Supplementary Material for

Spark: Phylogenetic analysis using series-parallel resistor-derived features from k-mer and substring positional accumulation sum

Zhifeng Xiao[1], Jingjing Zhang[1], Jianwen Huang[1], and Runbin Tang[1]
[1]*School of Mathematical Sciences, Chongqing Normal University, Chongqing, China.*
[*]*Corresponding author*

1 Data Description

To comprehensively evaluate the performance of the proposed method, we selected six genomic datasets with varying lengths and complexities for experimentation. The detailed information is as follows:

Supplementary Table S1: The detail description of HIV-1, including type and accession number of se-quence.

Type	Accession Number	Type	Accession Number	Type	Accession Number
A1(M)	AB098332	C(M)	AB254143	G(M)	AB287003
A1(M)	AB098333	C(M)	AB254144	G(M)	AB287004
A1(M)	AB253421	C(M)	AB254145	G(M)	AB485662
A1(M)	AB253422	C(M)	AB254146	G(M)	AB485663
A1(M)	AB253428	C(M)	AB254147	G(M)	AF061640
A1(M)	AB253429	C(M)	AB254148	G(M)	AF061641
A1(M)	AF004885	D(M)	AB485648	G(M)	AF061642
A1(M)	AF069669	D(M)	AB485649	H(M)	AF005496
A2(M)	AF286237	D(M)	AB485650	H(M)	AF190127
A2(M)	AF286238	D(M)	AB485651	J(M)	AF082394
B(M)	A04321	D(M)	AF457090	J(M)	AF082395
B(M)	A07867	F1(M)	AB485656	N	AJ006022
B(M)	AB078005	F1(M)	AB485657	N	AJ271370
B(M)	AB097870	F1(M)	AB485658	N	AY532635
B(M)	AB221005	F1(M)	AB485659	O	AB485668
C(M)	AB097871	F1(M)	AF005494	O	AB485669
C(M)	AB023804	F2(M)	AF377956	O	AF407418
C(M)	AB254141	F2(M)	AJ249236	P	GQ328744
C(M)	AB254142	G(M)	AB231893	P	GU111555

Supplementary Table S2: The detail description of HCV, including type and accession number of sequence.

Type	Accession Number	Type	Accession Number	Type	Accession Number
type1	EF407478	type2	HB455599	type4	DQ418785
type1	EU155220	type2	HB455600	type4	DQ988073
type1	EU155261	type2	HC510319	type4	DQ988074
type1	EU155302	type2	HM049503	type4	DQ988075
type1	EU155308	type2	HQ852454	type4	DQ988076
type1	EU155328	type2	HQ852455	type4	DQ988077
type1	EU155337	type2	HQ852456	type4	DQ988078
type1	EU155369	type3	GQ275355	type4	DQ988079
type1	EU155370	type3	JN714194	type4	HQ537008
type1	EU255961	type3	JQ717255	type4	HQ537009
type1	EU256000	type3	JQ717257	type4	JX227973
type1	EU256092	type3	JQ717260	type4	MH119606
type1	EU482886	type3	KF035123	type5	JA212565
type1	EU862837	type3	KF035124	type5	JA212566
type1	HQ639947	type3	KY620473	type5	JA212567
type1	KU180728	type3	KY620662	type6	JE984232
type2	HB455523	type3	KY620664	type6	DQ835769
type2	HB455524	type3	KY620677	type6	EF424629
type2	HB455525	type3	KY620679	type6	EU408331
type2	HB455526	type3	KY620681	type6	KJ470624
type2	HB455527	type3	KY620710	type6	KM252782
type2	HB455528	type3	KY620782	type6	KM504112
type2	HB455538	type3	KY620828	type6	KM504113
type2	HB455539	type3	KY620829	type6	KM504115
type2	HB455540	type3	KY620831	type6	KM504117
type2	HB455597	type3	KY620833	type6	KM504118
type2	HB455598	type3	KY620860	type6	KM504121
				type6	KM504124

Supplementary Table S3: The detail description of HBV, including type and accession number of sequence.

Type	Accession Number	Type	Accession Number	Type	Accession Number
A	EU594388	D	JN040762	F	KJ843212
A	EU859930	D	KF679994	F	KJ843213
A	EU859935	D	KF679996	F	KJ843220
A	EU859936	D	KM524353	F	KJ843225
A	EU859942	D	KX276856	F	KJ843226
A	KT749840	D	MF925358	F	KX264499
B	AB033555	D	MF925363	F	KY476326
B	AB100695	E	AB091256	G	AB056513
B	AB219429	E	AB205189	G	AB056515
B	EF473977	E	EU239221	G	AB064310
B	EU439020	E	EU239224	G	AB064311
B	EU881997	E	FN594758	G	AB064312
B	GQ358143	E	FN594759	G	AB064313
B	GQ924638	E	FN594762	G	AB375165
B	HM011467	E	GQ161755	G	AB375168
B	JQ040138	E	GQ161761	G	AF160501
B	JQ429082	E	GQ161787	G	AF405706
B	KJ803796	E	GQ161793	G	GU563556
B	KX276825	E	GQ161804	G	HE981174
C	DQ089769	E	GQ161817	G	KF414679
C	JF899336	E	GQ161827	G	KJ194508
C	JN827416	E	GQ161835	G	KM998716
C	JQ801497	E	GQ161836	G	KX264500
C	JQ801502	E	HM363591	G	KY004111
C	JQ801503	E	KU736894	H	AB179747
C	KJ410489	E	KU736895	H	AB205010
C	KT364718	E	KU736896	H	AB266536
C	KT987426	E	KU736899	H	AB275308
C	KU051425	F	AB214516	H	AB353764
C	KX765824	F	DQ776247	H	AB375159
C	KX765843	F	DQ823088	H	AB375160
C	KX765845	F	DQ823090	H	AB375161
C	MF674403	F	FJ657519	H	AB375162
C	MF674504	F	FJ657528	H	AB375164
C	MF925359	F	HE981181	H	EF157291
C	MF925394	F	KJ843175	H	EU498228
C	MF925409	F	KJ843185	H	FJ356715
D	AJ627224	F	KJ843189	H	FJ356716
D	AY090453	F	KJ843208	H	KX264501
D	EU594425	F	KJ843210		
D	GQ477453	F	KJ843211		

Supplementary Table S4: The detail description of HEV, including type and accession number of sequence.

No.	Strain name	Accession	Country	Group	Length (nt)
1	B1 (Bur-82)	M73218	Burma (Rangoon)	Ia	7207
2	B2 (Bur-86)	D10330	Burma (Rangoon)	Ia	7194
3	I2 (Mad-93)	X99441	India (Madras)	Ia	7194
4	I3	AF076239	India (Hyderabad)	Ia	7194
5	Np1 (TK15/92)	AF051830	Nepal (Kathamandu)	Ia	7199
6	P2 (Abb-2B)	AF185822	Pakistan (Abbottabad)	Ia	7134
7	Yam-67	AF459438	India (Yamuna Nagar)	Ic	7206
8	C1 (CHT-88)	D11092	China (Xinjiang Hetian)	Ic	7207
9	C2 (KS2-87)	L25595	China (Xinjiang Kashi)	Ic	7221
10	C3 (CHT-87)	L08816	China (Xinjiang Hetian)	Ic	7176
11	C4 (Uigh179)	D11093	China (Xinjiang Uighur)	Ic	7194
12	China Hebei	M94177	China (Hebei)	Ic	7200
13	P1 (Sar-55)	M80581	Pakistan (Sargodha)	Ic	7138
14	I1 (FHF)	X98292	India	Ib	7202
15	Morocco	AY230202	Morocco	Id	7212
16	T3	AY204877	Chad	Ie	7170
17	M1	M74506	Mexico (Telixtac)	II	7180
18	HE-JA10	AB089824	Japan (Tokyo)	IIIa	7262
19	JNI-Sap	AB074918	Japan (Sapporo)	IIIa	7256
20	JNY-HAW	AB074920	Japan (Sapporo)	IIIa	7240
21	swUS1	AP082843	USA	IIIa	7207
22	US1	AF060668	USA (Minnesota)	IIIa	7202
23	US2	AF060669	USA (Tennessee)	IIIa	7277
24	JBOARI-Hyo04	AB189070	Japan (Hyogo)	IIIb	7247
25	JDEER-Hyo03L	AB189071	Japan (Hyogo)	IIIb	7230
26	JIT-KAN	AB091394	Japan (Kanagawa)	IIIb	7218
27	JMO-Hyo03L	AB189072	Japan (Hyogo)	IIIb	7180
28	JRA1	AP003430	Japan (Tokyo)	IIIb	7230
29	JSO-Hyo03L	AB189073	Japan (Tokyo)	IIIb	7180
30	JTH-Hyo03L	AB189074	Japan (Tokyo)	IIIb	7180
31	JYO-Hyo03L	AB189075	Japan (Tokyo)	IIIb	7180
32	swJ570	AB073912	Japan (Tochigi)	IIIb	7257
33	Kyrgyz	AF455784	Kyrgyzstan	IIIc	7239
34	Arkell	AY115488	Canada (Ontario Guelph)	IIIa	7255
35	HE-JA1	AB097812	Japan (Hokkaido)	IVa	7258
36	HE-JK4	AB099347	Japan (Tochigi)	IVa	7250
37	HE-JI4	AB080575	Japan (Tochigi)	IVa	7186
38	JAK-Sai	AB074915	Japan (Saitama)	IVa	7236
39	JKK-Sap	AB074917	Japan (Sapporo)	IVa	7235
40	JSN-Sap95	AB161717	Japan (Hokkaido)	IVa	7202
41	JSN-Sap-FH	AB091395	Japan (Hokkaido)	IVa	7234
42	JSN-Sap-FH02C	AB200239	Japan (Hokkaido)	IVa	7251
43	JTS-Sap02	AB161718	Japan (Hokkaido)	IVa	7202
44	JYW-Sap02	AB161719	Japan (Hokkaido)	IVa	7202
45	swJ13-1	AB097811	Japan (Hokkaido)	IVa	7258
46	swCH25	AY594199	China (Uighur)	IVc	7270
47	T1	AJ272108	China (Beijing)	IVc	7232
48	CCC220	AB108537	China (Changchun)	IVb	7193

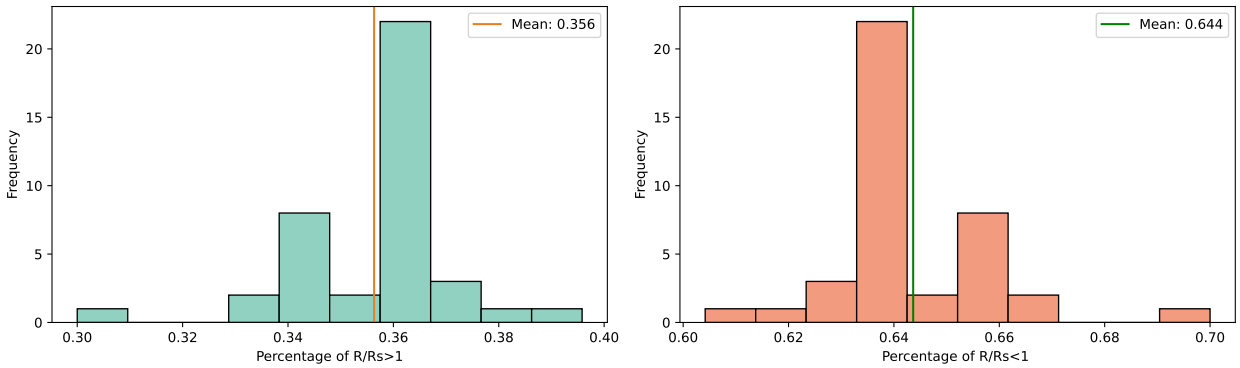
Supplementary Table S5: The detail description of HRV, including type and accession number of sequence.

Type	Accession Number	Type	Accession Number	Type	Accession Number
A(HRV)	AY751783	A(HRV)	FJ445136	A(HRV)	FJ445190
A(HRV)	DQ473491	A(HRV)	FJ445138	A(HRV)	L24917
A(HRV)	DQ473492	A(HRV)	FJ445139	A(HRV)	X02316
A(HRV)	DQ473493	A(HRV)	FJ445140	B(HRV)	DQ473485
A(HRV)	DQ473494	A(HRV)	FJ445141	B(HRV)	DQ473486
A(HRV)	DQ473496	A(HRV)	FJ445142	B(HRV)	DQ473488
A(HRV)	DQ473497	A(HRV)	FJ445143	B(HRV)	DQ473489
A(HRV)	DQ473499	A(HRV)	FJ445144	B(HRV)	DQ473490
A(HRV)	DQ473500	A(HRV)	FJ445145	B(HRV)	EF173420
A(HRV)	DQ473504	A(HRV)	FJ445146	B(HRV)	EF173423
A(HRV)	DQ473505	A(HRV)	FJ445147	B(HRV)	EF173425
A(HRV)	DQ473506	A(HRV)	FJ445148	B(HRV)	FJ445112
A(HRV)	DQ473507	A(HRV)	FJ445149	B(HRV)	FJ445124
A(HRV)	DQ473508	A(HRV)	FJ445152	B(HRV)	FJ445130
A(HRV)	DQ473510	A(HRV)	FJ445154	B(HRV)	FJ445137
A(HRV)	DQ473511	A(HRV)	FJ445156	B(HRV)	FJ445151
A(HRV)	EF173414	A(HRV)	FJ445157	B(HRV)	FJ445153
A(HRV)	EF173415	A(HRV)	FJ445158	B(HRV)	FJ445155
A(HRV)	FJ445111	A(HRV)	FJ445159	B(HRV)	FJ445161
A(HRV)	FJ445113	A(HRV)	FJ445160	B(HRV)	FJ445162
A(HRV)	FJ445114	A(HRV)	FJ445163	B(HRV)	FJ445164
A(HRV)	FJ445115	A(HRV)	FJ445165	B(HRV)	FJ445168
A(HRV)	FJ445116	A(HRV)	FJ445166	B(HRV)	FJ445169
A(HRV)	FJ445117	A(HRV)	FJ445167	B(HRV)	FJ445172
A(HRV)	FJ445118	A(HRV)	FJ445170	B(HRV)	FJ445174
A(HRV)	FJ445119	A(HRV)	FJ445171	B(HRV)	FJ445186
A(HRV)	FJ445121	A(HRV)	FJ445173	B(HRV)	FJ445187
A(HRV)	FJ445122	A(HRV)	FJ445175	B(HRV)	FJ445188
A(HRV)	FJ445123	A(HRV)	FJ445176	B(HRV)	L05355
A(HRV)	FJ445125	A(HRV)	FJ445177	C(HRV)	EF077279
A(HRV)	FJ445126	A(HRV)	FJ445178	C(HRV)	EF077280
A(HRV)	FJ445127	A(HRV)	FJ445179	C(HRV)	EF186077
A(HRV)	FJ445128	A(HRV)	FJ445180	C(HRV)	EF582385
A(HRV)	FJ445129	A(HRV)	FJ445181	C(HRV)	EF582386
A(HRV)	FJ445131	A(HRV)	FJ445182	C(HRV)	EF582387
A(HRV)	FJ445132	A(HRV)	FJ445183	outgroup	AF499637
A(HRV)	FJ445133	A(HRV)	FJ445184	outgroup	AF546702
A(HRV)	FJ445134	A(HRV)	FJ445185	outgroup	V01149
A(HRV)	FJ445135	A(HRV)	FJ445189		

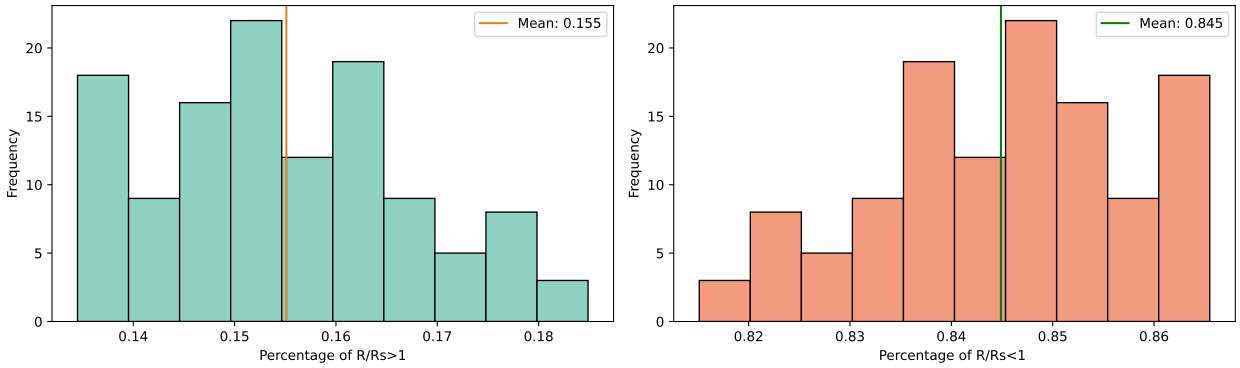
Supplementary Table S6: The detail description of 18SRNA, including type and accession number of sequence.

Name	Accession Number	Name	Accession Number
Birds_Ciconia	AF173636	Birds_Coturnix	AF173611
Birds_Coracias	AF173625	Birds_Gallus	AF173612
Birds_Trogon	AF173623	Birds_Apteryx	AF173609
Birds_Apus	AF173619	Birds_Dromaius	AF173610
Birds_Chordeiles	AF173622	Reptiles_Alligator	AF173605
Birds_Grus	AF173632	Reptiles_Sphenodon	AF115860
Birds_Cuculus	AF173628	Mammals_Oryctolagus	X00640
Birds_Turnix	AF173631	Mammals_Rattus	V01270
Birds_Larus	AF173637	Mammals_Rattus	X01117
Birds_Charadrius	AF173638	Mammals_Oryctolagus	X06778
Birds_Columba	AF173630	Mammals_Mus	X00686
Birds_Tockus	AF173626	Mammals_Rattus	M11188
Birds_Tyrannus	AF173616	Mammals_Homo	M10098
Birds_Gallirex	AF173618	Mammals_Homo	K03432
Birds_Urocolius	AF173617	Mammals_Homo	U13369
Birds_Upupa	AF173627	Mammals_Homo	X03205
Birds_Galbula	AF173624	Mammals_Mus	NR_003278
Birds_Picoides	AF173615	Amphibians_Ranodon	AJ279506
Birds_Ortalis	AF173613	Amphibians_Xenopus	X04025
Birds_Anas	AF173614	Amphibians_Xenopus	X02995

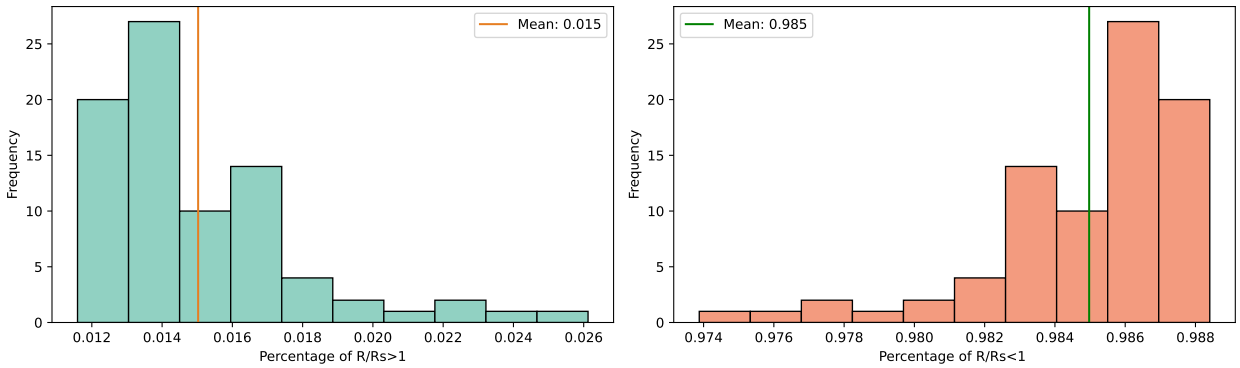
2 The results of comparative analysis of the proportion of $R(\zeta, s)/Rs(\zeta, s)$



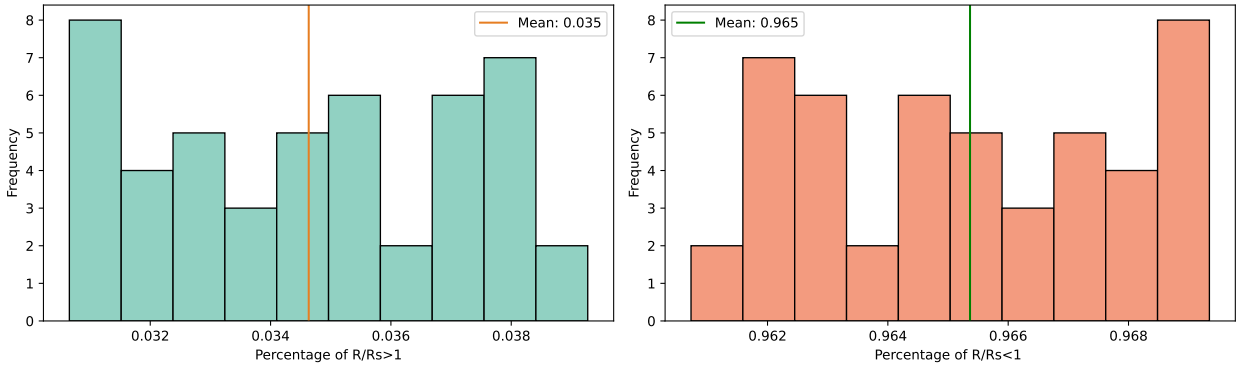
Supplementary Figure S1: Histogram of $R(\zeta, s)/Rs(\zeta, s)$ greater than 1 (left) and less than 1 (right) in 18SRNA dataset. The solid vertical line indicates the data mean.



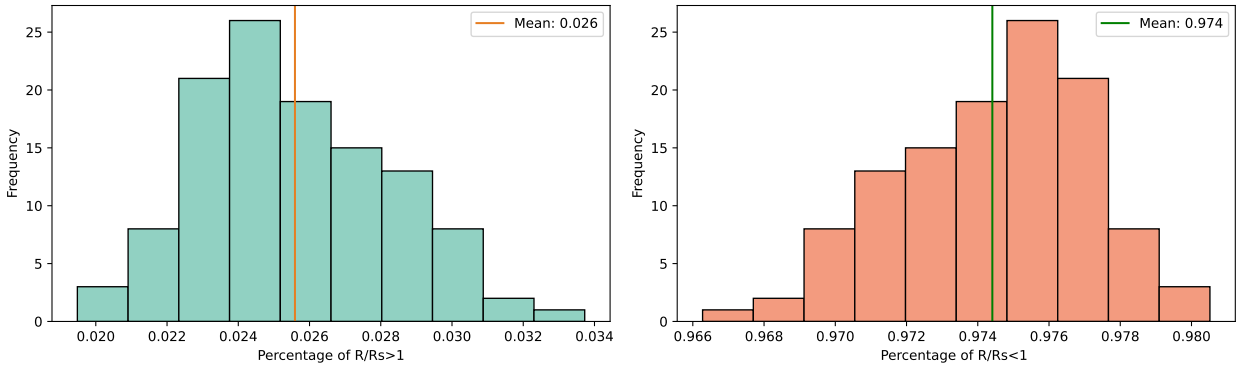
Supplementary Figure S2: Histogram of $R(\zeta, s)/Rs(\zeta, s)$ greater than 1 (left) and less than 1 (right) in HBV dataset. The solid vertical line indicates the data mean.



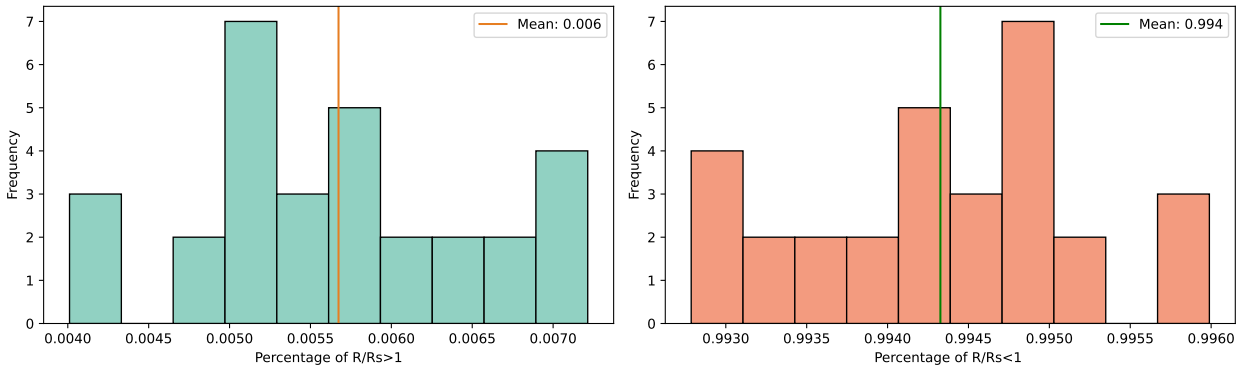
Supplementary Figure S3: Histogram of $R(\zeta, s)/Rs(\zeta, s)$ greater than 1 (left) and less than 1 (right) in HCV dataset. The solid vertical line indicates the data mean.



Supplementary Figure S4: Histogram of $R(\zeta, s)/Rs(\zeta, s)$ greater than 1 (left) and less than 1 (right) in HEV dataset. The solid vertical line indicates the data mean.

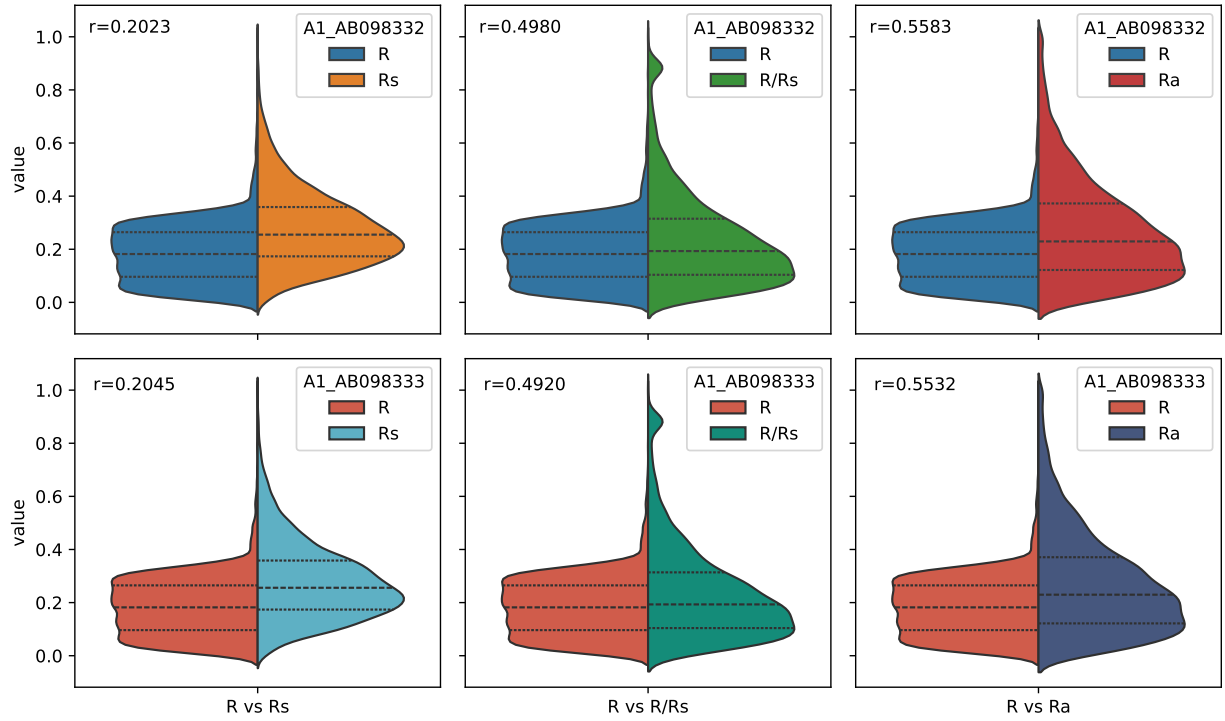


Supplementary Figure S5: Histogram of $R(\zeta, s)/Rs(\zeta, s)$ greater than 1 (left) and less than 1 (right) in HRV dataset. The solid vertical line indicates the data mean.

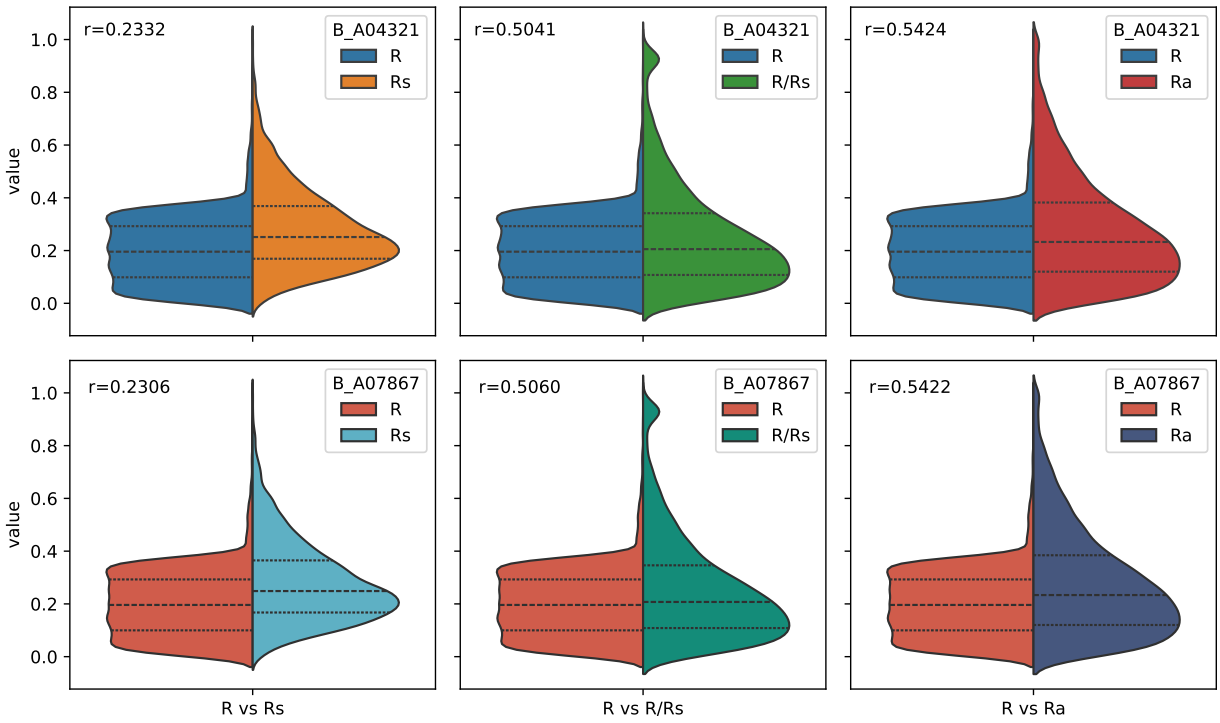


Supplementary Figure S6: Histogram of $R(\zeta, s)/Rs(\zeta, s)$ greater than 1 (left) and less than 1 (right) in mammalian genomes. The solid vertical line indicates the data mean.

3 The effects of the features of the simulated signal

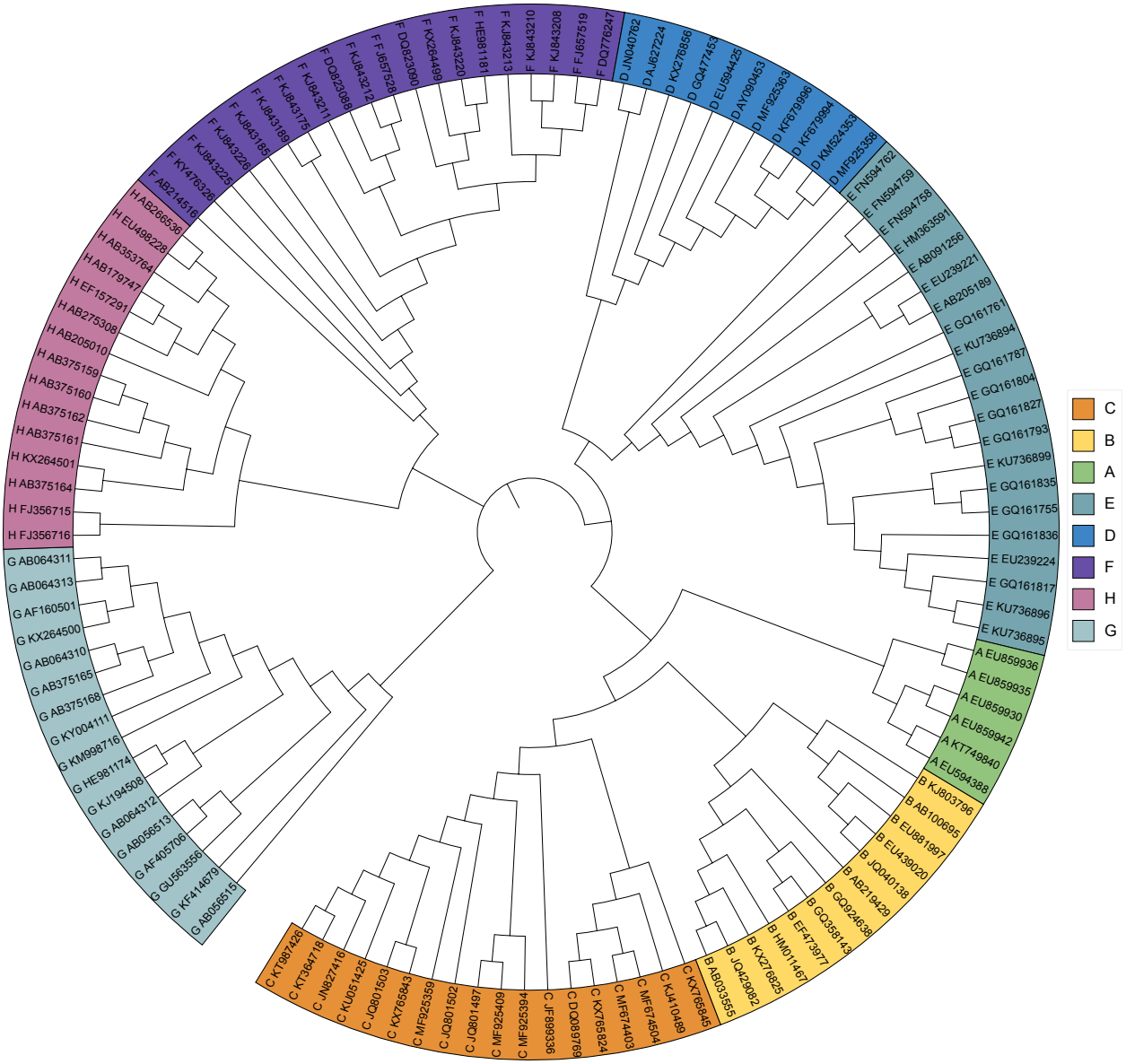


Supplementary Figure S7: Comparison of four feature vectors for sequences A1_AB098332 and A1_AB098333 from the identical type in HIV-1 dataset ($k = 9$, $t = 3$).

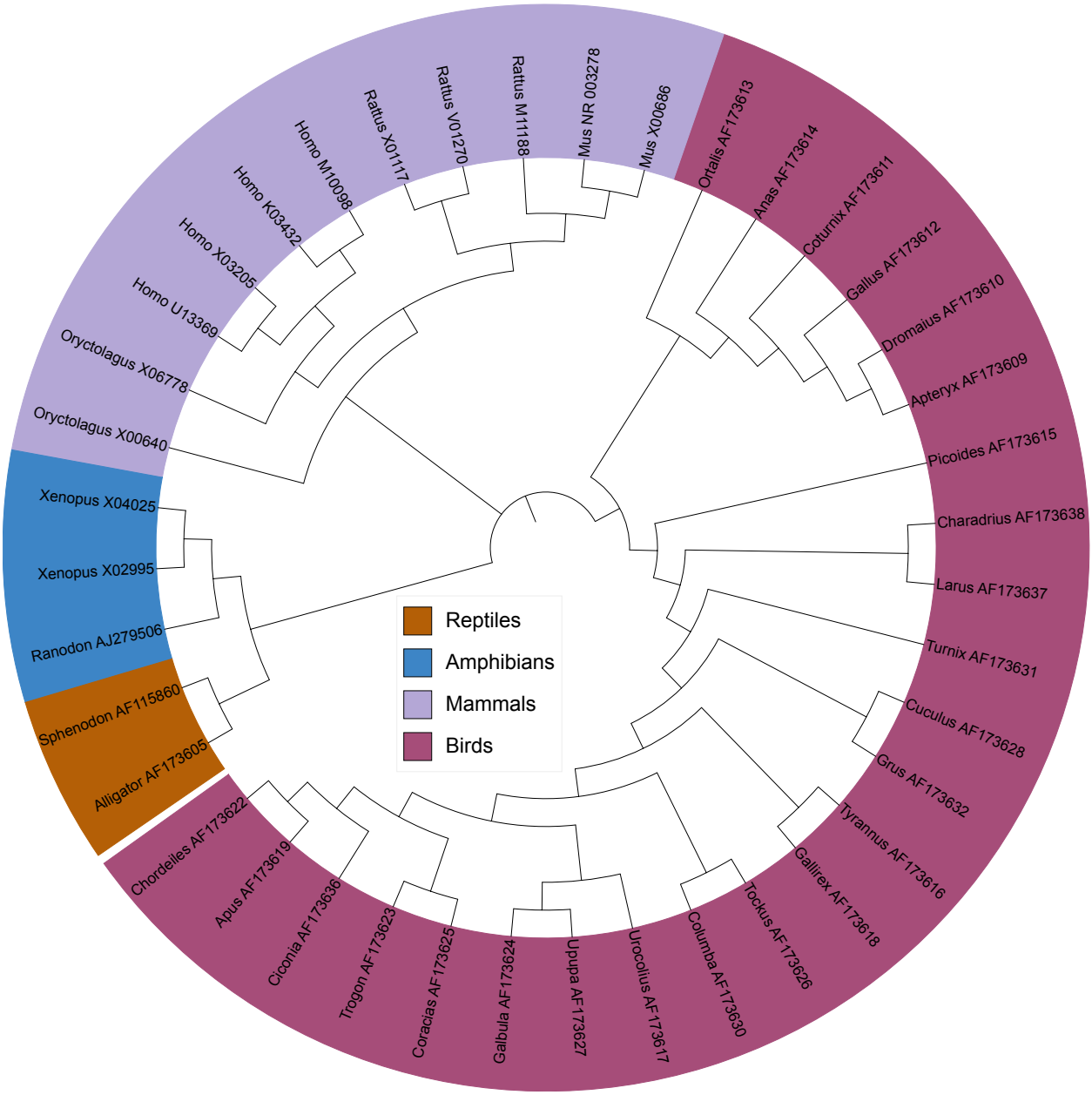


Supplementary Figure S8: Comparison of four feature vectors for sequences B_A04321 and B_A07867 from the identical type in HIV-1 dataset ($k = 9$, $t = 3$).

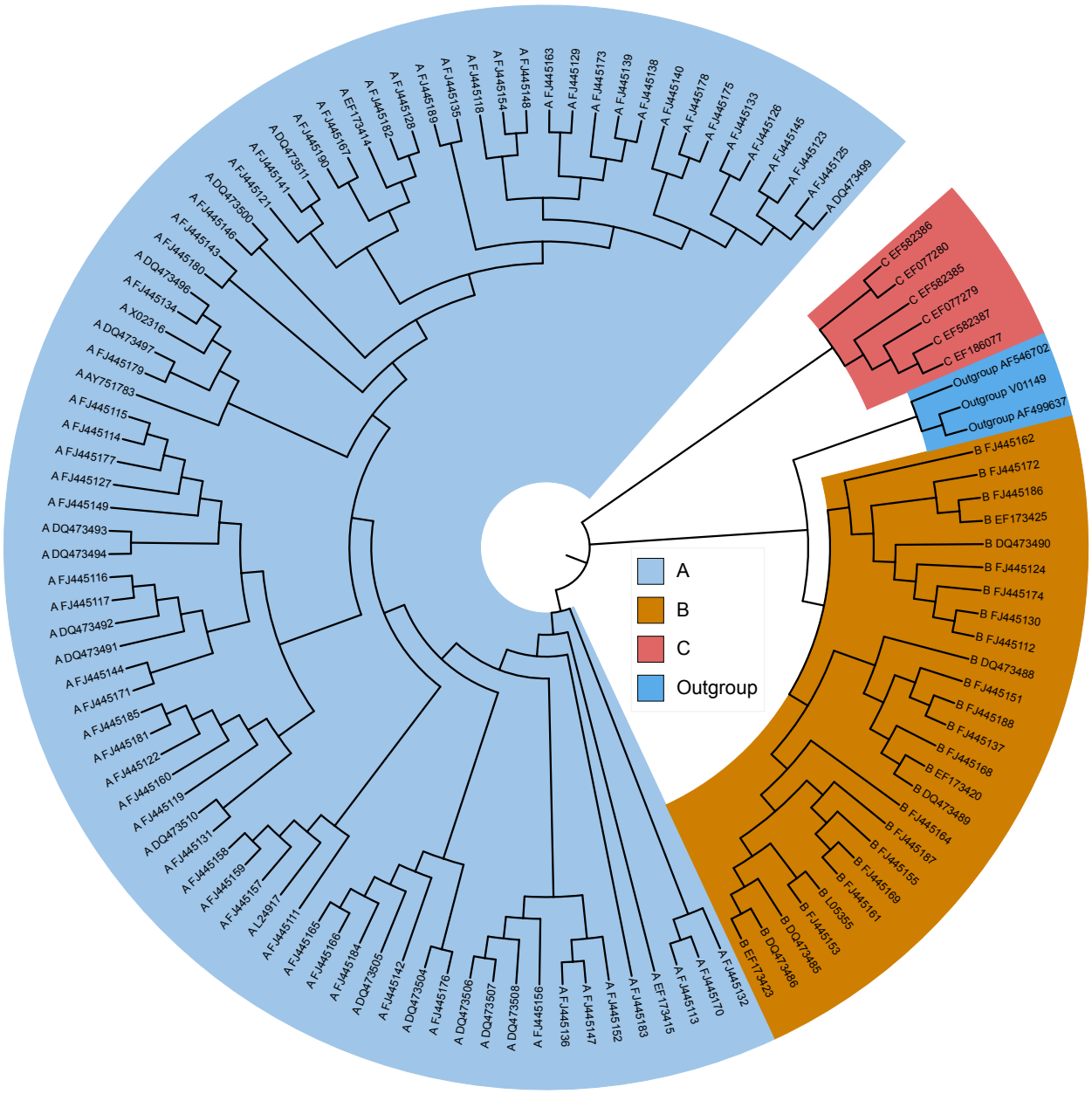
4 Comparative visualization of Spark-based species phylogenetic trees



Supplementary Figure S9: The phylogenetic tree of HBV was constructed using Spark, with parameters k and t set to 9 and 3, respectively.



Supplementary Figure S10: The phylogenetic tree of 18SrRNA was constructed using Spark, with param-eters k and t set to 9 and 3, respectively.



Supplementary Figure S11: The phylogenetic tree of HRV was constructed using Spark, with parameters k and t set to 9 and 3, respectively.