Heterogeneous periodicity of drosophila mtDNA: new refutations of neutral and nearly neutral evolution

Carlos Y Valenzuela

Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile. Independencia 1027, Casilla 70061, Santiago, Chile.

ABSTRACT

We found a consistent 3-site periodicity of the χ^2_9 values for the heterogeneity of the distribution of the second base in relation to the first base of dinucleotides separated by 0 (contiguous), 1, 2, 3 … 17 (K) nucleotide sites in Drosophila mtDNA. Triplets of χ^2_9 values were found where the first was over 300 and the second and third ranged between 37 and 114 (previous studies). In this study, the periodicity was significant until separation of 2011K, and a structure of deviations from randomness among dinucleotides was found. The most deviant dinucleotides were G-G, G-C and C-G for the first, second and third element of the triplet, respectively. In these three cases there were more dinucleotides observed than expected. This inter-bases correlation and periodicity may be related to the tertiary structure of circular DNA, like that of prokaryotes and mitochondria, to protect and preserve it. The mtDNA with 19.517 bp was divided into four equal segments of 4.879 bp. The fourth sub-segment presented a very low proportion of G and C, the internucleotide interaction was weaker in this sub-segment and no periodicity was found. The maintenance of this mtDNA structure and organization for millions of generations, in spite of a high recurrent mutation rate, does not support the notion of neutralism or near neutralism. The high level of internucleotide interaction and periodicity indicate that every nucleotide is co-adapted with the residual genome.

Key terms: DNA organization, non-protein-coding evolution, ordered nucleotide sequences, inter-base associations, refutation of neutralism.

INTRODUCTION

Studies of evolution have assumed that most or all evolutionary processes are directly or indirectly related to protein synthesis and regulation (Nei, 2005; Valenzuela, 2009, 2010a; Nei et al., 2010). The acquisition and maintenance of the genetic code (as a whole) have been accepted non-critically as an out-of-evolution process. However, if the genetic code was acquired and is maintained by selective processes, all the other processes founded on the code are also selective. This non-critical position occurs with other aspects related to the structure, size and shape of the hereditary material (chromosomes, DNA or RNA segments), non-protein-coding functions and structures, replication functions (velocity, structural restrictions, etc.), number of bases, tandem repeat segments, isochores, signatures and several other properties of the hereditary material. For example, no systematic evolutionary studies have been performed on the acquisition and maintenance of isochores and signatures present in the biotic world for more than a thousand million years (Valenzuela, 2007, 2009, 2010a). Recurrent forward and backward point and chromosome mutations occurring equally (neutrally) at any nucleotide site destroy inexorably any chromosome or gene organization. These non-protein-synthesis and regulation-coding functions cannot be studied by the statistics of mutations, substitutions or fixations in relation to protein-coding functions (occurring at the first, second and third position of the codon), or in relation to associated functions as synonymous or non-synonymous fixations, or to biases of codon usage. The evolution of the chromosome structure (its constitution in centromeres, arms and telomeres) cannot be studied according to the genetic code for most chromosome segments have non-protein-coding functions. It is necessary to realize that protein-coding functions are a part of all the coding functions of the hereditary material. We can mention among these functions the folding-coding-functions for putting DNA or RNA viruses into their capsids or envelopes; coiling or hypercoiling-coding functions of prokaryotes or mitochondrion DNA; information for the relationships of DNA with histones or other associated proteins, information for the constitution of telomeres and centromeres, etc. These non-protein-coding characters and functions behave with non-Mendelian inheritance. Here, Mendelian inheritance is synonymous with particulate inheritance (Mendel's laws are cases of particulate segregation), in opposition to diffuse inheritance (Darwin's belief that the paternal and maternal "genetic factors" fuse in descendants). Thus, mitochondrial, prokaryote and virus inheritance, as we study them at present as far as genes are concerned, are fully Mendelian. The examination of a point mutation may help us to understand. When guanine (G) mutes to thymine (T) in a protein-coding segment, several kinds of phenotypes (pleiotropy) are produced (a mutation is always pleiotropic). I) a gene change leading to a synonymous or a non-synonymous mutation with Mendelian behavior. II) A structural change in the DNA because G (two chemical rings of purines) is larger than T (one ring of pyrimidines). This change is inherited as a Lamarckian character. III) A change in the velocity of replication and transcription because G-C implies 3 hydrogen bonds and T only 2 (a non-Mendelianly inherited character). IV) A change in the interactions with the residual genome (the remaining genome that is not this particular mutated site); this is partially a non-Mendelianly inherited trait and is the subject of this article. Other characters may be produced.

We analyzed these non-protein-coding functions depending on the nucleotide sequences by studying the relations and correlations (not only statistical) among all

* Independencia 1027, Casilla 70061, Independencia, Chile. FAX (56-2) 7373158; Phone (56-2) 9786302 E. Mail < cvalenzu@med.uchile.cl >

the nucleotides of a RNA or DNA segment, excluding any reference to protein-coding functions. Our study seeks to know the sequence information for tertiary or quaternary DNA structures (in relation to non-DNA molecules). We seek to affirm or refute neutral, nearly neutral or selective evolution (Valenzuela and Santos, 1996, Valenzuela, 1997, 2000, 2002, 2007, 2009, 2010a; Valenzuela et al., 2010). Unexpectedly, we found a very high correlation between both bases of a dinucleotide separated by 0, 1, 2 …K (K= 35) nucleotide sites in the whole genome of the HIV-1 virus, and in one *env* gene of this virus (Valenzuela, 2009). This correlation has nothing to do with coding-functions (it occurs between any nucleotide separated from the others by a number of sites that is or is not a multiple of 3). It is probably related to the tertiary structure needed to fold the RNA virus into its capsid. In a second study we found similar correlations until K= 21, in *Drosophila melanogaster* mtDNA, Gene Torso, and in human beta globin (βHb) gene (Valenzuela, 2010a). However, the significance of the interaction was observed until K = 309 in HIV-1 and until K = 609 in mtDNA (Valenzuela 2010b). These correlations between nucleotides separated by more than 4 sites do not support neutral or nearly neutral evolution and indicate that evolution of HIV-1, mtDNA, gene Torso and βHb are rather panselective. This conclusion follows from the fact that a huge number of dinucleotides, significantly more frequent than randomly expected, have been positively selected over hundreds of millions of cell generations. Dinucleotides that are less frequent than randomly expected have been negatively selected over and over again over hundreds of millions of cell generations. The maintenance of this strong interaction over hundreds of millions of DNA replication cycles can be only achieved by a widespread selective process, because recurrent mutation destroys any non-random nucleotide association.

Moreover, a three nucleotide periodicity was found only in mtDNA among these strong non-random distributions which is not related to protein synthesis coding functions (Valenzuela, 2010a, 2010b, this article). This periodicity was seen in the series of χ^2_9 values [9 degrees of freedom due to four rows (less one) for the first and fourth columns (less one) for the second nucleotide of the dinucleotide] that measure the distance to randomness (neutrality). It should be emphasized that this periodicity is not related to protein-coding processes due to definite conditions: I) Among gene segments, both strands of the mtDNA code for tRNA, rRNA and mRNA in the opposite 3'-5' sense; however, in the present study the analysis of correlations is performed in one strand. A few genes overlap a small part of their sequences, but most of them are separated by a few nucleotide sites. Thus most if not all correlations of nucleotides separated by K sites calculated on the whole mtDNA do not coincide with any long series of codon positions. II) Within a gene segment the periodicity (for small K) is found in dinucleotides whose bases belong to the same or to another codon (see APPENDIX 1). III) When K is large and, since the largest gene segment (ND5) in this mtDNA has 1723 sites, the two bases belong mostly to different genes. IV) Since mtDNA codes for tRNA, rRNA and mRNA the correlations among bases of a dinucleotide separated by large K occur often between nucleotides belonging to these three kinds of DNA (coding positions are defined only for mRNA). V) Some correlations are found between bases belonging to coding and non-coding segments. VI) We have found that the significant

periodicity extends to more than 600K (Valenzuela, 2010b) or 2000 K (this article), a distance beyond any protein-coding mtDNA segment.

The extension of this analysis to other mtDNA or genomes showed the same result (Valenzuela 2010a, 2010b, 2011). To test our program we examined the collagen type I alpha 2 gene (it is a periodical gene) and as was expected, a highly significant association of the bases of a dinucleotide was found every 3 and 9 nucleotide sites. Significant non-random associations and periodicities were found in long prokaryote genes, but not in non-periodical eukaryotes genes (Valenzuela 2011) or in short prokaryote genes (Valenzuela unpublished). Significant deviations from randomness and periodicities were present in these genomes until K = 1007 or more (Valenzuela 2011, this article). It is possible to think that these non-random interactions and periodicities are due to mathematical or statistical artifacts from trivial properties of polymers; or that the highly significant association between contiguous nucleotides generates the others. The present study intends to show that these are not the case and that there is a systematic genetic structure underlying the base associations in dinucleotides separated by 0, 1, 2 … K sites, in the *Drosophila melanogaster* mtDNA (19,517 bp) and in four equal consecutive segments of 4,879 bp.

RATIONALE, DATA AND METHOD

RATIONALE

The expected internucleotide correlation under mutation alone

In the present disciplinary matrix of evolution, mutation occurs independently of the following fate of the mutant allele or base and independently of the processes of natural selection or genetic drift (Prevosti, 2000; Valenzuela and Santos, 1996; Valenzuela, 2000, 2002a, 2011a). The mechanisms of mutation and repair occur with their own matter-energy characteristics. Thus, the occurrence of mutation at any site is mostly independent on the occurrence of mutation at any other site. This does not mean that mutation occurs at random, because it is known the variation of the mutation rate (cold-, hot-, normal-spots; Valenzuela and Santos, 1996; Li, 1997; Valenzuela, 2000), and mutation seems to be influenced by the neighborhood, at least in laboratory conditions with mutagens acting on viral RNA (Koch, 1971). The mutation rate varies enormously from organism to organism, but it is similar in similar organisms with some exceptions (Drake, 1993,1999, 2009; Drake et al., 1998; Mackwan et al., 2008). It is assumed that equal neighbors have equal mutation rates; that is mutation rates occur with isotropy in DNA or RNA nucleotide sites (Valenzuela, 1997). If we consider only mutations, under neutral evolution the 4 bases are expected to be in a site with equal probability along with evolution during a number of generations larger than the inverse of the mutation rate (Valenzuela and Santos, 1996; Valenzuela, 1997, 2000, 2002). Then, the expected historical correlation of two neutral bases located in two different sites is zero. If the neighbor influence operates depending on one upstream and one downstream base, there are, for every base, 16 different contexts with 16 different influences on mutation rates that yield an average mutation rate for all these contexts (with 2 sites of influence 256 contexts are produced). However, these contexts should

be, in turn, influenced by their contexts of each upstream base and each downstream base, and these second sets of contexts should be influenced by the bases that are 3 up- and 3 down-stream sites, and so on. Since all the bases are continuously mutating (fixation is impossible), it is expected that the average difference in neighbor influence on every mutation rate should be small or zero. Independently of these factors, mutation and its possible neighbor influence cannot generate the meaning (a wide internucleotide correlation among all the sites needed to code a protein or a biotic hermeneutics) of the DNA segments by neighbor influence of bases on the mutation rate, because it occurs independently of the environmental requirements for the living being. Mutation is, from a biotic viewpoint, hermeneutically powerless (Valenzuela, 2009, last paragraph). Thus the expectation for a correlation between the bases of a dinucleotide that are separated by 0, 1, 2 …K sites is zero or near zero. Gatlin (1976) found non-random distribution of longitudinal nucleotide sequences and proposed this sequential order was a proof for non-neutral evolution. Neutralists answered fast (Jukes, 1976; Kimura and Ohta, 1977) proposing that the neighbor influence of bases on mutation rates could explain this order. However, this is an intuitive, ad hoc hypothesis that was never demonstrated and, as we saw, mutation plus the neighbor influence cannot produce the meaningful order we see in life (Valenzuela 2009, 2010a; Valenzuela et al., 2010). It is very often proposed that the genetic code implies a constraint that explains the order. This argument falls into rational circularity, because the cause that originated and maintains the genetic code is sent to the unexplained, non-analyzable or un-debatable set of evolutionary processes (constraints). Recurrent mutation inexorably destroys any nucleotide sequence (also sequential constraints), as is evidenced regularly by the cancers, aging and genetic diseases of living beings (Valenzuela, 2007, 2009; Valenzuela et al., 2010).

The expected internucleotide correlation under mutation and random drift

Random fluctuations of genetic frequencies (genetic drift) could result in frequencies of alleles in a locus or base in a site reaching frequency 1.0 (substitution), 0.0 (elimination or loss) or between 0.0 and 1.0 (polymorphism). However, and by constitution and definition, genetic drift occurs independently and equally in all the nucleotide sites. Thus, it cannot generate a stable internucleotide correlation and is hermeneutically powerless. It may move up or down with the same average magnitude, but its final contribution is zero. A widespread error equalizes substitution (a turn-over process) with fixation (a permanent state). Thus, in early articles neutralists calculated the probability of what they named fixation, but it was substitution instead because it was the probability to attain the frequency 1.0 by random frequency fluctuations (Kimura 1962; Nei et al., 2010). To calculate the probability of fixation we need the number of generations over which the allele or base has remained fixed. The probability to remain at frequency 1.0 is completely different for alleles or bases that remain at this frequency for one million generations. With recurrent forward and backward mutation fixation, it is physically, logically, mathematically and biologically impossible (Wright, 1931; Feller, 1951; Valenzuela and Santos, 1996; Valenzuela, 2000, 2002, 2007, 2009, 2010a, 2011). Neither mutation nor drift can give sense to a DNA or RNA segment.

Thus, the expected average internucleotide correlation under recurrent forward and backward mutation and genetic drift is very small or stochastically zero. Gene mutations and genetic drift are hermeneutically empty biotic processes; they are similar to Brownian motion. This does not mean that they cannot give rise to biotic processes, but if they do, they would present random distributions of their elementary components (nucleotide or amino-acid sequences), as we shall see.

Only selection gives meaning to nucleotide sequences. Our logic of demonstration

The only process that can produce permanent biotic functional sense to sequences is selection, because it is a process of co-variation between biotic sequences and environmental requirements (adaptation). Our search for internucleotide correlations is founded in this feature of the evolutionary process. If we do not find significant non-random association between the two bases of dinucleotides, neutral or nearly neutral evolution is affirmed, but selective evolution is not refuted. However, if we find significant non-random associations between the two bases, neutral and nearly neutral evolution are refuted and selective evolution affirmed. Neutralists and nearly-neutralists included selection in their models but stated that their models were not related to adaptation (Ohta, 1992, 2002; Nei, 2005). They do not accept the pan-adaptationist condition of the Synthetic Theory of Evolution (Gould, 2002). Our position emphasizes that thermodynamically non-random nucleotide sequences cannot be maintained unless selection operates to do it. Thus, non-random sequences are really adaptive sequences that remain in spite of the strong tendency to entropic distributions. Dynamic non-random processes are physical conditions for life production and maintenance; thus, they are synonymous with adaptation (see also Introduction).

DATA

The *Drosophila melanogaster* mtDNA (GenBank accession NC 001709, with 19,517 nucleotide sites) was studied.

METHODS

The heterogeneity of the distribution of the second base in relation to the first base in dinucleotides, whose bases are separated by 0 (contiguous), 1, 2, … K, nucleotide sites, was determined by a χ^2 test. The total deviation from randomness of the 16 possible dinucleotides (pairs) was determined by a χ^2_9 test [9 = degrees of freedom, 3 independent rows (4 possible bases less 1) times 3 independent columns (4 possible bases less 1)], and the particular deviation of a pair by a χ^2_1 test (its contribution to the total χ^2_9 value). These χ^2 tests and their associated probabilities directly measure the distance to neutrality (randomness). The first purpose is to know whether positive (more dinucleotides than expected, ↑) or negative (fewer dinucleotides than expected, ↓) associations (in relation to random dinucleotide distribution) are present between the two bases of dinucleotides (χ^2 tests) and the extension in terms of the number of nucleotide sites (separation of K sites) and of the neighborhood influence. A second aim is to examine whether these base associations are homogeneously distributed along with the whole mtDNA or if they are different in four

equal and consecutive sub-segments. For this aim, the total mtDNA with 19,517 nucleotide sites was divided into four consecutive and equal segments with 4,879 sites each. A previous study showed highly heterogeneous composition of bases and dinucleotides when the mtDNA was divided into ten segments (unpublished). Abbreviations: A = Adenine, T = Thymine, G = Guanine, C = Cytosine.

More specific details of the methods are published (Valenzuela 2009, 2010a). The most important methodological features are presented with help of APPENDICES 1 and 2. APPENDIX 1 describes the method to calculate the association for one coding segment. This appendix also shows that the correlations do not depend on protein-coding functions. APPENDIX 2 shows the analysis for dinucleotides whose bases are separated by 17 nucleotides (an example). Bases found in the 1st and 19th nucleotide sites constitute the first dinucleotide, the second includes bases at the 2nd and 20th sites, the third with bases at the 3rd and 21st sites, and so on, until the last dinucleotide whose bases are at the sites 19th, 499th and $19,517^{th}$. There are then 19,499 dinucleotides whose bases are separated by 17 sites. The expected number of dinucleotides is obtained by the frequency of the first base, times the frequency of the second base, times the total number of dinucleotides. Taking the observed frequencies of bases is the best estimate of the expected historical action of mutation rates and all the possible neighbor influences (as average) because these frequencies should be considered as the expected equilibrium frequencies under neutral evolution (see Valenzuela et al., 2010). APPENDIX 2 presents dinucleotides ordered according to the significance of their deviation from randomness. A sign indicates whether there is more ($\uparrow$) or fewer ($\downarrow$) dinucleotides than expected. Significance at the 0.05 level is found for a χ^2_1 value equal to 3.84 and for a χ^2_9 with a value equal to 17. There were 14 significantly deviated dinucleotides among the 16 possible pairs, 6 with excess and 8 with deficiency. The total excess was 825.6 and the deficiency added up to 825.5. The excesses are fewer, but larger than the deficiencies. Excesses are produced by positive selective processes, and deficiencies by negative selective processes that have been maintained over million of mitochondria generations to the present (see Rationale). The selective process occurs because mutations happen continuously and destroy any non-random associations. The facts that: i) 14 pairs among 16 are significantly distant to the random (neutral) expected distribution and i) the incommensurable value of the χ^2_9 test = 322.2 (P<10^{-50}) refute neutral evolution definitively. We observe that most (14/16 = 87.5%) pairs of bases chosen at random and separated by 17 nucleotide sites are distributed enormously far from the expected neutral distribution. Furthermore, and considering that I) this distance to neutrality has been maintained by millions of mitochondrion generations, and II) major significances were found with separations from 0 to more than 2000, we can only conclude that these evolutionary conditions are impossible under neutral and nearly neutral evolution.

RESULTS

Table 1 shows the total χ^2_9 value found in the entire Drosophila mtDNA, when the bases of dinucleotides are separated by 0 (contiguous or consecutive), 1, 2, 3 … 17, nucleotide sites, and

the χ^2_1 contribution to the χ^2_9 value by the five most significant pairs, ordered according to the value of their significance. The total significance for the χ^2_9 at the 0.05, 0.025, 0.01, 0.005 and 0.001 critical probability levels is found with 16.9 (17), 19.0 (19), 21.7 (22), 23.6 (24) and 27.9 (28) χ^2_9 values (rounded to integer number in parentheses), respectively. The significance of the deviation from the expected randomness of a particular pair may be evaluated by its contribution to the total χ^2_9 by the χ^2_1 values that are 3.84 (4), 5.02 (5), 6.64 (7), 7.88 (8) and 10.83 (11) for the same critical probability levels, respectively. The χ^2 values of Table 1 have been rounded to integer figures. The most significant figure for contiguous (0 separation) bases was the excess of G-G, followed by excess of C-C, depression of G-T, excess of G-C and excess of T-T pairs (χ^2_1=27.8); other deviations were less significant, even though six of them had χ^2_1 over 4 or a probability of less than 0.05 [(A-C)$\downarrow$ 25; (T-C)$\downarrow$ 11; (T-G)$\downarrow$ 11; (A-A)$\uparrow$ 11; (C-A)$\downarrow$ 10; (T-A)$\downarrow$ 5]. The spectrum of significances changes when K increases, as can be seen in the table (and in Table 2). The most significant pairs (1st pairs) showed more observed pairs than expected ($\uparrow$); the other pairs showed both possibilities ($\uparrow$, $\downarrow$).

We see a clear periodicity in the χ^2_9 value after the separation by 1 site. There are triplets of χ^2_9 values, the first over 300 that we named the head figure, followed by two consecutive smaller figures between 37 and 114 (tail 1 and tail 2 figures). The largest χ^2_1 contribution for the head figure was always given by an excess of G-G pairs, while for tail 1 and tail 2 figures it was given by excesses of G-C and C-G, respectively. There are other ordered distributions in the 2º, 3º, 4º and 5º pairs, but they are not as exclusive as those found in the most significant pair. Their analysis is left to the reader.

The periodicity in triplets of the χ^2_9 values indicates the deviation from the expected random distribution of the second base in relation to the first base of dinucleotides separated by 0, 1, 2 …17 was observed until 609 K (Valenzuela 2010b). It is also evident that the most significant pair shows a periodicity in these triplets: G-G, G-C and C-G, which could be followed until separation 36, with only one exception (separation 33 where G-C was replaced by G-G). However, it must be noted that the χ^2 test is rough for finding fine nucleotide associations because its high degree of variance may lead to variable hierarchical orders. Table 2 presents the analysis for two sets of 10 K: 1000-1009K and 2002-2011K. In both sets the periodicity is evident, with head values between 48 and 65 and tail values between 13 and 39 in the range 1000-1009K and head values between 38 and 49 and tail values between 7 and 21 in the range 2002-2011K. It is remarkable that in the range 1000-1009K, 9 among 10 χ^2_9 values are significant with a highest probability equal to 0.0062 (χ^2_9 = 23), and there are 6 significant χ^2_9 values with the highest probability 0.0179 (remember the inverse relationship between significance and probability) among 10 in the range 2002-2011K. As well, it is remarkable that the most significant χ^2_1 value was always given by more observed pairs than expected ($\uparrow$) and was mostly G-G as head, G-C and C-G as 1st and 2nd tail, respectively in the 1000-1009K range. In the 2002-2011K range these last relationships holds, but it is necessary to consider the five χ^2_1 values to reconstruct them, given that other pairs appeared as the most deviated from randomness.

These pair structures are averages found when scanning large DNA segments. The existence of micro-isochores (sub-segments with different base composition; Valenzuela 1997,

2009, this article) and micro-signatures (sub-segments with different dinucleotides compositions) (Valenzuela 2009, this article) could change the structure of base associations of dinucleotides. Figure 1 shows that the *Drosophila melanogaster* mtDNA has different composition of bases in its segments. Thus, an analysis of the mtDNA divided into four equal segments was performed. Tables 3, 4, 5 and 6 show this analysis for the 1º (sites 1º-4,879º), 2º (4,880º-9,758º), 3º (9,759º-14,637º) and 4º (14,637º-19,516º) segments, respectively. The base composition of the segments were: 1º (A = 1,663, 34.08%; T = 2,029, 41.59%; G = 567, 11.62%; C = 620, 12.71%); 2º (A = 2,126, 43.57%; T = 1,712, 35.09%; G = 428, 8.77%; C = 613, 12.56%); 3º (A = 1,972, 40.42%; T = 1,899, 38.92%; G = 397, 8.14%; C = 611, 12.52%); 4º (A = 2,391, 49.01%; T = 2,242, 45.95%; G = 87, 1.78%; C = 159, 3.26%). The high degree of heterogeneity of base composition of these four segments is evident (micro-isochores); particularly the G and C proportions decay around 5 and 3.8 times, respectively in the fourth segment, with the corresponding increase of A and T proportions. The χ^2_9 value for the heterogeneity of base composition of these 4 segments resulted 855.4 (P<10^{-100}). The pair structure found in the total mtDNA is partially valid for these segments (especially for the excess of G-G as the most significant pairs), except segment 4º where non G-G pairs were the most significant. A detailed analysis and comparison of the four segments are left to the reader.

DISCUSSION

Highly significant (statistical) interactions were found between any nucleotide and nucleotides separated as far as 2011 nucleotide sites. This demonstrates that in the mtDNA organization any nucleotide maintains strong non-random associations with the whole mtDNA. Moreover, these interactions include a high significant periodicity between the bases of dinucleotides, when the bases are separated by at least by 0, 1, 2 … 2011 sites. The internucleotide correlations we have just described can be seen as processes of internucleotide co-adaptation. Those dinucleotides whose frequencies are over the random expected frequency were positively selected and are now maintained by positive selection; those that are below their expected frequencies were and are negatively selected. These strong associations, maintained over several million mitochondrion generations (the time during which *Drosophila melanogaster* has had this mtDNA) refute the neutral theory, the nearly neutral theory and the neighbor influence of a base on mutation rates of its neighborhood as main factors of evolution. It is impossible to maintain this organization during that time by random mutation, genetic drift and weak natural selection (nearly-neutral evolution). On the contrary, forward and backward recurrent mutation and drift are processes that should inexorably destroy this organization (Valenzuela and Santos 1996, Valenzuela 1997,

TABLE 1

Total χ^2_9 values for heterogeneity of the distribution of dinucleotide bases separated by 0, 1, 2 …17 nucleotide sites, and their χ^2_1 contribution of the 5 most significant dinucleotides

Sep	χ^2_9	1st Pair		2nd Pair		3rd Pair		4th Pair		5th Pair	
		Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co
0	485	[(G-G)↑	124]	[(C-C)↑	113]	[(G-T)↓	91]	[(G-C)↑	50]	[(T-T)↑	28]
1	94	[(C-G)↑	36]	[(C-C)↑	25]	[(C-T)↓	12]	[(A-G)↓	6]	[(T-C)↓	4]
2	405	[(G-G)↑	116]	[(C-C)↑	106]	[(G-C)↑	33]	[(C-G)↑	25]	[(T-G)↓	23]
3	114	[(G-C)↑	23]	[(A-A)↑	22]	[(T-T)↑	15]	[(T-A)↓	11]	[(C-C)↑	9]
4	47	[(C-G)↑	20]	[(A-G)↓	8]	[(A-T)↑	6]	[(C-T)↓	4]	[(G-T)↓	2]
5	381	[(G-G)↑	139]	[(C-C)↑	51]	[(A-G)↓	32]	[(G-C)↑	32]	[(C-G)↑	30]
6	87	[(G-C)↑	38]	[(T-A)↑	14]	[(T-C)↓	12]	[(G-A)↓	6]	[(A-A)↓	5]
7	37	[(C-G)↑	17]	[(C-T)↓	7]	[(T-G)↓	3]	[(G-G)↑	3]	[(T-T)↑	2]
8	375	[(G-G)↑	149]	[(C-C)↑	45]	[(C-G)↑	36]	[(G-C)↑	29]	[(T-G)↓	24]
9	76	[(G-C)↑	34]	[(G-A)↓	16]	[(T-C)↓	12]	[(G-G)↑	8]	[(C-G)↓	2]
10	49	[(C-G)↑	28]	[(A-G)↓	6]	[(C-T)↓	6]	[(A-T)↑	4]	[(T-A)↑	1]
11	367	[(G-G)↑	144]	[(G-C)↑	45]	[(C-C)↑	35]	[(C-G)↑	26]	[(A-G)↓	23]
12	65	[(G-C)↑	34]	[(G-A)↓	13]	[(T-C)↓	8]	[(C-G)↓	3]	[(C-A)↑	2]
13	70	[(C-G)↑	38]	[(A-G)↓	13]	[(A-A)↑	5]	[(G-C)↓	3]	[(C-A)↓	2]
14	310	[(G-G)↑	78]	[(C-G)↑	48]	[(G-C)↑	44]	[(C-C)↑	32]	[(G-A)↓	22]
15	60	[(G-C)↑	34]	[(G-A)↓	10]	[(C-G)↓	6]	[(T-C)↓	3]	[(C-T)↑	2]
16	52	[(C-G)↑	27]	[(A-G)↓	12]	[(A-A)↑	3]	[(C-T)↓	2]	[(C-A)↓	2]
17	322	[(G-G)↑	91]	[(C-G)↑	45]	[(C-C)↑	40]	[(G-C)↑	22]	[(A-A)↑	22]

Sep = number of separation sites; χ^2_1Co = χ^2_1 contribution of this pair to the total χ^2_9 value; ↑ = more pairs observed than expected; ↓ = fewer pairs observed than expected.
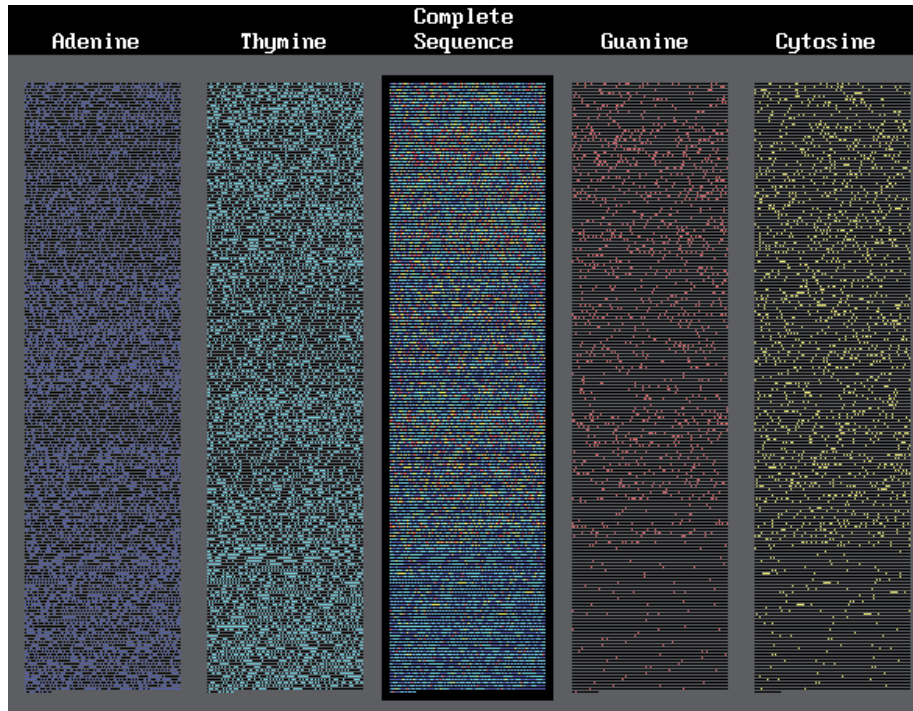
Figure 1. Distribution of bases, A in dark blue, T in light blue, G in red, C in yellow, and non-bases in black in the *Drosophila melanogaster* mtDNA. The composition with the four bases is at the center.

TABLE 2

Total χ^2_9 values for heterogeneity of the distribution of dinucleotide bases separated by 1000-1009 and 2002-2011 nucleotide sites, and their 5 most significant dinucleotides χ^2_1 contribution

Sep	χ^2_9	1st Pair		2nd Pair		3rd Pair		4th Pair		5th Pair	
		Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co
1000	52	[(G-G)↑	12]	[(C-C)↑	9]	[(A-G)↓	8]	[(G-C)↑	8]	[(C-A)↑	4]
1001	24	[(G-C)↑	14]	[(T-C)↓	5]	[(C-C)↑	2]	[(G-A)↓	1]	[(C-G)↓	1]
1002	39	[(C-G)↑	10]	[(C-C)↑	8]	[(A-C)↓	6]	[(G-G)↑	3]	[(C-A)↓	3]
1003	48	[(G-G)↑	13]	[(C-C)↑	12]	[(A-G)↑	5]	[(T-C)↓	4]	[(G-C)↑	3]
1004	36	[(G-C)↑	10]	[(G-T)↓	7]	[(A-C)↓	5]	[(G-G)↑	4]	[(C-G)↑	3]
1005	13	[(C-G)↑	4]	[(A-G)↓	3]	[(G-C)↑	2]	[(G-G)↑	2]	[(G-A)↓	1]
1006	65	[(G-C)↑	24]	[(G-G)↑	8]	[(C-G)↑	6]	[(G-A)↓	6]	[(A-G)↓	5]
1007	23	[(G-C)↑	9]	[(C-C)↑	3]	[(A-C)↓	2]	[(A-A)↑	2]	[(G-G)↑	1]
1008	31	[(C-G)↑	10]	[(A-G)↓	10]	[(G-G)↑	2]	[(A-C)↓	2]	[(A-T)↑	2]
1009	59	[(C-C)↑	10]	[(C-G)↑	10]	[(A-G)↓	10]	[(G-G)↑	9]	[(G-C)↑	5]
2002	49	[(C-G)↑	12]	[(G-G)↑	11]	[(C-A)↓	9]	[(G-T)↓	4]	[(T-G)↓	4]
2003	20	[(G-G)↑	7]	[(C-A)↓	2]	[(G-T)↓	2]	[(C-G)↑	2]	[(C-C)↑	2]
2004	21	[(C-C)↑	8]	[(C-G)↑	4]	[(C-T)↓	4]	[(A-G)↓	2]	[(A-C)↓	1]
2005	38	[(G-G)↑	9]	[(C-C)↑	8]	[(G-C)↑	4]	[(A-G)↓	3]	[(C-G)↑	3]
2006	12	[(G-A)↓	4]	[(G-C)↑	3]	[(G-G)↑	1]	[(T-G)↓	1]	[(G-T)↑	1]
2007	13	[(C-C)↑	3]	[(G-G)↑	2]	[(G-A)↓	2]	[(G-C)↑	1]	[(C-T)↓	1]
2008	41	[(G-C)↑	9]	[(G-G)↑	8]	[(C-G)↑	5]	[(G-C)↑	3]	[(C-T)↓	3]
2009	7	[(T-C)↓	2]	[(G-G)↑	1]	[(A-G)↓	1]	[(G-C)↑	1]	[(G-A)↓	1]
2010	12	[(G-C)↑	3]	[(T-C)↓	2]	[(C-C)↑	2]	[(A-G)↓	1]	[(C-G)↑	1]
2011	41	[(C-G)↑	12]	[(G-C)↑	8]	[(C-C)↑	4]	[(T-G)↓	4]	[(G-G)↑	3]

Nomenclature as in Table 1.

TABLE 3

χ^2_9 values for heterogeneity of the distribution of dinucleotide bases separated by 0, 1, 2 …17 nucleotide sites, and their χ^2_1 contribution of the 5 most significant pairs of bases, in the 1º segment of mtDNA (site 1º- 4879º)

Sep	χ^2_9	1st Pair		2nd Pair		3rd Pair		4th Pair		5th Pair	
		Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co
0	136	[(G-T)↓	48]	[(G-G)↑	28]	[(C-G)↓	13]	[(T-T)↑	9]	[(G-A)↑	9]
1	63	[(A-A)↓	12]	[(C-G)↑	9]	[(A-C)↑	7]	[(G-A)↑	7]	[(A-T)↑	6]
2	118	[(C-C)↑	29]	[(G-G)↑	18]	[(C-A)↓	17]	[(A-C)↓	17]	[(A-A)↑	10]
3	44	[(A-G)↑	15]	[(C-G)↓	13]	[(C-A)↑	5]	[(A-T)↓	3]	[(G-C)↑	2]
4	28	[(G-C)↓	5]	[(C-A)↑	4]	[(C-C)↓	3]	[(T-A)↓	3]	[(A-G)↓	2]
5	93	[(G-G)↑	19]	[(A-C)↓	16]	[(C-A)↓	12]	[(A-G)↓	11]	[(A-A)↑	7]
6	52	[(A-G)↑	12]	[(C-G)↓	10]	[(G-C)↑	10]	[(T-C)↓	5]	[(G-G)↓	4]
7	25	[(C-C)↓	7]	[(G-C)↓	5]	[(C-G)↑	4]	[(G-A)↑	2]	[(T-C)↑	2]
8	115	[(G-G)↑	30]	[(C-A)↓	15]	[(C-G)↑	13]	[(A-A)↑	11]	[(G-T)↓	10]
9	39	[(C-G)↓	11]	[(G-C)↑	8]	[(T-C)↓	5]	[(C-A)↑	5]	[(A-G)↑	2]
10	42	[(G-A)↑	7]	[(C-G)↓	5]	[(G-G)↓	5]	[(C-G)↑	5]	[(C-T)↑	4]
11	69	[(G-G)↑	25]	[(A-G)↓	10]	[(G-C)↑	7]	[(G-T)↓	6]	[(A-A)↑	5]
12	39	[(C-G)↓	10]	[(C-A)↑	10]	[(T-C)↓	3]	[(A-G)↑	3]	[(A-C)↑	3]
13	43	[(G-C)↑	9]	[(A-G)↓	7]	[(T-G)↑	5]	[(C-G)↑	4]	[(C-C)↓	4]
14	76	[(A-C)↓	12]	[(C-A)↓	10]	[(G-G)↑	10]	[(G-A)↓	10]	[(C-C)↑	9]
15	44	[(C-G)↓	9]	[(A-G)↑	8]	[(G-G)↓	5]	[(G-C)↑	4]	[(C-C)↓	3]
16	19	[(G-C)↓	5]	[(A-C)↑	3]	[(G-G)↓	3]	[(T-G)↑	2]	[(G-T)↑	1]
17	77	[(C-G)↑	13]	[(C-C)↑	12]	[(G-G)↑	10]	[(A-G)↓	9]	[(C-A)↑	9]

Nomenclature as in Table 1.

TABLE 4

χ^2_9 values for heterogeneity of the distribution of dinucleotide bases separated by 0, 1, 2 …17 nucleotide sites, and their χ^2_1 contribution of the 5 most significant pairs of bases, in the 2º segment of mtDNA (4880º- 9758º site)

Sep	χ^2_9	1st Pair		2nd Pair		3rd Pair		4th Pair		5th Pair	
		Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co
0	78	[(C-C)↑	22]	[(G-T)↓	14]	[(A-C)↓	14]	[(C-G)↓	8]	[(G-G)↑	7]
1	21	[(G-C)↓	5]	[(A-A)↓	3]	[(C-T)↓	3]	[(A-C)↑	1]	[(G-G)↓	1]
2	73	[(G-G)↑	30]	[(C-C)↑	13]	[(T-G)↓	8]	[(T-T)↑	5]	[(G-A)↓	5]
3	17	[(T-G)↑	4]	[(C-G)↓	3]	[(G-C)↑	2]	[(A-A)↓	2]	[(G-G)↓	2]
4	12	[(C-G)↑	3]	[(G-A)↑	2]	[(G-C)↓	1]	[(A-G)↓	1]	[(G-G)↓	1]
5	74	[(G-G)↑	32]	[(G-T)↓	8]	[(C-G)↑	6]	[(C-C)↑	6]	[(T-T)↑	6]
6	16	[(C-G)↑	5]	[(C-T)↑	4]	[(C-C)↓	3]	[(T-G)↑	1]	[(T-T)↓	1]
7	11	[(C-T)↓	3]	[(C-A)↑	3]	[(G-C)↓	2]	[(T-A)↓	1]	[(T-G)↑	1]
8	78	[(G-G)↑	34]	[(T-T)↑	7]	[(T-G)↓	7]	[(G-A)↓	6]	[(A-A)↑	5]
9	18	[(T-G)↑	5]	[(G-C)↑	3]	[(C-G)↓	3]	[(G-A)↓	2]	[(T-C)↓	1]
10	14	[(C-G)↑	5]	[(G-C)↓	2]	[(G-T)↓	2]	[(C-C)↓	1]	[(A-C)↑	1]
11	77	[(G-G)↑	27]	[(T-T)↑	8]	[(G-T)↓	7]	[(C-C)↑	7]	[(T-C)↓	7]
12	24	[(G-C)↑	8]	[(C-G)↓	5]	[(G-A)↓	3]	[(T-C)↓	2]	[(C-T)↑	2]
13	22	[(C-G)↑	4]	[(T-C)↑	4]	[(G-G)↓	3]	[(C-C)↓	2]	[(G-T)↑	2]
14	93	[(G-G)↓	18]	[(C-G)↑	13]	[(G-C)↑	10]	[(T-T)↑	9]	[(A-A)↑	8]
15	23	[(C-G)↓	13]	[(G-C)↑	3]	[(C-T)↑	2]	[(G-T)↓	1]	[(A-G)↑	1]
16	15	[(C-G)↑	5]	[(T-C)↑	3]	[(G-C)↓	1]	[(T-G)↓	1]	[(T-A)↓	1]
17	69	[(G-G)↑	21]	[(T-T)↑	9]	[(G-A)↓	6]	[(T-G)↓	6]	[(G-C)↑	6]

Nomenclature as in Table 1.

TABLE 5
χ^2_9 values for heterogeneity of the distribution of dinucleotide bases separated by 0, 1, 2 …17 nucleotide sites, and their χ^2_1 contribution of the 5 most significant pairs of bases, in the 3º segment of mtDNA (site 9,759º- 14,637º)

Sep	χ^2_9	1st Pair		2nd Pair		3rd Pair		4th Pair		5th Pair	
		Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co
0	128	[(G-G)↑	33]	[(C-C)↑	19]	[(G-T)↓	17]	[(T-T)↑	16]	[(T-G)↓	8]
1	13	[(A-T)↑	2]	[(C-G)↑	2]	[(A-G)↑	2]	[(G-G)↑	1]	[(T-T)↓	1]
2	46	[(G-G)↑	11]	[(C-C)↑	9]	[(G-T)↓	4]	[(T-C)↓	4]	[(C-G)↑	3]
3	15	[(C-C)↑	5]	[(C-G)↓	3]	[(T-G)↑	1]	[(G-C)↑	1]	[(T-C)↓	1]
4	13	[(G-T)↓	3]	[(G-A)↑	2]	[(T-A)↓	2]	[(T-T)↑	2]	[(G-G)↑	1]
5	63	[(G-G)↑	27]	[(G-C)↑	8]	[(A-G)↓	5]	[(G-T)↓	4]	[(C-C)↑	4]
6	20	[(G-C)↑	7]	[(T-C)↓	4]	[(G-A)↓	2]	[(C-A)↑	1]	[(G-G)↓	1]
7	11	[(C-A)↑	3]	[(C-T)↓	2]	[(G-T)↑	2]	[(G-A)↓	1]	[(A-G)↑	1]
8	40	[(G-G)↑	16]	[(G-A)↓	7]	[(T-G)↓	5]	[(C-C)↑	2]	[(C-A)↓	2]
9	19	[(G-A)↓	4]	[(G-C)↑	2]	[(A-G)↑	2]	[(T-G)↓	2]	[(T-C)↓	2]
10	7	[(C-C)↓	2]	[(T-C)↑	2]	[(C-A)↑	1]	[(G-C)↓	0]	[(C-T)↓	0]
11	60	[(G-G)↑	28]	[(G-A)↓	6]	[(T-G)↓	4]	[(C-C)↑	4]	[(G-C)↓	4]
12	17	[(C-A)↑	4]	[(G-C)↓	4]	[(G-A)↓	3]	[(C-T)↓	2]	[(G-G)↓	1]
13	20	[(C-G)↑	7]	[(G-C)↓	5]	[(G-T)↑	2]	[(T-T)↓	1]	[(T-C)↑	1]
14	20	[(G-G)↑	8]	[(C-G)↑	4]	[(G-A)↓	3]	[(T-G)↓	1]	[(A-G)↓	1]
15	8	[(G-C)↑	2]	[(G-A)↓	2]	[(C-G)↓	1]	[(G-G)↓	1]	[(A-G)↑	1]
16	5	[(A-G)↑	1]	[(C-T)↓	1]	[(C-G)↑	1]	[(T-G)↑	0]	[(T-A)↓	0]
17	27	[(G-G)↑	15]	[(C-C)↑	4]	[(A-G)↓	2]	[(G-G)↑	1]	[(T-G)↓	1]

Nomenclature as in Table 1.

TABLE 6
χ^2_9 values for heterogeneity of the distribution of dinucleotide bases separated by 0, 1, 2 …17 nucleotide sites, and their χ^2_1 contribution of the 5 most significant pairs of bases, in the 2º segment of mtDNA (site 14,638º- 19,516º)

Sep	χ^2_9	1st Pair		2nd Pair		3rd Pair		4th Pair		5th Pair	
		Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co	Pair	χ^2_1Co
0	100	[(C-C)↑	61]	[(A-G)↓	7]	[(T-G)↑	6]	[(G-T)↓	5]	[(A-C)↓	4]
1	216	[(C-C)↑	119]	[(T-T)↓	18]	[(T-A)↓	14]	[(C-G)↑	13]	[(A-A)↑	13]
2	62	[(C-C)↑	19]	[(T-G)↓	10]	[(A-G)↑	6]	[(C-T)↓	5]	[(A-A)↓	5]
3	85	[(A-A)↑	21]	[(A-T)↓	20]	[(T-A)↓	19]	[(T-T)↑	19]	[(A-C)↓	1]
4	62	[(A-T)↑	13]	[(T-T)↓	13]	[(T-A)↑	13]	[(A-A)↓	13]	[(C-C)↑	5]
5	29	[(C-C)↑	15]	[(A-C)↓	3]	[(G-T)↑	3]	[(G-A)↓	1]	[(T-T)↑	1]
6	118	[(T-A)↑	24]	[(T-T)↓	23]	[(A-T)↓	15]	[(A-A)↓	12]	[(C-A)↓	11]
7	33	[(C-G)↑	13]	[(C-C)↑	9]	[(C-A)↓	4]	[(T-G)↓	2]	[(G-G)↑	1]
8	83	[(A-T)↑	18]	[(T-T)↓	17]	[(T-A)↑	16]	[(C-C)↑	15]	[(A-A)↓	14]
9	11	[(C-G)↑	6]	[(C-T)↓	1]	[(G-C)↓	1]	[(C-A)↑	0]	[(G-A)↑	0]
10	23	[(T-T)↓	4]	[(T-A)↑	4]	[(A-T)↑	4]	[(A-A)↓	4]	[(G-A)↓	2]
11	6	[(G-C)↑	2]	[(G-A)↓	1]	[(G-T)↓	1]	[(T-C)↓	1]	[(C-G)↑	0]
12	17	[(C-G)↓	3]	[(T-T)↓	2]	[(A-T)↑	2]	[(A-A)↓	2]	[(G-T)↓	1]
13	13	[(A-T)↓	2]	[(A-A)↑	2]	[(T-A)↓	2]	[(T-T)↑	2]	[(C-G)↓	1]
14	31	[(T-C)↓	9]	[(G-C)↑	6]	[(A-C)↑	6]	[(G-T)↓	2]	[(G-G)↑	1]
15	21	[(G-C)↑	10]	[(C-T)↑	3]	[(C-A)↓	3]	[(G-G)↓	2]	[(T-G)↑	1]
16	47	[(G-C)↑	19]	[(C-G)↑	9]	[(A-G)↓	8]	[(T-G)↑	4]	[(C-C)↓	2]
17	39	[(A-A)↑	6]	[(A-T)↓	5]	[(T-A)↓	4]	[(T-T)↑	4]	[(G-C)↑	4]

Nomenclature as in Table 1

2000, 2002, 2007, 2009; Valenzuela et al., 2010). Recurrent forward and backward mutations make neutral fixation impossible (Wright 1931; Feller 1951; Valenzuela 2000, 2002a; Valenzuela et al., 2010). Mutations do occur in mtDNA of eukaryote organisms during their life and are a main factor in aging and death (Gredilla et al., 2010). As well, hundreds of human mtDNA mutations are known that produce lethal or sub-lethal conditions (Tuppen et al., 2010). Thus, mtDNA is almost always destroyed (depending only on the life span of the individual) during the life of eukaryote individuals, but it is much more stable in phylogeny. The individual instability and phylogenetic stability are only possible if there are strong selective mechanisms (DNA repair and protection; Gredilla et al. 2010) acting on mitochondria from one generation to the next, especially in females among sexually reproductive species. The HIV-1 virus has a high correlation among its genome bases, but not periodicity (Valenzuela 2009, 2011). The HIV-1 virus may need this correlation to fold the RNA chain into the capsid, which is not needed by the mtDNA. However, mtDNA needs coiling and hypercoiling to protect itself and locate functionally in the mitochondrion matrix. It is fascinating that this structure may be present in prokaryotes, as we found preliminarily (Valenzuela 2010a, 2011).

The strong non-random association between a nucleotide (and its complementary one) in a DNA site and the residual genome, maintained over millions of generations, convinced us that the main selector (selection factor) for this nucleotide site is not the environment, but the residual genome. Once hereditary polymers are acquired in biotic systems, evolution goes on mainly by polymer interactions and recombination (Valenzuela, 2002a, 2002b), either in the endogenous or exogenous plane. Horizontal evolution, sex and symbiogenesis are positive selective mechanisms of inter-polymer selection in evolution. However, if other polymers are the best "friends" for a particular polymer, they may also be its worst enemies (negative selection) leading to polymer destruction, competition, diseases and extinction. We see only those organisms that reached a resilient equilibrium after the intra-individual polymers' coexistence or fusion.

The different structures of deviations from randomness according to significances found in the four segments with equal numbers of nucleotides indicate that the nucleotide association in dinucleotides is rather flexible and depends on the base composition of the segment (a relationship among purines and pyrimidines and base complementariness). The strong G-G, G-C and C-G associations found in the head, tail 1 and tail 2 periodicities, respectively, in the whole mtDNA represent the most frequent deviation from randomness, but, their frequencies vary in the four segments. Perhaps these statistical attractions and repulsions indicate physical attractions and repulsions that are necessary to accomplish important non-protein-synthesis functions that were and are crucial in evolutionary development. The high degree of heterogeneity of the base composition of the four segments maintained for millions of mitochondrion generations conclusively refutes neutral and nearly neutral evolution and the neighbor influence hypothesis (and independently of the internucleotide interactions described above). These theories and hypothesis predict a homogeneous distribution instead.

This article could finish here; however there are fundamental misconceptions in the neutralist and nearly-neutralist position that need special treatment (Valenzuela

2000, 2002a, 2010a, 2010b, Valenzuela et al., 2010). Perhaps the reader, habituated to phylogenetic analyses based on protein-coding features such as coding positions, synonymous or non-synonymous mutations, thinks our conclusive refutation of neutral and nearly neutral evolution is rather unsupported. We have received the critical position from the Neutral Theory of Evolution (made from neutralist and non-neutralist colleagues) that neutralism accepts "purifying selection" in the form of lethal and sub-lethal mutations as an important part of the theory. However, this proposition is not true as far as mtDNA is concerned, because the clinical genetic practice (a method with very low sensitivity to study selectors) shows that mtDNA mutations (in humans and animals) are lethal, sub-lethal, and compatible with life with and without impairment of reproduction (Tuppen et al., 2010). As mentioned above, neutralists protected the theory by including "constraints" such as the genetic code, the restriction to four nucleotide bases, some invariant functional parts of proteins or DNA, biases of codon usage, etc. However, neutralists or nearly neutralists (and selectionists or neo-Darwinists) never advanced a proportion of evolution that is due to lethal or sub-lethal nucleotide mutations or to invariant constraints, so as to put these hypotheses to the test (epistemologically this proposition is a negative heuristic protective belt). If the proportion of evolution in every nucleotide site is mostly (over 50%) due to lethal and sub-lethal bases, then evolution is, by definition, selective and not neutral or nearly-neutral. But in this case life is impossible. The same occurs with constraints. First, studies of protein constraints were focused on the active site of enzymes. Then allosteric sites, receptor sites, signal sequences, attachment to membrane sequences, and several other amino-acid functional sequences that conferred an invariant function to almost every amino-acid of a protein were described. Now, it is difficult to conceive of any amino acid of a protein without a function that severely constraints it. Constraints proposed without precision are also negative heuristic protective belts of neutralism or near neutralism. A simple question shows that: what is a constraint, in the evolutionary process? Or, how were constraints acquired and maintained in evolution during paleontological eras? To propose that the genetic code, eukaryote, prokaryote, unicellular, multicellular, vertebrate (and so on) organizations were acquired and are maintained by mutation and random drift is simply madness. Evolution is mostly conservation not variation (Valenzuela, 2007, 2009; Valenzuela et al., 2010). That the genetic code or any organization was acquired and is maintained mostly by selection implies that evolution is selective. Constraints are mostly produced by non-random nucleotide sequences; thus, they were included in the present analysis. The nearly neutral theory of evolution added selection to mutation and drift with a positive selection coefficient (Ohta, 1992, 2002) making this theory almost undistinguishable from the Synthetic Theory of Evolution.

The present analysis is a trans- supra- or non-protein-coding study. Moreover, these base-to-base interactions and periodicities occur between bases separated by 0 to more than 2,000 sites, and mitochondrial genes have less than 1,750 nucleotides. On the contrary, it is expected that a gene sequence disturbs or destroys non-protein-coding periodicities to specify its own coding message, which is seldom periodical. Furthermore, there are no G-G, G-C or

C-G periodical dinucleotide associations in mtDNA, where the two bases are separated by 0 to more than 2,000 sites involved in protein-coding processes. Now, if we add to these widespread non-protein-coding co-adaptive interactions those due to protein-coding adaptive processes, we will have a more complete picture of pan-adaptive evolutionary processes. The reader interested in protein-coding functions will find the total nucleotide information for this mtDNA through the accession number in Genbank. The fourth segment includes most of the control region with several TA tandem repeat regions (see it in Fig 1). However, the information about the gene organization is, at present, not necessary for our study.

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

ACKNOWLEDGEMENT

LITERATURE CITED

DRAKE JW (1993) Rates of spontaneous mutation among RNA viruses. Proc Natl Acad Sci USA 90:4171-4175.

DRAKE JW (1999) The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Ann N Y Acad Sci 870: 100-107.

DRAKE JW (2009) Avoiding dangerous missense: Thermophiles display especially low mutation rates. PloS Genetics 5, Issue 6, e1000520.

DRAKE JW, CHARLESWORTH B, CHARLESWORTH D, CROW JF (1998) Rates of spontaneous mutation. Genetics 148: 1667-1686.

FELLER W (1951) Diffusion processes in genetics. Proc Second Berkeley Symp Math Stat Prob. Pp 227-246.

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

GOULD SJ (2002) The structure of evolutionary theory. The Belknap Press of Harvard University Press, Cambridge, MA, USA. pp: 518-524.

GREDILLA R, BOHR VA, STEVNSNER T Mitochondrial DNA repair and association with aging – An update. Exp Gerontol (2010) doi:10.1016/j.exger.2010.01.017

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

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

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

KOCH RE (1971) The influence of neighboring base pairs upon base-pair substitution mutation rates. PNAS (USA) 68: 773-776.

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

MACKWAN RR, CARVER GT, KISSLING GE, DRAKE JW, GROGAN DW (2008) The rate and character of spontaneous mutation in Thermus thermophilus. Genetics 180: 17-25.

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.

PREVOSTI A (2000) La selección natural 30 años después. Memorias de la Real Academia de Ciencias y Artes de Barcelona. Tercera Época Nº 964; 58(9): 349-397.

TUPPEN HAL, BLAKELY EL, TURNBULL DM, TAYLOR RW (2010) Mitochondrial DNA mutations and human diseases. Biochim Biophys Acta 1797:113-128.

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

VALENZUELA CY (2002b) Does biotic life exist?. In: Palyi G, Zucchi C & Caglioti L (eds) Fundamentals of Life. Paris: Elsevier. pp: 331-334.

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 conditions for neutral evolution. Genet Mol Biol 32: 159-169.

VALENZUELA CY (2010a) Internucleotide correlation and nucleotide periodicity in *Drosophila* mtDNA: New evidence for panselective evolution. Biol Res 43:497-502.

VALENZUELA CY (2010b) Periodicidades e interacciones del DNA. El fin del neutralismo y del casi neutralismo (Textbook, in press)

VALENZUELA CY (2011) Nucleotide Correlation and Periodicity. End of Neutral and Nearly-Neutral Evolution (to be sent).

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.

APPENDIX 1

METHOD TO STUDY THE ASSOCIATION BETWEEN TWO NUCLEOTIDES
SEPARATED (SEP) BY 0, 1, 2 ... K NUCLEOTIDE SITES (VALENZUELA, 2010b)

SITE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Base	A	A	T	G	G	C	T	T	A	C	G	G	C	T	C	A

Sep						PAIR OF BASES					
0	1-2, A-A	2-3, A-T	3-4, T-G	4-5, G-G	5-6, G-C	6-7, C-T	7-8, T-T	8-9, T-A	9-10, A-C	10-11, C-G	etc.
1	1-3, A-T	2-4, A-G	3-5, T-G	4-6, G-C	5-7, G-T	6-8, C-T	7-9, T-A	8-10, T-C	9-11, A-G	etc.	
2	1-4, A-G	2-5, A-G	3-6, T-C	4-7, G-T	5-8, G-T	6-9, C-A	7-10, T-C	8-11, T-G	9-12, A-G	etc.	
3	1-5, A-G	2-6, A-C	3-7, T-T	4-8, G-T	5-9, G-A	6-10, C-C	7-11, T-G	8-12, T-G	9-13, A-C	etc.	

And so on, until K Sep. The distribution of pairs is compared to a random distribution by a χ^2_9 test (4 rows for the first and 4 columns for the second base yield 9 degrees of freedom). The expected and observed values are compared for each of the 16 pairs and the χ^2_1 contribution is obtained. The addition of these contributions increases the χ^2_9 value.

These associations have no relation to the expected protein-coding functions. With 0Sep, in a protein-coding segment a significant distribution will imply non-random associations between the 1° and 2° codon positions, between the 2° and 3° positions, between the 3° position of a codon and the 1° position of the next codon, and so on. The 1Sep implies association between the 1° and 3 ° codon positions, between the 2° and the first position of the next codon, between the 3° and the 2° codon position of the next codon, and so on.

APPENDIX 2

Deviations from randomness for the 16 dinucleotides
whose bases are separated by 17 nucleotide sites, ordered by the χ^2_1 value (significance) contribution to the total χ^2_9 value.

Pair	Expected	Observed	D Sign	Obs-Exp	χ^2_1 cont
G-G	111.9	213	↑	101.1	91.4
C-G	151.6	234	↑	82.4	44.8
C-C	205.5	296	↑	90.5	39.8
G-C	151.7	210	↑	58.3	22.4
A-A	3401.0	3677	↑	276.0	22.4
A-G	616.4	516	↓	-100.4	16.4
G-A	617.3	518	↓	- 99.3	16.0
T-T	3181.7	3399	↑	217.3	14.8
C-A	836.6	732	↓	-104.6	13.1
T-G	596.1	513	↓	- 83.1	11.6
A-C	835.7	749	↓	- 86.7	9.0
G-T	597.1	537	↓	- 60.1	6.1
C-T	809.3	741	↓	- 68.3	5.8
T-C	808.1	746	↓	- 62.1	4.8
A-T	3289.9	3201	↓	- 88.9	2.4
T-A	3289.1	3217	↓	- 72.1	1.6
Total	19499.0	19499		χ^2_9	= 322.2

D Sign = the sign of the Observed-Expected difference; ↑ = more dinucleotides observed than expected; ↓ = fewer dinucliotides observed than expected.