www.directivepublications.org

Journal of DNA Research

ISSN 3068-3831

DIRECTIVE PUBLICATIONS

Research Article

Selective Values Of Dinucleotides Indicate The Evolutionary Double Or Single Stranded Nature Of An Organism's Genome.

Carlos Y Valenzuela[1] **M.D., Ph.D.**

[1] *Interdisciplinary Nucleus of Biology and Genetics, ICBM, Faculty of Medicine, University of Chile, Santiago, Chile.* **Email:** cvalenzu@med.uchile.cl.

Abstract

I intend to confirm the use of selection coefficients of dinucleotides whose bases are separated by 0, 1, 2… K nucleotide sites to determine whether a genome is single or double stranded DNA or RNA. The previous studies determined that prokaryotes and eukaryotes have categorically double stranded DNA, but viruses had and intermediate composition of double and single stranded DNA or RNA. These studies refer not to the state of the nucleic acid of viruses in the virion particle but to the state of the viruses along their life cycle. Previous studies used the selective profile that include the distance to (selective) neutrality, the sign and value of selection and the place in the scale of selective values. Now, I use the simple correlation between the selection coefficients of dinucleotides found in each set of separation between both bases of the dinucleotides to determine the stranded conditions of nucleic acids. I examined human, archaea, bacteria and viruses' nucleic acids. Results confirm that Eukaryote and prokaryote nucleic acids are double stranded and viruses are in between double and single stranded nucleic acid. However, the large *H. cytomegalovirus* (229,000 bp) was detected as a double stranded virus.

Keywords: selective evolution; dinucleotides; single double stranded DNA-RNA; correlation of selective values; strand condition detection.

INTRODUCTION

Studying neutral and selective evolution by the analyses of the distribution of bases in genomes and in dinucleotides whose bases were separated by 0, 1, 2, K nucleotides sites, I discovered that neutral evolution was impossible [1-7]. The studies of dinucleotides constructed a matrix where columns describe the distance to randomness of the distribution of both bases of a dinucleotide (measured by a chi-square test) and selective values, and rows the number of nucleotide sites separations between both bases of a dinucleotide. The study of selective profiles of the 16 possible dinucleotides shows in procaryotes (double stranded DNA) a great selective similitude between an Index (Ind) dinucleotide with its 5'-3' Antiparallel (a-Par) dinucleotide, but a large difference between this Index and its Parallel (Par) dinucleotide [8,9]. In single stranded virus there is a small or none similitude between the index dinucleotide and their a-Par dinucleotide and a smaller similitude with its Par dinucleotide. Now I present a new analysis founded in a matrix where columns show selective values of the 16 dinucleotides and rows the number of site separations.

GENOMES, METHOD AND RATIONALE

Genomes and Method

As in previous studies [2,9], we collected from Genbank genomes of the Human chromosome 21, a bacterium, one Archaea and some double and single stranded viruses. In the present Study genomes are: H. sapiens Chromosome 21 (HCh21) NC_000021.9 GRCh38.p7, 46,709,983 bp, double stranded (ds) *DNA; M. Smithii* (Archaea) NC_009515.1, strain ATCC 35061, 1,853,160 bp, dsDNA; Human cytomegalovirus *(also Human herpesvirus 5*, HHV-5), X17403.0 strain AD169, 229,354 bp, dsDNA; *Phi X 174* phage of *E. coli*, NC_001422, 5,386 b, ssDNA; *SARS-CoV-2* virus, LR757998.1Wuhan, 29,903 b, ssRNA: HIV-1 MN691959.1, isolate ACH2-NFLMDA13 B1 (USA), 9,493 bp, ssRNA; *E. coli* genome, NZ_CP124398.1, strain AVSO967, 5,097,505 bp, dsDNA (processed for this program, I could not find this reference in the present information).

***Corresponding Author: Carlos Y Valenzuela.** Núcleo Interdisciplinario De Biología Y Genética. Instituto De Ciencias Biomédicas (Icbm), Facultad De Medicina, Universidad De Chile. Independencia 1027, Código Postal 8380453, Independencia, Chile. **Phone:** (56-2) 29786302. **Email:** cvalenzu@med.uchile.cl,
Orcid ID: 0000-0002-2235-1050.

Received: 28-Jan-2026, Manuscript No. JODNAR- 5387 **; Editor Assigned:** 29-Jan-2026 **; Reviewed:** 10-Feb-2026, QC No. JODNAR- 5387 **;**
Published: 18-Feb-2026.**DOI:** 10.52338/jodnar.2026.5387.
Citation: Carlos Y Valenzuela. Selective Values Of Dinucleotides Indicate The Evolutionary Double Or Single Stranded Nature Of An Organism's Genome. Journal of DNA Research. 2026 February; 15(1). doi: 10.52338/jodnar.2026.5387.

Note: the number of nucleotides is approximate, it varies with strains and time since some variants are replaced; the collection of these genomes began more than 15 years ago. Also, there are few clerical errors (in relation to previous articles). The number of actual bases that are analyzed is less than the total number of sites that GenBank provides (sites without assigned bases). However, if the number and the chi-square value may vary, the structure of data is maintained almost equal. In single stranded nucleic acids, I used b instead of bp. For these genomes I constructed the set of dinucleotides separated by 0, 1, 2, ...100 dinucleotide sites. In these set of dinucleotides, I estimated the distance to randomness of the distribution of their both bases by a chi-square (χ^2_9) test with 9 degrees of freedom (d.f.) for the total value, and with a χ^2_1 test for each dinucleotide. This is the result of the matrix of four columns and four rows given by the four nucleotides: adenine (A), thymine (T), guanine (G) and cytosine (C). The selective value for each dinucleotide is given by the expected (random) and observed values $(Obs_i\text{-}Exp_i)/Exp_i$ where i goes from 1 to 16, the ordered dinucleotides are: AA, AT, AG, AC, TA, TT, TG, TC, GA, GT, GG, GC, CA, CT, CG, and CC. This conforms a table where columns are the dinucleotides with their respective selective values and their distances to randomness measure by the χ^2_1 (described in **Table 1** but not analyzed here) and rows given by the separation of bases by K nucleotide sites from 0 to 100 as shown in **Table 1**. The χ^2_1 is obtained by $(Obs_i\text{-}Exp_i)^2/Exp_i$ where i goes from 1 to 16. This table was shortened and shows, in columns, the first 2 dinucleotides and the last 2 dinucleotides, and in rows the first 16 separations and the last 9 separations.

Rationale

Dinucleotides, in double stranded DNA or RNA, present obligatory complementary and concatenated behavior with their related dinucleotides given the double stranded and the complementary constitution of nucleic acid. An index dinucleotide in the 5′-3′ strand have complementary parallel (Par) and anti-parallel (a-Par) dinucleotides in the complementary strand and a reverse (Rev) dinucleotide in the same strand. Thus, the coefficient of selection of, for example, AG as the index dinucleotide with 5′-3′ sense (the published strand) is expected correlated with its Par TC (3′-5′ sense) or the a-Par CT (5′-3′ sense) dinucleotides, respectively, in the complementary strand of the original DNA. These dinucleotides are tested in the index strand only in the 5′-3′ sense. In the single stranded DNA or RNA none of these relationships are possible. Thus, it is necessary to restrict the comparisons to those more discriminant ones. I presented in previous articles three categories of dinucleotides in double stranded DNA or RNA. I) those (Index) dinucleotides with the same Par and a-Par dinucleotides: AA-TT and GG-CC = 4 dinucleotides; II) those dinucleotides with a different Par and equal a-Par: AT (or vice versa TA) with TA (Par) and AT (a-Par), and GC (or vice versa CG) with GC (Par) and CG (a-Par) = 4 dinucleotides; those dinucleotides with different Par and a-Par dinucleotides (the reader will easily construct their Par and a-Par dinucleotides): AG, AC, TG, TC, GA, GT, CA, CT = 8 dinucleotides. The rationale of the following analyses is founded in the assumption of similarity in selection of dinucleotides according to the complementary rule of bases and the direction of the synthesis 5′- 3′ or 3′-5′ of nucleic acids. I assume that both strands are equally selected. Given that we use a 5′-3′ strand of DNA (index dinucleotides from GenBank) or RNA, if mutation occurs preferentially in the transcription process and is selected in the following life processes it is expected that either the 5′-3′ (A-Par) dinucleotide or the 3′-5′ (Par) dinucleotide to the index (5′-3′) has more similarities with the index dinucleotide. It is expected that in single stranded viruses this relationship be less evident or does not exist. I must prevent the reader this study is not based in the single or double stranded condition of viruses in their virion state, which is probably the weakest evolutionary instance. The rationale is based in the idea that viruses exist, during their reproductive cycle, either in the single or double stranded form. The typification here of a single or double stranded virus depends on their relative existence of these two stranded forms during the life cycle of the virus. The time is evolutionary time because some stages have a larger density of evolutionary events such as mutation or selection. Double stranded DNA organisms have single stranded stages, but their existence is so ephemeral that they do not have known evolutionary transcendence.

RESULTS

The nuclear data were collected in the matrix of the 16 dinucleotides where columns have the dinucleotide with its deviation from the random expectancy of the distribution of the two bases (χ^2_1), and its selection coefficient. The description is presented in **Table 1** for H. cytomegalovirus, a big double stranded DNA virus with near 230,000 bp, that can insert in the host DNA. In this table the initial 2 dinucleotides and the terminal 2 dinucleotides, and the initial 15 and the terminal 9 separations are described. The minimal significant value for a χ^2_9 (all the dinucleotides) is 16.92 ($P = 0.05$) and it is 3.84 for the individual (dinucleotide) χ^2_1, thus a great deal of tests is significant. Each dinucleotide has 100 selective values corresponding to the 100 separations (from 0 to 99). It is remarkable the periodicity of the value of the χ^2_9 and also that of the individual χ^2_1 values. Even though only four dinucleotides are presented they are not completely homogeneous for their periodicity. This periodicity continues until separation 99 (not the aim of this article).

Table 1. Selective parameters according to separation in H. cytomegalovirus. First two and last two dinucleotides

Adenine-Adenine					Adenine-Thymine			…	Cytosine-Guanine			Cytosine-Cytosine		
Sep	χ^2_9	Din	χ^2_1	Sel	Din	χ^2_1	Sel	…	Din	χ^2_1	Sel	Din	χ^2_1	Sel
0	2981.2	AA[+]	183.8	0.131	AT[-]	83.2	-0.089	…	CG[+]	666.97	0.189	CC[-]	348.97	-0.138
1	1581.4	AA[+]	419.7	0.198	AT[-]	515.0	-0.221	…	CG[+]	32.72	0.042	CC[-]	0.04	-0.001
2	3808.1	AA[+]	591.0	0.235	AT[+]	0.2	0.004	…	CG[-]	9.28	-0.022	CC[+]	656.13	0.189
3	379.9	AA[+]	50.1	0.069	AT[-]	0.9	-0.009	…	CG[+]	94.01	0.071	CC[-]	22.15	-0.035
4	308.8	AA[+]	5.0	0.022	AT[-]	127.8	-0.110	…	CG[-]	12.5	-0.026	CC[-]	0.67	-0.006
5	2660.8	AA[+]	365.7	0.185	AT[+]	19.3	0.043	…	CG[-]	1.04	-0.007	CC[+]	389.65	0.146
6	112.8	AA[+]	4.9	0.022	AT[+]	1.5	0.012	…	CG[+]	1.86	0.010	CC[+]	11.86	0.025
7	61.6	AA[+]	2.1	0.014	AT[-]	12.7	-0.035	…	CG[-]	3.84	-0.014	CC[+]	5.92	0.018
8	2419.4	AA[+]	306.7	0.170	AT[+]	31.5	0.055	…	CG[+]	0.59	0.006	CC[+]	364.06	0.141
9	47.1	AA[+]	1.5	0.012	AT[+]	2.8	0.016	…	CG[+]	4.88	0.016	CC[+]	0.13	0.003
10	112.0	AA[+]	2.1	0.014	AT[-]	15.3	-0.038	…	CG[-]	11.92	-0.025	CC[+]	2.42	0.011
11	2530.4	AA[+]	341.2	0.179	AT[+]	36.6	0.059	…	CG[-]	1.18	-0.008	CC[+]	348.95	0.138
12	107.7	AA[+]	20.4	0.044	AT[+]	0.0	0.001	…	CG[+]	4.54	0.016	CC[+]	3.4	0.014
13	51.7	AA[+]	8.2	0.028	AT[-]	11.9	-0.034	…	CG[-]	0.92	-0.007	CC[+]	0.18	0.003
14	1978.6	AA[+]	258.8	0.156	AT[+]	53.2	0.071	…	CG[+]	0.69	0.006	CC[+]	266.25	0.120
.	.	.	.	.	.	.	.	…	.	.	.	.	.	.
.	.	.	.	.	.	.	.	…	.	.	.	.	.	.
91	87.3	AA[+]	0.1	0.003	AT[-]	4.2	-0.020	…	CG[-]	22.13	-0.034	CC[+]	2.13	0.011
92	1174.2	AA[+]	167.7	0.125	AT[+]	20.2	0.044	…	CG[+]	11.5	0.025	CC[+]	114.31	0.079
93	75.7	AA[+]	1.9	0.013	AT[+]	5.9	0.024	…	CG[+]	7.59	0.020	CC[+]	2.75	0.012
94	43.9	AA[+]	1.0	0.010	AT[-]	15.0	-0.038	…	CG[-]	2.37	-0.011	CC[+]	0.1	0.002
95	1052.8	AA[+]	125.4	0.108	AT[+]	46.0	0.066	…	CG[+]	10.75	0.024	CC[+]	86.99	0.069
96	69.3	AA[+]	1.7	0.013	AT[+]	2.8	0.016	…	CG[+]	0.53	0.005	CC[+]	2.01	0.010
97	53.9	AA[+]	0.0	0.002	AT[-]	8.4	-0.028	…	CG[-]	0.65	-0.006	CC[-]	0.2	-0.003
98	1104.4	AA[+]	147.9	0.118	AT[+]	28.1	0.052	…	CG[+]	11.06	0.024	CC[+]	94.48	0.072
99	57.9	AA[+]	0.3	0.005	AT[+]	3.5	0.018	…	CG[+]	5.21	0.017	CC[+]	2.41	0.011

Sep = separation; χ^2_9 = total chi square with 9 d.f.; Din = dinucleotide with its sign of selection; Sel = selection coefficient.
χ^2_1 = individual (for each nucleotide) chi square with 1 d.f.

The following analysis was by means of the Pearson's correlation coefficient between the selection coefficients of dinucleotide couples: AA-AA (r = 1.0); AA-AT; AA-AG; AA-AC; … CT-CG; CT-CC; CG-CC; CC-CC (r = 1.0) = 128 pairs of correlations between selection coefficients of dinucleotides from which 16 pairs are of the same dinucleotide (with correlation 1.0). For example (**Table 1**), excluding the first five separations, the selection coefficients of AA are highly correlated with the selection coefficients of CC; this is directly seen by the chi-square values that present equal periodicities, this indicates a high correlation as it is seen in **Table 2**.

Table 2. correlations between selection coefficients of dinucleotides. H. cytomegalovirus

	AA	AT	AG	AC	TA	TT	TG	TC	GA	GT	GG	GC	CA	CT	CG	CC
AA	1,000	0,460	-0,791	-0,868	0,542	0,991	-0,939	-0,792	-0,798	-0,848	0,814	0,373	-0,920	-0,744	0,403	0,803
AT		1,000	-0,781	-0,718	0,582	0,458	-0,619	-0,472	-0,448	-0,728	0,700	0,031	-0,639	-0,744	0,219	0,700
AG			1,000	0,720	-0,361	-0,779	0,766	0,521	0,519	0,722	-0,687	-0,137	0,772	0,974	-0,642	-0,676
AC				1,000	-0,817	-0,865	0,957	0,873	0,861	0,983	-0,951	-0,339	0,946	0,712	-0,096	-0,947
TA					1,000	0,537	-0,728	-0,881	-0,863	-0,826	0,806	0,438	-0,719	-0,317	-0,269	0,785
TT						1,000	-0,940	-0,799	-0,789	-0,854	0,817	0,364	-0,914	-0,794	0,387	0,809
TG							1,000	0,835	0,829	0,951	-0,940	-0,278	0,983	0,757	-0,214	-0,928
TC								1,000	0,975	0,874	-0,814	-0,625	0,808	0,509	-0,006	-0,795
GA									1,000	0,843	-0,788	-0,667	0,777	0,509	-0,045	-0,754
GT										1,000	-0,963	-0,313	0,947	0,686	-0,066	-0,949
GG											1,000	0,114	-0,940	-0,653	-0,051	0,992

GC												1,000	-0,210	-0,151	0,260	0,072
CA													1,000	0,743	-0,200	-0,940
CT														1,000	-0,668	-0,658
CG															1,000	-0,060
CC																1,000

These correlations are presented in **Table 2** for *H. cytomegalovirus*. With 100 pairs, correlations are significantly different from 0.0 (index r), for a two tailed z test ($P \leq 0.05$), when r = ± 0.275. The comparison between two correlations is significant depending on the values of r. If one correlation is ± 0.5, significant correlations with larger absolute value of r, are with $r \geq 0.681$ for positive r and ≤ -0.681 for negative r; and with smaller absolute value of r with $r \leq 0.261$ for positive r and $r \geq -0.261$ for negative r. If the index correlation to compare is r = ± 0.75, significant values are found with $r \geq 0.85$ and $r \leq 0.599$ for positive r, and with $r \leq -0.85$ and $r \geq -0.599$ for negative r. With index r = 0.25 significances occur with $r \geq 0.495$ and $r \leq -0.026$, for larger and smaller r values, respectively. With index r -0.25 significant values occur with $r \leq -0.495$ and $r \geq 0.026$. These figures allow interpret **Table 2**.

Let us remember that double stranded DNA are: *M. smithi* (archaea), *H. cytomegalovirus* (virus), *H. sapiens* and *E. coli* (bacterium). The phage of *E. coli*, *Phi X* 174 is a single stranded DNA virus that may integrate the host DNA after a transformation in a double stranded DNA; *SARS-CoV-2* is a single stranded lytic RNA virus and *HIV-1* is a single stranded RNA retrovirus that can integrate into the host DNA.

The most critical correlations are presented in **Table 3. I**) Among the dinucleotides with equal Par and a-Par dinucleotides, AA-TT and GG-CC, the discriminant power of the test is evident. Double stranded DNA present a very high correlation (over 0.9) for both type of dinucleotide pairs. The other correlations are significant with lower significance than those found in organisms with dsDNA with the exception of *Phi X 174* for AA-TT. II) Among dinucleotides with different Par and a-Par, with a-Par equal to the index dinucleotide (AT-TA; GC-CG) dsDNA presented higher correlations (in absolute values) with the exception of *SARS-CoV-2* in both AT-TA (0.66) and GC-CG (-0.55) pairs. HIV showed non-significant correlations. III) Among dinucleotides with Par and a-Par different from the Index dinucleotide. In general, with few exceptions a-Par dinucleotides shows higher correlations than Par dinucleotides with index dinucleotides. The values of correlations among ssDNA or ssRNA are significatively lower than correlations among dsDNA with the exception of Par-Index correlation (AG-TC, and GA-CT) in Phi X 174. It is remarkable the positive and high correlations of H. Cytomegalovirus either with the Par or a-Par dinucleotides. It is interesting that *Phi X 174* presented higher correlations between Par-Index pairs and negative correlations between a-Par-Index pairs; this contrasts with the dsDNA organisms particularly with *M. smithi* (at its left side).

Table 3. Critical selective correlations in double and single stranded nucleic acids.

	M. smithi	**Coli pX174**	**H Cytomeg**	***SARS-CoV-2***	***HIV-1***	***H. sapiens***	***E. coli***
	dsDNA	**ssDNA**	**dsDNA**	**ssRNA**	**ssRNA**	**dsDNA**	**dsDNA**
Equal Parallel and Anti-Parallel Dinucleotide; both different to Index dinucleotide							
AA-TT	0.98245	0.85556	0.99069	0.38501	0.35952	0.94605	0.99899
GG-CC	0.99934	0.40174	0.99170				
	0.69554	0.46313	0.91909	0.99927			
Different Parallel and Anti-Parallel Dinucleotide; Anti-Parallel equal to Index dinucleotide							
AT-TA	0.54821	0.30705	0.58173	0.66171	-0.25841	0.78944	-0.76046
GC-CG	-0.70511	-0.06223	-0.05052	-0.55469	-0.24668	0.08654	0.31400
Index, Parallel and Anti-Parallel Dinucleotides are different							
AG-CT	0.94486	-0.37320	0.97364	-0.32252	0.42950	0.87042	0.99826
AG-TC	-0.71539	0.56141	0.52069	0.45469	-0.45748	0.35616	-0.09087
AC-GT	0.98432	-0.38395	0.98330	0.04884	0.35999	0.86769	0.99878
AC-TG	0.57925	-0.06202	0.95679	0.28188	0.34009	-0.31221	0.20667
TG-CA	0.98664	-0.28970	0.98287	0.14781	-0.16202	0.95469	0.99893
TC-GA	0.95956	-0.32028	0.97490	-0.33411	0.32342	0.77378	0.99767
GA-CT	-0.76752	0.63999	0.50948	0.38763	-0.12491	0.37300	-0.09300
GT-CA	0.59263	-0.13609	0.94696	0.05596	0.03653	-0.25719	0.40646

Coli pX174 = phi X174 phage of E. coli; H Cytomeg = H. cytomegalovirus.

I found unexplainable correlations, among them, those of non-related bases by complementarity. **Table 4** shows these correlations. The AA-TT and GG-CC are repeated from **Table 3** to contrast these comparisons. It is remarkable that all these correlations are positive.

Table 4. Correlations between dinucleotides of one base: AA, TT, GG, CC

	M. smithi	**Coli fX174**	**H Cytomeg**	***SARS-CoV-2***	**HIV**	***H. sapiens***	***E. coli***
	dsDNA	**ssDNA**	**dsDNA**	**ssRNA**	**ssRNA**	**dsDNA**	**dsDNA**
AA-TT	0.98245	0.85556	0.99069	0.38501	0.35952	0.94605	0.99899
AA-GG	0.96572	0.74596	0.81403	0.29006	0.42631	0.76498	0.71051
AA-CC	0.96787	0.36017	0.80274	0.19690	0.23965	0.77114	0.70991
TT-GG	0.91177	0.88266	0.81723	0.95409	0.45717	0.73310	0.70405
TT-CC	0.91677	0.46229	0.80887	0.70175	0.41910	0.79453	0.70325
GG-CC	0.99934	0.40174	0.99170	0.69554	0.46313	0.91909	0.99927

Nomenclature as in Table 3.

It is remarkable that all correlations in double stranded DNA are over 0.70. Several correlations among single stranded viruses, all in HIV, are under 0.5. Correlations AA-CC are very critical, while dsDNA show correlations over 0.7, ssDNA or ssRNA show correlations under 0.37; all these differences are highly significant.

DISCUSSION

Data show a clear discrimination between double and single stranded DNA or RNA. It is also clear that single stranded viral DNA or RNA are not so single as they are found in the virion particle. Along their life cycle these viruses exist in double or single stranded forms that undergo mutational or selective events in different frequencies, and very probable, mostly during the replication process. During replication they are mostly in double stranded state. These results confirm those found in previous articles [8,9]. It is remarkable the results found in Index dinucleotides with different Par and a-Par complementary dinucleotides that is in agreement with the 5'-3' sense of replication. In present results the a-Par dinucleotide has a high correlation with its Index dinucleotide as it was found in previous studies [8,9]. However, in those studies and in this one, *Phi X 174* shows that Par dinucleotides are selectively more similar to their Index dinucleotides than the a-Par dinucleotide. This is very probably due to the existence of a negative or reverse strand needed in replication in this bacteriophage of *E. coli* [10]. The biochemical and molecular study is out of the scope of this article but it is in the current virological knowledge. These studies may be a useful tool for molecular studies of viral replication and evolution and open a big road of research in several fields of biology. By the way, one of the most significant selection coefficients was found in *H cytomegalovirus*. It was CpG with positive selection. This is, contradictory, because CpG was negatively selected in *H sapiens* and its parasitic viruses *HIV-1* and *SARS-CoV-2*, and in the archaea *M. smithi* that is a commensal of *H. sapiens*. The negative selection is very probable due to their epigenetic condition of shutting down DNA [11]. In H cytomegalovirus it seems positive CpG protects the virus from the host immune system (out of the scope of this article) [12]. This is another property of this analysis, to discover functional evolutionary hidden relationships. Genetic drift or any random process are not mentioned since the demonstration that the found deviations are incompatible with any drift process. In the literature there is a confusion between contingent and random processes. According to our results contingent processes that are always causal processes do exist; random processes, as multicausal processes leading to fluctuation of frequencies of genetic forms of unknown origin, do exist in our analyses or models, but at last all of them are produced by precise causal events. As the synthetic evolutionary theory proposes drift is not a directional (causal) factor of evolution [5-7,9,13]

CONCLUSIONS

These analyses confirm that the study of selection coefficients of dinucleotides whose bases are separated by 0, 1, 2 … K nucleotide sites allows to determine the proportion of the existence of a nucleic acid in single or double stranded condition along its life cycle. The analyses open a great deal of new properties of genomes hidden to previous current analyses.

Acknowledgments

I am indebted to my professors of population genetics Francesco Scudo and Ching Chung Li, and to my colleagues.

Funding

No extraordinary funds were used in this research

Conflicts of Interest

The author declares no conflicts of interest

Availability of data and methods

Genomes are freely found in GenBank, even though most of the original genomes may not be found at present, because they were replaced by actualized genomes. The methods and analyses can be generated by any researcher or student that know statistics, genomes, and a computational language.

REFERENCES

1. Valenzuela CY, Santos JL. A model of complete random molecular evolution by recurrent mutation. Biol Res. 1996; 29: 203-212. PMID: 9278711
2. Valenzuela CY. Non-random pre-transcriptional evolution in HIV-1. A refutation of the foundational conditions for neutral evolution. Genet Mol Biol. 2009; 32: 159-169. PMID: 21637663.https://doi.org/10.1590/S1415-47572009005000025
3. Valenzuela CY, Flores SV, Cisternas J. Fixations of the HIV-1 env gene refute neutralism: new evidence for pan-selective evolution. Biol Res. 2010; 43: 149-163. https://dx.doi.org/10.4067/S0716-97602010000200002
4. Valenzuela CY. Internucleotide correlation and nucleotide periodicity in Drosophila mtDNA: New evidence for panselective evolution. Biol Res. 2010; 43: 481-486. https://dx.doi.org/10.4067/S0716-97602010000400014
5. Valenzuela CY. Heterogeneous periodicity of drosophila mtDNA: new refutations of neutral and nearly neutral evolution. Biol Res. 2011; 44: 283-293. http://dx.doi.org/10.4067/S0716-97602011000300010
6. Valenzuela CY. The structure of selective dinucleotide interactions and periodicities in D.melanogaster mtDNA. Biol Res. 2014; 47: 1-12. PMID: 25027717https://doi.org/10.1186/0717-6287-47-18
7. Valenzuela CY. Selective intra-dinucleotide interactions and periodicities of bases separated by K sites: a new vision and tool for phylogeny analyses. Biol Res. 2017; 50: 3-16. https://doi.org/10.1186/s40659-017-0112-0
8. Valenzuela CY. Selective Profiles among Single or Double Stranded DNA or RNA Viruses Detect their Double or Single Stranded Condition. Arch Microbiol Immunol. 2024; 8: 84-95. https://doi.org/10.26502/ami.936500155
9. Valenzuela CY. Differences in Selective Profiles Between H. Sapiens and SARS-Cov-2 Genomes Confirm Double or Single Stranded DNA or RNA. J Biotech Biomed. 2025; 8(1): 24-38. https://www.doi.org/10.26502/jbb.2642-91280181
10. Logel DY, Jaschke PR. A high-resolution map of bacteriophage ΦX174 transcription. Virology 2020; 547: 47-56. https://doi.org/10.1016/j.virol.2020.05.008
11. Jang HS, Shin WJ, Lee JE, Do JT. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. Genes (Basel). 2017; 8(6): 148. https://www.doi.org/10.3390/genes8060148
12. Lin YT, Chiweshe S, McCormick D, Raper A, Wickenhagen A, DeFillipis V, et al.Human cytomegalovirus evades ZAP detection by suppressing CpG dinucleotides in the major immediate early 1 gene. PLoS Pathog. 2020; 16(9): e1008844. https://www.doi.org/10.1371/journal.ppat.1008844.
13. Valenzuela CY. Foundational errors in the Neutral and Nearly-Neutral Theories of evolution in Relation to the Synthetic Theory. Is it necessary a new evolutionary paradigm? Biol Res. 2013; 46: 101–119. https://www.doi.org/10.4067/S0716-97602013000200001