Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500155

of freedom due to fixed marginal (rows or columns) totals. Distances to neutrality resulted enormous. The significance at the 0.05 level for χ^2_1 corresponds to 3.84; I found 377.3 (see Table 1) in *Sars-Cov-2* virus, for dinucleotide CG, with separation 0 (contiguous bases); the probability (P) is out of tables and programs; with Gaussian approximation (5), $P < 10^{-50}$. The significance at the 0.05 level for χ^2_9 corresponds to 16.9; I found in human chromosome 21 χ^2_9 = 1,885,266.8 [5] for Sep 0 ($P < 10^{-222,180}$). Unexpectedly I found periodicity with period 3 in the χ^2 value of distance to neutrality in virus and prokaryotes and period 2 and 6 in human chromosomes. This periodicity of the distance to neutrality is incompatible with neutral and nearly neutral evolution [2-5].

Selective values allow the characterization of selective profiles of dinucleotides (hereafter, pair and dinucleotide will be synonymous unless I state other precision). The selective profile of a dinucleotide includes: 1) its selective value, 2) the sign of this selection value, 3) its χ^2 value and 4) the order of significance of this value (1° to 16°). The χ^2 value is calculated by $(Obs_i - Exp_i)^2/Exp_i$, where Obs_i and Exp_i are the observed and expected numbers of the i_{th} dinucleotide, respectively. The total χ^2_9 value is the sum of the 16 values of the 16 nucleotides. The selection coefficient and its sign is the quotient $(Obs_i - Exp_i)/Exp_i$, where positive coefficients (selectively advantageous) apply to dinucleotides with more observed than expected pairs and negative coefficients (selectively disadvantageous) apply to dinucleotides with less observed than expected pairs. Any dinucleotide has its selective profile and this profile (a profile of an index dinucleotide or I profile) compared with the other 15 profiles produces 15 distances (to the I profile).

In double stranded (ds) RNA or DNA mutation or selection (and eventually genetic drift) of one or the two bases of one dinucleotide occur synchronously among the four dinucleotides that a pair implies due to base complementariness and 5'-3' sense involved in duplication or transcription processes. This does not happen in a single stranded (ss) DNA or RNA. Thus, the test of comparison among selective profiles can discriminate whether the nucleic acid of the organism in study is ds or ss DNA or RNA. In ds DNA as in the human genome, the Index (I) 5'G…A3' dinucleotide has a contra-sense (C-S) pair 3'A…G5' (in the same DNA strand), a parallel (Par) pair 3'C…T5', and an anti-Parallel (a-Par) pair 5'T…C3' pair in the complementary strand. These four dinucleotides evolve together and synchronously. This coordinated evolution is not possible (physically but not evolutionarily during the viral life cycle) in ss viruses. Since there is only one published strand (the index strand from GenBank) of the nucleic acid, I study the selective profile of these four dinucleotides in the index strand (all pairs in the sense 5'…3') and determine the similitude or difference among their selective profiles. Regardless (a priori) the reproductive processes of the organism, if there are evolutionary differences in replication or transcription between the senses of the copy, or differences due to the specific pair, the test will detect these differences. In ss DNA or RNA, these differences should be impossible because there are not a complementary strand to generate those differences.

This study is on RNA or DNA viruses, however, the differences I study are not on the ss or ds physical constitution of viruses well stablished from the infective viral particle (virion). This study deals with on the evolutionary constitution and dynamic behavior of viruses. Either ds or ss viruses have different ds or ss processes along their life cycles. Mutation, selection, genetic drift, evolutionary contingencies occur at any stage of the reproductive or vital viral cycle with different intensities and in mono or bi-stranded constitutions. As for example, ss retroviruses having a retro-duplication of its nucleic acid and existence as a temperate or inserted virus in the DNA of the host may have rather ds (evolutionary) nucleic acid selective profile. A ds virus whose stage of mutation is mainly at the ss replication process may have ss selective profile (this is very improbable).

Genomes and Methods

Complete genomes taken from GenBank (search in PubMed, Nucleotide, FASTA).

Double stranded DNA: Human adenovirus C strain KF268127.1 Human USA CL 42 1988 (35,931 bp); Human cytomegalovirus strain AD169 X17403 (229,354 bp); Escherichia phage 2 vB EcoM PhAPEC2 KF562341 (167,318 bp); His2 virus NC_007918.1 (16,067 bp).

Single stranded DNA: Penaeus monodon hepandensovirus 4 NC_011545.2 (6,310 b); Escherichia phage phi-X174 NC_001422 (5,386 b); Phage M13 Enterobacteria NC_003287.2 (6,407 b).

Double stranded RNA: Marmot picobirnavirus strain HT1 KY855428.1 (4,579 bp); Gentian Kobu virus NC_020252.1 (22,711 bp).

Single stranded RNA: *SARS-Cov-2* LR757998.1 Wuhan (29,866 b); Sindbis virus NC_001547.1 (11.703 b); HIV MN691959.1 isolate ACH2-NFLMDA13 B1 (USA) (9,943 b). *E coli* genome strain AVS0967 NZ_CP124398 (5,097,505 bp). The number of nucleotides used in the analyses may be less than the number in GenBank because the base identification is not complete. I did not study positive and negative condition of the viral strand to restrict the analysis.

Method

- The nucleotide sequence of a genome, chromosome or nucleic acid segment constitute a primary file obtained from GenBank.

Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500151

)LJXUH GLVSOD\V WKH DDH
separated by 0, 1, 2, … K sites are constructed.

- Each of these sets yield a matrix where rows are the four EDVHV $7*& RU WKH ILUVW EDVH GRZQVWUHDP columns are the four bases A, T, G, and C for the second base (upstream). This is a matrix of 4x4 for the 16 possible dinucleotides. For each of the 16 dinucleotides with bases VHSDUDWHG E\ VLWHV WKH VHW $ describes the distance to neutrality. The expected number of each dinucleotides (Exp) is calculated by (fB_u)x(fB_d)xN, where f is the REVHUYHG IUHTXHQF\ RI RFFXUUHQFHV $ RU & and d denote upstream and downstream, respectively and N is the total number of observed (Obs) dinucleotides for WKLV .7KH $ (one degree of freedom = df) values for each dinucleotide is [(Obs_i-Exp_i)^2/ Exp_i), i going from 1 WR 7KH WRWDO $ with 9 df is the sum of the 16 values SOXV D PLQLPDO YDOXH FRUUHVSRQGLQJ WR WKH GLuHUHQFH RI the Obs number of each dinucleotide with the total. This VSHFLDO GLuHUHQFH LV VPDOO IRU HDFK GLQXFOHRWLGH DQG neglected in present calculations.

7KH VHOHFWLRQ FRHIILFLHQW ZLWKLQ VDOJHEUDLF VLJQ UHODWLRQVKLS from ((Obs_i-Exp_i)/ Exp_i)

- Data are presented in a matrix where the rows are K (Sep = separations) and columns are selective values ordered DFFRUGLQJ WR WKH LGLVWDQFH WR $ VLJQDO FDH6LJVLJQ IWKHVHOHFWLRQ YDOXH RI WKH 33 separations, including 0 Sep. These selective values FRQWLWXWHV WKH VHOHFWLYH SURILOH RI WKLV SDLU XVHG RQ four Seps to short the analysis.

7KH HVWLPDWHG WKH YDOXHV RI WKH SURFOLV GLUHFWLRQ JXUHVMXVWGHVFULEHG)URP WKHVHHVWLPDWHV WKH GLuHUHQFH EHWZHHQ WKH RUGHU RI VLJQFDH VHOHFWLRQ RI and the sign of selection between the parallel (Par) or the anti-Parallel (a-Par) and the Index dinucleotide follows by GLUHFWV XEWUDFWLRQ GLGRQWLE($)XGH WKH HVWLPDWLRQ IVHOHFWLYHGLuHUHQFHVEHFDXVHJHQHPHVG in the number of dinucleotides, and this value depends JUHDWO\RI WKLV QPEHUQU , LQ OXGHG WKH SURFOHRYHG contra-sense pair.

- In the present study separations go from 0 (contiguous bases) to 3 sites of separations.

- Study these traits in Table 1.

6RPH DG KRF VWDWLVWLFDO FRQLGHUDWLRQ DQ WRROV PRGLHGIURP

I analyzed all the genome dinucleotides of an organism, VRSDUDPHWHUVDUHNRZQWDWLVWLFDOWHVWVVDUHRWHHVVU\ applied statistical tests to show the robustness of the analyses DQVHQH RI GLuHUHQFHV 7KH GLVWDQFH RI D VHOHFWLYH SSURQH between two types of dinucleotides has mean (M), variance (V) and standard deviation (SD). For example, the variable distance of Par or a-Par dinucleotides to the Index (ID) dinucleotide. These parameters consider all the possible GLVWDQFHV RFFXUULQZLWK HTXDO SUREDELOLW\ QLJQ 6LJ FRPSDULVRQ RQ\DVROXWH GLuHUHQFHV ZLWKRXW signs) distances are used. An algebraic full demonstration of estimates of M, V and SD is out of the scope of this article 26$RuHUHGDQWXLWLYHXWFRPSOHWHFDOFXODWLRQ Valenzuela, submitted for publication). The expected mean GLuHUHQFH LGLJV EHWZHHQ WKH ,QH[SURQH DQ 3DU RU D3DUSURQHVLV0 DQLWV6LV:LWKWKHVH SDUDPHWHUV,WHVWHGWKHREVHUYHGJXUHVZLWKRQWD ZLWKWKHXVHG VDPSOH VL]HV LW GRHV RQW GLuHUIURP DQHTXDO YDULDQFHV 7KHLQWLDOGLuHUHQFHEHWZHHQ FRHvFLHQV FRVO LV FDOFXODWHG ZLWK GLuHUHFW JXUH GLuHUHQFHV WR OLPLWV VRXQG EHWZHHQ WKH VHWRI,GLQFOHRWL is the same of the set of Par and a-Par dinucleotides). The SD must approximate that of a uniform distribution theoretically EHWZHHQ WR OLPLWV RI FRVHOV EXW VLQH VHOHFWLRQ is strongly dependent on evolutionary contingencies, I used the observed variance to construct the test because variances SURFHHG IURP WKH VDPH VRXUFH 7KH GHQWLYH FRPSDUL WHVW PXVW EH RQ DEVROXWH GLuHUHQFHV WR DYRLG WKH EHWZHHQ HDV WKDW LV ZKDW LPSOLHV GLVWDQFH HTXDO 7KH $ value is always positive and hides the direction of selection. I ordered those values according to a cline from the most positively to the most negatively selected or vice versa. The previous analysis (C. Y. Valenzuela, submitted for publication) considered only the magnitude of the deviations to neutrality the present one includes the sense of selection; that is why the Table in reference (C. Y. Valenzuela, VXEPLWWHGIRUSXEOLFDWLRQVGLuHUHQWZLWKWKHSUH is irrelevant for the order of dinucleotides whether they begin from the most negative or the most positive pair. The matrix RUGHUHG DFFRUGLQJWKH VLJQI VHOHFWLRQDNHV XQFHV the comparison of signs of dinucleotides, because this is implicit.

7KH $ values of these studies are often enormous and out of tables and programs. I approximated the probability YWKHVHVWKHRUHWLFDORUGXVVLDQHTXLYDOHQWWKLVVDQW for small degrees of freedom (df), however 9 df may be approximated with a moderate error. The mean and variance RID$^{2} distribution are the degrees of freedom (df) and twice GI UHVSHFWLYHO\, PDGH HTXLYDOHQW RQ SRLQW GHFLPDO each two SD from the expected mean (df). With one df the application is no as valid as with nine df, but with enormous JXUHV LW LHOGV DQSSUR[LPDWLRQV R DTXDOLWDWLYH evaluation.

5HVXOWV

7DEOH VKRZV WKH PDWUL[6HOHFWLYH SURILHV FROX 16 Separations (Sep) for this strain of the *SARS-CoV-2* virus.

Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500155

selection coefficient followed the same pattern as Significance, with a higher probability of significance (a less significant result). It is remarkable that this result is different from a previous study (C.Y. Valenzuela, submitted for publication) with distances of the chi-squared value without specifying the sign of selection (OSA).

Before presenting the study of twelve viruses, I present the dinucleotide analysis of the *E. coli* genome to establish clear differences with a rather big (5 million pb) ds DNA. This is necessary because there is not another researcher working in this field and the reader must know what happens with big ds genomes. Table 3 presents the dinucleotide spectrum of the genome analysis of *E coli* with 4 Seps.

Table 2: *SARS-CoV-2*. Index dinucleotides with different Par and a-Par pairs. Four Separations

	Indexes		Par		a-Par		D Ind-Par		D Ind-a-Par		I Par	I a-Par
Sep	dinuc[s]	co-sel	dinuc[s]	co-sel	dinuc[s]	co-sel	D Si	D Se	D Si	D Se	AB Se	AB Se
0	AG[-] 8°	-0.00747	TC[-] 4°	-0.198	CT[+] 13°	0.1809	4	-0.1905	5	0.1884	0.1905	0.1884
1	AG[+] 5°	-0.01681	TC[+] 2°	0.0673	CT[-] 9°	-0.0145	3	0.0841	4	0.0023	0.0841	0.0023
2	AG[+] 4°	0.03148	TC[+] 7°	0.0044	CT[-] 11°	-0.0316	3	-0.0271	7	-0.0631	0.0271	0.0631
3	AG[+] 6°	0.02928	TC[-] 13°	-0.0349	CT[+] 2°	0.0627	7	-0.0641	4	0.0334	0.0641	0.0334
0	AC[+] 14°	0.2312	TG[+] 16°	0.377	GT[+] 10°	0.0573	2	0.1458	4	-0.1739	0.1458	0.1739
1	AC[-] 8°	-0.00692	TG[+] 6°	0.0158	GT[+] 1°	0.1232	2	0.0227	7	0.1302	0.0227	0.1302
2	AC[-] 9°	-0.01397	TG[-] 16°	-0.1342	GT[-] 15°	-0.0677	7	-0.1202	6	-0.0538	0.1202	0.0538
3	AC[+] 7°	0.00987	TG[+] 1°	0.0907	GT[+] 3°	0.0498	6	0.0809	4	0.0399	0.0809	0.0399
0	TG[+] 16°	0.37697	AC[+] 14°	0.2312	CA[+] 15°	0.269	2	-0.1458	1	-0.108	0.1458	0.108
1	TG[+] 6°	0.01581	AC[-] 8°	-0.0069	CA[+] 4°	0.0424	2	-0.0227	2	0.0266	0.0227	0.0266
2	TG[-] 16°	-0.1342	AC[-] 9°	-0.014	CA[+] 6°	0.0107	7	0.1202	10	0.1449	0.1202	0.1449
3	TG[+] 1°	0.09072	AC[+] 7°	0.0099	CA[-] 11°	-0.0136	6	-0.0809	10	-0.1043	0.0809	0.1043
0	TC[-] 4°	-0.19796	AG[-] 8°	-0.0075	GA[-] 6°	-0.0805	4	0.1905	2	0.1175	0.1905	0.1175
1	TC[+] 2°	0.06726	AG[+] 5°	-0.0168	GA[-] 14°	-0.0547	3	-0.0841	12	-0.122	0.0841	0.122
2	TC[+] 7°	0.00437	AG[+] 4°	0.0315	GA[-] 10°	0.015	3	0.0271	3	0.0106	0.0271	0.0106
3	TC[-] 13°	-0.03486	AG[+] 6°	0.0293	GA[-] 9°	-0.0028	7	0.0641	4	0.0321	0.0641	0.0321
0	GA[-] 6°	-0.08048	CT[+] 13°	0.1809	TC[-] 4°	-0.198	7	0.2614	2	-0.1175	0.2614	0.1175
1	GA[-] 14°	-0.05474	CT[-] 9°	-0.0145	TC[+] 2°	0.0673	5	0.0402	12	0.122	0.0402	0.122
2	GA[-] 10°	0.01495	CT[-] 11°	-0.0316	TC[+] 7°	0.0044	1	-0.0466	3	-0.0106	0.0466	0.0106
3	GA[-] 9°	-0.00278	CT[+] 2°	0.0627	TC[-] 13°	-0.0349	7	0.0654	4	-0.0321	0.0654	0.0321
0	GT[+] 10°	0.05733	CA[+] 15°	0.269	AC[+] 14°	0.2312	5	0.2116	4	0.1739	0.2116	0.1739
1	GT[+] 1°	0.12324	CA[+] 4°	0.0424	AC[-] 8°	-0.0069	3	-0.0808	7	-0.1302	0.0808	0.1302
2	GT[-] 15°	-0.06772	CA[+] 6°	0.0107	AC[-] 9°	-0.014	9	0.0784	6	0.0538	0.0784	0.0538
3	GT[+] 3°	0.04977	CA[-] 11°	-0.0136	AC[+] 7°	0.0099	8	-0.0634	4	-0.0399	0.0634	0.0399
0	CA[+] 15°	0.26897	GT[+] 10°	0.0573	TG[+] 16°	0.377	5	-0.2116	1	0.108	0.2116	0.108
1	CA[+] 4°	0.04243	GT[+] 1°	0.1232	TG[+] 6°	0.0158	3	0.0808	2	-0.0266	0.0808	0.0266
2	CA[+] 6°	0.01072	GT[-] 15°	-0.0677	TG[-] 16°	-0.1342	9	-0.0784	10	-0.1449	0.0784	0.1449
3	CA[-] 11°	-0.01358	GT[+] 3°	0.0498	TG[+] 1°	0.0907	8	0.0634	10	0.1043	0.0634	0.1043
0	CT[+] 13°	0.1809	GA[-] 6°	-0.0805	AG[-] 8°	-0.0075	7	-0.2614	5	-0.1884	0.2614	0.1884
1	CT[-] 9°	-0.01453	GA[-] 14°	-0.0547	AG[+] 5°	-0.0168	5	-0.0402	4	-0.0023	0.0402	0.0023
2	CT[-] 11°	-0.0316	GA[-] 10°	0.015	AG[+] 4°	0.0315	1	0.0466	7	0.0631	0.0466	0.0631
3	CT[+] 2°	0.06265	GA[-] 9°	-0.0028	AG[+] 6°	0.0293	7	-0.0654	4	-0.0334	0.0654	0.0334
	Mean	0.031		0.031		0.031	4.94	0	5.31	0	0.099	0.0844
	SD	0.1099		0.1099		0.1099	2.36	0.1196	3.06	0.102	0.0672	0.0572
	P t test							P t Si	0.2952		P t Se	0.3626

Par = parallel; a-Par = anti-parallel; dinuc[s] = dinucleotide; co-sel = selection coefficient; D Ind-Par = distance Index-Par; D Ind-a-Par = distance Index a-Par; D Si = difference in Significance; D Se = difference in co-sel; AB Si = absolute difference in D Si; AB Se = absolute difference in D Se; SD =standard deviation; P t test = probability under a one tailed t test. with equal variance for the comparisons of D Si and D Se. respectively

Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500155

Table 3: Significance vs Separations of dinucleotides from *E. coli*. Four Separations

Sep	dinuc(s)	χ^2_1	co-sel	dinuc(s)	χ^2_1	co-sel	dinuc(s)	χ^2_1	co-sel	dinuc(s)	χ^2_1	co-sel
	1° Significance			2° Significance			3° Significance			4° Significance		
0	GC[+]	24073.2	0.2715	TT[+]	12324.3	0.1998	AA[+]	11833.5	0.1946	CG[+]	6535.8	0.1415
1	CG[+]	27665.5	0.2911	AC[+]	4121	0.1136	GT[+]	3967	0.1117	TA[+]	2204	0.0842
2	GC[-]	8072.1	-0.1572	TG[-]	6440.1	-0.1424	CA[-]	6434.2	-0.142	GT[-]	1547.7	-0.0698
3	TA[-]	7439.1	-0.1547	GC[-]	2552	-0.0884	AG[-]	526.3	-0.0406	CT[-]	462.2	-0.0382
	5° Significance			6° Significance			7° Significance			8° Significance		
0	CA[+]	4858.4	0.1234	TG[+]	4746.1	0.1222	AT[+]	3154.5	0.1008	GA[-]	1757.2	-0.0742
1	TT[+]	292.3	0.0308	AA[+]	279.4	0.0299	GA[+]	0.4	0.0012	TC[-]	4	-0.0034
2	AC[-]	1418.3	-0.0666	AT[-]	1245.2	-0.0633	TA[-]	876.4	-0.0531	CG[-]	167.9	-0.0227
3	CG[-]	195.2	-0.0245	CC[-]	53.4	-0.0128	GG[-]	38.2	-0.0108	AT[-]	5.7	-0.0043
	9° Significance			10° Significance			11° Significance			12° Significance		
0	TC[-]	1818.6	-0.0757	GG[-]	2513.4	-0.0877	CC[-]	2557.2	-0.0885	AC[-]	4003.1	-0.112
1	GC[-]	83.6	-0.016	AT[-]	805.5	-0.0509	CT[-]	2682.8	-0.0919	CC[-]	2756.3	-0.0919
2	AG[+]	245.4	0.0277	CT[+]	252.7	0.0282	TC[+]	2621.8	0.0909	GA[+]	2766.3	0.093
3	AA[+]	13.4	0.0066	TT[+]	25.4	0.0091	GT[+]	356.2	0.0335	AC[+]	470	0.0384
	13° Significance			14° Significance			15° Significance			16° Significance		
0	GT[-]	4147.5	-0.1143	AG[-]	9941.2	-0.1764	CT[-]	10128.4	-0.1786	TA[-]	18837.6	-0.2462
1	AG[-]	2778.5	-0.0932	GG[-]	2872.4	-0.0937	TG[-]	3769.3	-0.1089	CA[-]	4032	-0.1124
2	AA[+]	3295.6	0.1027	TT[+]	3501.9	0.1065	GG[+]	5864.6	0.1339	CC[+]	5873.5	0.1342
3	TC[+]	1361.2	0.0655	GA[+]	1481.5	0.0681	CA[+]	1844.5	0.076	TG[+]	1887.2	0.0771

Nomenclature as in Table 1

The difference between *SARS-CoV-2* and *E. coli* is evident at the first inspection and comparison between Table 1 and Table 3. Significances in *E. coli* are 50 or more times those of *SARS-CoV-2*. The χ^2_1 value of the first dinucleotide GC+ is 24,073.2, this implies a probability (Gaussian approximation; no as valid as in χ^2_9 test) less than $10^{-4,250}$. Dinucleotides with equal Par and a-Par complementary dinucleotides shows sharp differences. In *SARS-CoV-2* AA and TT, and GG and CC separate at the places of significances (Sigs). For example at Sep 0, AA is at 12° Sig while TT is at 9°, the difference in Sig is 3°, at Sep 1 the difference is 7° (3° and 10°, respectively), at Sep 2 is 3° (5° and 2°, respectively), and at Sep 3 it is 3° (12° and 15°, respectively). In *E. coli* AA and TT are contiguous (distance 1°) in the four Seps. GG and CC behave equally, while in *SARS-CoV-2* the distances between both are 2°, 4°, 3° and 4° for Seps 0, 1, 2 and 3, respectively, in *E. coli* they are 1°, 2°, 1° and 1°, respectively. It is evident that in a ds genome *(E. coli)* complementary dinucleotides behave different as in *SARS-CoV-2*, a ss genome, where complementary nucleic acid do not exist. (However, it is remarkable that the difference is not as big as expected; I shall discuss that). This occurs equally with the human genomes (C. Y. Valenzuela, submitted for publication). I remember that the expected mean difference in Sig profile is 5.667 and the standard deviation is 3.636. The following analysis in *E. coli* is similar to that presented in Table 2 for *SARS-CoV-2* on the distance of selective profiles from Par and a-Par dinucleotides to their Index; these Index pairs have different Par and a-Par dinucleotides. Table 4 presents it. In this Table the mean Sig distance I-a-Par is 1.06 (contiguous dinucleotides) while in I-Par is 6.06 (t test $P < 10^{-9}$) and in Sel distance the mean is 0.0679 for I-Par and for I-a-Par is 0.0013 ($P < 10^{-13}$). Selective profile of the 5'-3' a-Par dinucleotide is practically the same as the selective profile of the Index. The 5'-3' parallel dinucleotide differs from the Index. Remark that it is not possible to analyze the true 3′-5′ parallel dinucleotide, because we have only one strand from GenBank. The correction of this possible error is, at present OSA. Statistical tests where done as compliance of statistical norms, because values of differences in Sigs and co-sel are significant by the simple inspection. Differences in co-sel in the I-Par comparisons are at the order of tenths or hundredths, and the differences in I-a-Par are at the order of thousandths or ten thousandths. In Table 4 only absolute distances I-Par and I-a-Par are considered.

Table 5 for DNA and Table 6 for RNA viruses present the comparisons for 12 viruses with the same format of *SARS-CoV-2* and *E. coli*. Only mean, SD and t test probability appear. The single or double stranded DNA or RNA condition

Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500155

Table 4: *E. coli*. Index dinucleotides with different Par and a-Par pairs, four Separations

	Indexes			Par			a-Par			D Ind - Par		D Ind - a-Par	
Sep	dinuc[s]	co-sel	Sig	dinuc[s]	co-sel	Sig	dinuc[s]	co-sel	Sig	D Si	D Se	D Si	D Se
1	AG[-]	-0.1764	14	TC[-]	-0.0757	9	CT[-]	-0.1786	15	5	0.1007	1	0.0023
2	AG[-]	-0.0932	13	TC[-]	-0.0034	8	CT[-]	-0.0919	11	5	0.0899	2	0.0013
3	AG[+]	0.0277	9	TC[+]	0.0909	11	CT[+]	0.0282	10	2	0.0632	1	0.0005
4	AG[-]	-0.0406	3	TC[+]	0.0655	13	CT[-]	-0.0382	4	10	0.1061	1	0.0024
1	AC[-]	-0.112	12	TG[+]	0.1222	6	GT[-]	-0.1143	13	6	0.2342	1	0.0023
2	AC[+]	0.1136	2	TG[-]	-0.1089	15	GT[+]	0.1117	3	13	0.2225	1	0.0019
3	AC[-]	-0.0666	5	TG[-]	-0.1424	2	GT[-]	-0.0698	4	3	0.0757	1	0.0032
4	AC[+]	0.0384	12	TG[+]	0.0771	16	GT[+]	0.0335	11	4	0.0387	1	0.0049
1	TG[+]	0.1222	6	AC[-]	-0.112	12	CA[+]	0.1234	5	6	0.2342	1	0.0011
2	TG[-]	-0.1089	15	AC[+]	0.1136	2	CA[-]	-0.1124	16	13	0.2225	1	0.0034
3	TG[-]	-0.1424	2	AC[-]	-0.0666	5	CA[-]	-0.142	3	3	0.0757	1	0.0004
4	TG[+]	0.0771	16	AC[+]	0.0384	12	CA[+]	0.076	15	4	0.0387	1	0.0011
1	TC[-]	-0.0757	9	AG[-]	-0.1764	14	GA[-]	-0.0742	8	5	0.1007	1	0.0015
2	TC[-]	-0.0034	8	AG[-]	-0.0932	13	GA[+]	0.0012	7	5	0.0899	1	0.0045
3	TC[+]	0.0909	11	AG[+]	0.0277	9	GA[+]	0.093	12	2	0.0632	1	0.0022
4	TC[+]	0.0655	13	AG[-]	-0.0406	3	GA[+]	0.0681	14	10	0.1061	1	0.0026
1	GA[-]	-0.0742	8	CT[-]	-0.1786	15	TC[-]	-0.0757	9	7	0.1045	1	0.0015
2	GA[+]	0.0012	7	CT[-]	-0.0919	11	TC[-]	-0.0034	8	4	0.0931	1	0.0045
3	GA[+]	0.093	12	CT[+]	0.0282	10	TC[+]	0.0909	11	2	0.0648	1	0.0022
4	GA[+]	0.0681	14	CT[-]	-0.0382	4	TC[+]	0.0655	13	10	0.1062	1	0.0026
1	GT[-]	-0.1143	13	CA[+]	0.1234	5	AC[-]	-0.112	12	8	0.2376	1	0.0023
2	GT[+]	0.1117	3	CA[-]	-0.1124	16	AC[+]	0.1136	2	13	0.2241	1	0.0019
3	GT[-]	-0.0698	4	CA[-]	-0.142	3	AC[-]	-0.0666	5	1	0.0722	1	0.0032
4	GT[+]	0.0335	11	CA[+]	0.076	15	AC[+]	0.0384	12	4	0.0425	1	0.0049
1	CA[+]	0.1234	5	GT[-]	-0.1143	13	TG[+]	0.1222	6	8	0.2376	1	0.0011
2	CA[-]	-0.1124	16	GT[+]	0.1117	3	TG[-]	-0.1089	15	13	0.2241	1	0.0034
3	CA[-]	-0.142	3	GT[-]	-0.0698	4	TG[-]	-0.1424	2	1	0.0722	1	0.0004
4	CA[+]	0.076	15	GT[+]	0.0335	11	TG[+]	0.0771	16	4	0.0425	1	0.0011
1	CT[-]	-0.1786	15	GA[-]	-0.0742	8	AG[-]	-0.1764	14	7	0.1045	1	0.0023
2	CT[-]	-0.0919	11	GA[+]	0.0012	7	AG[-]	-0.0932	13	4	0.0931	2	0.0013
3	CT[+]	0.0282	10	GA[+]	0.093	12	AG[+]	0.0277	9	2	0.0648	1	0.0005
4	CT[-]	-0.0382	4	GA[+]	0.0681	14	AG[-]	-0.0406	3	10	0.1062	1	0.0024
M		-0.0178	9.41		-0.0178	4.41		-0.0178	9.41	6.06	0.1172	1.06	0.0022
SD		0.0938	4.48		0.0938	4.48		0.0938	4.48	3.67	0.0679	0.24	0.0013
P t												1.10E-10	7.00E-14

Nomenclature as in Table 2. D Ind-Par = absolute distance Index-Par; D Ind –a-Par = absolute distance Index-a-Par

Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500155

is in the first row of each set of viruses. Names are those used in current language and in GenBank and not necessarily scientific species nomenclature (excepting *SARS-CoV-2* and *E.coli*, see accession numbers of genomes).

These results were expected; ds DNA viruses showed differences between I-Par larger than differences between I-a-Par. Cytomegalovirus had the largest difference very probably because it is the largest virus. On the contrary, ss DNA viruses disagreed with this tendency. Penaeus monodon did not present significant difference between I-Par and I-a-Par

Table 5: Means, standard deviations and t test probability for distances Index-Par and Index-a-Par. DNA viruses

	Indexes	Par	a-Par	D Ind - Par		D Ind - a-Par		Ind Par	Ind a-Par
	co-sel	co-sel	co-sel	D Si	D Se	D Si	D Se	AB Se	AB Se
Double stranded DNA viruses									
Human adenovirus C strain KF268127.1; 35,931 bp									
Mean	-0.0414	-0.0414	-0.0414	4.7188	0	3.5	0	0.0752	0.0417
SD	0.0649	0.0649	0.0649	2.2533	0.1002	1.7854	0.0456	0.0662	0.0184
Prob t					P t Si =	0.0107		P t Se =	0.0043
Human cytomegalovirus; 229,354 bp									
Mean	-0.0304	-0.0304	-0.0304	4.1875	0	11,875	0	0.0491	0.0088
SD	0.058	0.058	0.058	2.7207	0.0582	0.3903	0.0127	0.0312	0.0091
Prob t					P t Si =	4.10E-08		P t Se =	1.70E-09
Escherichia phage 2; 167,318 bp									
Mean	0.0089	0.0089	0.0089	4.75	0	2.625	0	0.0555	0.0346
SD	0.0554	0.0554	0.0554	3.0923	0.0745	1.9325	0.0512	0.0498	0.0378
Prob t					P t Si =	0.0009		P t Se =	0.0339
His2 virus; 16,067 bp									
Mean	0.05	0.05	0.05	3.5313	0	1.5313	0	0.07576	0.0368
SD	0.1099	0.1099	0.1099	2.1935	0.1026	0.8286	0.0441	0.0692	0.0244
Prob t					P t Si =	6.30E-06		P t Se =	0.0022
Single stranded DNA Viruses									
Penaeus monodon hepandensovirus 4; 6,310 b									
Mean	0.0186	0.0186	0.0186	5.25	0	4.875	0	0.1034	0.1016
SD	0.1034	0.1034	0.1034	3.4369	0.1507	3.2763	0.1341	0.1096	0.0875
Prob t					P t Si =	0.3308		P t Se =	0.4731
Escherichia phage phi-X174; 5,386 b									
Mean	0.0038	0.0038	0.0038	3.4375	0	7.125	0	0.1038	0.1721
SD	0.1174	0.1174	0.1174	2.0606	0.1343	2.966	0.2016	0.0852	0.1051
Prob t					P t Si =	1.90E-07		P t Se =	0.0033
Phage M13 Enterobacteria; 6,407 b									
Mean	-0.0271	-0.0271	-0.0271	4.3125	0	5.9375	0	0.0952	0.1062
SD	0.0939	0.0939	0.0939	3.0765	0.1309	3.7992	0.1442	0.0899	0.0975
Prob t					P t Si =	0.0345		P t Se =	0.3234

Nomenclature as in Table 2

Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500155

Table 6: Means. standard deviations and t test probability for distances Index-Par and Index-a-Par. RNA viruses

	Indexes	Par	a-Par	D Ind - Par		D Ind - a-Par		Ind Par	Ind a-Par
	co-sel	co-sel	co-sel	D Si	D Se	D Si	D Se	AB Se	AB Se
Double stranded RNA Viruses									
Marmot picobirnaviridae; 4,579 bp									
Mean	0.0257	0.0257	0.0257	4.6875	0	4.0625	0	0.0837	0.0717
SD	0.0679	0.0679	0.0679	2.9521	0.1041	2.8607	0.0842	0.0619	0.0442
Prob t					P t Si =	0.2003		P t Se =	0.1907
Gentian Kobu virus; 22,711 bp									
Mean	0.0251	0.0251	0.0251	4.9375	0	2.3125	0	0.0758	0.0445
SD	0.0768	0.0768	0.0768	3.3998	0.1162	1.2609	0.0592	0.0881	0.039
Prob t					P t Si =	0.00005		P t Se =	0.0378
Single stranded RNA Viruses									
SARS-CoV-2 LR757998.1 Wuhan; 29,866 b									
Mean	0.0309	0.0309	0.0309	4.9375	0	5.3125	0	0.099	0.0844
SD	0.1099	0.1099	0.1099	2.3577	0.1196	3.0561	0.102	0.0672	0.0572
Prob t					P t Si =	0.2952		P t Se =	0.3626
Sindbis virus; 11,703 b									
Mean	0.0122	0.0122	0.0122	5.125	0	3.4375	0	0.081	0.0583
SD	0.0752	0.0752	0.0752	3.3889	0.0752	2.5365	0.1001	0.0588	0.0375
Prob t					P t Si =	0.0151		P t Se =	0.0376
HIV; 9,488 b									
Mean	-0.0021	-0.0021	-0.0021	6.875	0	4.6563	0	0.1994	0.0879
SD	0.112	0.112	0.112	3.8548	0.1904	2.9005	0.1111	0.118	0.0679
Prob t					P t Si =	0.0064		P t Se =	0.0073

Nomenclature as in Table 2

differences. However and unexpectedly Coliphage phi-X174 presented an inverse difference of the differences: I-Par difference is significant smaller than the I-a-Par difference.

Table 6 shows RNA viruses. In ds RNA viruses Marmot picobirnaviridae virus presented the expected larger difference in I-Par pair, but no significantly, while Gentian Kobu virus did. In this case, the small size of the viruses may explain the non-significant result. As expected the ss RNA virus *SARS-CoV-2* did not presented difference between I-Par and I-a-Par, while Sindbis virus presented a small but significant difference and HIV virus a large and significant difference.

I mentioned these distances to neutral evolution were periodical. It seems important to examine whether these viromes present periodicity because this is the first time the method applies to a set of viruses. *SARS-CoV-2* is periodical (C. Y. Valenzuela. submitted for publication) the 12 viromes and *E coli* are in Table 7. The pivot distance (the largest) of the period is in italics it follows and precedes smaller values. Asterisks show the element of the series that do not accomplish partially or completely this criterion. The first three values and sometimes the six ones are influenced by the first and second values that have regulatory dinucleotides. Those values less than 17 are not significant and do not alter necessarily the periodicity in this set of values. This is because the variance of the test is 2xdf = 18.0 and the SD is 4.243; random fluctuation near those values blurs the periodicity. Besides the χ^2_9 value is less significant in ss DNA or RNA and smaller viruses. Regardless these restrictions the periodicity is impressive. Moreover, we see that single or double stranded viruses that may incorporate to the host dsDNA present a clear periodicity similar as the host one without exceptions (HIV presents only one exception at Sep 23).

Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500155

Table 7: Periodicities of distances to evolutionary neutrality in viromes and *E. coli* genome

Sep	Adenov	Cytome	Ph 2	His2	phiX	hepand	ph M13	marmot	Gentian	SARS	Sindbis	HIV	*E. coli*
0	509.1	2981.2	1237.4	306.6	132	196.2	144.9	41.5	684.3	1236.9	107.4	464.5	123230
1	140.3	1581.4	888.9	318.1	92.8	46.9	61.9	32.3	83.3	75.7	95.8	83.7	58312.6
2	*424.2*	*3808.1*	*488.1*	*150.6*	*61*	*33.6*	*33.2*	*26.7*	*76.1*	*144*	*55.2*	*61.3*	*50623.7*
3	44.3	379.9	432.3	73.5	53.6	10.9	24	10.8	38.7	69.5	24.9	22.4	18711.1
4	13.4	308.8	254.3	66.6	25.8	18.7	20.9	11.9	14.4	88.8	5.8	34.7	5443.7
5	*413.2*	*2660.8*	*392*	*59.3**	*47.6*	*18.4**	*32.3*	*22.2*	*24.7*	*72.1**	*15.7**	*23.0**	*8927.9*
6	22	112.8	292.8	46	18.7	8.4	20.4	5.4	18.7	72.7	16.2	31.2	1363.4
7	23.6	61.6	247.3	78.9	29.9	17.7	11.4	7.8	8.2	44.9	16.3	11.4	2052.5
8	*356.3*	*2419.4*	*523.6*	*120*	*87.2*	*30.8*	*50.8*	*8.5**	*42.9*	*127.6*	*12.0**	*45.9*	*10597*
9	13.4	47.1	320.6	33.4	28.3	12.6	20.9	12.1	15.6	56	3.3	10	2138.6
10	9.3	112	309.2	17.3	40	13.7	26.5	15.5	29.4	40.4	14.7	18.9	2346
11	*311*	*2530.4*	*583.1*	*95.1*	*68.4*	*29.7*	*60.6*	*20.6*	*53.2*	*131.6*	*24.3*	*47.2*	*16223.7*
12	25.8	107.7	243.6	44.5	44.8	27	32.4	18.5	40.8	85.8	18.1	35.2	1986.7
13	13.5	51.7	213.1	36.4	28.1	6.1	26.1	13.1	12.4	70.3	19.8	15.1	1676.3
14	*278.7*	*1978.6*	*468.1*	*64.6*	*47.9*	*32.8*	*72*	*18*	*9.4**	*101.1*	*13.4**	*23.9*	*9563.7*
15	11.8	69.4	157.9	15.4	38.4	7.7	28.5	14.9	29.9	57.1	7.2	14.8	1647.4
16	18.6	88.9	200.7	25.5	46.9	15.5	35.2	9	25.9	61.4	31.7	17.2	2235.3
17	*287.6*	*1989.8*	*423*	*54*	*49.6*	*22.6*	*52.4*	*12.4*	*42.7*	*86.4*	*18.1**	*36.8*	*9725.6*
18	3.4	38.2	155.4	16.8	25.7	9.2	39.9	7.1	15.5	46.5	9.3	10	1566.9
19	18.3	74.4	176.4	25.3	26.8	11.8	44.1	9.7	13.4	57.3	12.5	28	1940.5
20	*254.3*	*1935*	*534.5*	*70.3*	*69.6*	*34.8*	*54.7*	*20.8*	*30*	*93.9*	*21.8*	*34.8*	*13582.1*
21	24.4	82.5	183.1	28.9	20.7	17.4	20.8	18.2	27.5	87.4	11.2	7.4	2220.1
22	12	61.3	235.3	23	29.2	14.4	31.8	30.7	26.4	63.3	20.8	18	1661
23	*225.4*	*1821.7*	*500.8*	*78.9*	*42.6*	*39.3*	*61.9*	*14.8**	*25.6**	*124.8*	*31.3*	*15.6**	*12420.6*
24	22.3	79.9	153.9	9.4	27	16.6	37.1	17	16.4	67.8	15.9	17.7	2320.9
25	15.1	57.1	190.1	12.5	31.3	19.6	20.4	7.3	17.1	54.7	20.9	8.9	1684.5
26	*234.6*	*1651.8*	*482*	*68*	*42.6*	*37.9*	*58.6*	*25*	*20.2**	*74.1*	*18.7**	*18.4*	*10400.1*
27	24.2	79.8	209	10.7	20.9	11.8	28.7	17.1	27.5	59.1	6.7	13.6	2078.5
28	22.4	51.8	240.3	14.6	27	11.3	20.3	18.1	5	45	15.5	9.1	1926.3
29	*189.4*	*1805.1*	*456.9*	*72.5*	*59.4*	*10.6**	*55.1*	*12.5**	*34.2*	*64.2*	*18.5*	*22.4*	*10553.1*
30	23.8	45.4	206	25.6	28.1	13.4	17.1	13.6	4.7	42.5	9.7	11.2	2215.2
31	15.7	44.4	180.8	19.9	27.6	17.4	14.4	15.9	14.1	40.2	6.9	7	1817.7
32	*172.3*	*1593*	*534.8*	*62.2*	*65*	*15.8**	*32.3*	*42.6*	*35.4*	*124.9*	*11.3*	*44.3*	*12264.9*

In italics the largest distance among the triplet of the period measured by the total χ29. For names of viruses see Genomes. *figure that does not fit partially or completely the condition of the largest distance

Valenzuela CY., Arch Microbiol Immunology 2024
DOI:10.26502/ami.936500155

Discussion

This is the first time this statistical tool is used to search for the single or double stranded nature of RNA and DNA viruses and their implications on evolution. Results are according to the theoretical expected behavior of ds or ss viruses with some important exceptions. Escherichia phage phi X174, a ss DNA virus presented a more similar selective profile between Indices and Par pairs than between Indices and a-Par pairs. A plausible explanation is the fact that this virus replicates by using a negative DNA strand after constructing a dsDNA [6, 7]. The Marmot Picobirnaviridae a dsRNA virus [8] shows a non-significant larger selective profile difference between I-Par than I-a-Par, this is perhaps it is a very small virus (4.579 bp). However, Gentian Kobu sho-associated virus also a dsRNA virus [9] according to the Oxford database shows a big significant difference as expected; however, other authors present it as an ssRNA virus [10]. According to my analysis, this virus behaves as an evolutionary or replicative dsRNA virus. Arguments in favor or against are present in these analyses and Tables but their study is OSA. In the group of ssRNA, viruses *SARS-CoV-2* did not presented differences between I-Par and I-a-Par pairs selective profiles, while Sindbis virus and HIV showed a significant difference where I-Par differences were larger than I-a-Par differences. This was more significant in HIV than in Sindbis virus perhaps for HIV is a retrovirus that exists as a double stranded DNA segment incorporated in the host DNA; however, Sindbis virus needs ds RNA phase in its replication cycle [11]. These data, methods and analyses were presented in international and national congresses [12, 13].

The present analyses indicate that the classification in single or double stranded DNA or RNA virus based on the virion physical constitution is useful but it does not include evolutionarily functional double or single stranded constitution mostly produced at the replicative or transcription stage of the viral vital cycle. This new conceptualization seems useful and offers a different understanding of the viral behavior and evolution, which is also useful in epidemiological studies. However, to analyze more precisely these hypotheses need a deep research (OSA).

As for the impressive periodicities, two hypotheses may be advanced. The structural hypothesis proposes these periodicities of the distance to neutrality originated in a basic property of nucleic acids that has different structural stability in triplets of nucleotides; triplets for the genetic code are a result of adaptation to this primary periodicity. The functional hypothesis proposes the periodicity has evolved as the genetic code installed. During and after this installation the genome segments (coding or non-coding for proteins) optimize an organization that dances at the rhythm of trinucleotides. Arguments in favor or against of these hypotheses are OSA.

Conclusion

The comparisons of differences in selective profiles between an index (I) dinucleotide with its complementary parallel (Par) and anti-parallel (a-Par) dinucleotide informs on the double or single stranded (ds or ss) nature of the nucleic acid of any organism. The foundation of this discrimination happens because ds DNA or ds RNA have parallel and antiparallel strands that evolve together, but ss DNA or ss RNA do not. The single and double stranded evolutionary functional constitution of viruses appears as a useful tool to understand the life cycle of viruses and their epidemiological properties. The article demonstrates that this methods works without exceptions in the 13 examined genomes.

Ethics approval and consent to participate

Not applicable.

Consent for publication: Not Applicable.

Availability of data and Materials: The genomes' information is available in GenBank as cited. Any geneticist, student or statistician who knows the current informatics languages can construct these computer programs and perform these analyses.

Competing Interests: I do not have any competing interest.

Funding: This work and the article did not have extraordinary funds.

Author's Contributions: CYV carried out all the work described in this article.

Acknowledgements: To Francesco M Scudo and Ching Chung Li my first teachers of population genetics. My students Franccesca Soriano, Francisco Rosas and Matías Rozas helped me in processing data.

References

1. Valenzuela CY. Non-random pre-transcriptional evolution in HIV-1. A refutation of the foundational conditions for neutral evolution. Genet Mol Biol 32 (2009): 159-169.
2. Valenzuela CY. Internucleotide correlation and nucleotide periodicity in Drosophila mtDNA: New evidence for panselective evolution. Biol Res 43 (2010): 481-486.
3. Valenzuela CY. Heterogeneous periodicity of drosophila mtDNA: new refutations of neutral and nearly neutral evolution. Biol Res 44 (2011): 283-293.
4. Valenzuela CY. The structure of selective dinucleotide interactions and periodicities in *D melanogaster* mtDNA. Biol Res 47 (2014):1-12.
5. Valenzuela CY. Selective intra-dinucleotide interactions