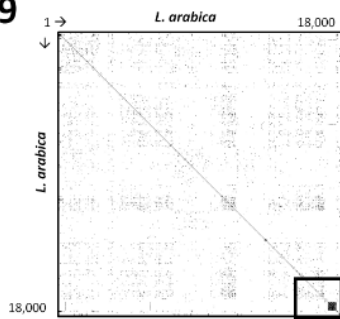
7



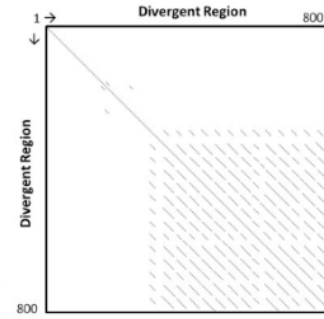
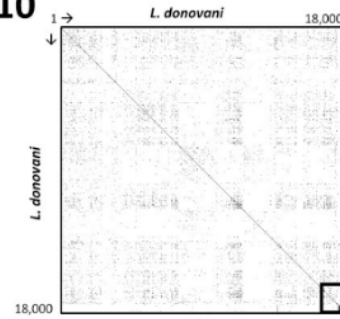
8



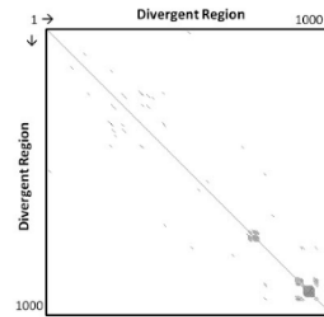
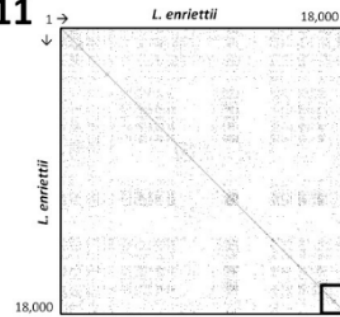
9



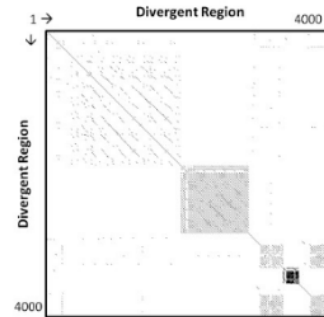
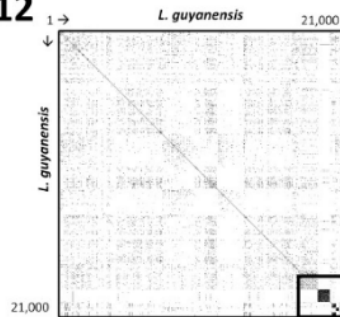
10



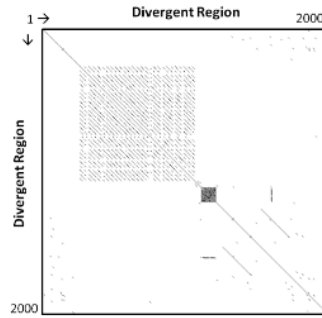
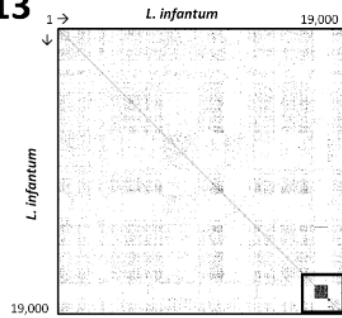
11



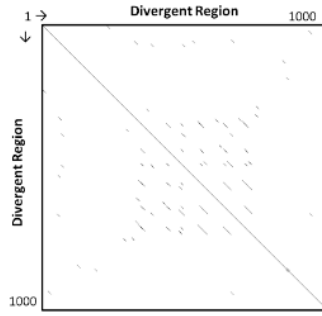
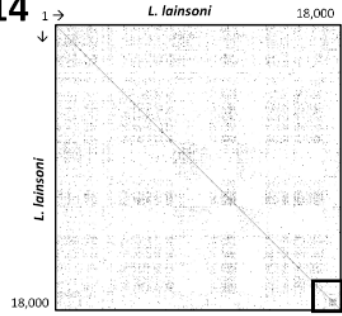
12



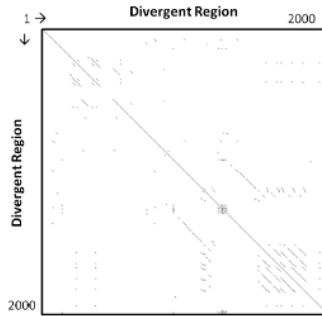
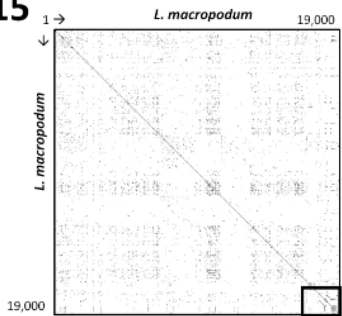
13



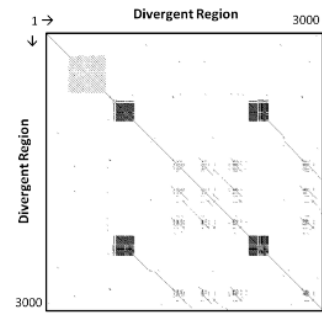
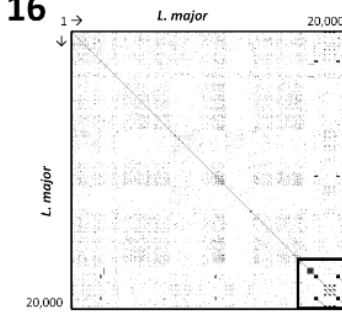
14



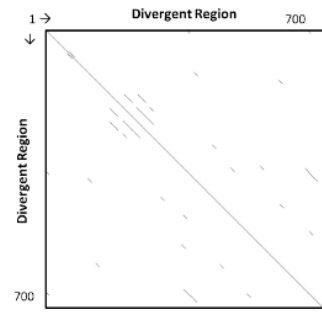
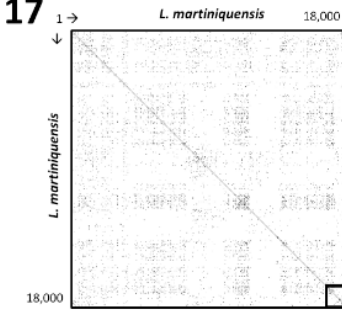
15



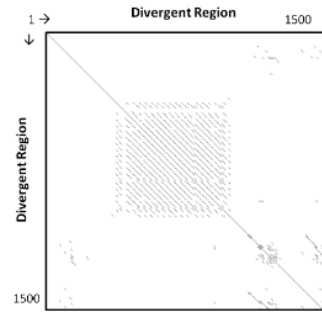
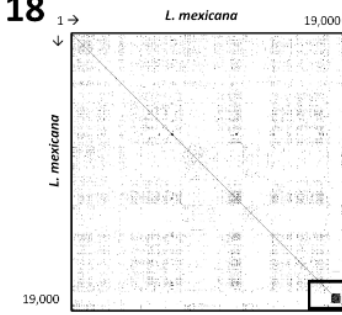
16

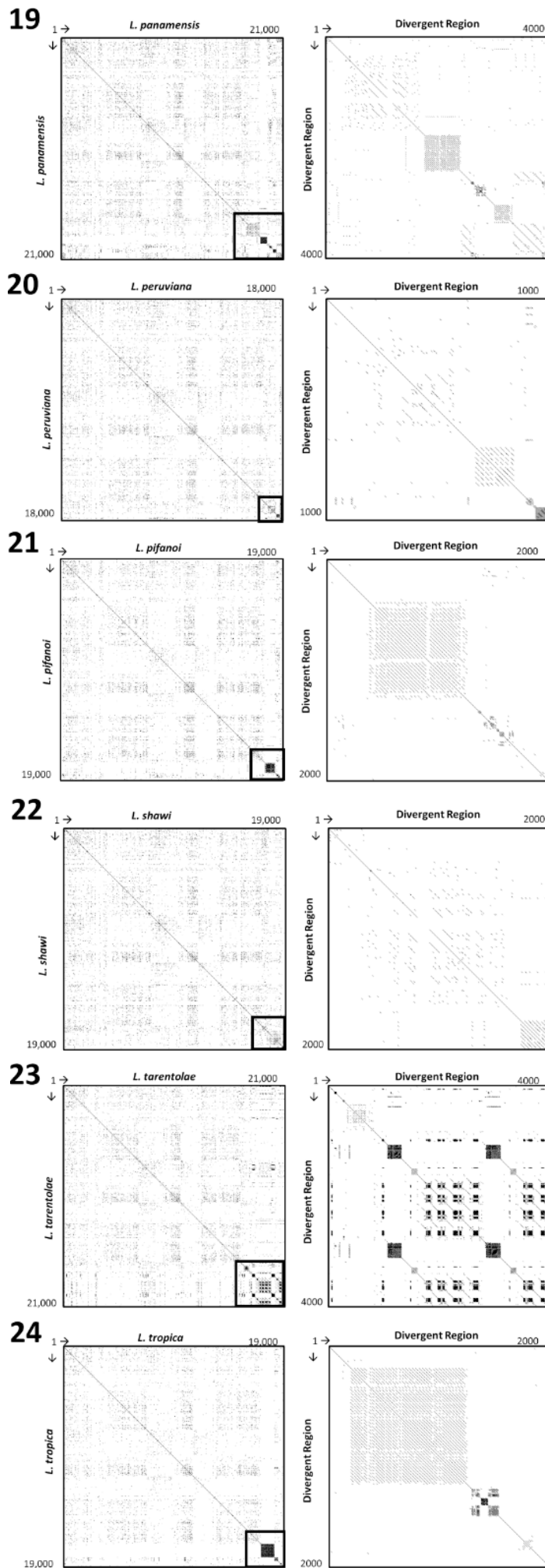


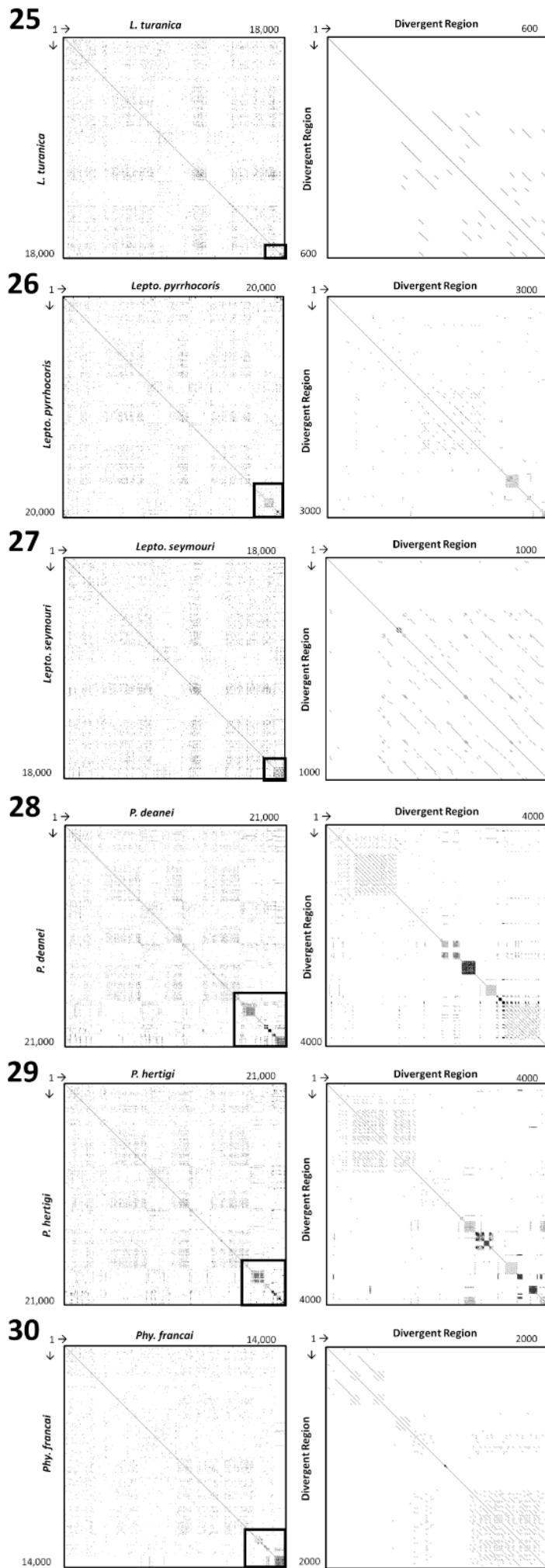
17

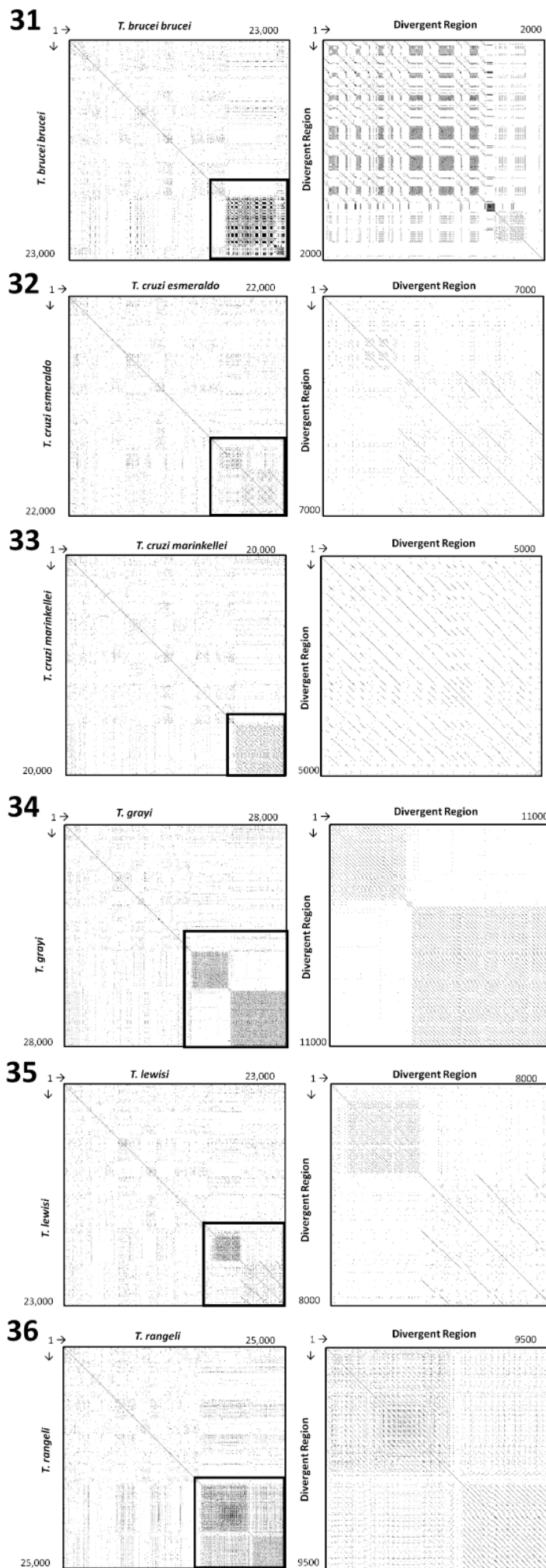


18

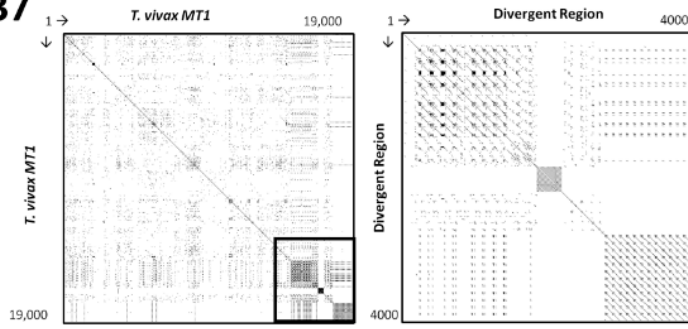




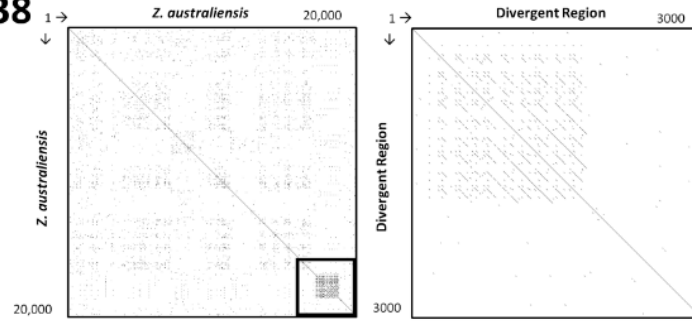




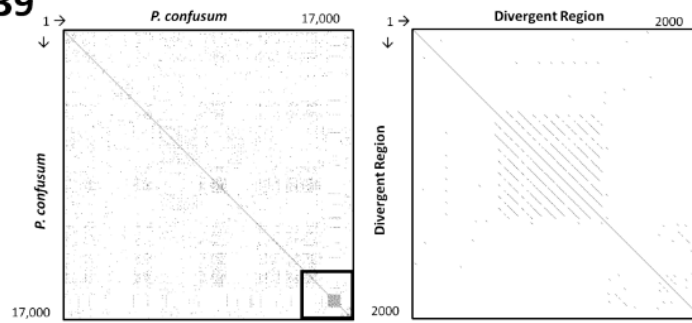
37



38



39



S4 - Analysis of repeated sequences in the maxicircle divergent region

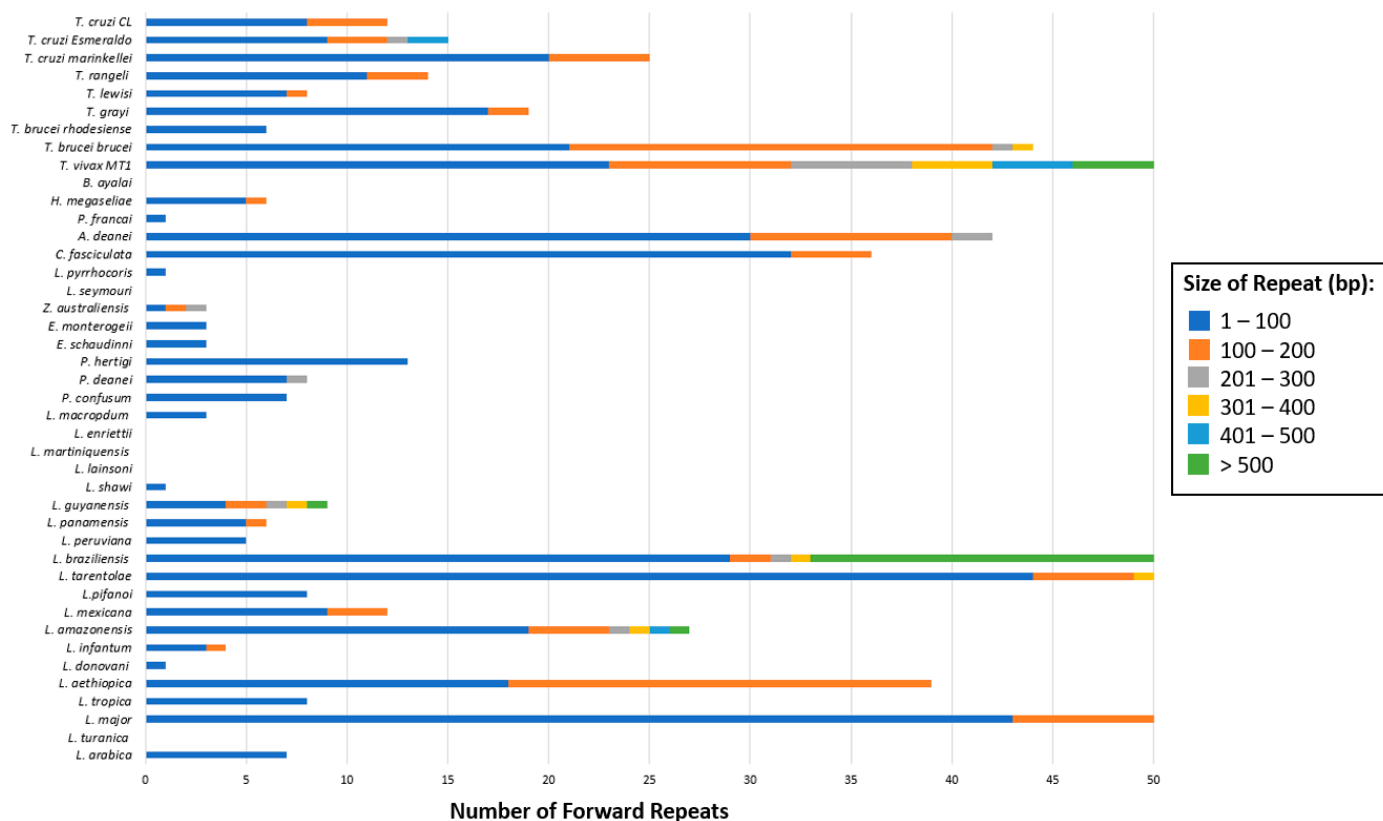


Figure S4-A: Analysis of forward repeats in the maxicircle divergent region 5 trypanosomatid species.

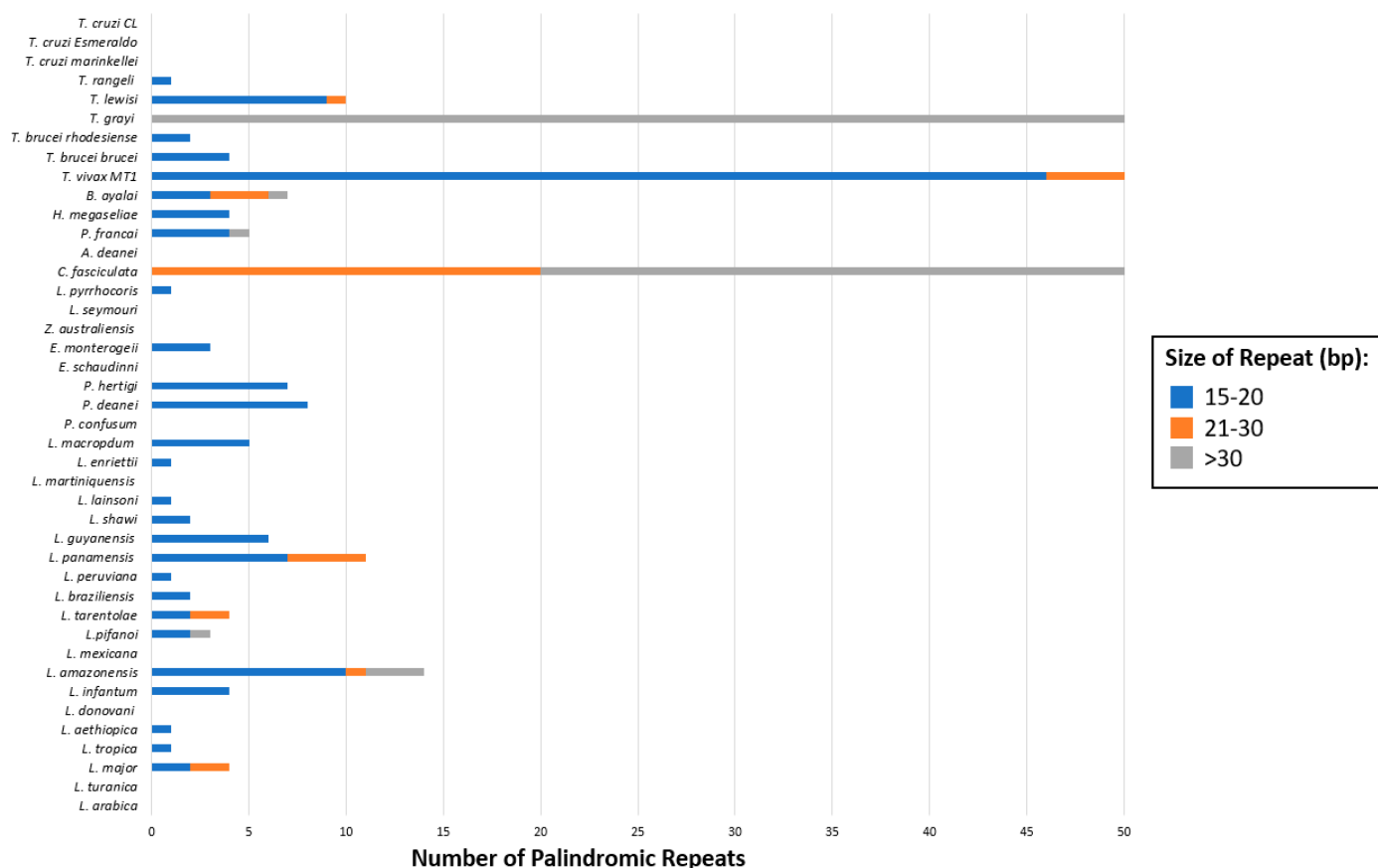


Figure S4-B: Analysis of palindromic repeats in the maxicircle divergent region 5 trypanosomatid species.

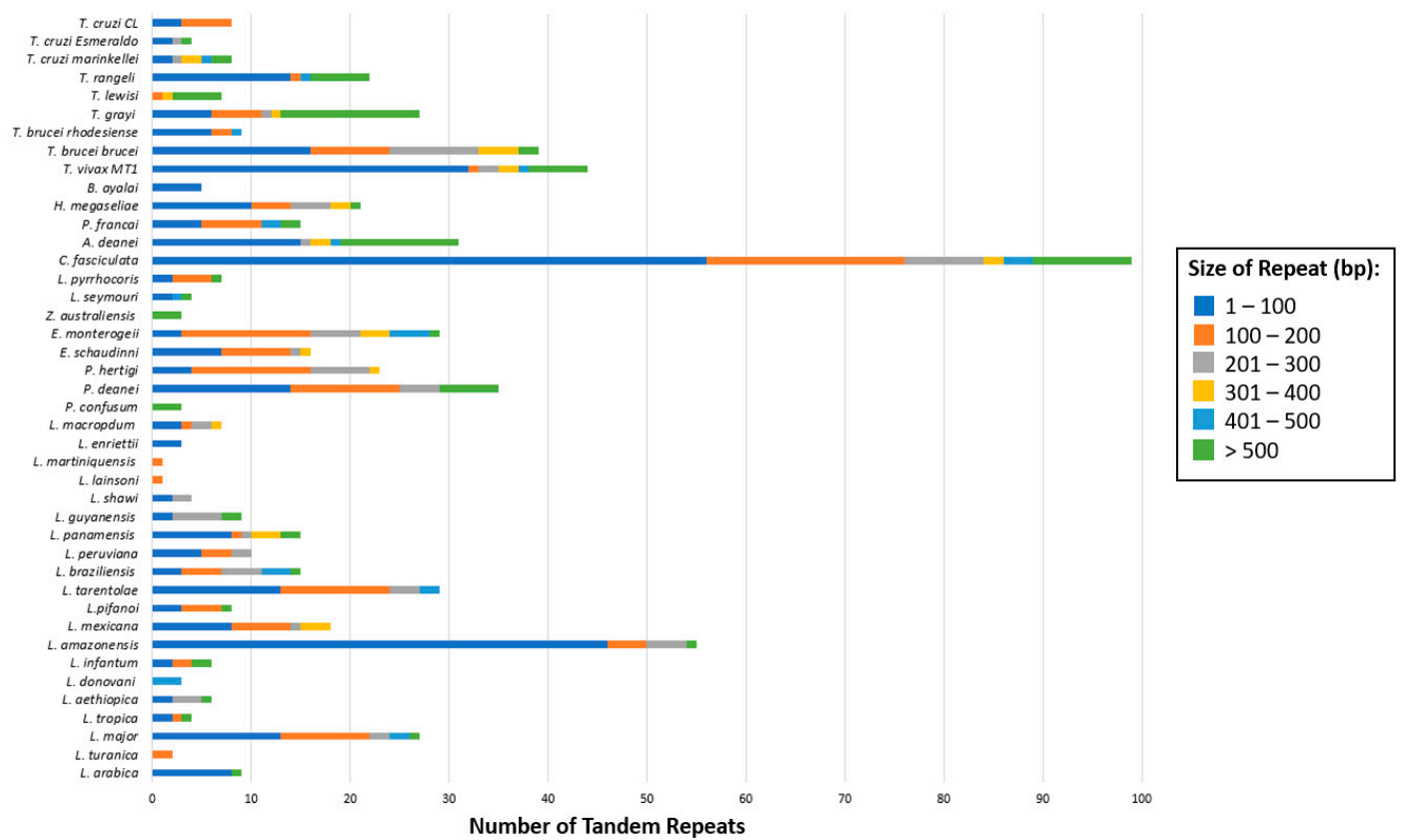
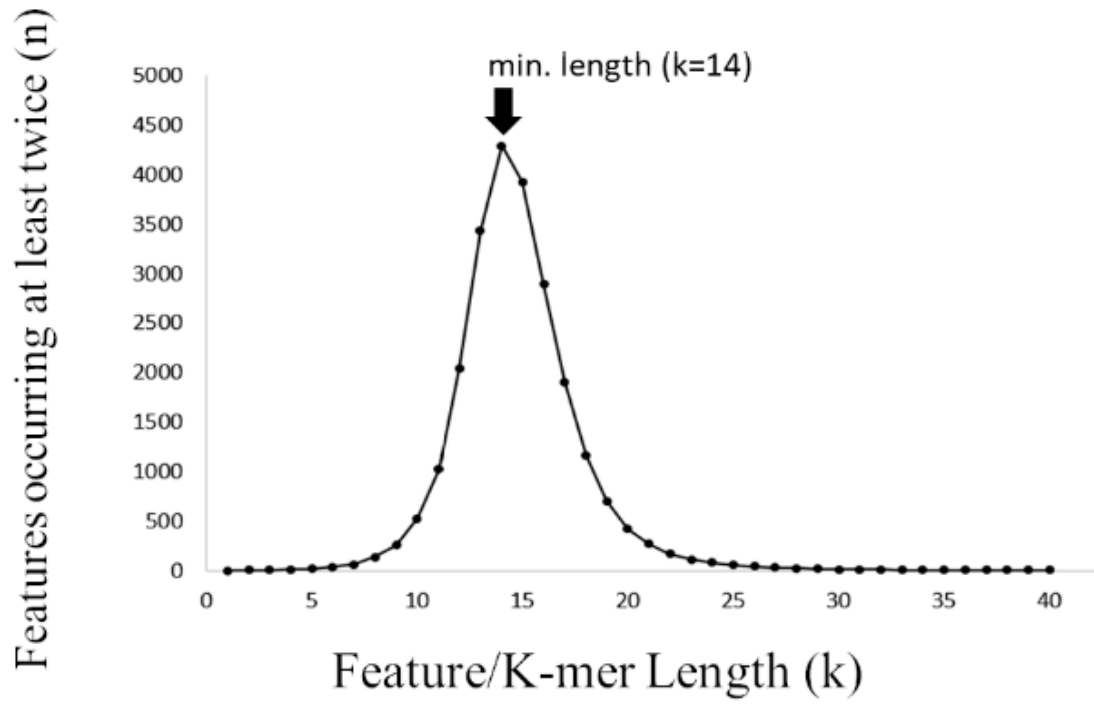


Figure S4-C: Analysis of tandem repeats in the maxicircle divergent region 5 trypanosomatid species.

A



B

