BR

Internucleotide correlations and nucleotide periodicity in *Drosophila* mtDNA: New evidence for panselective evolution.

Carlos Y Valenzuela

Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile, Independencia 1027, Santiago, Chile

ABSTRACT

Analysis for the homogeneity of the distribution of the second base of dinucleotides in relation to the first, whose bases are separated by 0, 1, 2,… 21 nucleotide sites, was performed with the VIH-1 genome (cDNA), the *Drosophila* mtDNA, the *Drosophila* Torso gene and the human β-globin gene. These four DNA segments showed highly significant heterogeneities of base distributions that cannot be accounted for by neutral or nearly neutral evolution or by the "neighbor influence" of nucleotides on mutation rates. High correlations are found in the bases of dinucleotides separated by 0, 1 and more number of sites. A periodicity of three consecutive significance values (measured by the χ^2_9) was found only in *Drosophila* mtDNA. This periodicity may be due to an unknown structure or organization of mtDNA. This non-random distribution of the two bases of dinucleotides widespread throughout these DNA segments is rather compatible with panselective evolution and generalized internucleotide co-adaptation.

Key terms: nucleotide heterogeneity, DNA periodicity, dinucleotide analysis, neutral evolution, panselective evolution.

INTRODUCTION

Most, if not all studies on molecular evolution have been made assuming that evolution is related, directly or indirectly, to protein synthesis and to the genetic code (Nei, 2005, Nei et al., 2010). The evolutionary mechanisms that gave rise to the present genetic code among several other, the selection of four bases, the length and shape of chromosomes, the secondary and tertiary organization of nucleic acids, non-protein-coding DNA (i.e. most of eukaryote DNA), base isochores and signatures and several other pre-transcriptional or non-transcriptional traits are seldom or never studied (Valenzuela, 2007, 2009; Valenzuela et al., 2010). In a previous article the non-random association of a dinucleotide base pair, where the first base was separated by 0, 1, 2, and 3 nucleotide sites from the second base, showed an almost pan-selective evolution in the HIV-1 complete genome and in the corresponding segment of the *env* GP120 gene of its envelope (Valenzuela, 2009). Our conclusion was that selection made HIV-1 genomes, whose nucleotide bases were correlated with those of the neighborhood, along with the whole genome, have a higher probability to remain in the population; this internucleotide non-random distribution indicated a internucleotide co-adapted organization of the HIV-1 genome. This could occur because of the need of the virus' RNA to be folded and put into the envelope. Thus, the tertiary RNA or DNA structure by itself may be regarded as having been submitted to strong selection pressure, beside selection due to transcriptional or protein functions. We also found that in 103 HIV-1 strains, chosen from very separate regions in 35 countries from 5 continents, fixation at every site of the GP120 env gene, distributed as if each site had its own pattern of mutation. This independent evidence also affirms the condition of pan-selective evolution (Valenzuela et al., 2010). The present study analyses the complete HIV-1 genome, the complete *Drosophila* mtDNA genome and two eukaryote nuclear genes in search of internucleotide correlations, along with whole genomes or DNA segments.

RATIONALE, DATA AND METHODS

RATIONALE

Base frequency expectancies under evolutionary models

An allele in a gene locus or a base in a nucleotide site does exist in frequency p = 1.0 (fixation), 1.0>p>0.0 (polymorphism) or 0.0 (elimination or loss); these are the exhaustive classes of the state of alleles or bases in a locus or site, respectively, in a population. Our work deals with bases in sites, thus we shall refer to bases Adenine (A), Thymine (T), Guanine (G) and Citosine (C) in a nucleotide site. More precisely, we study the genetic equilibrium of bases in nucleotide sites (Valenzuela et al., 2010). The present theories of evolution, namely the Synthetic Theory of Evolution (STE), the Neutral Theory of Evolution (NTE) and the Nearly-Neutral Theory of Evolution (NNTE) agree in that the genetic variation in evolution emerges by mutation and that selection or drift (random fluctuations of base frequencies) are the main factors (less important factors are migrations, assortative mating in sexual species) that lead mutations to substitution, fixation, loss or to remain polymorphic. However, the different theories disagree in the importance of these factors. The STE proposes that fixations, losses or polymorphism are mostly due to selection (adaptive processes) and drift rarely contributes to evolution, except in extinction processes; drift is not a directional evolutionary force, its expected contribution (by the nature of random processes) to evolutionary processes is zero, on average. The NTE proposes that most of evolution has occurred at random (by drift) and selection rarely contributes to evolution through purifying selection (lethal

Corresponding author: Carlos Y Valenzuela, Programa de Genética Humana. Instituto de Ciencias Biomédicas (ICBM), Facultad de Medicina, Universidad de Chile, Independencia 1027, Casilla 70061, Independencia, CHILE, FAX (56-2) 7373158; Phone (56-2) 9786302, E. Mail < cvalenzu@med.uchile.cl >

or semi-lethal base conditions); positive selection is infrequent. The NNTE proposed the same as the NTE, but adds some selection processes with selection coefficients similar in value to mutation rates; the inclusion by the NNTE of a selection coefficient similar to $1/N$ (N is the population size) lead to inconsistencies when populations change their N, making evolutionary processes fluctuate between the STE and NNTE models (Kimura, 1991a, 1991b, 1993; Ohta, 1992; Ohta and Gillespie, 1996; Nei, 2005; Leigh, 2007; Valenzuela, 2007, 2009, 2010a, 2010b; Nei et al., 2010, Valenzuela et al., 2010). The value of rare or small contributions to specify the quantitative participation of selection or drift in evolution has never been given so as to test them with data. NNTE models have included all the values of selection coefficients (even positive one) in variable proportion, interacting with drift at various levels, so that they cannot be differentiated from STE models (Ohta and Gillespie, 1996; Ohta, 2002; Valenzuela, 2010a). Fortunately, these theoretical models lead to sharp expectancies in relation to base frequencies in a site and can be tested with data (Valenzuela, 2009; Valenzuela et al., 2010).

NTE expectancies for equilibrium frequencies

The neutral (random, NTE) expectancy of equilibrium frequencies for the four bases in a site, due to equal random forward and backward mutation rates, is (by definition of neutrality and randomness) $1/4$ A, $1/4$ T, $1/4$ G, $1/4$ C (Li 1997, Nei, 1987, Valenzuela et al., 2010). This expectancy changes according to the actual mutation rates given by the matrix of mutation from one base to the others (Nei, 1987, Valenzuela et al., 2010), but the expected equality of frequencies between A and T and between G and C due to base complementariness is a strong constraint (Sueoka, 1995; Valenzuela, 1997). Thus, the expected neutral situation for a locus or a site is a polymorphism of alleles or bases, respectively, and not fixation or loss, regardless of the population size. With recurrent forward and backward mutation, fixation is impossible (Wright, 1931; Feller, 1951; Jacquard, 1970; Valenzuela and Santos, 1996; Valenzuela, 1997, 2000, 2002, 2007, 2009; Valenzuela et al., 2010). The fact that gene or base fixation is impossible has not been accepted because there is a regrettable confusion between replacement or substitution and fixation (Valenzuela, 2000, 2002, 2007, 2009; Valenzuela et al., 2010). This confusion originated since the founding articles of NTE (Kimura, 1957, 1962, 1968, 1993; King and Jukes, 1969). Substitution is the arrival of an allele or base at frequency 1.0 (a turnover process); fixation is the permanence of an allele or base at frequency 1.0 (a fixated state); they are antithetical physical processes. Both are dimensionally different: substitutions (sub) occur per site (s) or locus, per generation (g) or year (sub/s/g); fixations (fix) occur per site, but not per generation (fix/s, it may be expressed per set of data that are under analysis), because the number of generations during which a base remains fixated is very different from taxon to taxon (or biotic group) (Valenzuela, 2000, 2002, 2007, 2009; Valenzuela et al., 2010). All the articles on phylogenies use fixations (taxonomic characters), but they are presented as substitutions or replacements; this error is widespread. The error is also included in models, such as that of Nei et al. (2010), when defining the neutral theory

following the studies of Fisher and Wright, state that "the probability of fixation (u) of a new mutant allele (A_2) in the population is …", however, this is not the probability of fixation, it is the probability for a new mutant allele to reach the frequency 1.0 (substitution). Once an allele or base has reached the frequency 1.0, the probability to remain at this frequency (fixation) is 0, because in the next generation its frequency should be (1-m), where m is the mutation rate (Wright, 1931, Feller, 1951; Valenzuela and Santos, 1996; Valenzuela, 2000, 2002, 2007, 2009). In the n_{th} generation, the probability that this allele or base or their copies remain at frequency 1.0 is $(1-m)^n$; as $1>(1-m)>0$, this probability tends to 0 as n increases (Valenzuela and Santos, 1996, Valenzuela, 2000). At the time of Wright and Fisher, the distinction between fixation and substitution was not made, however it is clear from the Wright (1931) article that fixation (a permanent state as we understand it at present) is 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". This agrees with the expected polymorphism of neutral frequency of bases in a site. To demonstrate conclusively that fixation is impossible independently of population size, a population of one bacterium was studied (Valenzuela, 2000); in this population (where drift is maximal) only monomorphism is possible at any site, but at in any generation a mutation can substitute the monomorphic base at this site, making fixation impossible. Fixation is factually impossible as is demonstrated in our daily life; we age, become ill, die by mutation; our genome is unstable during our life (Valenzuela, 2007, 2009).

NNTE expectancies for equilibrium frequencies

The expectancy of the equilibrium base frequencies in a site is in the NNTE, assuming a positive selection coefficient for a base, say A, equal to the forward mutation rate to T, G, and C, is near 0.43A, 0.19T, 0.19G, 0.19C [this is in haploid organisms; the demonstration is beyond the scope of this article, see Wright (1931)]. That is, the expectancy for a site according to NNTE is also a polymorphism of the four bases and not fixation or loss, regardless of the population size.

STE expectancies for equilibrium frequencies

The STE expectancy is also a polymorphism that depends on the mutation rate and the selection coefficient (sc). For a highly negative selected base in a site its expected equilibrium frequency is near m/s (Li, 1976). For example (working with a haploid organism), if $m = 10^{-8}$ and sc = -0.1 the expected frequency of this base is 10^{-7}; for a highly positive selected base (s = +0.1) its frequency is 1-m/s (complement of m/s), in the example it is near 0.9999999 (Wright, 1931). These highly positive selected bases appear fixated (some of them are used as taxonomic characters in phylogenies), but they are not fixated, they remain in an equilibrium frequency near 1, but mutations do occur at these sites inexorably. STE expected equilibrium frequencies may be any that sc and m stochastically determine.

Non-overlapping conditions of evolutionary models

Moreover, all these equilibrium frequencies (for the three models) are stable and resilient; they do not overlap nor are there overlapping situations among these models (except factual sample variations). If random fluctuations change frequencies from the equilibrium, the evolutionary factors tend to re-establish them immediately (equilibriums are resilient). The equilibrium frequencies, for the three models, are determined by the mutation rate and selection coefficient independently of N and drift. Neutral (random) fluctuations may move (transitorily) these equilibrium frequencies up or down, but their evolutionary contribution in thousands or millions of generations is finally zero on average.

Expectancies for nucleotide correlations

Now we provide a rational for the correlation among sites. As mutations occur independently (not necessarily at random) at each site, the correlation of bases among sites, in long DNA segments, is expected to be zero. Moreover, mutations occur without evolutionary "purpose", that is the causal mechanisms of mutation are different from the processes of selection and drift (they are independent). As mutations (in a wide sense including chromosome mutation, duplication, etc.) are accepted for all the evolutionary theories as the base for evolution, the fundamental expectation for internucleotide correlations, due to mutation, is zero. The NTE and the NNTE propose that, besides mutation, evolution (fixation, loss or polymorphism) in a nucleotide site is mostly determined by random genetic drift and selection or adaptive processes have limited importance. The expectancy for internucleotide correlation due to mutation is zero, now if we add random processes to this expectation, it continues to be zero; moreover, if the expectancy of correlations due to mutations was not exactly zero because of some small correlations among sites due to the hypothetical neighbor influence of a base on neighboring mutation rates, drift contributes to blur these small correlations and bring them nearer zero. Gatlin (1976) found by studying longitudinal nucleotide or amino acid series that there was a high correlation among nucleotides in genomes. This was assumed as a demonstration of non-neutrality of evolution. Neutralists counter-argued: this was not a refutation of neutralism because a base could influence mutation rates of the neighbor sites (Jukes, 1976; Kimura and Ohta, 1977). But neutralists have never demonstrated this neighborhood influence nor shown that Gatlin's order of nucleotides was produced by a factual neighbor influence. This debate ended without solution, and the battle was assumed to be won by neutralists. Moreover, even accepting the neighbor influence, the expected frequencies of bases at sites are homogeneously distributed among sites and continue to be mostly an independent polymorphism for each site. Historically, and since the expected situation for a site is a polymorphism of the 4 bases, the expected influence on neighbor sites is this historical influence of the 4 bases, that is, as an average, equal for all the sites, regardless of the actual base in each site. The neighbor influence proposed by neutralists is only partially possible for small neighborhoods (2 or 3 sites up and downstream) if, and only if, non-selective (neutral or random) fixation for millions of generations can occur. Since fixation is impossible, finding a high correlation between both bases of a dinucleotide or pair of bases separated by more than 2 sites is a strong refutation of neutralism, near-neutralism and the neighbor influence hypothesis. This is equivalent to searching for co-adaptive processes among alleles of two or more loci. Finding non-random associations among nucleotides separated by K sites demonstrates co-nucleotide-adaptation. There are 4 bases for the first and second position in a dinucleotide, thus 16 dinucleotides are possible. If NTE or NNTE are true, the expected frequency of these 16 dinucleotides is obtained by the product of the total observed frequencies of the 4 bases in the analyzed DNA segment. If, as for example, the frequency of G is 0.2, the expected frequency of G-G pairs, whose bases are separated by 100 (or K) sites, is 0.04 and we found 0.12. This indicates that G-G pairs have been positively selected (and adapted to life conditions) and other pairs were negatively selected and did not remain during evolution in this DNA segment. Thus, these G-G pairs whose bases are separated by 100 sites are co-adapted. Mutation and drift alone cannot construct a meaningful (adapted) DNA segment (by the nature of both processes); only selection by environmental requirements can give the necessary prevalence of correlations among nucleotides sites, randomly produced, to obtain and maintain better adapted (to these environments) biased (statistically) sequences. The present study is aimed to study these biased sequences (association of bases of a dinucleotide). This is not a quantitative problem that can be solved by adjusting N, coefficients, types of mutations, historical processes, and so on; this is a qualitative (factual and conceptual) problem that needs to be rigorously handed. Of course, we should be cautious with statistical significances and work with a small type I error (alpha = 0.01 or less).

DATA

The total HIV-1 cDNA [Genbank accession number (GbAN) AF005495, number of sites (Ns) = 8954], the total mitochondrial DNA (mtDNA,) of *Drosophila melanogaster* (GbAN NC 001709, Ns = 19517), the DNA segment of the Torso gene of *Drosophila melanogaster* (Sprenger et al., 1989, Ns = 4744; Valenzuela, 1997) and the DNA segment of human beta globin (βHb) gene (Nei, 1987, Ns = 2051, Valenzuela, 1997) were studied. For nomenclature, the names of segments are those used in cited references.

METHODS

The analysis of dinucleotides (longitudinal pairs of mono-nucleotides) whose two bases were separated by 0 (consecutive bases), 1, 2, 3 … 21 nucleotide sites were performed. The first nucleotide of the pair was chosen consecutively from the first to the S-K$_{th}$ (S = total number of nucleotides of the genome or DNA segment; K = number of nucleotide sites that separates the first and the second nucleotide of the pair) nucleotide of the genome or DNA segment; the second nucleotide is chosen 0, 1, 2 … K nucleotide sites downstream from the first one. Two nucleotides, with 4 alternatives each, yield 16 dinucleotides whose frequency expectancies are obtained by the product of the frequencies of both mononucleotides calculated by the

marginal totals of the 4x4 matrix for the possible bases of the first and second nucleotide. In this analysis, the homogeneity or heterogeneity of the distribution of the second nucleotide base according to the first was tested by a χ^2_9 test for homogeneity distribution with 9 degrees of freedom (df, 4 rows minus 1, times 4 columns minus 1), noted as subscript. The critical values of the χ^2_9 for significance levels of probabilities 0.05, 0.025, 0.01, 0.005, 0.001 are 16.9 (17), 19.0 (19), 21.7 (22), 23.6 (24) and 27.9 (28), respectively. For very large χ^2_9 values, a conservative z test approximation was used, taking 9 (df) as the mean and 18 (2xdf) as the variance of the normal distribution. More specific details of the method are given in a previous article (Valenzuela, 2009) and results are presented.

RESULTS

Tables I and II present the distribution of dinucleotides separated by 0, 1, 2 and 3 sites for the HIV-1, *Drosophila* mtDNA, *Drosophila* Torso gene and human βHb. The analyses for the HIV-1 have been presented previously (Valenzuela, 2009), isolated and in a different format. The expected (Exp) values are those of the dinucleotides with 0 site separation, because there are minimal variations with higher separations (the proportion of bases remain almost the same). For any observed (Obs) value at a row, a χ^2_1 test can be estimated by the known formula $\Sigma[(Obs-Exp)^2/Exp]$. The χ^2_1 critical values are 3.84 (4), 5.02 (5), 6.64 (7), 7.88 (8) and 10.83 (11) for the probabilities 0.05, 0.025, 0.01, 0.005 and 0.001, respectively. Results are highly significant for HIV-1 and mtDNA for separations 0 and 1; their χ^2_9 values are beyond any table or computer program. In HIV-1, probabilities of 2 and 3 separations can be estimated by tables and computer programs, while in mtDNA they continue to be extremely low. The steady decay in significance from 0 to 3 separations was evident in HIV-1, while mtDNA significances did not decay steadily.

Table III presents the χ^2_9 values from 0 to 21 separations for the four DNA segments. We can easily read this table helped by the significant levels for the χ^2_9 given in methods. As for the example, the expected value for a χ^2_9 is 9 and significances at the 0.05 level are obtained with a χ^2_9 value equal to 17; the 0.01 level implies $\chi^2_9 = 22$. HIV-1 showed highly significant χ^2_9 values (by far larger than 22) for 17 of the 21 separations (excepting separations 7, 13, 18 and 21). All the separations of the *Drosophila* mtDNA were enormously significant. Most separations were not significant in the *Drosophila* Torso gene, however, the six

Table I

Expected and observed dinucleotides frequencies whose bases are separated by 0, 1, 2 and 3 sites. HIV-1 and *Drosophila* mtDNA

Pair Expected		HIV-1 genome*				*Drosophila* mtDNA				
			Observed			Expected		Observed		
		0S	1S	2S	3S		0S	1S	2S	3S
A - A	1169.3	1107	1233	1256	1190	3404.7	3594	3436	3521	3678
A - T	709.7	680	738	697	672	3292.8	3268	3334	3343	3078
A - G	785.2	941	703	750	791	617.8	599	557	560	630
A - C	570.9	607	561	522	582	836.7	691	825	727	765
T - A	709.9	663	601	688	738	3292.0	3168	3295	3397	3103
T - T	430.8	501	472	458	426	3183.7	3481	3265	3304	3399
T - G	476.7	507	529	458	479	597.3	518	566	480	600
T - C	346.6	293	361	358	319	809.0	715	755	700	778
G - A	785.4	727	887	789	719	617.7	646	630	513	564
G - T	476.7	379	395	430	520	597.4	364	573	517	580
G - G	527.4	646	495	568	527	112.1	230	130	226	124
G - C	383.5	421	396	386	412	151.8	239	146	223	211
C - A	571.4	739	515	503	593	836.6	743	789	719	805
C - T	346.8	404	358	378	345	809.1	770	711	718	825
C - G	383.7	79	446	386	375	151.8	132	226	213	124
C - C	279.0	359	262	314	267	205.6	358	277	353	249
Tot	χ^2_9	446	87	38	20	485	94	405	114	
Prob		<10^{-50}	<10^{-20}	<10^{-4}	0.017		<10^{-50}	<10^{-20}	<10^{-50}	<10^{-20}

0S, 1S, 2S, 3S = Separation of 0 (consecutive), 1, 2 and 3 nucleotide sites, respectively; χ^2_9 = Chi-squared test (9 df) rounded to the integer figure; Prob = probability. * = published in Valenzuela (2009).

significant separations at 0, 1, 5, 11, 14 and 20 nucleotide sites were so high (except at 14) as to render the whole analysis highly significant. In Human βHb there are 14 significant (most of them highly significant) and 8 non-significant values. Table III showed a remarkable structure of significances in *Drosophila* mtDNA, which is not present in the other DNA segments. From the separation of 2 sites, until 21, mtDNA showed a highly significant periodicity (several hundred) followed by two less significant ones (with the exception of separation 3, less than one hundred and close to 60). This periodicity was found between 300 and 309 separations and between 600 and 609 separations (Valenzuela, 2010a). In a further analysis, it was found between 1000 and 1009 and between 2002 and 2011 separations (Valenzuela, 2010c). Systematic significant values of the χ^2_9 were found for the HIV-1 until 320 K, for the Torso gene until 25 K and for the βHb gene until 70K.

DISCUSSION

These correlations between mononucleotides of a dinucleotide separated by 0, 1, 2 … K nucleotide sites make neutral and nearly neutral evolution untenable, as well as the hypothesis that the nucleotide neighborhood

systematically changes the mutation rate (see Rationale and Valenzuela, 1997, 2000, 2007, 2009; Valenzuela et al., 2010). As stated in the Rationale, these correlations are non-random (biased) frequencies of dinucleotides; thus they imply that there are excesses of dinucleotides (over the random expectation) that were positively selected and maintained by positive selection, and deficiencies of dinucleotides whose frequencies are less than randomly expected that were negatively selected and are maintained in lower frequencies by negative selection. Even in the case of Torso, the least significant DNA segment, any site correlates significantly with several sites that are separated until 25 sites upstream in the genome. These correlations are probably due to pre-transcriptional or non-transcriptional processes. The lower significance found in Torso and human βHb genes, may be due to their small number of sites; however, these segments have sufficiently significant results to consider them equally significant qualitatively, and the βHb gene shows a higher significance than Torso, in spite of it is half its length. The periodicity in the level of values of three consecutive χ^2_9 tests found only in *Drosophila* mtDNA was unexpected and deserves a more detailed and comparative analysis. It shows that there are other levels of organization of nucleic acids, at least, as far as mtDNA is concerned. Our analyses scanned

Table II

Expected and observed dinucleotides frequencies whose bases are separated by 0, 1, 2 and 3 sites.
Drosophila torso and human βHb genes

Pair Expected		*Drosophila* Torso				Human β globin gene				
		Observed				Expected	Observed			
		0S	1S	2S	3S		0S	1S	2S	3S
A - A	410.6	456	399	406	416	138.5	160	154	148	141
A - T	367.4	362	361	388	364	177.8	164	162	164	188
A - G	313.2	307	319	296	317	109.4	113	124	115	101
A - C	303.8	270	315	303	296	107.3	96	93	105	102
T - A	367.9	266	301	387	371	177.5	148	174	164	160
T - T	329.2	370	367	344	341	227.8	244	253	252	260
T - G	280.7	337	281	249	261	140.2	174	110	129	133
T - C	272.2	277	301	270	277	137.5	117	146	138	130
G - A	313.5	323	352	296	300	109.4	92	110	116	135
G - T	280.5	255	284	241	290	140.4	103	125	126	104
G - G	239.1	234	224	308	231	86.4	125	97	93	86
G - C	232.0	253	205	222	243	84.8	101	89	86	95
C - A	304.0	351	344	307	308	107.6	133	95	105	97
C - T	272.0	262	237	276	254	138.1	173	144	142	131
C - G	232.0	187	241	214	256	85.0	9	90	84	101
C - C	225.0	233	211	236	215	83.4	99	84	82	84
Tot χ^2_9		75	40	35	8		142	22	9	30
Prob		$<10^{-10}$	$<10^{-5}$	$<10^{-4}$	0.520		$<10^{-20}$	0.009	0.425	0.005

Nomenclature as in Table I.

these DNA segments completely, so those non-random associations are widespread throughout the four segments; this indicates that evolution is panselective, not neutral or nearly neutral. Any base is selected in the whole context of the genome of the individual to which it belongs. Thus, there is always a co-selective process that implies a co-adapted condition. DNA sequences (with their internucleotide correlations) that do not fit the environmental requirements are negatively selected and do not remain in the population.

Note: These ideas were presented at the Annual Meeting of the Chilean Society of Evolution and the Chilean Society of Genetics, in Concepción, Chile, October 21-23 2009.

TABLE III

χ^2_9 for homogeneity found in dinucleotides separated from 0 to 21 nucleotide sites in HIV-1, *Drosophila* mtDNA, *Drosophila* Torso gene and human βHb gene

Separation	HIV-1	mtDNA	Torso	Human βHb
0	445.55	485.10	74.91	141.75
1	86.98	94.27	40.36	22.10
2	38.28	405.03	5.98	9.13
3	20.09	114.23	8.14	29.73
4	30.66	46.51	9.47	9.97
5	21.29	380.50	27.21	37.36
6	24.83	86.76	12.40	22.15
7	9.42	36.73	13.13	18.38
8	40.74	375.33	16.19	14.80
9	17.87	75.66	3.07	26.10
10	21.96	49.35	8.83	10.19
11	56.49	367.13	35.60	42.00
12	31.09	65.28	12.67	17.30
13	10.22	69.95	10.70	14.42
14	22.98	310.01	17.68	23.10
15	19.58	60.10	4.95	9.70
16	17.06	51.75	13.71	18.18
17	17.28	322.18	8.56	9.47
18	13.85	65.78	1.69	17.03
19	19.23	40.60	5.86	21.60
20	23.28	387.99	25.89	18.89
21	7.41	53.16	10.74	15.58

REFERENCES

FELLER W (1951) Diffusion processes in genetics. Proceedings of the 2nd Berkeley Symp Math Stat Prob. Berkeley CA: University of California Press. pp: 227-246.

GATLIN LL (1976) Counter-examples to a neutralist hypothesis. J Mol Evol 7:185-195.

JACQUARD A (1970) The genetic structure of populations. Biomathematics Volume 5. New Cork, Springer -Verlag, pp 388-419

JUKES TH (1976) Comments on Counter-Examples to a Neutralist Hypothesis. J Mol Evol 8:295-297.

KIMURA (1957) Some problems of stochastic processes in Genetics. Ann Math Stat 28: 882-901.

KIMURA M (1962) On the probability of fixation of mutant genes in a population. Genetics 47: 713-719.

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

KIMURA M (1991a) The neutral Theory of molecular evolution: A review of recent evidence Jpn J Genet 66: 367-386.

KIMURA M (1991b) Recent development of the neutral theory viewed from the Wrightian tradition of theoretical populations genetics. Proc Nat Acad Sci (USA) 88:5969-5973.

KIMURA M (1993) Retrospective of the last quarter century of the neutral theory. Jpn J Genet 68:521-528.

KIMURA M, OHTA T (1977) Further Comments on "Counter-Examples to a Neutralist Hipótesis". J Mol Evol 9:367-368.

KING JL, JUKES TH (1969) Non-Darwinian evolution. Science 64:788-798.

LEIGH EG (2007) Neutral theory: a historical perspective. J Evol Biol 20:2075-2091.

LI CC (1976) First course in population genetics. Pacific Grove, CA, The Boxwood Press, pp 505-521.

LI WH (1997) Molecular Evolution. Sunderland: Sinauer Associates.

NEI M (1987) Molecular Evolutionary Genetics. New York, NY: Columbia University Press. p: 21.

NEI M (2005) Selectionism and neutralism in molecular evolution. Mol Biol Evol 22:2318-2342.

NEI M, SUZUKI Y, NOZAWA M (2010) The neutral theory of molecular evolution in the genomic era. Annu Rev Genomics Hum Genet. Jun 21 (on line) PMID:20565254.

OHTA T (1992) The Nearly Neutral Theory of molecular evolution. Annu Rev Ecol Syst 23:263-86

OHTA T (2002) Near-neutrality in evolution of genes and gene regulation. Proc Nat Acad Sci 99:16134-16137.

OHTA T, GUILLESPIE JH (1996) Development of Neutral and Nearly Neutral Theories. Theor Popul Biol 49:128-142.

SPRENGER F, STEVENS LM, NÜSSLEIN-VOLHARD C (1989) The *Drosophila* gene torso encodes a putative receptor tyrosine kinase. Nature 338:478-483.

SUEOKA N (1995) Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J Mol Evol 40:318-325.

VALENZUELA CY (1997) Non random DNA evolution. Biol Res 30:117-123.

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

VALENZUELA CY (2002) A biotic Big Bang. In: Palyi G, Zucchi C & Caglioti L (eds) Fundamentals of Life. Paris: Elsevier. pp: 197-202.

VALENZUELA CY (2007) Within selection. Rev Chil Hist Nat 80:109-116.

VALENZUELA CY (2009) Non-random pre-transcriptional evolution in HIV-1. A refutation of the foundational condition for neutral evolution. Genet Mol Biol 32:159-169.

VALENZUELA CY (2010a) Periodicidades e interacciones del DNA. El fin del neutralismo y del casi neutralismo (Editorial Universitaria, Santiago, Chile, textbook, in press).

VALENZUELA CY (2010b) Neutral illusions, selective nightmares and autopoietic madness (Rev Chil Hist Nat, in press)

VALENZUELA CY (2010c) Heterogeneous periodicity of drosophila mtDNA indicates generalized selective evolution (Biol Res in revision).

VALENZUELA CY, SANTOS JL (1996) A model of complete random molecular evolution by recurrent mutation. Biol Res 29:203-212.

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

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