fortune Journals

Research Article

JOURNAL OF BIOTECHNOLOGY AND BIOMEDICINE

ISSN: 2642-9128

Differences in Selective Profiles Between H. Sapiens and SARS-Cov-2 Genomes Confirm Double or Single Stranded DNA or RNA

Carlos Y Valenzuela*

Abstract

Background: The debate between selective and neutral evolution is endless, even though the neutral theory of evolution cannot fully account for neutral evolution. I developed a method to study neutral evolution by the distance to neutrality (randomness) that the two bases of dinucleotides have between them.

Method: 1) the sets of dinucleotides whose bases are separated by 0, 1, 2, …K sites were obtained from genomes. 2) The chi-squares tests of the distance to neutrality of the first base in relation to the second base were calculated for each set and dinucleotide. 3) This allows construct the matrix of significance (chi-squares values) vs separations (K) of the deviations from neutrality of dinucleotides. 4) From this matrix the selective profile (significance order, sign of selection and selection coefficient) was calculated and compared between parallel (Par) and antiparallel (a-Par) dinucleotides with their index dinucleotide. 5) The distances between the index-Par or Index-a-Par dinucleotides within the human chromosome 21 and *SARS-CoV-2* were obtained and compared.

Results: In HCh21, the Index and a-Par dinucleotides present almost equal selective profile, while the Par dinucleotides differ from the Index profiles. In *SARS-CoV-2*, a-Par and Par dinucleotides differ from the Index dinucleotides and differ one another.

Conclusions: The almost equality of the selective profile between a-Par and the Index indicates that both strands of DNA of double stranded DNA (human) evolve together; this cannot occur in single stranded RNA (*SARS-CoV-2* virus).

Keywords: Dinucleotides; Distance to neutrality; Selective evolution; Selective differences; Double or single stranded nucleic acids

Introduction

The neutral theory of evolution (NTE) and the nearly neutral theory of evolution (NNTE) intended, at first, to study and describe the process of neutral evolution. Neutralists defined it as evolution occurring mostly by mutation-genetic drift with a marginal role for selection [1,2]. The synthetic theory of evolution (STE) proposes evolution as occurring mostly by mutation-selection processes with a marginal contribution of genetic drift [3]. Neutralists and nearly-neutralists added selection to neutral models. I) Purifying selection to explain genetic monomorphisms [4-6] Small selection coefficients to explain some polymorphisms [2,7]. III) They accepted that selectively neutral means selectively equivalent alleles or genes [6] regardless the values of their selection coefficients. Kimura [6] stated “I would like

Affiliation:
Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile

***Corresponding author:**
Carlos Y Valenzuela, Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile.

Citation: Carlos Y Valenzuela. Differences in Selective Profiles Between H. Sapiens and SARS-Cov-2 Genomes Confirm Double or Single Stranded DNA or RNA. Journal of Biotechnology and Biomedicine. 8 (2025): 82-96.

Received: February 23, 2024
Accepted: February 27, 2024
Published: March 26, 2025

to add here that by 'selectively neutral' I mean selectively equivalent …" This is a contradiction with the concept of neutral fitness, which is always equal to 1.0 (see Discussion). This mixture of drift and several types of selective factors does not allow study neutral evolution and makes neutral and nearly neutral models indistinguishable from STE models. Besides this confounding process, NTE and NNTE include some theoretical insufficiencies [8] who referred to several texts [9-18]. These two theories coexist since the Wright´s generalized model of evolution [9] and adaptive peaks [19,20], by one part, and the elaboration of NTE and NNTE [1,2,4-7,11] by the other, without any intention of integration. There are also criticism and refutations of NTE and NNTE [8,21-30] often overlooked. However, regardless these conclusive studies, the selective [31] - neutral [32] evolution discussion is at present an endless debate (see Discussion). That is why I decided to study neutral-selective evolution by their direct expected distribution of bases on genomes or chromosomes, forgetting all genetic approaches [16,18,33]. This new research is evolutionary physical research not genetic research it considers only the physical sequences of bases without their genetic properties, coding-non-coding, repeated or non-repeated sequences, telomere or centromere sequences, heterogeneous or homogeneous composition (Bernardi's isochores), etc. If evolution is neutral or at random the distribution of nucleotides on chromosome should be neutral or at random, this implies homogeneity of sequences along the genome. The random expected nucleotide frequency in a nucleotide site is for any of the four base 0.25 (ideal position) or that given by the mutation matrix of the four bases into the four bases. This random expectancy holds for all the nucleotide sites [13-17,28,29]. Bases should distribute randomly along any DNA segment, because we must assume a primary homogeneous rate of base mutation along the genome, that is the covariance of distribution of bases in their sites is zero [33]. Heterogeneous segments as isochores show their acquisition and maintenance is by selective evolution [34]. We did not find the equilibrium given by the mutational matrix, based on the observed base frequencies [15,17,25,33,35]. Furthermore, the distribution of the four bases along chromosomes is also far from a random distribution [25,28,33,35]. Thus, the random equilibrium of bases and their expected homogeneous random distribution does not occur [25,35]. The observed heterogeneity of base sequences is produced by a highly selective process; thus, we found a distribution of bases in chromosomes far from a random one [25,26,33,35]. To take these factors out (non-zero covariances), I went further and studied the distribution of the bases of dinucleotides, separated by 0, 1, 2, K nucleotide sites in chromosomes or genomes [29]. I emphasize this is not a genetic analysis but a physical analysis intended to estimate the covariance between bases separated by K nucleotide sites. If we consider a dinucleotide … A …… G … with K nucleotide sites between A and G, both bases may be independently located at any site of the genome, coding or non-coding, highly, medium or lowly repeat sequences, in homogeneous or heterogeneous segments. The sequence of K nucleotides between A and G is also irrelevant for they belong to all type of sequences of the genome or segment between both nucleotides.

There are previous studies performed in this sense to find permanent properties of dinucleotide associations called signatures [23]. These are nonrandom associations found in dinucleotides whose bases are contiguous (K = 0), and in general were not related to evolutionary theories, but to study phylogenies or evolutionary tendencies. I studied all the possible dinucleotides a genome (chromosome or genome segment) produces by defining the first base upstream, B_i, i from 1 to (N-K), N = total number of nucleotides, and the second base downstream running from (B_i+K) to B_N. (K goes from 0 to K). Other studies on periodicities with Fourier series, autocorrelations, latent periodicities in sequences of bases have nothing to do with my analyses, because both bases and bases between both are irrelevant and not taken into account for my studies [28]. Criticisms based on DNA or RNA structure or function, homogeneous or heterogeneous sequences are non-pertinent because the first base may belong to any of these sequences and the second may belong to equal or different sequences. A definitive proof occurs for bases separated by thousand or million sites that show highly significant distance to neutrality and both bases mostly belong to different types of DNA or RNA. See figures 6 and 7 in [29]. Moreover, the highly distant distribution of bases from randomness within dinucleotides and the periodicity of this distance found from single or double stranded DNA or RNA viruses until complex eukaryotes show that this structuration holds for any nucleic acid regardless its composition [29,30, the present study]. The number of sites in *SARS-Cov-2* is 29,866; a contrast with the number of sites of the human chromosome 21: 46,709,983. If at least one of the four bases in a site is not highly selectively advantageous, life is simply impossible (see discussion): a cornerstone of STE. Some cautions are necessary. To well understand these studies the researcher or reader must leave thinking in coding or noncoding segments, sense or anti-sense in transcription or translation, repetitive or unique DNA, telomeres, centromeres, isochores and studies based on DNA or RNA sequences. In short, the reader must consider DNA or RNA genomes as physical polymers without genetic meaning. The sense or anti-sense indicate the direction of DNA or RNA transcription or translation, but here these conceptualizations lack of meaning, with the exception that in double stranded nucleic acid a dinucleotide involves 4 dinucleotides that participate together in evolution. The nucleotides between both bases do not have any meaning, except they are in K consecutive nucleotide sites. The 5'-3' direction is also a

chemical and physical direction without any other meaning. A second caution comes from the use of the word interaction. As we know in formal genetics interactions are physiological interactions that leads to modification of the gene action by the gene action of other genes (epistasis, etc.). Here, interaction is simply non-randomness of the distribution of one base in relation to another base of the genome; it is a statistical interaction, a non-zero covariance between the distributions of both bases.

Unexpected initial results

In eukaryotes, I found chi-square values over 1,000,000 with 9 degrees of freedom yielding probabilities of occurrence at random less than $10^{-1,000,000}$. Prokaryotes show less significant interactions even when separations were 6,000 nucleotides sites, and in human chromosomes with separations over 10,000,000. Surprisingly, this distance or the value of the Chi-square test had periodicity with a 3K period in small genomes, and 2K and 6K in the human genome. To understand better this selective system the idea of a physical stability-instability of DNA or RNA with a 3K period is useful, because the periodicity occurs in the 16 dinucleotides. I have analyzed more than 150 species of pro and eukaryotes, mtDNA and chloroplast DNA from which near 40 are published [Valenzuela 25,26,29,30]. I found few significant exceptions to interactions and periodicities. No researchers are working in this subject. In studies with the HIV genome, I found some selective functional similarities to the human genome. I expected that, because both genomes suffer the same host selective pressure. For example, both genomes show a low frequency of CpG dinucleotides due to its frequent inactivation by methylation of cytosine [25]. Proteins of both species suffer the same negative selection of epitopes, a process leading to share epitopes not recognized by the host immune system [25,30,35]. Virus genomes may greatly differ from mammal genome and particularly with the *H. sapiens* genome, because they may be single or double stranded RNA or DNA genomes and code for different proteins. In this article, I explore the detection of differences and similarities in the *H. sapiens* and *SARS-CoV-2* genomes by interactions and periodicities of the distance to neutrality of the two bases of dinucleotides. I also examine, by their selective profiles, whether these tools can uncover the double or single stranded condition of DNA or RNA. A third caution is necessary. Here, double or single stranded condition, especially in viruses, does not mean the nucleic acid condition in the virion particle. This is evolutionary research, thus the most important conditions are the instances where mutation, selection and drift occur and lead to different evolutionary results. These events can occur along the life cycle of the virus when they have double or single stranded transitory genome structures. All the stages of a virus suffer these events but not all have equal evolutionary transcendence [30]. Prokaryotes and Eukaryotes suffer equal processes but the single stranded condition, in Eukaryotes, is so ephemeral that it is not relevant for the present analyses as it is for viruses' analyses [30]. The reader may think differences between the genome of viruses and humans are obvious, but humans do not have single stranded DNA that is solely present in viruses. Moreover *SARS-CoV-2* infect humans, thus the comparison is necessary to show at least that the method may ascertain differences (others than trivial). However, I have shown a critical difference between *E. coli* (double stranded DNA) and *SARS-CoV-2* [30].

Genomes and Summary on the Method and its Aim

Genomes and Assumptions

I obtained *SARS-CoV-2* genome and *Homo sapiens* chromosome 21 genome (HCh21) from Genbank, PubMed, Nucleotide (*SARS-CoV-2*, LR757998.1 Wuhan; 29,866 nucleotides; HCh21 NC_000021.9; 46,709,983 bp, respectively). I presented HCh21 previously [29] but with a very different analysis and *SARS-Cov-2* [30] but in comparisons with other viruses and not with human. I assumed this is the 5'-3' genome strand and named it the index strand. It is important to remark that in these analyses the condition of double or single stranded found in the virion disappears. These analyses show that viruses have always both conditions along with their life cycles, so the differentiation of double or single stranded is exclusively by statistical tests of the difference in their selective profiles [30]. Really, the tests show rather the proportion of single or double stranded phases viruses have during their life cycles. I have chosen a kind of gold standards for eukaryotes with human chromosome 21, for prokaryotes with *E coli* for bacteria and *M. smithi* for archaea [29], and *SARS-CoV-2* for single stranded RNA non-lysogenic lytic viruses [30]. They will appear in following analyses when necessary.

Method

Obtaining the basic matrix of significance vs separations.

I obtained the sets of dinucleotide pairs whose bases are separated by 0 (contiguous), 1, 2…K sites (until K=32, in this study), with a program in Python. These sets have the total observed number (Obs) of dinucleotides for each K and for each of the 16 dinucleotides or pairs. The program obtains the observed frequency of bases (A, T, G, C), that I assumed was their neutral expectancy: fA1, fT1, fG1, fC1 for the first nucleotide (upstream) and fA2, fT2, fG2, fC2 for the second one (downstream). The expected number (Exp_i) of a $dinucleotide_i$, (i from 1 to 16) considering B as a generic base is fB1 x fB2 times the total number of dinucleotides for this K. From Obs and Exp numbers the individual chi-square (χ^2_1) value is $[(Obs_i-Exp_i)^2/Exp_i]$ with one degree of freedom, where i goes from 1 to 16 for each particular dinucleotide. This chi-square value measures the distance to neutrality. The sum of

Valenzuela CY., J Biotechnol Biomed 2025
DOI:10.26502/jbb.2642-91280181

these figures is the total χ^2_9 value, with 9 degrees of freedom. For each dinucleotide the program calculates the selection coefficient by $[(Obs_i-Exp_i)/Exp_i]$ with the corresponding positive sign (+) if $Obs_i>Exp_i$ or negative sign if $Obs_i<Exp_i$. These values generate a matrix whose columns denote the order of significance, (χ^2_1) value, or distance to neutrality and whose rows denote the separation between both bases of the dinucleotide (K). In columns, the selection coefficient with its sign constitutes one element of the selective profile of this dinucleotide that includes the order of significance (1 to 16), the χ^2_1 value, the sign of selection and the selection coefficient. The minimum significant value (P = 0.05) for the distance to randomness of each pair is $\chi^2_1 = 3.84$ and 16.9 for the total χ^2_9. Details are in previous articles [25-30], see Table 1 and Table 2.

Studying the difference of selective profiles in the index dinucleotide and its parallel and anti-parallel pairs

In double stranded nucleic acids, evolution occurs together in four dinucleotides

We must first mind what evolution implies for a double stranded nucleic acid. To compare selective profiles, first we must consider an index dinucleotide with bases separated by K sites, in double stranded DNA. For example, …A….G.., in the … physical sense or direction 5'-3' in the Index strand. It includes four dinucleotides 5'A …. G3' (Index); 3'G ….A5' (anti-Index; 3'T….C5' [a parallel dinucleotide (Par)] in the complementary strand; and 5'C….T3' [an anti-parallel dinucleotide (a-Par)] in the complementary strand. These four dinucleotides evolve together, that is suffer mutation, selection and random fluctuations (dinucleotide drift) of their frequencies. However, we can study the evolutionary behavior of the four dinucleotides only in the index strand (with the assumed 5'-3' direction). Here I study the index dinucleotides and its Par and a-Par pair, but not anti-Index pair, because the GA pair has 5'-3' direction in this chromosome from Genebank. The reader may study other possibilities; this is an open study. The difference or distance in a selective profile between two dinucleotides is direct. The absolute value of the difference between their order of Significance (Sig) between 1 and 16, the difference between their selection coefficients and the difference in the sign of selection. The difference in the χ^2 value cannot yield a useful figure because they correspond to very different genome sizes and its statistic is very complicated.

Correction for the sign of selection. Since the χ^2 value is always positive but includes positive and negative values of selection, I ordered the dinucleotide significance according to their χ^2 considering their sign of selection, regardless that the most significant value was positive or negative (see tables 3 and 4). This treatment increase the discrimination of similarities or differences among dinucleotides.

Some ad hoc statistical considerations and tools.

The analyses take all of genome and dinucleotides, so we work with parameters and we do not need statistical tests. However, I applied statistical tests to show the robustness of the analyses. The mean (M), variance (V) and standard deviation (SD) of the distance of Par or a-Par dinucleotides to the Index (ID) dinucleotide were obtained by considering all the possible distances; in Significance (Sig) and Selection (Sel) comparisons, absolute (differences without + - signs) distances are used. An algebraic full demonstration with formulae is out of the scope of this article.

I offer an intuitive but complete calculation, using the Par dinucleotide (PD). We imagine the square where the positions of ID are the columns and the positions of PD are rows; in the diagonal, only ID is present (distance 0). Let us examine the situation when PD moves towards the right of ID (the triangle upper the diagonal). ID can occupy the 16 (G) places from left to right and PD can move on the remaining 15 (G-1) places. There are G-1 positions when ID is in the 1° position, G-2 positions when ID is in the 2° position…. G-15 = 1 position when ID is in the 15° position. The number of distances is the sum of (G-i) values, i from 1 to G-1. This is (G-1)(G)/2. However, for each distance towards the right of ID there is one distance towards the left (or bottom) of ID (the triangle under the diagonal), then the total number is twice that figure = (G-1)(G), that is the number of combinations of two elements when they can move on G positions, considering the order between both elements (right and left). These enumeration of all the possible distances allow calculate the parameters, given that the distance between ID and PD is exactly the index i. M or E(x) = sum 2(G-i)i/[(G-1)G] = (G+1)/3 = 17/3 = 5.667. The variance (V) is obtained by $E(x^2) - [E(x)]^2$; the sum of squared values is: $E(x^2)$ = sum $2(G-i)i^2/[(G-1)G] = G(G+1)/6$; $V = G(G+1)/6 - [(G+1)/3]^2 = (G+1)(G-2)/18 = 13.22$; SD = 3.6362. With these parameters, I tested the observed figures with one tailed z tests. Another statistic distribution that must be mind is the simple variable of numbers from 1 to 16 whose mean is 8.5 and SD = 4.610. However, this is not the true variable for a rigorous analysis because this variable comes from a discretization in the order of significance of the continuous χ^2 variable that has changed to consider their negative and positive values. The analysis with the true variable is out of the scope of this article. The calculated mean and SD are sufficient for the present study.

Results

Table 1 (for HCh21) and Table 2 (for *SARS-CoV-2*) present the matrix of significance (columns) vs separations (rows) of dinucleotides with signs of selection, distances to neutrality given by their χ^2_1 (that determines the significance order among the 16 dinucleotides) and their selection values (elements of the selective profile); the three most distant dinucleotides from neutrality (significance) are described. The

Valenzuela CY., J Biotechnol Biomed 2025
DOI:10.26502/jbb.2642-91280181

total chi-square is also included (χ^2_9). Both genomes showed enormous deviations from neutrality and periodicities of the value of these deviations. Since present data and analyses of the HCh21 and *SARS-CoV-2* overlap with those of previous articles [29,30], I shall refer mostly to comparative analyses between these genomes.

In both genomes there is a large lack of CG pairs (CG[-]), less than 70% of the expected pairs in HCh21 and less than 50% in *SARS-CoV-2*. However, the spectrum of pairs within the three most significant dinucleotides is different in the 33 separations. While in HCh21 most pairs are AA[+](24 pairs), TT[+](21p), CC[+](19p), GG[+](15p), that is complementary pairs, in *SARS-CoV-2* most of dinucleotides are GG[-] (19p), GG[+] (11p), GT[+](11p), TG[+](11p), TG[-](9p), CG[+] (8p), TT[+](7p), GT[-](6p), non-complementary pairs, and others in lower frequencies. The most important difference is the high frequency of AA and CC in HCh21 and their absence in *SARS-CoV*-2.

I remember that selective profiles include: 1) the order of significance (1 to 16) given by 2) the χ^2_1 value, 3) the selection coefficient, with its 4) sign of selection. We can compare three of these four traits, but not χ^2_1 values between both species due to their huge difference in their genome sizes and complication of its statistics.

In Table 2, italics mark the significant head of a periodicity (3K), not indicated in Table 1 (periodicity 2K and 6K). The following analyses are the comparisons of selective profiles within and between HCh21 and *SARS-CoV-2* viral genome. We need first, a matrix with the 16 dinucleotides ordered according to their significance and separations. I chose 4 separations: 0, 1, 2, 3. The analysis begins by choosing one index dinucleotide in the GenBank published strand (I assumed it is in the 5'-3' sense) and compare the selective profile of the index dinucleotide with the selective profile of the parallel (Par, 3'-5') and anti-parallel (a-Par, 5'-3') dinucleotides.

Table 1: Distribution of dinucleotides according to separation and significance. Human chromosome 21 (HCh21). The three most significant dinucleotides.

		1° Significance			2° Significance			3° Significance		
Sep	χ^2_9	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	4503675	CG[-]	2089279	-0.717	TA[-]	559068.3	-0.2984	CA[+]	245130.3	0.2212
1	956940.3	AA[+]	205531.4	0.1818	TT[+]	197571.3	0.1765	CA[-]	104237.8	-0.1442
2	387656.7	GG[+]	56265	0.1172	CC[+]	56241.5	0.118	TT[+]	49062.9	0.088
3	572371.6	TT[+]	103500.3	0.1278	AA[+]	99691.1	0.1266	CC[+]	55055.7	0.1168
4	348768	AA[+]	63295	0.1009	TT[+]	60640.7	0.0978	CC[+]	49725.9	0.111
5	606929.9	GG[+]	108856.7	0.1631	CC[+]	107063.9	0.1629	AA[+]	85650.8	0.1173
6	322720.2	GG[+]	73416	0.1339	CC[+]	65079.8	0.127	AA[+]	36353.3	0.0765
7	511155.1	GG[+]	91225.2	0.1493	CC[+]	85992.4	0.146	AA[+]	79122	0.1128
8	410408.8	GG[+]	90611.6	0.1488	CC[+]	82753.5	0.1432	AA[+]	54608.4	0.0937
9	397319.7	GG[+]	67399.9	0.1283	CC[+]	64505.7	0.1264	AA[+]	55657.2	0.0946
10	223757.7	CG[+]	34915.4	0.0927	CT[-]	33738.6	-0.0817	AG[-]	29901.7	-0.077
11	384475.9	CC[+]	64105.6	0.126	GG[+]	56262.8	0.1172	TT[+]	49716.6	0.0886
12	235804.1	AA[+]	49396.8	0.0891	TT[+]	45058.8	0.0843	GG[+]	32131.8	0.0886
13	249975.7	AA[+]	50276	0.0899	TT[+]	43569.9	0.0829	CC[+]	30827.6	0.0874
14	195965.4	CC[+]	25522.7	0.0795	AA[+]	24647.7	0.063	TT[+]	22254.7	0.0592
15	291127.3	AA[+]	47778.7	0.0876	TT[+]	41766.2	0.0812	GT[-]	35719.9	-0.0837
16	144295.5	CC[+]	18815.2	0.0683	AA[+]	16410.6	0.0514	TT[+]	15174.8	0.0489
17	220784.6	GG[+]	38476.8	0.097	CC[+]	35890.5	0.0943	AA[+]	23575.7	0.0616
18	135281.7	TT[+]	22219.3	0.0592	AA[+]	21082	0.0582	GT[-]	17936.3	-0.0593
19	169384.5	AA[+]	26424.5	0.0652	TT[+]	22279.3	0.0593	GA[-]	18229.5	-0.0601
20	140010	GC[+]	20009.8	0.0702	AA[+]	16036.9	0.0508	TT[+]	15582.6	0.0496
21	166745.5	AA[+]	33968.7	0.0739	TT[+]	32852.5	0.072	GT[-]	15635	-0.0554
22	131823.7	AA[+]	27282.7	0.0662	TT[+]	25551.8	0.0635	TG[-]	14209.8	-0.0528
23	315685.1	TT[+]	56902.3	0.0947	AA[+]	54388.2	0.0935	CC[+]	43905.4	0.1043
24	185470.1	TT[+]	36755.6	0.0761	AA[+]	32448.2	0.0722	GG[+]	25298	0.0786
25	176532.3	TT[+]	35745.3	0.0751	AA[+]	30008.3	0.0695	GT[-]	18507.5	-0.0603
26	155254.6	GG[+]	30524.8	0.0864	CC[+]	27379.1	0.0824	TT[+]	22698.1	0.0598

Valenzuela CY., J Biotechnol Biomed 2025
DOI:10.26502/jbb.2642-91280181

27	161677.4	GG[+]	19823.3	0.0696	AC[-]	17791.1	-0.0596	CC[+]	17411.5	0.0657
28	154223.6	AA[+]	22316	0.0599	AC[-]	19950.7	-0.0631	GT[-]	19903.4	-0.0625
29	201569.2	AA[+]	36243.9	0.0763	TT[+]	31441.8	0.0704	AC[-]	28480	-0.0754
30	169023.5	AC[-]	26445.6	-0.073	CC[+]	26201.9	0.0806	GG[+]	22360.6	0.0739
31	217227	CC[+]	43398.9	0.1037	GG[+]	41946.8	0.1012	TT[+]	32101.6	0.0712
32	117890.5	GG[+]	19749.6	0.0695	CC[+]	19350.8	0.0692	GT[-]	11251.8	-0.047

Sep = n° sites of separation; din[s] = dinucleotide with its sign of selection; χ^2_9 = total chi square value; χ^2_1 = chi-square value for the dinucleotide at the left hand; c-se = selection coefficient.

Table 2: Distribution of dinucleotides according to separation and significance. SARS-CoV-2 Genome. The three most significant dinucleotides

		1° Significance			2° Significance			3° Significance		
Sep	**χ^2_9**	**din[s]**	**χ^2_1**	**c-se**	**din[s]**	**χ^2_1**	**c-se**	**din[s]**	**χ^2_1**	**c-se**
0	1236.9	CG[-]	377.3	-0.592	TG[+]	267.2	0.377	CA[+]	118.8	0.269
1	75.7	GT[+]	28.6	0.123	AT[-]	8.3	-0.054	TC[+]	8	0.067
2	*144*	GG[+]	*55.5*	0.22	TG[-]	*33.9*	-0.134	TT[+]	*21.2*	0.083
3	69.5	TG[+]	15.5	0.091	GG[-]	14.8	-0.114	TT[-]	11.1	-0.06
4	88.8	GT[+]	33.4	0.133	GG[-]	14.2	-0.111	CG[+]	11.6	0.104
5	72.1	GG[+]	24.7	0.146	TG[-]	*14.4*	-0.087	TT[+]	*12.9*	0.065
6	72.7	TG[+]	22.9	0.11	GG[-]	12.9	-0.106	GC[+]	11.8	0.105
7	44.9	GT[+]	11.7	0.079	CG[+]	10.7	0.1	GG[-]	10.3	-0.095
8	*127.6*	GG[+]	*43.7*	0.195	TT[+]	*22.9*	0.086	TG[-]	*21*	-0.106
9	56	TG[+]	20.1	0.103	GG[-]	14.9	-0.114	GC[+]	5	0.068
10	40.4	CG[+]	9.5	0.094	GT[+]	8.3	0.067	GG[-]	7.2	-0.079
11	*131.6*	TT[+]	*35.9*	0.108	GT[-]	*34.5*	-0.136	GG[+]	*25.8*	0.15
12	85.8	GG[-]	34.8	-0.174	TG[+]	25.5	0.116	GC[+]	9.5	0.094
13	70.3	CG[+]	24.2	0.15	GT[+]	17.6	0.097	GG[-]	6.9	-0.077
14	*101.1*	GG[+]	*45.6*	0.199	TG[-]	*21.5*	-0.107	GT[-]	*13.8*	-0.086
15	57.1	GG[-]	16.6	-0.12	TG[+]	10.1	0.073	TT[-]	8.9	-0.054
16	61.4	GT[+]	16.9	0.095	CT[-]	12	-0.082	GG[-]	6.9	-0.077
17	*86.4*	GG[+]	*23.6*	0.143	GT[-]	*18.7*	-0.1	TG[-]	*17.1*	-0.095
18	46.5	GG[-]	16.4	-0.119	GC[+]	9	0.092	TG[+]	6.6	0.059
19	57.3	GT[+]	14.2	0.087	CG[+]	10.7	0.1	GG[-]	8.7	-0.087
20	*93.9*	GG[+]	26.6	0.152	TG[-]	*25.1*	-0.115	TT[+]	*18*	0.076
21	87.4	GC[+]	28.9	0.164	TG[+]	18.7	0.1	GG[-]	14.3	-0.111
22	63.3	CG[+]	19.9	0.136	GT[+]	17.3	0.096	CT[-]	7.7	-0.066
23	*124.8*	GG[+]	*34.8*	0.174	TG[-]	*28.9*	-0.124	GT[-]	*28.5*	-0.123
24	67.8	TG[+]	16.1	0.092	CT[+]	13	0.086	GG[-]	11.7	-0.101
25	54.7	CG[+]	13.6	0.113	GT[+]	11.8	0.079	GG[-]	10.3	-0.095
26	*74.1*	TG[-]	*22.2*	-0.109	GG[+]	*18.5*	0.127	GT[-]	*16.7*	-0.094
27	59.1	GC[+]	12.9	0.11	TG[+]	12.8	0.082	GG[-]	11.6	-0.101
28	45	GG[-]	11.3	-0.099	GT[+]	8.5	0.067	TT[-]	5.5	-0.042
29	*64.2*	GT[-]	*16.6*	-0.094	TT[+]	*13.2*	0.065	GG[+]	*11*	0.098
30	42.5	TG[+]	11.3	0.078	GG[-]	7.9	-0.083	GC[+]	6	0.075
31	40.2	GT[+]	13.1	0.083	CG[+]	9.4	0.093	GG[-]	3.3	-0.053
32	*124.9*	GG[+]	*38.6*	0.183	TT[+]	*27.7*	0.095	TG[-]	*26.9*	-0.12

Nomenclature as in Table 1. Italics indicate the head of the periodicity

Valenzuela CY., J Biotechnol Biomed 2025
DOI:10.26502/jbb.2642-91280181

I did not perform other fascinating analyses, as for example the anti-index dinucleotide, simply, to stop this study somewhere. Tables 3 (for humans) and Table 4 (for *SARS-CoV-2*) have the data for the analyses of selective profiles. For example, in Table 3, Sep 0, 4° we found 5'AC3'[-] that I assume an index pair. Its Par pair is 3'TG[+]5' pair (in the complementary strand) at the 15° Sig (in the Index strand with 5'-3' direction); its a-Par pair is the 5'GT[-]3' pair (in the Index and complementary strands) found at the 3° Sig. We see the a-Par contiguous (1 Sig) to the index but the Par is farer (11 Sigs) from the Index. Moreover, the Index and a-Par show positive selection while the Par has negative selection.

Table 3: Distribution of the 16 dinucleotides according to their significance and four separations. H. sapiens Chromosome 21. Dinucleotides according to their significance order.

	***H. sapiens* Chromosome 21**											
	1° significance			**2° significance**			**3° significance**			**4° significance**		
Sep	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	CG[-]	2089279.1	-0.717	TA[-]	559068.3	-0.298	GT[-]	156176.7	-0.175	AC[-]	153458.7	-0.175
1	AA[+]	205531.4	0.182	TT[+]	197571.3	0.177	GG[+]	68488.5	0.129	CC[+]	65803.5	0.128
2	GG[+]	56265	0.117	CC[+]	56241.5	0.118	TT[+]	49062.9	0.088	AA[+]	47970.9	0.088
3	TT[+]	103500.3	0.128	AA[+]	99691.1	0.127	CC[+]	55055.7	0.117	GG[+]	52617.4	0.113
	5° significance			6° significance			7° significance			8° significance		
	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	AT[-]	107803.8	-0.131	TC[-]	1076.1	-0.015	GA[-]	444.8	-0.009	GC[+]	19.6	0.002
1	CG[+]	35443.3	0.093	GC[+]	781.1	0.014	GA[-]	39.2	-0.003	TC[-]	203.5	-0.006
2	CG[+]	23840.4	0.077	AT[+]	12	0.001	GA[-]	187.7	-0.006	TC[-]	1095.6	-0.015
3	CG[+]	39237.4	0.098	GC[+]	3226.8	0.028	TA[-]	165.4	-0.005	AT[-]	2138.1	-0.018
	9° significance			10° significance			11° significance			12° significance		
	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	AA[+]	106213.3	0.131	TT[+]	114109.6	0.134	CT[+]	147627.9	0.171	AG[+]	155881.9	0.176
1	CT[-]	5864.6	-0.034	AG[-]	9231.7	-0.043	AT[-]	22416.6	-0.06	TA[-]	24706.3	-0.063
2	TA[-]	1622.9	-0.016	AC[-]	4257.1	-0.029	GT[-]	7324	-0.038	GC[-]	16118.9	-0.063
3	AC[-]	13690.6	-0.052	GT[-]	15656.2	-0.055	GA[-]	17388.1	-0.059	TC[-]	20754.9	-0.064
	13° significance			14° significance			15° significance			16° significance		
	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	GG[+]	211479.5	0.227	CC[+]	219708.2	0.233	TG[+]	236197.2	0.215	CA[+]	245130.3	0.221
1	AC[-]	58118.1	-0.108	GT[-]	64171.6	-0.112	TG[-]	94331.7	-0.136	CA[-]	104237.8	-0.144
2	CT[-]	27540.8	-0.074	TG[-]	28693.4	-0.075	AG[-]	33133.4	-0.081	CA[-]	34290	-0.083
3	AG[-]	33259.1	-0.081	CT[-]	33516.2	-0.081	TG[-]	40424.8	-0.089	CA[-]	42049.6	-0.092
Nomenclature as in Table 1.												

I assumed uniformity of selection in both strands (this is not necessarily true). Let us perform a complete calculation. Also, from Table 3 we found for the AG[+] index pair, Sep 0: 12° Sig, positively selected, selection coefficient (c-se) = 0.176; its Par-pair is TC[-] pair: 6° Sig, negatively selected, c-se = -0.015. its a-Par-pair is CT[+] pair: 11° Sig, positively selected, c-se = 0.171. With these figures we calculate the distance Index-Par (pair) and the distance Index-a-Par (pair). The Index-Par distance of Sig (absolute values) is 6° (12°-6°), the Index have different selection sign with the Par (+ and -, respectively) and their c-se are 0.176 and -0.015, respectively, their distance or difference is 0.191. The Sig distance Index-a-Par is 1° (12°-11°); they have equal selection sign (+), their c-se difference is 0.176-0.171 = 0.005. We see the a-Par pair has practically the same selective profile as the index pair;

Valenzuela CY., J Biotechnol Biomed 2025
DOI:10.26502/jbb.2642-91280181

Table 4: Distribution of the 16 dinucleotides according to their significance and four separations. SARS-CoV-2. Dinucleotides according to their significance order and sign of selection.

***SARS-CoV-2* Wuhan**												
	1° Significance			**2° Significance**			**3° Significance**			**4° Significance**		
Sep	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	CG[-]	377.3	-0.592	AT[-]	110	-0.196	TA[-]	85.7	-0.173	TC[-]	69	-0.198
1	GT[+]	28.6	0.123	TC[+]	8	0.067	AA[+]	5.6	0.046	CA[+]	3	0.042
2	GG[+]	55.5	0.22	TT[+]	21.2	0.083	CC[+]	8.7	0.093	AG[+]	1.7	0.031
3	TG[+]	15.5	0.091	CT[+]	6.9	0.063	GT[+]	4.7	0.05	TA[+]	2	0.026
	5° Significance			6° Significance			7° Significance			8° Significance		
	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	CC[-]	15	-0.122	GA[-]	11.4	-0.08	GG[-]	3	-0.051	AG[-]	0.1	-0.007
1	AG[+]	1.1	0.025	TG[+]	0.5	0.016	CG[+]	0	0.005	AC[-]	0.1	-0.007
2	AA[+]	0.6	0.015	CA[+]	0.2	0.011	TC[+]	0	0.004	TA[-]	0	-0.004
3	GC[+]	1.6	0.039	AG[+]	1.5	0.029	AC[+]	0.2	0.01	CC[+]	0	0.003
	9° Significance			10° Significance			11° Significance			12° Significance		
	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	TT[+]	6	0.044	GT[+]	6.2	0.057	GC[+]	7.8	0.085	AA[+]	14.3	0.073
1	CT[-]	0.4	-0.015	TT[-]	0.9	-0.017	CC[-]	2.5	-0.049	GC[-]	3.1	-0.053
2	AC[-]	0.3	-0.014	GA[-]	1.3	-0.027	CT[-]	1.8	-0.032	AT[-]	1.8	-0.025
3	GA[-]	0	-0.003	AT[-]	0.1	-0.007	CA[-]	0.3	-0.014	AA[-]	0.9	-0.018
	13° Significance			14° Significance			15° Significance			16° Significance		
	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se	din[s]	χ^2_1	c-se
0	CT[+]	57.6	0.181	AC[+]	87.7	0.231	CA[+]	118.8	0.269	TG[+]	267.2	0.377
1	TA[-]	3.2	-0.034	GA[-]	5.3	-0.055	GG[-]	5.4	-0.068	AT[-]	8.3	-0.054
2	CG[-]	2.8	-0.051	GC[-]	5.7	-0.073	GT[-]	8.6	-0.068	TG[-]	33.9	-0.134
3	TC[-]	2.1	-0.035	CG[-]	7.8	-0.085	TT[-]	11.1	-0.06	GG[-]	14.8	-0.114
Nomenclature as in Table 3												

the Par pair has a different profile. The values to calculate distances are in Table 3 (HCh21) and Table 4 (*SARS-CoV-2*).

Table 5 presents the distances in selective profiles for HCh21 for those Index-Pairs dinucleotides that have equal Index-Par and Index-a-Par pairs (AA-TT, GG-CC) and for those Indexes that have different Index-Par and equal Index-a-Par-pairs (AT-TA, GC-CG, Index is equal to a-Par). Expected Selective distances between Indexes and Par and a-Par dinucleotides are obviously equal. For equal Par and a-Par pairs we found for Sig: mean = 1, and SD = 0, that is, all the pairs were contiguous, a result similar in selection values (mean = 0.0027, SD = 0.0020). The comparisons Index-a-Par and Index-Par are equal, however, in double stranded DNA they imply different 5'-3' senses that do not exist in single stranded DNA or RNA. The mean distance in Sig (Equal Par and a-Par pairs) can be tested with the expected mean (5.667, SD = 3.636; standard error = 1.29), the result is z = 3.63, P = 0.00014. When Par and a-Par pairs are different, we observe mean = 3 and SD = 2.45 and z = -2.075, P = 0.0192. Table 6 shows the same comparisons in *SARS-CoV-2.* For the I-Par distance in the equal I-Par and I-a-Par set of dinucleotides, the mean Significance was 4 (against the expected mean 5.667, z = -1.297, P < 0.199). For different I-Par and I-a-Par the mean difference in Significance was 4.875, z = 0.6161, P = 0.2676. While in *Homo*

The mean of the significance site indicates a similar result of

Valenzuela CY., J Biotechnol Biomed 2025
DOI:10.26502/jbb.2642-91280181

Table 5: Distance Index-Par and a-Par dinucleotides when pairs has the same Par and a-Par pairs and different Par and a-Par pairs. *H. sapiens*

Equal Par and a-Par pair								
	Indexes			Par			D Ind - Par	
Sep	din[s]	c-se	Sig	din[s]	c-se	Sig	D Sig	D Sel
Equal Index-Par and Index a-Par pairs								
0	AA[+]	0.1307	9	TT[+]	0.1342	10	1	0.0035
1	AA[+]	0.1818	1	TT[+]	0.1765	2	1	0.0052
2	AA[+]	0.0878	4	TT[+]	0.088	3	1	0.0002
3	AA[+]	0.1266	2	TT[+]	0.1278	1	1	0.0012
0	GG[+]	0.2273	13	CC[+]	0.2333	14	1	0.006
1	GG[+]	0.1294	3	CC[+]	0.1277	4	1	0.0017
2	GG[+]	0.1172	1	CC[+]	0.118	2	1	0.0008
3	GG[+]	0.1134	4	CC[+]	0.1168	3	1	0.0034
	Mean	0.1393	4.625		0.1403	4.875	1	0.0027
	SD	0.0414	3.967		0.042	4.314	0	0.002
				t test P for Sig and Sel			0.4559	0.4823
Different Par and a-Par pairs								
0	AT[-]	-0.131	5	TA[-]	-0.2984	2	3	0.1673
1	AT[-]	-0.0597	11	TA[-]	-0.0627	12	1	0.003
2	AT[+]	0.0014	6	TA[-]	-0.0161	9	3	0.0175
3	AT[-]	-0.0185	8	TA[-]	-0.0051	7	1	0.0133
0	GC[+]	0.0022	8	CG[-]	-0.7169	1	7	0.7191
1	GC[+]	0.0139	6	CG[+]	0.0934	5	1	0.0795
2	GC[-]	-0.063	12	CG[+]	0.0766	5	7	0.1396
3	GC[+]	0.0282	6	CG[+]	0.0982	5	1	0.0701
	Mean	-0.0283	7.75		-0.1039	5.75	3	0.1512
	SD	0.0495	2.3848		0.261	3.3448	2.4495	0.2218
				t test P for Sig and Sel			0.1093	0.2321

Nomenclature as in Table1. D Ind-Par = distance between the Index and Par dinucleotides; D Sig = distance of Significance; D Sel = distance of selection; P = probability; SD = standard deviation.

Table 1 and Table 2. While AA-TT-GG-CC are in the first places of significance in HCh21 they are in the middle or the extreme right of Table 2 in *SARS-CoV-2*. The expected mean in a scale of 16 places is 8.5 (SD = 4.61). In HCh21 Indexes means were 4.6 (z = 2.38, P = 0.009) and 7.75 (z =0.46, P = 0.323) for equal Par and a-Par and different Par and a-Par, respectively. The respective figures for Par pairs were 7.75 and 5.75 for equal and different Par and a-Par pairs, respectively (under the expected mean). In *SARS-CoV-2* we found 8.875 and 7.875 for Equal Par and a-Par pairs (over and under the expected mean), respectively and 10.25 and 7.875 in different Par and a-Par pairs, respectively (over and under the mean).

Here I present only the distances in significances because distances in selection coefficients are highly correlated with them, and had the same structure of results; the distance D sel is in HCh21 at the level of thousandths; in *SARS-CoV-2* it is at the level of hundredths and tenths (tests are unnecessary).

Tables 7 and 8 show the same comparison for dinucleotides with different a-Par and Par pairs (AG, AC, TG, TC, GA, GT, CA and CT). There are 32 comparisons thus the expected standard error is SD/√32 = 0.6428. Dinucleotides in comparisons are repeated three times but as Indexes, Pars and a-Pars. In HCh21 the mean I-Par distance (Sig) was 5.0 (z = -1.038; P < 0.149) and the mean I-a-Par distance was 1.125 (z = -7.066; P < 10^{-9}). The comparison between I-Par and I-a-Par pairs proceeds from a t test. The t probabilities for the differences in Sig and Sel were $4x10^{-9}$ and $2x10^{-5}$, respectively.

Table 6: Distance Index-Par and a-Par dinucleotides when pairs have the same Par and a-Par pairs and different Par and a-Par pairs. *SARS-Cov2*

Equal Par and a-Par pairs								
	Indexes			Par			D Ind - Par	
Sep	din[s]	c-se	Sig	din[s]	c-se	Sig	D Sig	D Sel
0	AA[+]	0.073	12	TT[+]	0.044	9	3	0.029
1	AA[+]	0.0459	3	TT[-]	-0.0168	10	7	0.0627
2	AA[+]	0.015	5	TT[+]	0.083	2	3	0.068
3	AA[-]	-0.0182	12	TT[-]	-0.0601	15	3	0.042
0	GG[-]	-0.0508	7	CC[-]	-0.1218	5	2	0.071
1	GG[-]	-0.0683	15	CC[-]	-0.0494	11	4	0.0189
2	GG[+]	0.2197	1	CC[+]	0.0927	3	2	0.127
3	GG[-]	-0.1136	16	CC[+]	0.0033	8	8	0.1169
	Mean	0.0128	8.875		-0.0031	7.875	4	0.0669
	SD	0.0969	5.2782		0.0694	4.0755	2.1213	0.0362
				t test P for Sig and Sel			0.3488	0.364
Different Par and a-Par pairs								
0	AT[-]	-0.1959	2	TA[-]	-0.1728	3	1	0.023
1	AT[-]	-0.0539	16	TA[-]	-0.0336	13	3	0.0202
2	AT[-]	-0.0252	12	TA[-]	-0.0037	8	4	0.0215
3	AT[-]	-0.0066	10	TA[+]	0.0264	4	6	0.033
0	GC[+]	0.0852	11	CG[-]	-0.5921	1	10	0.6774
1	GC[-]	-0.0532	12	CG[+]	0.0053	7	5	0.0585
2	GC[-]	-0.0726	14	CG[-]	-0.0514	13	1	0.0212
3	GC[+]	0.0389	5	CG[-]	-0.0849	14	9	0.1238
	Mean	-0.0354	10.25		-0.1134	7.875	4.875	0.1223
	SD	0.0781	4.3229		0.1902	4.7021	3.14	0.2124
				t test P for Sig and Sel			0.171	0.1664
Nomenclature as in Table 5								

For selection, we do not have the parameter estimates (exact expected mean and SD) because this is a biased sample from the total dinucleotide population. In *SARS-Cov-2* the distance in Sig I-Par was 4.938 (z = -1.134; P = 0.129) and the I-a-Par was 5.313 (z = -0.551; P = 0.291). Both distances between I-Par and I-a-Par in Sig and Sel are not significant. The comparisons between HCh21 and *SARS-CoV-2* are direct. While I-Par is almost equal, between HCh21and *SARS-CoV-2*, in Sig (5,000 and 4.938, respectively) and similar in Sel (0.101 and 0,065, respectively), I-a-Par disagree in Sig (1.125 and 4, respectively) and Sel (0.005 and 0.033, respectively). These figures save any statistical test. Again, while the distance in selection coefficient for the I-a-Par distance in humans is at the level of thousandths or ten thousandths, in *SARS-CoV-2* it is at the level of tenths or hundredths.

Discussion

Results confirm the enormous non-neutral interaction between the bases of dinucleotides and the periodicity of the distance to neutrality found in more than 150 genomes, also in *SARS-CoV-2,* a single stranded RNA virus [30]. The estimates of significance, with so large figures of the χ^2 value are out of tables and programs. I approximate them as follows: if we assimilate the χ^2_9 to a Gaussian distribution and considering 2 standard deviations over the mean for each decimal figure (a conservative criterion) we can calculate the probability for so huge values. The probability for the random occurrence of the total χ^2_9 = 1,236.9, Sep 0, in *SARS-CoV-2* is $P < 10^{-145}$, a value sufficient to consider neutral evolution of *SARS-CoV-2* genome definitively refuted, as far

Valenzuela CY., J Biotechnol Biomed 2025
DOI:10.26502/jbb.2642-91280181

Table 7: Distances Index-Par and Ia-Par dinucleotides when they are different pairs. *H sapiens* Chrom 21.

	Indexes			Par			a-Par			D Ind - Par		D Ind - a-Par	
Sep	din[s]	c-se	Sig	din[s]	c-se	Sig	din[s]	c-se	Sig	D Si	D Se	D Si	D Se
0	AG[+]	0.176	12	TC[-]	-0.015	6	CT[+]	0.171	11	6	0.19	1	0.005
1	AG[-]	-0.043	10	TC[-]	-0.006	8	CT[-]	-0.034	9	2	0.036	1	0.009
2	AG[-]	-0.081	15	TC[-]	-0.015	8	CT[-]	-0.074	13	7	0.066	2	0.007
3	AG[-]	-0.081	13	TC[-]	-0.064	12	CT[-]	-0.081	14	1	0.017	1	0
0	AC[-]	-0.175	4	TG[+]	0.215	15	GT[-]	-0.175	3	11	0.39	1	0
1	AC[-]	-0.108	13	TG[-]	-0.136	15	GT[-]	-0.112	14	2	0.028	1	0.005
2	AC[-]	-0.029	10	TG[-]	-0.075	14	GT[-]	-0.038	11	4	0.046	1	0.009
3	AC[-]	-0.052	9	TG[-]	-0.089	15	GT[-]	-0.055	10	6	0.037	1	0.003
0	TG[+]	0.215	15	AC[-]	-0.175	4	CA[+]	0.221	16	11	0.39	1	0.006
1	TG[-]	-0.136	15	AC[-]	-0.108	13	CA[-]	-0.144	16	2	0.028	1	0.008
2	TG[-]	-0.075	14	AC[-]	-0.029	10	CA[-]	-0.083	16	4	0.046	2	0.008
3	TG[-]	-0.089	15	AC[-]	-0.052	9	CA[-]	-0.092	16	6	0.037	1	0.003
0	TC[-]	-0.015	6	AG[+]	0.176	12	GA[-]	-0.009	7	6	0.19	1	0.005
1	TC[-]	-0.006	8	AG[-]	-0.043	10	GA[-]	-0.003	7	2	0.036	1	0.004
2	TC[-]	-0.015	8	AG[-]	-0.081	15	GA[-]	-0.006	7	7	0.066	1	0.009
3	TC[-]	-0.064	12	AG[-]	-0.081	13	GA[-]	-0.059	11	1	0.017	1	0.005
0	GA[-]	-0.009	7	CT[+]	0.171	11	TC[-]	-0.015	6	4	0.18	1	0.005
1	GA[-]	-0.003	7	CT[-]	-0.034	9	TC[-]	-0.006	8	2	0.031	1	0.004
2	GA[-]	-0.006	7	CT[-]	-0.074	13	TC[-]	-0.015	8	6	0.068	1	0.009
3	GA[-]	-0.059	11	CT[-]	-0.081	14	TC[-]	-0.064	12	3	0.023	1	0.005
0	GT[-]	-0.175	3	CA[+]	0.221	16	AC[-]	-0.175	4	13	0.396	1	0
1	GT[-]	-0.112	14	CA[-]	-0.144	16	AC[-]	-0.108	13	2	0.032	1	0.005
2	GT[-]	-0.038	11	CA[-]	-0.083	16	AC[-]	-0.029	10	5	0.045	1	0.009
3	GT[-]	-0.055	10	CA[-]	-0.092	16	AC[-]	-0.052	9	6	0.036	1	0.003
0	CA[+]	0.221	16	GT[-]	-0.175	3	TG[+]	0.215	15	13	0.396	1	0.006
1	CA[-]	-0.144	16	GT[-]	-0.112	14	TG[-]	-0.136	15	2	0.032	1	0.008
2	CA[-]	-0.083	16	GT[-]	-0.038	11	TG[-]	-0.075	14	5	0.045	2	0.008
3	CA[-]	-0.092	16	GT[-]	-0.055	10	TG[-]	-0.089	15	6	0.036	1	0.003
0	CT[+]	0.171	11	GA[-]	-0.009	7	AG[+]	0.176	12	4	0.18	1	0.005
1	CT[-]	-0.034	9	GA[-]	-0.003	7	AG[-]	-0.043	10	2	0.031	1	0.009
2	CT[-]	-0.074	13	GA[-]	-0.006	7	AG[-]	-0.081	15	6	0.068	2	0.007
3	CT[-]	-0.081	14	GA[-]	-0.059	11	AG[-]	-0.081	13	3	0.023	1	0
	Mean	-0.036	11.25		-0.036	11.25		-0.036	11.25	5	0.101	1.125	0.005
	S D	0.099	3.614		0.099	3.614		0.099	3.61	3.221	0.121	0.3307	0.003
									P_tSi =	4.00E-09		P_tSe =	2.00E-05

Nomenclature as in Table 3 and 5. D Ind-a-Par = distance between the index dinucleotide and the antiparallel dinucleotide; Si = significance; P_tSi = probability with the Student's test for significance I-Par vs I-a-Par; P_tSe = probability with the Student test for selection coefficients, I-Par vs I-a-Par

Citation: Carlos Y Valenzuela. Differences in Selective Profiles Between H. Sapiens and SARS-Cov-2 Genomes Confirm Double or Single Stranded DNA or RNA. Journal of Biotechnology and Biomedicine. 8 (2025): 82-96.

Valenzuela CY., J Biotechnol Biomed 2025
DOI:10.26502/jbb.2642-91280181

Table 8: Distance Index-Parallel and a-Par dinucleotides when they are different pairs.

***SARS-CoV-2* total genome**													
	Indexes			**Par**			**a-Par**			**D Ind - Par**		**D Ind - a-Par**	
Sep	din[s]	c-se	Sig	din[s]	c-se	Sig	din[s]	c-se	Sig	D Si	D Se	D Si	D Se
0	AG[-]	-0.007	8	TC[-]	-0.198	4	CT[+]	0.181	13	4	0.19	5	0.188
1	AG[+]	0.025	5	TC[+]	0.067	2	CT[-]	-0.015	9	3	0.043	4	0.039
2	AG[+]	0.031	4	TC[+]	0.004	7	CT[-]	-0.032	11	3	0.027	7	0.063
3	AG[+]	0.029	6	TC[-]	-0.035	13	CT[+]	0.063	2	7	0.064	4	0.033
0	AC[+]	0.231	14	TG[+]	0.377	16	GT[+]	0.057	10	2	0.146	4	0.174
1	AC[-]	-0.007	8	TG[+]	0.016	6	GT[+]	0.123	1	2	0.023	7	0.13
2	AC[-]	-0.014	9	TG[-]	-0.134	16	GT[-]	-0.068	15	7	0.12	6	0.054
3	AC[+]	0.01	7	TG[+]	0.091	1	GT[+]	0.05	3	6	0.081	4	0.04
0	TG[+]	0.377	16	AC[+]	0.231	14	CA[+]	0.269	15	2	0.146	1	0.108
1	TG[+]	0.016	6	AC[-]	-0.007	8	CA[+]	0.042	4	2	0.023	2	0.027
2	TG[-]	-0.134	16	AC[-]	-0.014	9	CA[+]	0.011	6	7	0.12	10	0.145
3	TG[+]	0.091	1	AC[+]	0.01	7	CA[-]	-0.014	11	6	0.081	10	0.104
0	TC[-]	-0.198	4	AG[-]	-0.007	8	GA[-]	-0.08	6	4	0.19	2	0.117
1	TC[+]	0.067	2	AG[+]	0.025	5	GA[-]	-0.055	14	3	0.043	12	0.122
2	TC[+]	0.004	7	AG[+]	0.031	4	GA[-]	-0.027	10	3	0.027	3	0.031
3	TC[-]	-0.035	13	AG[+]	0.029	6	GA[-]	-0.003	9	7	0.064	4	0.032
0	GA[-]	-0.08	6	CT[+]	0.181	13	TC[-]	-0.198	4	7	0.261	2	0.117
1	GA[-]	-0.055	14	CT[-]	-0.015	9	TC[+]	0.067	2	5	0.04	12	0.122
2	GA[-]	-0.027	10	CT[-]	-0.032	11	TC[+]	0.004	7	1	0.005	3	0.031
3	GA[-]	-0.003	9	CT[+]	0.063	2	TC[-]	-0.035	13	7	0.065	4	0.032
0	GT[+]	0.057	10	CA[+]	0.269	15	AC[+]	0.231	14	5	0.212	4	0.174
1	GT[+]	0.123	1	CA[+]	0.042	4	AC[-]	-0.007	8	3	0.081	7	0.13
2	GT[-]	-0.068	15	CA[+]	0.011	6	AC[-]	-0.014	9	9	0.078	6	0.054
3	GT[+]	0.05	3	CA[-]	-0.014	11	AC[+]	0.01	7	8	0.063	4	0.04
0	CA[+]	0.269	15	GT[+]	0.057	10	TG[+]	0.377	16	5	0.212	1	0.108
1	CA[+]	0.042	4	GT[+]	0.123	1	TG[+]	0.016	6	3	0.081	2	0.027
2	CA[+]	0.011	6	GT[-]	-0.068	15	TG[-]	-0.134	16	9	0.078	10	0.145
3	CA[-]	-0.014	11	GT[+]	0.05	3	TG[+]	0.091	1	8	0.063	10	0.104
0	CT[+]	0.181	13	GA[-]	-0.08	6	AG[-]	-0.007	8	7	0.261	5	0.188
1	CT[-]	-0.015	9	GA[-]	-0.055	14	AG[+]	0.025	5	5	0.04	4	0.039
2	CT[-]	-0.032	11	GA[-]	-0.027	10	AG[+]	0.031	4	1	0.005	7	0.063
3	CT[+]	0.063	2	GA[-]	-0.003	9	AG[+]	0.029	6	7	0.065	4	0.033
	Mean	0.031	8.281		0.031	8.281		0.031	8.281	4.938	0.094	5.313	0.088
	SD	0.11	4.488		0.11	4.488		0.11	4.488	2.358	0.071	3.056	0.053
									P_tSi =	0.295		P_tSe =	0.36
Nomenclature as in Tables 1, 5, 7													

as dinucleotides whose bases are contiguous are concerned. In addition, results show that any base interacts with any base of a genome, so there is a pervasive co-adaptive base-to-base behavior within genomes. Besides that, the *SARS-CoV-2* genome showed a categorical 3K periodicity of the distance to neutrality that is impossible for neutral or nearly neutral evolution (see Table 2). Chromosomes or genomes are co-adaptive structure-organizations in agreement with the Wright's adaptive (selective) peak concept [9,19,20], now I re-define as integrated co-adaptive peaks, a concept close to that of the last article of Wright [20]. In the case of HCh21, the χ^2_9 is 4,503,674.869 and the probability for the occurrence of this distance to the neutral expectancy is $P < 10^{-530,762.1}$ a so low probability as to think HCh21 was determined since the Big Bang or before. This is the smallest of the 23 human chromosomes. If this is the distance to neutrality or randomness, see figures 6 and 7 in [29], neutral and nearly neutral evolution (if they exist) do not produce a detectable effect and neutral and nearly neutral evolution are conclusively refuted. Moreover, the periodicity of the distance to neutrality by itself indicates that neutral and nearly neutral evolution are not possible. The discussion still present, on neutralism as a refuted theory [31] and a valid theory based on untenable conceptualizations such as neutral fixation or genetic drift taken as a directional evolutionary factor [32] may endlessly continue. However, this is a discussion based on wrong arguments. Thus, it is necessary for me to repeat my previous analyses [8]. This arose from a widespread misconception of neutralists who considered that Prof. Wright [9,19] gave a big importance to drift as to convert it in a directional evolutionary force that could fixate or eliminate genes [1,5,6,10,11,32]. As I stated [8] on the neutralist's belief "Wright in his later years, used to claim that he had never attributed any significance to random drift except as an agent to bring about shift of adaptive peaks… however, Wright in his papers of the early 1930s used to attach much more weight to random drift" [6]. This was a regrettable misreading of Wright (1931)'s article. Wright (1931) [9] not only did not establish this importance or weight to drift but he stated categorically that neutral fixation and elimination were impossible: "…if mutation is occurring, however low the rate, the decline in heterozygosis, following isolation of a relatively small group from a large population, cannot go on indefinitely. There will come a time when the chance elimination of genes will be exactly balanced by new genes arising by mutation" and "It only requires a very moderate mutation rate in a large population for the number of unfixed loci to become enormous". Prof. Wright described clearly the resilient equilibrium given by mutation-drift that occurs equally at any site, where fixation and elimination are impossible [9]. As I indicated, destiny was not a word used by neutralist once, the use of colloquial words for randomness continued until the death of Prof. Kimura (in 1994), who in 1993 [1] wrote "The term 'survival of the fittest' is often equated with Darwinian theory of natural selection. Paraphrasing this, I proposed… 'survival of the luckiest' … to emphasize the importance of good fortune …" [1]. Good, fortune or lucky do not exist in the scientific language or in biotic world, randomness or drift is not chance, fortune or lucky, regardless the outcome of a random process. The fittest, randomness, drift, probability can be scientifically defined, studied and tested; fortune, chance, the luckiest, destiny cannot. I remark that the STE works with loci or nucleotide sites and with resilient equilibria of alleles or bases in them; NTE and NNTE work with alleles or bases and their isolate behavior; the dialogue go on parallel lanes and agreement is impossible. Neutral fixations may be or correspond to adaptive peaks within the adaptive landscape in the Wright's language [9,19,20]. From the foundations of the present research based on the resilient equilibrium of bases in a site, neutral and nearly neutral evolution are logically impossible. If the four bases A, T, G and C in a site are either negatively or neutrally selected life is impossible, because these conditions lead inexorably to the extinction of the species. The four neutral bases (fitness = 1.0) cannot recover a population after a contingent reduction. At least one base must have a highly positively fitness for the species survives, a central condition for STE.

In face of this and other much more significant results, I cannot avoid a reference to the intelligent design debate (as a general subject, not as an ideological debate). In Spanish, this debate is non-sense because design has two different meanings with different words. Design means: 1) diseño (idea, conception, drawing, sketch, outline, an object of art) that is always intelligent and 2) designio (will, plan, objective, intention) that applied to the origin and maintenance of the universe is always a matter of faith. These results indicate that the information to construct or develop anything (chromosomes in this case) in the universe was present at its origin (Big Bang?). If this is not accepted, we are forcedly to accept a source of dark information that permanently introduces information into the universe (or the unfounded production of an emergent situation). We have two possibilities: either there is a supra universe of trans-matter-energy (light or dark) and trans-dimensional existence that originates and maintains the total existence, or the universe, or existence is self-producing and self-maintaining (a kind of autopoiesis of Maturana and Varela). Thus, the intelligent design debate disappears and the matter-energy deterministic vision of the universe absorbs it completely. The sequence of mutational, contingent, selective and random events of the evolutionary process was determined since the origin or it is eternally self-determined (Determinism of Einstein, the Deity of Newton, the Laplace vast Intelligence and other conceptions [36]). Thus, evolution is the most intelligent (adaptive) process (design)

The large deficiency of CpG pairs is the most significant deviation from neutrality, in both genomes; this is a genetic factor not produced by this dinucleotide physical analysis; it indicates a co-evolutionary process that occurs in the host (mammals, in this case, humans) and the virus. I found the same result with HIV [25,30,35]. Therefore, the mechanism of inactivation of DNA or RNA by methylation of cytosine operates for the three genomes. However, HIV behaves rather a double stranded virus [30]. I concluded that, as virus strains suffers selection, they and their products assimilate to the host DNA, RNA or epitopes of proteins. Is it possible to develop a therapy to attacks only the viral RNA (in this case) based on the host inactivation or destruction mechanisms?

The most important result is the difference between the small I-a-Par distances (mean = 1.125) vs the large I-Par distances (mean = 5.0) in HCh21, and the similar large distances of both comparisons in *SARS-CoV-2* (5.313 vs 4.938, respectively). Even though the four distances are under the expected theoretical one (5.667). This implies that the evolutionary behavior of these organisms is not neutral. The result confirms these tests, as showing the single or double stranded nature of nucleic acids [30] known in the virion state is a partial and often mistaken view of the viral life cycle. The difference in double stranded DNA of I-a-Par and I-Par distances may happen because the test explores the selective behavior of the Par dinucleotide, which has the 3'-5' direction in the complementary strand, and the 5'-3' direction in the index strand, while the I-a-Par has the 5'-3' direction in the complementary and in the index strand. This is a passionate open field of research. It is evident that in human double stranded DNA the Index selective profile of a dinucleotide is almost identical with its anti-parallel selective profile and different from its parallel selective profile. This indicates that the double stranded condition of DNA and the 5'-3' or 3'-5' polarity are among the most important selective evolutionary traits. Of course, single stranded viruses (or nucleic acids) cannot yield these differences.

These tool and analyses were developed to search for non-neutral co-evolution of bases in DNA, but now I found that they are also a powerful test to discriminate between single and double stranded DNA or RNA within its evolutionary and life cycle behavior; because in the virion the result is different. I indicated the utility of this test in phylogenetic and comparative analyses [29,30] with large taxa; however, they are complementary and different from the current analyses that include some neutral assumptions to calculate quantitative distances based on nucleotide sequences. The phyletic distances detected by the present analyses are rather qualitative and are not adequate for nucleotide sequence analyses. The synthesis needs a large and complex work.

Author's Contribution

CYV carried out all the work described in this article

Author's information

It is sufficient the information given in the title page, methods and references

Acknowledgements

My student Skender Xhemale helped me in processing data. Anonymous reviewers improved the article. Criticism of visitors of my presentations in congresses and conferences contributed to the final manuscript.

Competing Interests

I do not have any competing interest.

Availability of Data and Material

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

Consent for Publication

The author approves the publication of this manuscript.

Ethics approval and consent to participate

Not applicable.

Funding

This work and the article did not have extraordinary funds.

References

1. Kimura M. Retrospective of the last quarter century of the neutral theory. Jpn J Genet 68 (1993): 521-528.
2. Ohta T. The Nearly Neutral Theory of molecular evolution. Annu Rev Ecol Syst 23 (1992): 263-286.
3. Kutschera U, Niklas KJ. The modern theory of biological evolution: an expanded Synthesis, Naturwissenschaften 91 (2004): 255-276.
4. Kimura M, Ohta T. On some principles governing molecular evolution, PNAS 71 (1974): 2848-2852.
5. Kimura M. The neutral Theory of molecular evolution: A review of recent Evidence. Jpn J Genet 66 (1991a): 367-386.
6. Kimura M. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. PNAS 88 (1991b) 5969-5973.
7. Ohta T. Origin of the neutral and nearly neutral theories of evolution. J Biosci 28 (2003): 371-377.

8. Valenzuela CY. Foundational errors in the Neutral and Nearly-Neutral theories of evolution in relation to the Synthetic Theory. Is a new evolutionary paradigm necessary? Biol Res 46 (2013): 101-119.

9. Wright S. Evolution in Mendelian populations. Genetics 16 (1931): 97-159.

10. Kimura M. Evolutionary rate at the molecular level. Nature 217 (1968): 624-626.

11. M. Kimura. The neutral theory of molecular evolution. Sci Am 241 (1979) 94-104.

12. King JL, Jukes TH. Non-Darwinian evolution. Science 64 (1969): 788-798.

13. Jacquard A. The Combined Effects of Different Evolutionary Forces in The Genetic structure of populations. Biomathematics, Volume 5, (1970) (eds. Krickeberg K, Lewontin RC, Neyman J, Schreiber M.) 388-418. New York: Springer-Verlag.

14. Crow JF, Kimura M. An Introduction to Population Genetics Theory (1970) New York: Harper and Row.

15. Nei M. Molecular Evolutionary Genetics (1987) New York, NY. Columbia University Press.

16. Valenzuela CY. Santos JL. A model of complete random molecular evolution by recurrent mutation. Biol Res 29 (1996): 203-212.

17. Li WH. Molecular Evolution (1997). Sunderland: Sinauer Associates.

18. Valenzuela CY. Misconceptions and false expectations in neutral evolution. Biol Res 33 (2000): 187-195.

19. Wright S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings of the sixth international congress of genetics (1932):356–366.

20. Wright S. Surfaces of selective value revisited Am Nat 131 (1988): 115-123.

21. Kreitman M. The neutral theory is dead. Long live the neutral theory, Bioessays 18 (1996): 678-683.

22. Ayala FJ, Barrio E, Kwiatowski J. Molecular clock or erratic evolution? A tale of two genes. PNAS 93 (1996): 11729-11734.

23. Karlin S, Mrazek J. Compositional differences within and between eukaryotic Genomes. PNAS 94 (1997): 10227-10232.

24. Ayala FJ. Neutralism and selectionism: the molecular clock. Gene 261 (2000): 27-33.

25. 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.

26. Valenzuela CY. Internucleotide correlation and nucleotide periodicity in Drosophila mtDNA: New evidence for panselective evolution. Biol Res 43 (2010): 481-486.

27. Valenzuela CY. Heterogeneous periodicity of drosophila mtDNA: new refutations of neutral and nearly neutral evolution. Biol Res 44 (2011): 283-293.

28. Valenzuela CY. The structure of selective dinucleotide interactions and periodicities in D melanogaster mtDNA. Biol Res 47 (2014): 1-12.

29. Valenzuela CY. Selective intra-dinucleotide interactions and periodicities of bases separated by K sites: a new vision and tool for phylogeny analyses. Biol Res 50 (2017): 3-16.

30. Valenzuela CY. Selective Profiles among Single or Double Stranded DNA or RNA Viruses Detect their Double or Single Stranded Condition. Arch Microbiol Immunol 8 (2024): 84-95.

31. Kern AD, Hahn MW. The neutral theory in light of natural selection, Mol Biol Evol 35 (2018): 1366-1371.

32. Jensen JD. Payseur BA. Stephan W, et al. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. Evolution 73 (2019): 111–114.

33. Valenzuela CY. Non-random DNA evolution. Biol Res 30 (1997) 117-123.

34. Bernardi G. The genome: an isochore ensemble and its evolution, Ann N Y Acad Sci 1267 (2012): 31–34.

35. Valenzuela CY, Flores SV, Cisternas J. Fixations of the HIV-1 env gene refute neutralism: new evidence for pan-selective evolution. Biol Res 43 (2010): 149-163.

36. Valenzuela CY. Distancia al azar de dinucleótidos y diseño inteligente generalizado. Int J Biol Nat Sci 3 (2023): 1-8.