Synonymous mutations reduce genome compactness in icosahedral ssRNA viruses

Luca Tubiana*
Department of Theoretical Physics, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia

Anže Lošdorfer Božič
Department of Theoretical Physics, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia and
Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9b, D-50931 Cologne, Germany

Cristian Micheletti
SISSA, Via Bonomea 265, I-34136 Trieste, Italy

Rudolf Podgornik
Department of Theoretical Physics, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia
Department of Physics, Faculty of Mathematics and Physics,
University of Ljubljana, SI-1000 Ljubljana, Slovenia and
Department of Physics, University of Massachusetts, Amherst, MA 01003
(Dated: October 13, 2014)

* luca.tubiana@ijs.si; Corresponding author

I. FIT OF THE SHUFFLED RNA MLD

To obtain the power law for the MLD of random RNAs, we shuffled 12 RNA sequences of different lengths (1000 nt, 1500 nt, ..., 6000 nt), all having a viral-like nucleotide composition: 0.26 A, 0.28 U, 0.24 G, 0.22 C (obtained excluding Tymoviridae, which have a significantly different composition). For every sequence length, we produced 500 independent sequences over which we computed the expected (thermally averaged) $\langle \mathrm{MLD} \rangle$. The power law of Eq. (1) in the main text is then obtained by fitting the dependence of $\overline{\langle \mathrm{MLD} \rangle}$, further averaged over the 500 different mutations, on the sequence length.

As already mentioned in the main text, Tymoviridae differ notably from the other families in their nucleotide composition, and they were not considered when producing the averaged viral-like composition. Evaluating the average composition for the set of Tymoviridae viruses considered in the main text, we obtain 0.20 A, 0.24 U, 0.18 G, 0.38 C.

Using this alternative composition and adopting the same procedure used for the other families we obtain a scaling law describing the $\overline{\langle \mathrm{MLD} \rangle}$ dependence of Tymoviridae-like random RNA sequences:

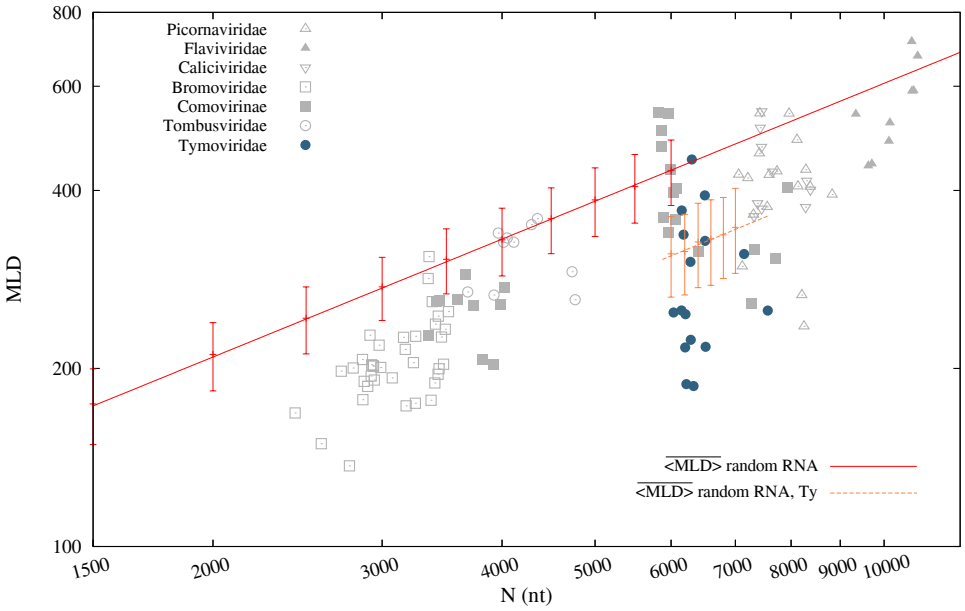
$$\overline{\langle \mathrm{MLD} \rangle}_{\mathrm{Ty}}(N) = (0.92 \pm 0.44) \times N^{(0.669 \pm 0.054)}. \tag{1}$$

Note that the exponent, 0.669 ± 0.054, is compatible with the one obtained for the other viral families, 0.662 ± 0.004. Both fits are shown in Fig. S1.

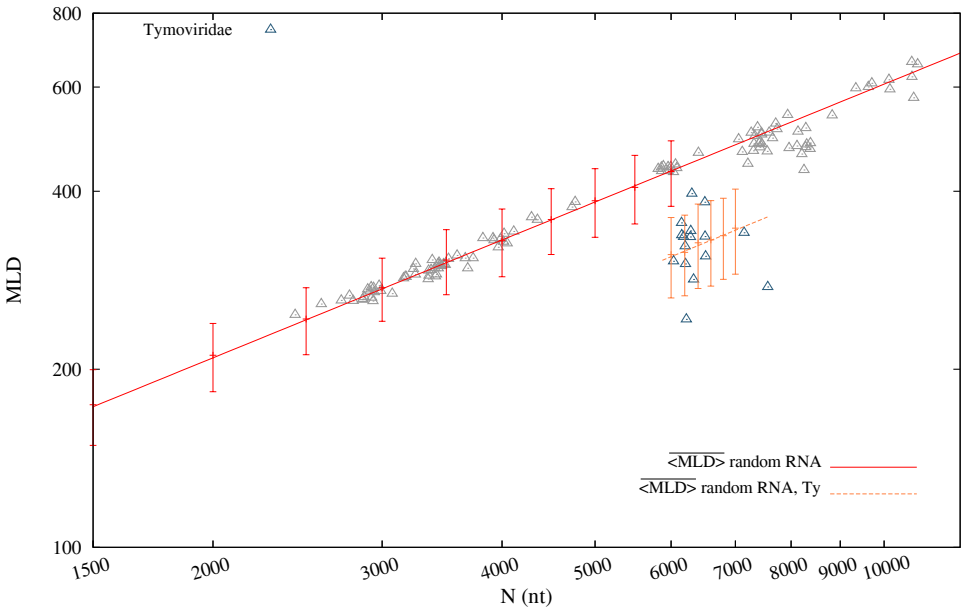
We further check the validity of the scaling laws for our viral families by randomly shuffling the WT RNA sequences themselves, without any further constraints. The results, shown in Fig. S2, show once again that the two scaling laws are a good reference for random RNAs with the viral-like composition considered in our sample. For Tymoviridae, we notice that a couple of viruses remain more compact than predicted by Eq. (S1). This is due to them having a composition which is substantially different from the Tymoviridae average composition.

II. MUTATIONS PRESERVING UTRS

As detailed in the main text, we further tested the robustness of our results by adding additional optional constraint as the preservation of Untranslated regions (UTRs) near the ends of the genome and the preservation of the codon



SI Fig. 1. $\langle \mathrm{MLD} \rangle$ values of WT RNA genomes are shown in gray for all families apart from Tymoviridae, which are highlighted in blue. $\overline{\langle \mathrm{MLD} \rangle}$ values of random sequences are shown with red and orange errorbars for viral-like and Tymoviridae-like nucleotide composition, respectively. The respective fitting lines are displayed with the same colors. The p-value of the fit parameters for the viral-like composition is below 10^{-10}, and the adjusted R^2 is 0.999948. For Tymoviridae-like composition the p-value of exponent is $\simeq 10^{-4}$ and the adjusted R^2 is 0.999968.



SI Fig. 2. $\langle \mathrm{MLD} \rangle$ values of randomly shuffled WT RNA genomes, shown in gray for all families apart from Tymoviridae, which are highlighted in blue. $\overline{\langle \mathrm{MLD} \rangle}$ values of random sequences are shown with red and orange errorbars for viral-like and Tymoviridae-like nucleotide composition, respectively.

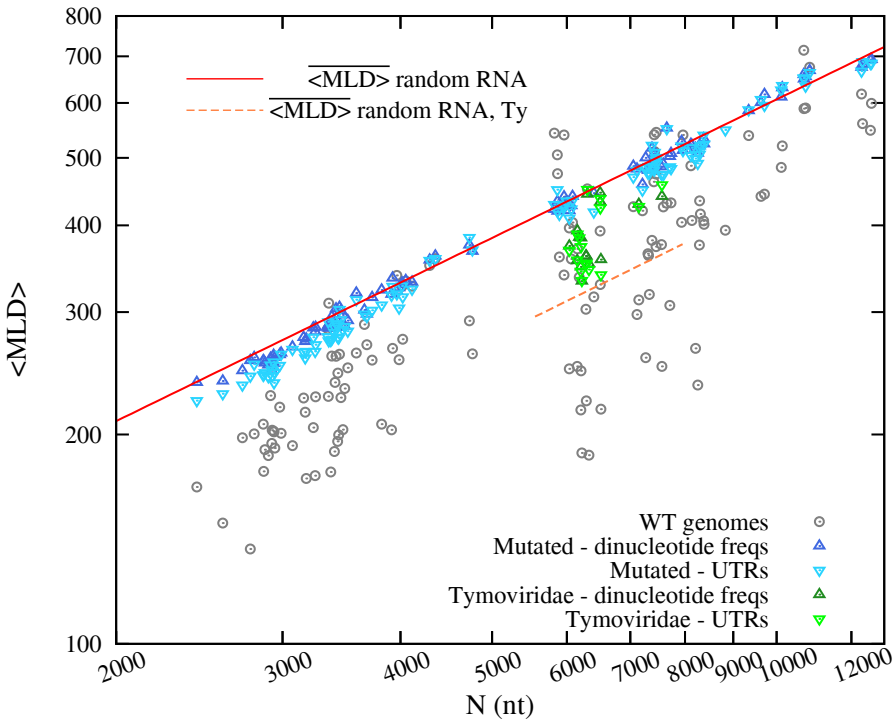
biases within each gene. The $\overline{\langle \mathrm{MLD} \rangle}$ values obtained with these additional constraint are compared with those obtained under the constraint of synonymous mutations only in Fig. 5 in the main text. Here, in Fig. S3 we extend the comparison for the additional constraint of preserving UTRs to include Tymoviridae. $\overline{\langle \mathrm{MLD} \rangle}$ values under the additional constraint of fixed codon bias were not calculated for this family since all the tymoviridae genomes in our set present overlapping genes.

III. MUTATIONS AT FIXED NUCLEOTIDE COMPOSITION

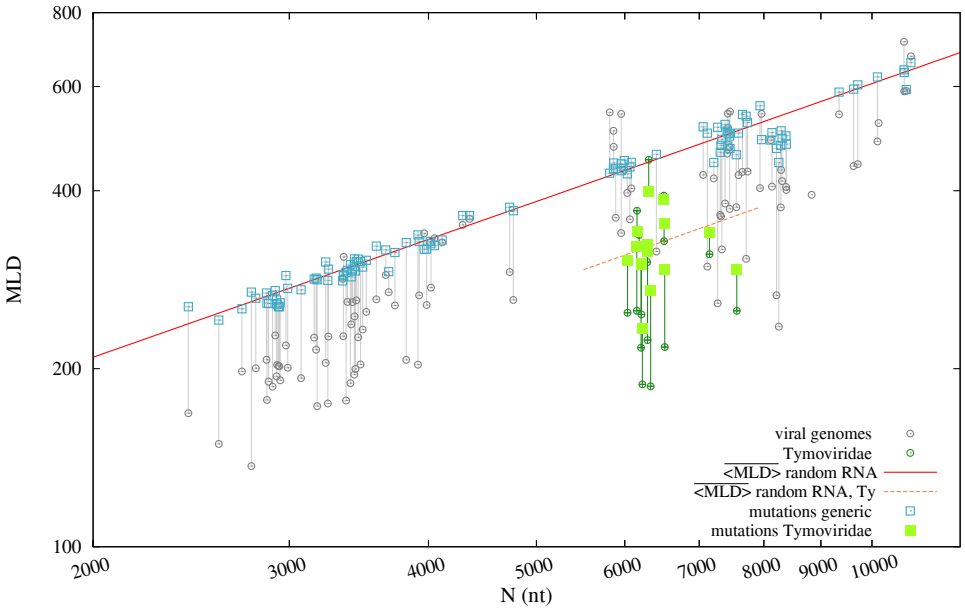
To test the robustness of the results reported in the main text, we implemented another mutation flow which conserves the nucleotide composition instead of the dinucleotide frequencies. This is achieved by using a Fisher-Yates algorithm where proposed shuffles are accepted or rejected on the basis of whether or not the resulting genome still encodes for the same proteins. The results of this different simulation setup are shown in Fig. S4.

Note that the values of $\overline{\langle \mathrm{MLD} \rangle}$ obtained in this way show a clear correlation with those obtained by unrestricted random shuffling of the WT RNA sequences, shown in Fig. S2.

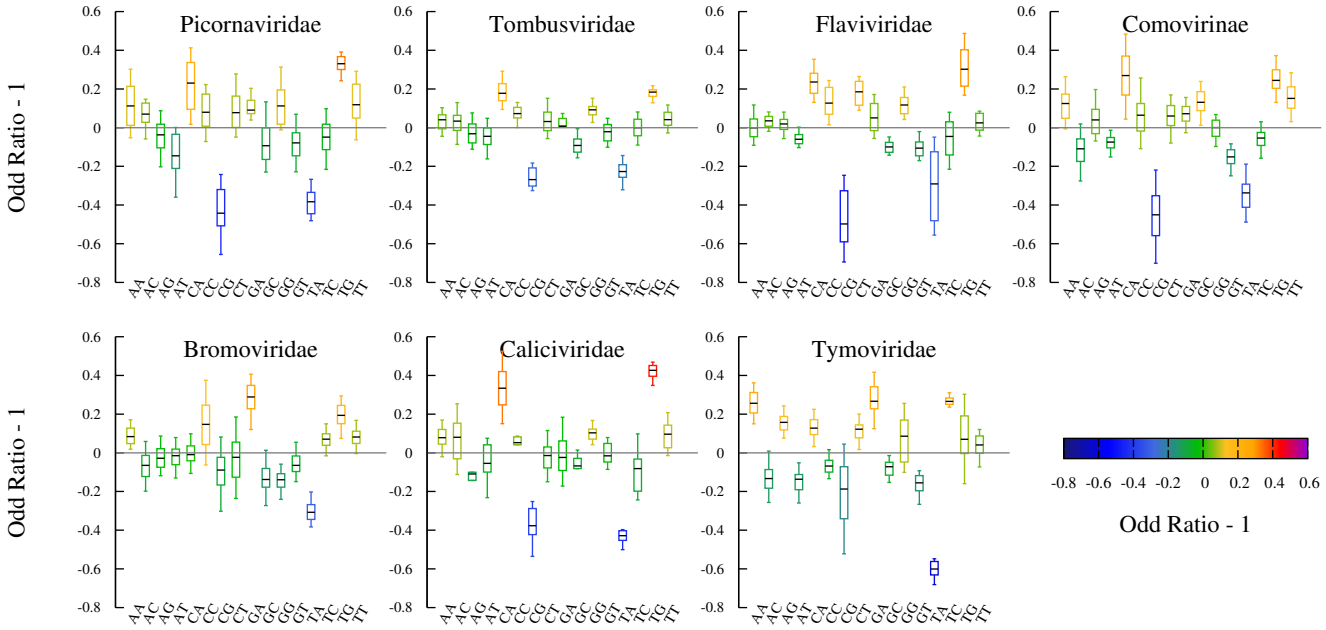
IV. DETAILS OF DINUCLEOTIDE AND NUCLEOTIDE COMPOSITIONS



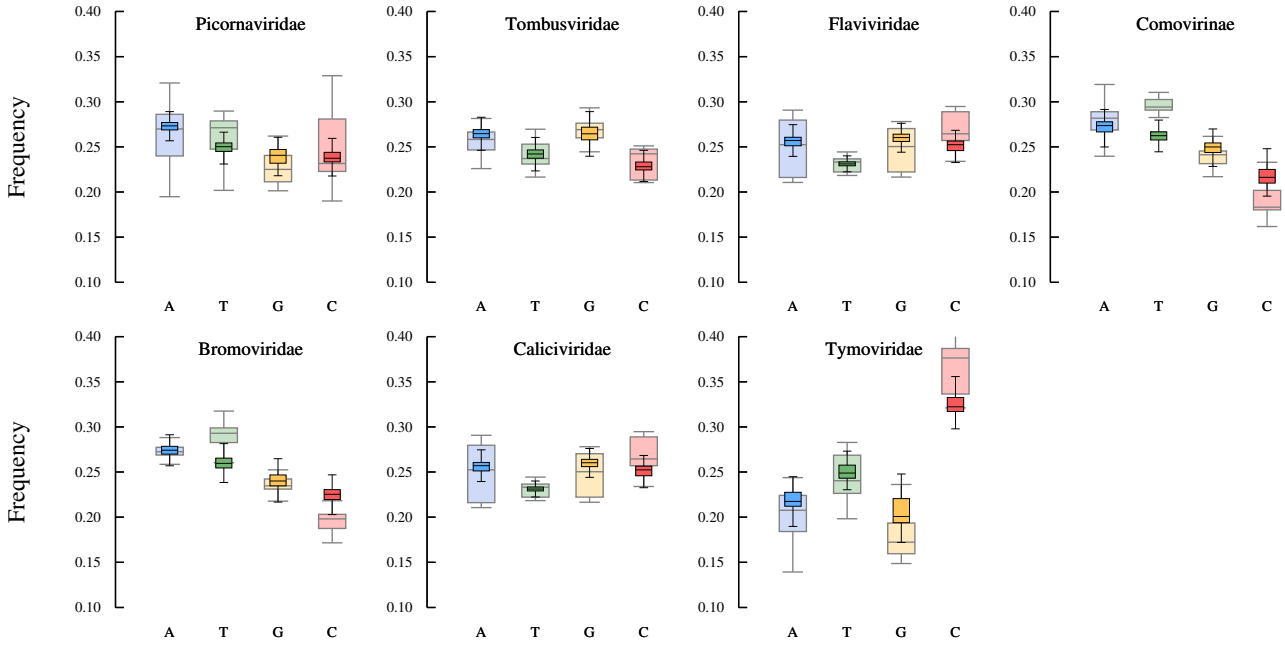
SI Fig. 3. Comparison between the $\overline{\langle \mathrm{MLD} \rangle}$ values for the synonymous constraint only (upward triangles) and for the additional constraints of preserving UTRs sequences (downward triangles), including Tymoviridae. In the latter case $\overline{\langle \mathrm{MLD} \rangle}$ have been evaluated over a set of 150 mutated sequences per virus.



SI Fig. 4. Mutations performed at fixed nucleotide composition. Note that the $\overline{\langle \mathrm{MLD} \rangle}$ of the mutated viral sequences approaches the random RNA values for viral-like and Tymoviridae-like nucleotide composition in both respective cases. We note that for Tymoviridae there are some viruses which remain more compact than the corresponding random RNAs. We argue that this is due to the fact that Tymoviridae show notable fluctuations in their nucleotide composition.



SI Fig. 5. Dinucleotide odd-ratios, rescaled to zero, for the viral families considered in our study. Boxes represent quartiles, and whiskers correspond to 1.5 of the interquartile distance. These values have been used to constrain the mutation flow and produce sequences with viral-like dinucleotide frequencies, see Materials and Methods section in the main text.



SI Fig. 6. Nucleotide frequencies for WT sequences (large boxes) and mutated sequences with constrained dinucleotide composition (small boxes), shown for each virus family considered in our study. Note that in most families the imposition of conserved dinucleotide frequencies results in conserved nucleotide frequencies as well, although for Tymoviridae, Comovirinae, and Bromoviridae the frequencies are not so well preserved, showing the effects of transversion changes.

V. DATASET OF VIRAL GENOMES

Taxon	**Family**	**Nuccore code**	**PDB code**	**length**	$\langle \mathrm{MLD} \rangle_{WT}$	$\overline{\langle \mathrm{MLD} \rangle_{mut}}$	$\overline{\langle \mathrm{MLD} \rangle_{UTRs}}$	$\overline{\langle \mathrm{MLD} \rangle_{CB}}$
Bromoviridae	Anulavirus	PeZSV_RNA1	–	3383	259 ± 19	295 ± 48	274 ± 45	280 ± 43
Bromoviridae	Anulavirus	PeZSV_RNA2	–	2435	168 ± 8	238 ± 39	224 ± 37	235 ± 38
Bromoviridae	Bromovirus	BMV_RNA1	1js9	3234	204 ± 15	285 ± 46	276 ± 43	278 ± 46
Bromoviridae	Bromovirus	BMV_RNA2	1js9	2865	177 ± 12	255 ± 42	244 ± 38	242 ± 40
Bromoviridae	Bromovirus	BrBMV_RNA1	–	3158	225 ± 21	276 ± 44	263 ± 43	264 ± 42
Bromoviridae	Bromovirus	BrBMV_RNA2	–	2799	200 ± 17	258 ± 43	252 ± 40	255 ± 43
Bromoviridae	Bromovirus	CaYBV_RNA1	–	3178	172 ± 14	274 ± 45	263 ± 40	270 ± 45
Bromoviridae	Bromovirus	CaYBV_RNA2	–	2720	197 ± 20	247 ± 41	236 ± 40	239 ± 39
Bromoviridae	Bromovirus	CCMV_RNA1	1cwp	3171	215 ± 27	272 ± 44	258 ± 40	267 ± 44
Bromoviridae	Bromovirus	CCMV_RNA2	1cwp	2774	136 ± 11	256 ± 43	243 ± 38	249 ± 40
Bromoviridae	Bromovirus	MeYFV_RNA1	–	3249	174 ± 19	281 ± 46	263 ± 44	274 ± 47
Bromoviridae	Bromovirus	MeYFV_RNA2	–	2862	207 ± 11	255 ± 40	243 ± 38	241 ± 37
Bromoviridae	Bromovirus	SpBLV_RNA1	–	3252	226 ± 25	285 ± 45	269 ± 43	281 ± 47
Bromoviridae	Bromovirus	SpBLV_RNA2	–	2898	186 ± 16	252 ± 41	242 ± 40	259 ± 39
Bromoviridae	Cucumovirus	GaMMV_RNA1	–	3350	283 ± 28	286 ± 47	277 ± 43	278 ± 44
Bromoviridae	Cucumovirus	GaMMV_RNA2	–	2935	202 ± 8	260 ± 43	254 ± 39	257 ± 42
Bromoviridae	Cucumovirus	PeSV_RNA1	–	3357	309 ± 16	287 ± 46	273 ± 44	273 ± 42
Bromoviridae	Cucumovirus	TAV_RNA1	1laj	3410	237 ± 18	287 ± 46	285 ± 44	281 ± 49
Bromoviridae	Cucumovirus	TAV_RNA2	1laj	3074	192 ± 18	267 ± 43	265 ± 42	–
Bromoviridae	Ilarvirus	ApMV_RNA1	–	3476	203 ± 36	297 ± 49	283 ± 45	292 ± 49
Bromoviridae	Ilarvirus	ApMV_RNA2	–	2979	218 ± 20	261 ± 43	251 ± 43	261 ± 45
Bromoviridae	Ilarvirus	CiLRV_RNA1	–	3404	189 ± 27	289 ± 46	281 ± 48	289 ± 4
Bromoviridae	Ilarvirus	CiLRV_RNA2	–	2990	200 ± 21	262 ± 43	261 ± 42	–
Bromoviridae	Ilarvirus	CiVV_RNA1	–	3433	245 ± 17	291 ± 48	287 ± 45	290 ± 42
Bromoviridae	Ilarvirus	CiVV_RNA2	–	2914	227 ± 29	257 ± 41	252 ± 40	–
Bromoviridae	Ilarvirus	ElMV_RNA1	–	3431	195 ± 11	285 ± 46	276 ± 41	279 ± 44
Bromoviridae	Ilarvirus	ElMV_RNA2	–	2874	190 ± 25	254 ± 41	246 ± 43	–
Bromoviridae	Ilarvirus	ParMV_RNA1	–	3518	249 ± 20	292 ± 48	282 ± 44	301 ± 50
Bromoviridae	Ilarvirus	ParMV_RNA2	–	2922	194 ± 21	247 ± 40	248 ± 39	–
Bromoviridae	Ilarvirus	PrDV_RNA1	–	3374	176 ± 24	285 ± 46	273 ± 42	275 ± 47
Bromoviridae	Ilarvirus	PrDV_RNA2	–	2593	149 ± 17	239 ± 39	229 ± 37	232 ± 39
Bromoviridae	Ilarvirus	SpLV_RNA1	–	3439	199 ± 19	291 ± 48	275 ± 44	291 ± 46
Bromoviridae	Ilarvirus	SpLV_RNA2	–	2939	201 ± 22	253 ± 40	237 ± 37	–
Bromoviridae	Ilarvirus	ToSV_RNA1	–	3491	232 ± 28	286 ± 46	286 ± 48	283 ± 46
Bromoviridae	Ilarvirus	ToSV_RNA2	–	2926	202 ± 15	253 ± 42	243 ± 40	–
Bromoviridae	Ilarvirus	TuAMV_RNA1	–	3459	226 ± 17	301 ± 48	292 ± 47	300 ± 48
Bromoviridae	Ilarvirus	TuAMV_RNA2	–	2944	191 ± 9	258 ± 41	246 ± 40	–
Caliciviridae	Nebovirus	caliciNB	–	7453	473 ± 48	502 ± 79	501 ± 77	498 ± 83
Caliciviridae	Nebovirus	newbury	–	7454	372 ± 18	495 ± 78	496 ± 82	498 ± 83
Caliciviridae	Norovirus	murineNoro1	–	7382	380 ± 36	517 ± 81	521 ± 82	491 ± 83
Caliciviridae	Norovirus	norwalk	1ihm	7654	430 ± 35	552 ± 84	551 ± 77	–
Caliciviridae	Sapovirus	porcineSapo	–	7320	361 ± 36	480 ± 77	486 ± 73	–
Caliciviridae	Sapovirus	sapoMc10	–	7458	544 ± 39	486 ± 78	491 ± 73	–
Caliciviridae	Sapovirus	saporo	–	7429	510 ± 33	508 ± 79	509 ± 79	–
Caliciviridae	Vesivirus	rabbitVV	–	8380	401 ± 20	524 ± 81	523 ± 82	–
Caliciviridae	Vesivirus	stellerVV	–	8305	415 ± 16	508 ± 79	521 ± 77	–
Caliciviridae	Vesivirus	VESV	–	8284	374 ± 41	516 ± 76	516 ± 78	–
Comovirinae	Comovirus	BPMV_RNA1	1bmv	5995	433 ± 40	430 ± 67	434 ± 72	443 ± 66
Comovirinae	Comovirus	BPMV_RNA2	1bmv	3662	288 ± 26	302 ± 49	298 ± 48	302 ± 50
Comovirinae	Comovirus	CowSMV_RNA1	–	5957	339 ± 28	427 ± 69	425 ± 63	430 ± 62
Comovirinae	Comovirus	CowSMV_RNA2	–	3732	255 ± 30	315 ± 51	302 ± 49	309 ± 54
Comovirinae	Comovirus	CPMV_RNA1	1ny7	5889	360 ± 24	423 ± 66	415 ± 66	439 ± 72
Comovirinae	Comovirus	RadMV_RNA1	–	6064	357 ± 21	427 ± 67	422 ± 63	431 ± 73
Comovirinae	Comovirus	RadMV_RNA2	–	4020	274 ± 20	329 ± 53	315 ± 52	323 ± 50
Comovirinae	Comovirus	RCMV_RNA1	rcmv	6033	396 ± 28	420 ± 65	410 ± 59	417 ± 61
Comovirinae	Comovirus	SquashMV_RNA1	–	5865	474 ± 27	419 ± 69	419 ± 70	436 ± 71
Comovirinae	Comovirus	SquashMV_RNA2	–	3354	226 ± 17	285 ± 48	291 ± 49	288 ± 45
Comovirinae	Comovirus	TurRV_RNA1	–	6082	403 ± 32	440 ± 70	434 ± 70	439 ± 63
Comovirinae	Comovirus	TurRV_RNA2	–	3985	256 ± 18	325 ± 52	304 ± 49	298 ± 46
Comovirinae	Fabavirus	BBWV_RNA1	–	5817	542 ± 37	422 ± 68	428 ± 64	444 ± 74
Comovirinae	Fabavirus	BBWV_RNA2	–	3446	260 ± 24	305 ± 49	303 ± 46	307 ± 49
Comovirinae	Fabavirus	mikaniaMMV_RNA1	–	5862	505 ± 46	433 ± 69	450 ± 67	443 ± 73

Comovirinae	Fabavirus	mikaniaMMV_RNA2	–	3418	259 ± 30	303 ± 49	289 ± 47	285 ± 51
Comovirinae	Fabavirus	patchMMV_RNA1	–	5956	539 ± 24	440 ± 70	428 ± 68	438 ± 62
Comovirinae	Fabavirus	patchMMV_RNA2	–	3591	262 ± 32	320 ± 51	313 ± 51	316 ± 55
Comovirinae	Nepovirus	arabisMV_RNA1	–	7334	318 ± 30	485 ± 74	475 ± 78	468 ± 73
Comovirinae	Nepovirus	arabisMV_RNA2	–	3820	207 ± 20	323 ± 53	307 ± 47	319 ± 45
Comovirinae	Nepovirus	blackCRV_RNA1	–	7711	306 ± 31	502 ± 75	486 ± 76	488 ± 78
Comovirinae	Nepovirus	blackCRV_RNA2	–	6405	315 ± 22	445 ± 67	418 ± 65	417 ± 71
Comovirinae	Nepovirus	raspRV_RNA1	–	7935	404 ± 39	528 ± 83	520 ± 81	539 ± 81
Comovirinae	Nepovirus	raspRV_RNA2	–	3914	203 ± 13	318 ± 50	317 ± 53	319 ± 48
Comovirinae	Nepovirus	TRSV_RNA2	1a6c	7271	257 ± 20	500 ± 80	484 ± 75	502 ± 84
Flaviviridae	Flavivirus	alkhurma	–	10685	714 ± 36	659 ± 10	651 ± 10	646 ± 96
Flaviviridae	Flavivirus	apoi	–	10116	484 ± 38	612 ± 96	626 ± 97	615 ± 88
Flaviviridae	Flavivirus	dengue	–	10735	589 ± 45	654 ± 99	634 ± 98	586 ± 96
Flaviviridae	Flavivirus	montana	–	10690	588 ± 34	649 ± 99	652 ± 10	628 ± 92
Flaviviridae	Flavivirus	powassan	–	10839	674 ± 52	668 ± 10	663 ± 97	656 ± 95
Flaviviridae	Flavivirus	rioBravo	–	10140	520 ± 79	631 ± 10	635 ± 98	618 ± 91
Flaviviridae	Hepacivirus	HepC2	–	9711	443 ± 36	617 ± 10	595 ± 86	604 ± 96
Flaviviridae	Hepacivirus	HepC5	–	9343	538 ± 88	585 ± 97	586 ± 91	580 ± 86
Flaviviridae	Hepacivirus	HepC6	–	9628	440 ± 25	601 ± 92	607 ± 93	599 ± 95
Flaviviridae	Pestivirus	border	–	12333	560 ± 38	681 ± 10	688 ± 10	–
Flaviviridae	Pestivirus	BVDV1	–	12573	547 ± 30	692 ± 10	684 ± 10	–
Flaviviridae	Pestivirus	classicalSFV	–	12301	617 ± 61	675 ± 10	667 ± 98	–
Flaviviridae	Pestivirus	pestiGiraffe	–	12602	598 ± 70	693 ± 11	684 ± 10	–
Picornaviridae	Aphthovirus	BovRBV	–	7556	375 ± 36	486 ± 76	474 ± 77	444 ± 68
Picornaviridae	Aphthovirus	ERAV	2wff	7734	430 ± 33	508 ± 81	483 ± 75	518 ± 82
Picornaviridae	Aphthovirus	FMDV_typeO	1zba	8134	406 ± 31	521 ± 81	517 ± 79	504 ± 86
Picornaviridae	Cardiovirus	saffold	–	8115	487 ± 36	523 ± 82	504 ± 78	485 ± 73
Picornaviridae	Cardiovirus	TMEVlike	–	7961	539 ± 37	513 ± 82	512 ± 84	494 ± 76
Picornaviridae	Enterovirus	BEV	1bev	7414	462 ± 47	497 ± 79	484 ± 76	495 ± 88
Picornaviridae	Enterovirus	Hentero107	–	7423	539 ± 31	487 ± 77	480 ± 77	474 ± 71
Picornaviridae	Enterovirus	Hrhino14	1d3i	7212	419 ± 17	458 ± 73	449 ± 71	437 ± 63
Picornaviridae	Erbovirus	ERBV1	–	8828	393 ± 27	548 ± 90	549 ± 86	538 ± 80
Picornaviridae	Kobuvirus	aichi	–	8251	235 ± 20	508 ± 79	491 ± 78	421 ± 63
Picornaviridae	Kobuvirus	bovineKV	–	8374	405 ± 28	533 ± 82	539 ± 79	470 ± 66
Picornaviridae	Kobuvirus	porcineKV	–	8210	266 ± 25	516 ± 79	499 ± 80	445 ± 75
Picornaviridae	Parechovirus	ljungan	–	7590	425 ± 36	490 ± 77	473 ± 75	478 ± 73
Picornaviridae	Sapelovirus	asapelo	–	8289	433 ± 38	520 ± 82	506 ± 77	506 ± 75
Picornaviridae	Senecavirus	SVV	3cji	7310	364 ± 24	480 ± 76	475 ± 76	454 ± 72
Picornaviridae	Teschovirus	ptescho1	–	7117	297 ± 23	482 ± 78	480 ± 73	498 ± 84
Picornaviridae	Tremovirus	AEV	–	7055	425 ± 43	487 ± 77	469 ± 74	488 ± 74
Tombusviridae	Aureusvirus	MaWLMV	–	4293	350 ± 15	357 ± 56	356 ± 55	–
Tombusviridae	Aureusvirus	pothos	–	4354	358 ± 18	361 ± 57	358 ± 56	–
Tombusviridae	Avenavirus	OCSV	–	4114	327 ± 18	331 ± 50	324 ± 51	–
Tombusviridae	Carmovirus	angelonia	–	3964	338 ± 16	322 ± 52	319 ± 49	–
Tombusviridae	Carmovirus	JapINRV	–	4014	326 ± 45	331 ± 52	327 ± 55	–
Tombusviridae	Carmovirus	PelFBV	–	3923	266 ± 13	336 ± 53	327 ± 52	–
Tombusviridae	Carmovirus	TuCrV	3zx8	4050	332 ± 25	333 ± 54	326 ± 53	–
Tombusviridae	Necrovirus	TNV_A	1tnv	3684	269 ± 6	298 ± 47	296 ± 49	–
Tombusviridae	Tombusvirus	GrALV	–	4731	291 ± 22	375 ± 59	384 ± 60	–
Tombusviridae	Tombusvirus	pearLV	–	4766	261 ± 12	367 ± 59	369 ± 57	–
Tymoviridae	Maculavirus	GFkV	–	7564	250 ± 20	440 ± 69	457 ± 69	–
Tymoviridae	Marafivirus	GVSV1	–	6506	392 ± 36	446 ± 68	437 ± 67	–
Tymoviridae	Marafivirus	MRFV	–	6305	451 ± 23	443 ± 67	450 ± 71	–
Tymoviridae	Marafivirus	OBDV	–	6509	328 ± 35	432 ± 72	424 ± 61	–
Tymoviridae	Marafivirus	OLV3	–	7148	312 ± 27	429 ± 67	426 ± 68	–
Tymoviridae	Tymovirus	AnVYV	–	6151	250 ± 17	356 ± 56	357 ± 55	–
Tymoviridae	Tymovirus	ChYMV	–	6517	217 ± 16	357 ± 58	339 ± 53	–
Tymoviridae	Tymovirus	DiYMV	–	6290	223 ± 26	361 ± 56	353 ± 55	–
Tymoviridae	Tymovirus	DuMV	–	6181	336 ± 44	384 ± 60	384 ± 59	–
Tymoviridae	Tymovirus	EgMV	–	6331	186 ± 18	352 ± 57	346 ± 53	–
Tymoviridae	Tymovirus	ErLV	–	6035	248 ± 24	373 ± 60	368 ± 59	–
Tymoviridae	Tymovirus	NeRNV	–	6285	302 ± 23	361 ± 56	351 ± 53	–
Tymoviridae	Tymovirus	OkMV	–	6223	188 ± 29	333 ± 52	333 ± 50	–
Tymoviridae	Tymovirus	OnYMV	–	6211	247 ± 31	384 ± 62	373 ± 57	–

Tymoviridae	Tymovirus	PlMV	–	6154	369 ± 26	393 ± 61	389 ± 64	–
Tymoviridae	Tymovirus	ScMV	–	6206	217 ± 18	348 ± 54	343 ± 53	–

SI Table I: **Set Viral genomes used in this study, including genome length and average MLD values.** $\langle MLD_{WT} \rangle$ refers to thermal average of the MLD obtained on WT sequences. $\overline{\langle MLD_{mut} \rangle}$, $\overline{\langle MLD_{UTRs} \rangle}$, $\overline{\langle MLD_{CB} \rangle}$, refer to average MLD values obtained on synonymously mutated sequences, synonymously mutated sequences with preserved UTRs, and synonymously mutated sequences with preserved UTRs and codon bias, respectively (see Material and Methods in the main text); in these cases an additional averaging over a wide set of possible mutations is performed. Errors are reported as standard deviations.