Research Article
Open Access
CrossMark

Maximum likelihood estimation of the evolutionary distance of complete genomes of mitochondrial DNA between human and 16 animals

Abstract

The mitochondrial DNA of vertebrates generally has same structure and functions, even share same numbers of genes, tRNAs, rRNAs and codon regions with different sequences in a narrow range of genome size. The resource of database was downloaded from the Genbank of National Center for Biotechnology Information (NCBI), and using a particular computational program to achieve the best results of alignment and statistical calculations to estimate the transition/transversion in the nucleotide and amino acid substitution, additionally, estimate the evolutionary distance rate between Human versus 16 animals. The results of maximum likelihood method show high rates of substitutions mainly from adenine and thymine to cytosine and guanine (A=>C; A=>G; T=>C; T=>G), respectively. Guanine (G) was the most conserved and stable nucleotide from changes in all over 17 organisms that may be related to the strength of the chemical bonds. The observation of evolutionary distance by the number of substitutions per site between sequences is shown for all three codon positions and non-codon regions. The scores put organisms in groups by comparing the numbers between pairs of sequence within difference and similarity, such as the human, chimpanzee, and gorilla had less distance among them, whilst, a remarkable likelihood between bison and the water buffalo were observed despite the historical and geographical distance between them. Additionally, studying the effect of substitution scores in nucleotide and amino acids on the synonymous/non-synonymous codon substitution in evolutionary distance bias was discussed.

Keywords: maximum likelihood, mitochondrial DNA, evolutionary distance, PERL, CPAN modules

Volume 5 Issue 6 - 2017

Omar Esmaill Hamad,[1] Ismail Akyol,[2] Mehmet Sait Ekinci,[1] Emin Ozkose[1]

[1]Department of Animal Science, Kahramanmaras Sutcu Imam University, Turkey
[2]Biotechnology and Gene Engineering Laboratory, Turkey

Correspondence: Mehmet Sait Ekinci, Kahramanmaras Sutcu Imam University, Agriculture Faculty, Animal Science Department 46100/ Kahramanmaras, Turkey, Tel+90 344 2802100, Fax+90 344 2802109, Email sekinci@ksu.edu.tr

Received: June 30, 2017 | **Published:** July 19, 2017

Introduction

Mitochondrial DNA (mtDNA) is a masterpiece of polynucleotide intelligence provided as a double stranded circular DNA, to be the spirit and manager of molecular activities in eukaryotic cells. The nucleic DNA activity and regulation depend on the signals and the levels of the tRNA and rRNA which mtDNA produced in the cell. The foremost attention-grabbing issue, mtDNA has the ability to adapt with each individual cell by modifying the sequence by slightly the initiation and termination points, likewise, the direction of transcription 5'=>3' or 3'=>5'.[1] Mitochondria generate most of the cellular energy within the form of adenosine triphosphate ATP, regulate cellular oxidation-reduction state and integrate several of the signals for initiating necrobiosis. By means of retrograde signaling, mitochondrial communicate of these events to the nucleus and thus modulate nuclear organic phenomenon and cell cycle. In human, mitochondrial pathology leads to a massive array of pathologies, and many diseases result from various defects of mitochondrial biogenesis and maintenance, metabolism chain complexes or individual mitochondrial proteins.[2]

Perhaps, the estimation of the distance between two sequences is the simplest phylogenetic analysis, because the calculation of pairwise distances is the first step in distance-matrix methods of phylogeny reconstruction. Cluster algorithms used to convert a distance matrix into a phylogenetic tree. Markov-process models of nucleotide substitution used in distance estimation form the basis of likelihood and Bayesian analysis of multiple sequences on a phylogeny.[3]

To estimate the number of substitutions, it is needed a probabilistic model to describe changes between nucleotides this purpose. Continuous-time Markov chains are commonly used for the nucleotide sites in the sequence are normally measured to be evolving independently of each other. Substitutions at any particular site are described by a Markov chain, with the nucleotides to be the states of the chain. The main advantage of a Markov chain is that it has no memory given the present, likewise, the future does not depend on the past. In other words, the probability with which the chain jumps into different nucleotide states depends on the current state, but not on how the current state is reached. This is referred to as the Markovian property. Besides this basic assumption, it is often placed further constraints on substitution rates between nucleotides, leading to variable models of nucleotide substitution.[4]

The first application of a maximum likelihood method to tree construction was made by Cavalli-Sforza and Edwards (1967) for estimation gene frequency data. Later, Felsenstein (1973, 1981) developed maximum likelihood algorithms for amino acid and nucleotide sequence data. Because this approach involves fairly sophisticated statistical theory, that presented only some basic principles of the method without any mathematical details.[5] A critical element is how the probabilities of the various changes are calculated. These probabilities depend on assumptions concerning the process of nucleotide substitution and the branch lengths, which in turn depend on the rate of substitution and the evolutionary time. These branch lengths are usually unknown and must be estimated as part of the process of computing the likelihood. The methods for

© 2017 Hamad et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and build upon your work non-commercially.

discovering the branch lengths that maximize the likelihood value usually involve an iterative approach also the likelihoods depend on the model of nucleotide substitution, a tree with the largest likelihood value under one substitution model. The maximum likelihood method is computationally extremely time-consuming, and so was not used often in the past. With the development of fast computers, the method is now used fairly often, although it is an exhaustive version it is still only applicable to a modest number of taxa.[6,7]

To outline some main points, as an observation from previous researchers have been documented about estimate the mitochondrial DNA evolutionary distance within maximum likelihood method among animals within various visions and scoring parameters. Started with the pronouncement by Irwin, Kocher, and Wilson (1991) studies on Evolution of the cytochrome-b gene of mammals that obtained 17 complete gene sequences representing three orders of hoofed mammals (ungulates) and dolphins (cetaceans). The fossil record of some ungulate lineages allowed estimation of the evolutionary rates for various components of the cytochrome DNA and amino acid sequences. The relative rates of substitution at first, second, and third positions within codons are in the ratio 10 to 1 to at least 33. For deep divergences (>5million years) it appears that both replacements and silent transversion in this mitochondrial gene can be used for phylogenetic inference. Phylogenetic findings include the association of (Drosophila 12 Genomes et al.) cetaceans, artiodactyls, and perissodactyls to the exclusion of elephants and humans, (2) pronghorn and fallow deer to the exclusion of bovids like cow, sheep, and goat, (3) sheep and goat to the exclusion of other pecans such as cow, giraffe, deer, and pronghorn, and (4) advanced ruminants to the exclusion of the chevrotain and other artiodactyls. Comparisons of these cytochrome sequences support current structure-function models for this membrane-spanning protein. Although there has been relatively a research results into mitochondrial DNA sequence divergence and diversity Chen & Li[8] about genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees, The average sequence divergence was only 1.24% 5 0.07% for the human-chimpanzee pair, 1.62% 5 0.08% for the human-gorilla pair, and 1.63% 5 0.08% for the chimpanzee-gorilla pair relation. More importantly, the modern hypothesis of the evolutionary relationship between human and pig that based on the assumption of similarity in some organs tissues like kidneys and eyes.[9,10]

All in all, the huge similarity in structure and functions appears in the genomic mitochondrial DNA (mtDNA) of vertebrates. That was an encouraging point to consider about having a chance to make a comparative view among seventeen organisms including the human mtDNA within maximum likelihood method to estimate the evolutionary distance and the substitution effects of the nucleotides and amino acids on the codons frequencies. Moreover, may be generating hypothesis about the evolutionary of the organisms in interest within the convenient algorithmic method to compute the sequences of the complete genomic mitochondrial DNA in alignment method and applying particular mathematical methods. To accomplish the aims of discovery the evolutionary distances between Human versus other sixteen vertebrate organisms.

Materials and methods

The sources of database

For Maximum likelihood estimation of the evolutionary distance of the complete genomes of Mitochondrial DNA (mtDNA) between Human's versus other 16 animals are investigated. The databases of all vertebrates for mitochondrial DNA (mtDNA) sequences were downloaded from the Genbank of National Center for Biotechnology Information NCBI[11] (www.ncbi.nlm.nih.gov/GENOME).[11,12] To find out the most trusted and proved sequences, by looking for same sequences could be found in International Nucleotide Sequence Database Collaboration (INSDC) (www.insdc.org). In this case, it is worth to mention the Human's mtDNA is the Cambridge reference sequence (isogg.org/wiki/Cambridge_Reference_Sequence), is count as the central sequence which all researchers on mitochondrial DNA of human need to use it for comparison and studying the variation rate from this sequence.[13]

The reason behind choosing these organisms as it mentioned in Table 1, being in interest to get the genomic mtDNA and apply them in the comparative study, is the historical observation in the similarity of morphological and physiological characteristics which known as related to each other, like, Arabian camel with Bactrian camel, so between sheep and goat, Likewise, some of these similarities between organisms were caused the most controversial and debatable issues among the biologists, for the evolutionary relationship between human and chimpanzee.[14]

Table 1 List of organisms which involved in the evolutionary study, with the information of database of the complete genome, mtDNA

S. no	Taxa	*Latin name*	Accession numbers	INSDC Number	References
1	Human	*Homo sapiens*	NC_012920	J01415.2	16
2	Chimpanzee	*Pan troglodytes*	NC_001643	D38113.1	62
3	Gorilla	*gorilla gorilla*	NC_011120	X93347.1	63
4	Cattle	*Bos taurus*	NC_006853	AY526085.1	(Chung HY, Ha JM.,2005) *
5	Water buffalo	*Bubalus bubalis*	NC_006295	AY702618.1	56*
6	Bison	*Bison bison*	NC_012346	EU177871.1	64
7	Arabian camel	*Camelus dromedarius*	NC_009849	EU159113.1	(Huang X et all, 2007)*
8	Bactrian camel	*Camelus bactrianus*	NC_009628	EF212037.2	65
9	Horse	*Equus caballus*	NC_001640	X79547.1	66
10	Sheep	*Ovis aries*	NC_001941	AF010406.1	67
11	Goat	*Capra hircus*	NC_005044	GU295658.1	68

Table Continued....

S. no	Taxa	*Latin name*	Accession numbers	INSDC Number	References
12	Pig	*Sus scrofa*	NC_000845	AF034253.1	69
13	Chicken	*Gallus gallus*	NC_001323	X52392.1	70
14	Rabbit	*Oryctolagus cuniculus*	NC_001913	AJ001588.1	71
15	Dog	*Canis lupus familiaris*	NC_002008	U96639.2	72
16	Domestic cat	*Felis catus*	NC_001700	U20753.1	73
17	House mouse	*Mus musculus*	NC_005089	AY172335.1	74

More importantly, including in the list some animals that considered as a high contrast with all, even out the cycle of mammals like chicken, then include the sequences in a parallel way with each other's for comparison view between them evenly. Lastly, the combination of these organisms actually put this study in a unique position as far as it is concerned.[15] As below Table 1, demonstrates the comparison of Human's mitochondrial DNA versus other 16 vertebrate organisms with the accession number of NCBI, and the INSDC number. Furthermore, with publication in the Medline database of references and abstracts on life sciences and biomedical (PubMed), but three references of these sequences were unpublished and they have NCBI Project numbers only. Firstly, cattle's project number is 13366 submitted in 22-February-2005(www.ncbi.nlm.nih.gov/nuccore/60101824/). Secondly, water buffalo with project number 13052 submitted in 02-Agust-2004 (www.ncbi.nlm.nih.gov/nuccore/NC_006295). Finally, Arabian camel with project number 20873 submitted in 17-September-2007 (www.ncbi.nlm.nih.gov/nuccore/NC_009849).

Computational approach

In the most trusted and depended websites which provide an open source bioinformatics tool services and databases resources. Practical extraction and report language known as Perl which is one of the major program applied in Bioinformatics for decades (https://www.perl.org/) supported by organization of Comprehensive Perl Archive Network(CPAN) (www.cpan.org) that provide thousands of modules shared from scientists and computer programmers studying on bioinformatics.[16–19] Nevertheless, needed to extract some mathematical functions from (megasoftware.net) which is academic open-public software for molecular evolutionary genetic analysis MEGA7-CC-Porto.[20]

Algorithm

A critical point is to decide choosing which algorithmic method would be used because it is related with the best way for interring data in a computer with choosing and designing the codes, then apply them to obtain the best results as it possible. The modules of Perl programming language which invented by the legendary computer programmer Larry Wall (en.wikipedia.org/wiki/LarryWall), were downloaded from CPAN (www.cpan.org& metacpan.org) also from (www.github.com). It is worth mentioning, that programming languages are easy to use but in the other hand, difficult to understand and learn, also it could not be used directly after downloaded from the open source access websites because they are designed for general purposes and need manipulating with adding the private data and the mathematical problems serve the particular study.[21]

The 17 sequences of mtDNA compiled and saved in a FASTA format (filename.fasta) then the codes were downloaded from the shell of CPAN by using the black window called command (CMD) in windows (http://www.bioperl.org/wiki/Installing_BioPerl_on_Windows) by using special codes for test and install in the computer, as an example (CPAN>test Bio:: Tools:: Run:: Alignment:: Muscle) and install the module if it works in this code (CPAN>install Bio:: Tools:: Run:: Alignment:: Muscle), next, open installed codes with a text editor like (ActiveState Komodo IDE8) and start to import the own data and mathematical problems by using some specific regular expressions to compile the all in one code like ($seq(x) =”<sequence (x)>”) and(use <module>;). Then save the code in Perl format (filename.pl).[22–24]

Another essential point is the best modules were served the research poses. Firstly, an object for the calculation of an iterative multiple sequence alignment from a set of unaligned sequences or alignments using the MUSCLE program (Bio:: Tools:: Run:: Alignment:: Muscle) authored by Christopher Fields in 2011, (metacpan.org/pod/Bio::Tools::Run:: Alignment:: Muscle). Secondly, the Representation for biological sequence alignment (Bio:: Tools:: Alignment:: Overview) announced by Felipe da Veiga Leprevost in 2014, (metacpan.org/pod/ Bio:: Tools:: Alignment:: Overview). Thirdly, the interface for evolving sequences (Bio::Seq Evolution::EvolutionI) reported by Christopher Fields in 2014, (metacpan.org/pod/Bio:: SeqEvolution::evolution). Finally, the module of Maximum likelihood methods (Bio::Tools::Run::Phylo::Molphy::ProtML) authored by Jason Stajich in 2011, (metal pan. org/ pod/Bio::Tools::Run::Phylo::Molphy::ProtML).

Alignment of 17 sequences

The 17 sequences of mtDNA were arranged in parallel depending on coding and noncoding regions of DNA even the proteins to distinguish regions of similarity and disparity. Consequently, get the distance and evolutionary relationships between the sequences. The dynamic programming algorithm of the multiple sequence alignment is by adding spaces (INDEL) or gaps in the sequences. Then calculate the highest scores of the alignment matrix were always being the diagonal arrows to yield an equal length sequences, in condition that obtain an optimum score value, then going to calculate the number of matches, mismatches and gaps, finally, apply the next model of maximum value.[25,26]

The computational multiple sequence alignment (MUSCLE) method used to provide high accuracy for creating different arrangements of high scale amino acids and nucleotide sequences.[27–29] The velocity and precision of MUSCLE were contrasted with other three methods. Firstly, Tree-based Consistency Objective Function For alignment Evaluation (T-Coffee). Secondly, multiple sequence alignment program for amino acid or nucleotide sequences (MAFFT). Finally, with Clustal is a series of widely used computer programs for multiple sequence alignment (CLUSTALW). The achievement of

most elevated or joint highest rank in precision in all tests. At the point when utilized without refinement its precision is the same as T-Coffee or MAFFT and is the speediest at adjusting extensive sequences.[30,31]

Relative synonymous codon usage (RSCU)

The numerous amino acids are coded by more than one codon, thus the several of multiple codons for a given amino acids are synonymous. Nevertheless, many genes display a nonrandom usage of synonymous codons for specific amino acids.[32,33] In addition, the codes of the mathematical problem extracted from program MEGA7-cc-Porto (www.megasoftware.net) in a particular file format (filename.mao).[34,35]

Maximum likelihood

The maximum likelihood method considered as the cornerstone of modern statistics depend on the parametric model of evolution appropriate for the characters and algorithm that will search through the trees The model depends essentially on the nature of the characters under study, among the many possible models of character evolution.[26] The statement of the problem, suppose when to have a random sample $x1,x2...xn$ whose assumed probability distribution depends on some unknown parameter θ. The primary goal here will be to find a point estimator $u(x1,x2...xn)$ such that $u(x1,x2...xn)$ is a "good" point estimate of θ, where $x1,x2...xn$ are the observed values of the random sample. for example, if planned to take a random sample $x1,x2...xn$ for which the xi are assumed to be normally distributed with mean μ and variance $\sigma 2$, then our goal will be to find a good estimate of μ , say, using the data $x1,x2...xn$ that we obtained from our specific random sample. The Basic Idea (onlinecourses.science.psu.edu; megasoftware.net).

Estimating the evolutionary distances between genomic sequences

The evolutionary distance between sequences usually is measured by the number of a polynucleotide or amino acid substitutions appear between them and the Alignment methods are used to compute evolutionary distances between DNA and protein sequences as a basis for phylogenetic reconstruction. It is calculated from the number of word matches between them, additionally; compute the substitutions of nucleotide, amino acids, and the synonymous-non-synonymous codes. Nucleotide sequences are compared nucleotide-by-nucleotide, these distances could be computed for protein coding and noncoding nucleotide sequences. Residue-by-residue for amino acid and Codon-by-codon for synonymous-non-synonymous codons with complete detection of gaps of missing data treatments and the substitution included the transition-transversion within maximum likelihood method.[36–40]

Results and discussion

Comparative view of the nucleotide and amino acids sequences in sizes

The Figure 1 provides a vision about the difference of the genome sizes in mitochondrial DNA between human and the other vertebrates' species, also the amino acid size numbers were around 5000 when the nucleotide sizes around 17000 bases, representing the total translation of the protein. However, the number of proteins is constant and same in all species that is 13, and even we have 22 tRNAs and 2 rRNAs, these numbers did not change between the 17 vertebrates. They have the same annotated structure with alternative lengths and sequences, to help to provide more functions for the same job, and this is what molecular evolution means.[41–43]

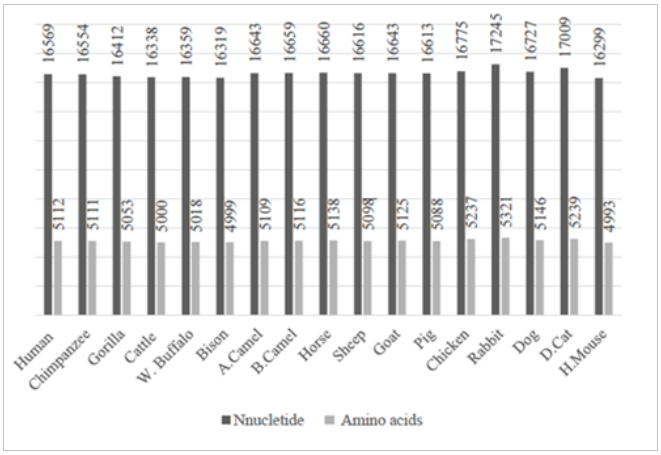

Figure 1 The length graph of nucleotide bases and amino acids number in Human's mitochondrial DNA with other 16 vertebrates.

Estimation of the codon usage bias

The results of the codon bias, Table 2 show a prejudice in codon frequencies has been used for the conformity with previous results between nucleotide composition within amino acid composition, it is shown the top scores in count for Leucine, Isoleucine, Proline and Serine, and even with relative synonymous codon usage.[44–48] The reason behind these results is due to tRNA corresponding to the codons CUA, UCA, AGC…..etc., are more abundant, because the translationary machinery tend to use abundant tRNA to produce proteins.[49,50]

Table 2 The frequency account of the codons and the relative synonymous codon usage (RSCU), in all over the 17 aligned mammalian mitochondrial sequences

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	109	0.95	UCU(S)	100	1.19	UAU(Y)	130	1.05	UGU(C)	28.4	0.82
UUC(F)	120	1.05	UCC(S)	117	1.38	UAC(Y)	118	0.95	UGC(C)	40.8	1.18
UUA(L)	132	1.32	UCA(S)	126	1.49	UAA(*)	135	1.38	UGA(*)	74.6	0.76
UUG(L)	44.3	0.44	UCG(S)	33.5	0.4	UAG(*)	84.4	0.86	UGG(W)	28.6	1
CUU(L)	90.8	0.91	CCU(P)	138	1.27	CAU(H)	126	1.07	CGU(R)	27.1	0.78
CUC(L)	101	1.01	CCC(P)	138	1.27	CAC(H)	110	0.93	CGC(R)	29.9	0.86
CUA(L)	185	1.85	CCA(P)	120	1.11	CAA(Q)	145	1.39	CGA(R)	35.4	1.02

Table Continued....

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
CUG(L)	47.4	0.47	CCG(P)	37.6	0.35	CAG(Q)	63.2	0.61	CGG(R)	15.9	0.46
AUU(I)	140	0.97	ACU(T)	130	1.1	AAU(N)	138	0.96	AGU(S)	45.6	0.54
AUC(I)	134	0.93	ACC(T)	143	1.21	AAC(N)	149	1.04	AGC(S)	85.5	1.01
AUA(I)	158	1.1	ACA(T)	154	1.31	AAA(K)	181	1.45	AGA(R)	62.4	1.79
AUG(M)	66.5	1	ACG(T)	44.9	0.38	AAG(K)	69.3	0.55	AGG(R)	37.9	1.09
GUU(V)	36.2	0.85	GCU(A)	69.5	1.1	GAU(D)	56.1	1	GGU(G)	32.8	0.76
GUC(V)	39.6	0.93	GCC(A)	92.4	1.46	GAC(D)	55.9	1	GGC(G)	46	1.06
GUA(V)	68.5	1.6	GCA(A)	74.7	1.18	GAA(E)	71.7	1.16	GGA(G)	66.6	1.54
GUG(V)	26.7	0.62	GCG(A)	17.1	0.27	GAG(E)	51.5	0.84	GGG(G)	27.6	0.64

*Termination codes of transcription.

Estimation of transition/transversion matrix by maximum composite likelihood (mcl)

The results obtained by estimating the Maximum Likelihood substitution patterns called transition (inside the purine group or the pyrimidine group) and the transversion (between the purine and pyrimidine groups), by observing the changes in the nucleotide through the 17 mitochondrial genome sequences, are illustrated vertically in columns of Table 3. The Guanine (G) was the most conservative nucleotide in spite of showing substitution changes, and the most changeable nucleotide to others was the Adenine (A) in general, calculating the total of substitutions from adenine to the other nucleotides was the highest 33. 7277 which came from (A=>T 5.9789+A=>C 11.5959+A=>G 16.1529) and the lowest total score of substitutions were from guanine to other nucleotides 8.3552 which came from (G=>A 7.0297+G=>T 0.6117+G=T 0.7138), another essential point is the highest substitution shown the transition inside the pyrimidine group between two nucleotides T=>C 19.9855 and C=>T 20.3332.

Table 3 Maximum likelihood estimation of transition/transversion bias

From\To	A	T	C	G
A	-	5.0463	9.9574	7.0297
T	5.9789	-	20.3332	0.6117
C	11.5959	19.9855	-	0.7138
G	16.1529	1.1863	1.4085	-

Additionally, the results agreed with the next following research in nucleotide behaviors that may be related to strength of the chemical bonds in spite of different area investigations also some of them called guanine an ancestral nucleotide as the most conserved nucleotide.[51–54]

Nucleotide pair frequencies from alignment of 17 sequences

The calculation of the transition/transversion in a maximum probable number of 16 nucleotide pairs that could obtain from four different nucleotides, through alignment of 17 sequences in the positions 1st, 2nd and 3rd respectively. The R ratio used as a parameter score that equal 1, between transition and transversion that show harmony in levels of exchanges in all positions. In other words, the number of transitions is semi equal the transversion in all 16 pairs of nucleotides within the codon positions. More importantly, in the second part of Table 4, which illustrate the frequencies of the nucleotide pairs as a genome map of mtDNA estimating the probability of codon frequencies in the three positions respectively, also help to predict the sequences of proteins by this map. For instance, the highest levels of AA exemplifies, a high ratio of Asparagine N and Lysine K because them codon contain the AA.

Furthermore, as demonstrated the top number of observations in Table 4, it could be noticed that the harmony in the numbers of observation through the three positions. The line chart seems to be one line in spite of there are three lines in all over the 17 mammalian mitochondrial DNA sequences aligned against each other's.[55–57]

Table 4 The transition/transversion calculated of 16 probable nucleotide pair frequencies by alignment of 17 sequences, in three codon positions

	*ii	si	sv	R	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG
Avg	12097	1975	1814	1	3188	730	322	65	619	3144	411	70	309	498	4047	316	59	80	310	1718
1st	4184	585	517	1	1087	209	97	20	179	1027	113	22	89	139	1389	99	16	21	98	680
2nd	4038	642	619	1	1124	239	106	22	206	1092	146	25	100	175	1297	100	19	28	97	524
3rd	3876	749	678	1	977	283	119	23	234	1025	153	23	120	184	1360	117	23	31	115	513

*ii, A total of 16 nucleotide pairs identical pairs; si, A total of 16 nucleotide pairs transition pairs; sv, transversion pairs; R, the ratio of transition/ transversion (R=si/sv) with a total of 16 nucleotide pairs

Nucleotide evolutionary distance

The parameter of results depended on the numbers of base substitutions per site from between sequences as are shown in. The analyses were conducted using the Maximum composite likelihood model. Also, the rate variation among sites was modeled with a gamma distribution shape parameter score value is equaled 1. Codon positions included 1st+2nd+3rd+Noncoding. Additionally, all positions containing gaps and missing data were eliminated. There were a total of 14430 positions in the final dataset.

Deeply, in details, the results of evolutionary distance as shown in, separated the 17 organisms in several groups depending on the value of the minimum score. Firstly, the nearest animals to human are chimpanzee and gorilla (0.0913, 0.1157) respectively. Secondly, is the biggest group of animals led by water buffalo following by cattle 0.2681, also with bison and 0.1329 then Arabian camel 0.2689. The water buffalo could lead the major group of relationships by the highest scores to appear diagonally in the lower matrix between cattle, bison Arabian camel, Bactrian camel, horse, sheep, and goat. Moreover, results demonstrate the lowest divergence between an Arabian camel and a Bactrian camel, with cattle, and also shown among dog, rabbit, and sheep.

The most interesting results that relate with highest evolutionary distance in the pig with all 16 organisms in contrast whilst, chicken also had high divergence scores with all but significantly lower than the mtDNA of a pig. That is mean in spite of the highly morphological contrast between chicken and other organisms even it is not a mammal but shows a considerable similarity in mtDNA with all other animals including human.

The evident about molecular evolutionary by distance estimation, from, applying the Markov model of maximum likelihood method between pairs of sequence alignment results. It could be observed the evolutionary distance among all mammals' organisms in spite of the variation in scores, the number of base substitutions per site from between sequences are shown for all three codon position and non-codon regions,[38,58] also these scores put the organisms in groups by comparing the numbers between pairs of sequence, for instance, the human, chimpanzee and gorilla, likewise, the discovery of likelihood between bison and the water buffalo despite the historical and geographical distance between them. The highest distance ever was observed pig compared to the all other organisms.[59,60]

Amino acid substitution evolutionary distance

The number of amino acid differences estimated with per sequence from other sequences is shown, in, which illustrates the results of involving 17 sequences of amino acid. The rate variation among sites was modeled with a gamma distribution score is equal to 1. Coding data translated assuming a vertebrate mitochondrial DNA genetic code table. All positions containing gaps and missing data were eliminated. Moreover, the coding data was translated assuming a Vertebrate Mitochondrial genetic code table. All positions containing gaps and missing data were eliminated. There was a total of 4091 positions in the final dataset. Similar to the previous results, the pig got the highest contrast distance against the others.[61]

The results as demonstrated in, came in the same way with the substitution of nucleotide, evolutionary distance in the alignment of 17 sequences of mtDNA. Furthermore, the chicken as a bird is less divergence than pig, dog, cat, and mouse in compare with a human. Similar results are observed between chicken, rabbit, dog, cat, and mouse with a pig.

Synonymous/non-synonymous codon substitution evolutionary distance

The aim behind estimation of codon-based evolutionary divergence between sequences is to see the effect of substitutions in nucleotides on the codons of amino acids if that cause any changes in protein sequences that may cause a difference in annotation or function in the genome of mitochondrial DNA. The number of synonymous differences per sequence from shown, sequences involved in all positions containing gaps and missing data which eliminated. A total of 4091 positions in the final dataset.[37,38,58]

The results in, demonstrate the effect of substitution levels on the frequencies of codon changes synonymously or synonymously. Firstly, the results between the human-chimpanzee pair were 445.33, human- gorilla pair 563.00 and chimpanzee-gorilla was 514.00. Secondly, the lowest score observed in Bactrian camel-horse pair 349.50. Finally, the huge change and divergence for a pig with all other organisms involved in this study Results in the respectively, explained and illustrated with results, the codons also changes show the harmony of the same rhythm with nucleotides and amino acids results, which provide same protein in another sequence. In fact, it was a shock if a study on mtDNA with 10 times more than nucleic DNA in substitution and could conserve itself through the time of evolution within animals. Since thousands of years mtDNA strict in the same function and annotation with keep changing its sequences. Actually, nowadays, this is a big foot step for human kind to explain or pretend understand the mechanism of mtDNA in evolution with all available sciences and Refutes all studies that talk about the evolutionary relationship between human and pig.[9,10,72–78]

Conclusion

The current study opened a gate of huge question like the similarity between human with chimpanzee and gorilla, also how could be the divergence of the pig greater than chicken as a bird with other mammals even with a human. Moreover, this study generates a motivation to study the phylogenetic in deep and the *de novo* annotation looking for some confused answers that help to discover and understand more about mitochondrial DNA.

Acknowledgements

None.

Conflict of interest

The author declares no conflict of interest.

References

1. Wilson JH, Hunt T. Molecular biology of the cell. 4th ed. *a problems approach*. USA: Garland Science; 2002.
2. Cízková A, Stránecký V, Ivánek R, et al. Development of a human mitochondrial oligonucleotide microarray (h–MitoArray) and gene expression analysis of fibroblast cell lines from 13 patients with isolated F1Fo ATP synthase deficiency. *BMC Genomics*. 2008;9:38.
3. Yang Z. *Molecular Evolution: A Statistical Approach*. UK: OUP Oxford; 2014.
4. Nielsen R. *Statistical methods in molecular evolution*. USA: Springer; 2005.
5. Li WH, Graur D. *Fundamentals of molecular evolution*. USA: Sunderland, Mass, Sinauer Associates; 1991.
6. Anderson JP, Rodrigo AG, Learn GH, et al. Substitution model of sequence evolution for the human immunodeficiency virus type 1 subtype B gp120 gene over the C2–V5 region. *J Mol Evol*. 2001;53(1):55–62.
7. Marjoram P, Molitor J, Plagnol V, et al. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A*. 2003;100(26):15324–15328.
8. Chen FC, Li WH. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet*. 2001;68(2):444–456.

9. Brown IH. The epidemiology and evolution of influenza viruses in pigs. *Vet Microbiol*. 2000;74(1):29–46.
10. Rettenberger G, Klett C, Zechner U, et al. Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. *Genomics*. 1995;26(2):372–378.
11. Coordinators NR. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2015;41(Database issue):D8–D20.
12. Aali M, Moradi–Shahrbabak M, Moradi–Shahrbabak H, et al. Detecting novel SNPs and breed–specific haplotypes at calpastatin gene in Iranian fat– and thin–tailed sheep breeds and their effects on protein structure. *Gene*. 2014;537(1):132–139.
13. Andrews RM, Kubacka I, Chinnery PF, et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999;23(2):147.
14. Jensen JV. Thomas Henry Huxley's address at the opening of the Johns Hopkins University in September 1876. *Notes Rec R Soc Lond*. 1993;47(2):257–269.
15. Beaz–Hidalgo R, Hossain MJ, Liles MR, et al. Strategies to avoid wrongly labeled genomes using as an example the detected wrong taxonomic affiliation for Aeromonas genomes in the GenBank database. *PloS One*. 2015;10(1):e0115813.
16. Li C, Guan F, Dai Z, et al. [Preparation and characterization of following the national standard anti–Brucella abortus serum, bovine]. *Sheng Wu Gong Cheng Xue Bao*. 2011;27(5):812–816.
17. Magwedere K, Bishi A, Tjipura–Zaire G, et al. Brucellae through the food chain:the role of sheep, goats, and springbok (Antidorcus marsupials) as sources of human infections in Namibia. *J S Afr Vet Assoc*. 2011;82(4):205–212.
18. Mohammed FU, Ibrahim S, Ajogi I, et al. Prevalence of bovine brucellosis and risk factors assessment in cattle herds in jigsaw state. *ISRN Veterinary Science*. 2011:132897.
19. Surendran N, Sriranganathan N, Lawler H, et al.Efficacy of vaccination strategies against intranasal challenge with Brucella abortus in BALB/c mice. *Vaccine*. 2011;29(15):2749–2755.
20. Stecher G, Liu L, Sanderford M, et al. MEGA–MD: molecular evolutionary genetics analysis software with a mutational diagnosis of amino acid variation. *Bioinformatics*. 2014;30(9):1305–1307.
21. http://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=5949&context=etd
22. Gao JF, Zhao Q, Liu GH, et al. Comparative analyses of the complete mitochondrial genomes of the two ruminant hookworms Bunostomum trigonocephaly and Bunostomum phlebotomy. *Gene*. 2014;541(2):92–100.
23. Leimeister CA, Boden M, Horwege S, et al. Fast alignment–free sequence comparison using spaced–word frequencies. *Bioinformatics*. 2014;30(14):1991–1999.
24. Qing J, Yan D, Zhou Y, et al. Whole–exome sequencing to decipher the genetic heterogeneity of hearing the loss in a Chinese family with deaf by deaf mating. *PloS One*. 2014;9(10):e109178.
25. Durbin R, Eddy S, Krogh A, et al. *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. UK: Cambridge University Press; 1998.
26. Yang Z. *Computational molecular evolution*. UK: Oxford University Press; 2006.
27. Hassanin A, An J, Ropiquet A, et al. Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals:Application to the tribe Bovini (Cetartiodactyla, Bovidae). *Mol Phylogenet Evol*. 2013;66(3):766–775.
28. Jia W, Yan H, Lou Z, et al. Mitochondrial genes and genomes support a cryptic species of tapeworm within Taenia taeniae formins. *Acta Trop*. 2012;123(3):154–163.
29. Martínez–Pérez JM, Robles–Pérez D, Rojo–Vázquez FA, et al. Comparison of three different techniques to diagnose Fasciola hepatica infection in experimentally and naturally infected sheep. *Veterinary Parasitology*. 2012;190(1–2):80–86.
30. Bonhomme F, Orth A, Cucchi T, et al. Genetic differentiation of the house mouse around the Mediterranean basin:matrilineal footprints of early and late colonization. *Proc Biol Sci*. 2011;278(1708):1034–1043.
31. Zhang W, Yue B, Wang X, et al. Analysis of variable sites between two complete South China tiger (Panthera tigris Amoy Ensis) mitochondrial genomes. *Mol Biol Rep*. 2011;38(7):4257–4264.
32. Chang BS, Campbell DL. Bias in phylogenetic reconstruction of vertebrate rhodopsin sequences. *Mol Biol Evol*. 2000;17(8):1220–1231.
33. Yang Z, Yoder AD. Estimation of the transition/transversion rate bias and species sampling. *J Mol Evol*. 1999;48(3):274–283.
34. Wang J, Yu X, Hu B, et al. Physicochemical evolution and molecular adaptation of the cetacean osmoregulation–related gene UT–A2 and implications for functional studies. *Sci Rep*. 2015;5:8795.
35. Xu CP, Lu YY, Yan JY, et al. [Molecular characteristics and its evolution of the complete genome of avian influenza H5N1 virus isolated in Zhejiang province from 2002 to 2006]. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2008;29(11):1114–1118.
36. Blair C, Davy CM, Ngo A, et al. Genealogy and Demographic History of a Widespread Amphibian throughout Indochina. *J Hered*. 2013;104(1):72–85.
37. Kari L, Hill KA, Sayem AS, et al. Mapping the space of genomic signatures. *PloS One*. 2015;10(5):e0119815.
38. Ross HA, Murugan S, Li WL. Testing the reliability of genetic methods of species identification via simulation. *Syst Biol*. 2008;57(2):216–230.
39. Soares I, Amorim A, Goios A. mtDNA office: a software to assign human mtDNA macro haplogroups through automated analysis of the protein coding region. *Mitochondrion*. 2012;12(6):666–668.
40. Soares I, Goios A, Amorim A. Sequence comparison alignment–free approach based on suffix tree and L–words frequency. *Scientific World Journal*. 2012;2012:450124.
41. Ma J, Coarfa C, Qin X, et al. mtDNA Haplogroup and single nucleotide polymorphisms structure human microbiome communities. *BMC Genomics*. 2014;15:257.
42. Pereira F, Soares P, Carneiro J, et al. Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. *Mol Biol Evol*. 2008;25(12):2759–2770.
43. Soto–Hermida A, Fernández–Moreno M, Oreiro N, et al. mtDNA haplogroups and osteoarthritis in different geographic populations. *Mitochondrion*. 2014;15:18–23.
44. D'Erchia AM, Atlante A, Gadaleta G, et al. Tissue–specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass, and respiratory activity. *Mitochondrion*. 2015;20:13–21.
45. Flensburg C, Kinkel SA, Keniry A, et al. A comparison of control samples for ChIP–set of histone modifications. *Front Genet*. 2014;5:329.
46. Kemper MF, Stirone C, Krause DN, et al. Genomic and non–genomic regulation of PGC1 isoforms by estrogen to increase cerebral vascular mitochondrial biogenesis and reactive oxygen species protection. *Eur J Pharmacol*. 2014;723:322–329.
47. Kenny NJ, Shen X, Chan TT, et al. The genome of the Rusty Millipede, Trigoniulus corallines, Illuminates Diplopod, Myriapod, and Arthropod Evolution. *Genome Biol Evol*. 2015;7(5):1280–1295.

48. Simpson L, Douglass SM, Lake JA, et al. Comparison of the Mitochondrial Genomes and Steady State Transcriptomes of Two Strains of the Trypanosomatid Parasite, Leishmania tarantulas. *PLoS Negl Trop Dis*. 2015;9(7):e0003841.

49. Moustafa IM, Uchida A, Wang Y, et al. Structural models of mammalian mitochondrial transcription factor B2. *Biochim Biophys Acta*. 2015;1849(8):987–1002.

50. Huttley GA, Wakefield MJ, Easteal S. Rates of genome evolution and branching order from whole genome analysis. *Mol Biol Evol*. 2007;24(8):1722–1730.

51. Mueller EE, Eder W, Ebner S, et al. The mitochondrial T16189C polymorphism is associated with coronary artery disease in Middle European populations. *PloS One*. 2011;6(1):e16455.

52. Romiguier J, Figuet E, Galtier N, et al. Fast and robust characterization of time–heterogeneous sequence evolutionary processes using substitution mapping. *PloS One*. 2012;7(3):e33852.

53. Woodhams MD, Fernández–Sánchez J, Sumner JG. A New Hierarchy of Phylogenetic Models Consistent with Heterogeneous Substitution Rates. *Syst Biol*. 2015;64(4):638–650.

54. Zaragoza MV, Brandon MC, Diegoli M, et al. Mitochondrial cardiomyopathies: how to identify candidate pathogenic mutations by mitochondrial DNA sequencing, MITOMASTER, and phylogeny. *Eur J Hum Genet*. 2011;19(2):200–207.

55. Boczonadi V, Horvath R. Mitochondria: impaired mitochondrial translation in human disease. *Int J Biochem Cell Biol*. 2014;48:77–84.

56. Chaffee BR, Shang F, Chang ML, et al. Nuclear removal during terminal lens fiber cell differentiation requires CDK1 activity: appropriating mitosis–related nuclear disassembly. *Development*. 2014;141(17):3388–3398.

57. Han MJ, Koc EC, Koc H. Post–translational modification and mitochondrial relocalization of histone H3 during apoptosis induced by staurosporine. *Biochem Biophys Res Commun*. 2014;450(1):802–807.

58. Massingham T, Goldman N. Statistics of the log–set estimator. *Mol Biol Evol*. 2007;24(10):2277–2285.

59. Faure E, Delaye L, Tribolo S, et al. Probable presence of a ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. *Biology Direct*. 2011;6:56.

60. Friesen VL, Burg TM, McCoy KD. Mechanisms of population differentiation in seabirds. *Mol Ecol*. 2007;16(9):1765–1785.

61. Paz A, Frenkel S, Snir S, et al. Implications of human genome structural heterogeneity: functionally related genes tend to reside in organizationally similar genomic regions. *BMC Genomics*. 2014;15:252.

62. Horai S, Hayasaka K, Kondo R, et al. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci U S A*. 1995;92(2):532–536.

63. Li X, He Y. Caspase–2–dependent dendritic cell death, maturation, and priming of T cells in response to Brucella abortus infection. *PloS One*. 2012;7(8):e43512.

64. Achilli A, Olivieri A, Pellecchia M, et al. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biol*. 2008;18(4):R157–R158.

65. Ji R, Cui P, Ding F, et al. Monophyletic origin of domestic Bactrian camel (Camelus bactrianus) and its evolutionary relationship with the extant wild camel (Camelus bactrianus virus). *Anim Genet*. 2009;40(4):377–382.

66. Heller MC, Watson JL, Blanchard MT, et al. Characterization of Brucella abortus infection of bovine monocyte–derived dendritic cells. *Vet Immunol Immunopathol*. 2012;149(3–4):255–261.

67. Kurar E, Splitter GA. Nucleic acid vaccination of Brucella abortus ribosomal L7/L12 gene elicits an immune response. *Vaccine*. 1997;15(17–18):1851–1857.

68. Splitter G, Oliveira S, Carey M, et al. T lymphocyte mediated protection against facultative intracellular bacteria. *Vet Immunol Immunopathol*. 1996;54(1–4):309–319.

69. Lin CS, Sun YL, Liu CY, et al. Complete nucleotide sequence of pig (Sus scrofa) mitochondrial genome and dating evolutionary divergence within Artiodactyla. *Gene*. 1999;236(1):107–114.

70. Newman MJ, Truax RE, French DD, et al. Evidence for genetic control of vaccine–induced antibody responses in cattle. *Vet Immunol Immunopathol*. 1996;50(1–2):43–54.

71. Baldelli R, Calistri P, Battelli G, et al. [Seroepidemiological studies on zoonoses in farm workers in Apulia]. *Ann Ig*. 1995;7(6):445–450.

72. Wyckoff JH 3rd, Howland JL, Confer AW. Comparison of Brucella abortus antigen preparations for in vitro stimulation of immune bovine T–lymphocyte cell lines. *Vet Immunol Immunopathol*. 1993;36(1):45–64.

73. Nemec M, Hidiroglou M, Nielsen K, et al. Effect of vitamin E and selenium supplementation on some immune parameters following vaccination against brucellosis in cattle. *J Anim Sci*. 1990;68(12):4303–4309.

74. Bayona–Bafaluy MP, Acín–Pérez R, Mullikin JC, et al. Revisiting the mouse mitochondrial DNA sequence. *Nucleic Acids Res*. 2003;31(18):5349–5355.

75. Di Rocco P. [Epidemiologic study on the ascertained cases of brucellosis in the hospital at Castel di Sangro in the last 6 years in connection with animal enzootics]. *Ig Mod*. 1969;62(3):158–167.

76. Drosophila 12 Genomes Consortium1, Clark AG, Eisen MB, et al. Evolution of genes and genomes on the Drosophila phylogeny. *Nature*. 2007;450(7167):203–218.

77. Wang F, Christiansen T, Orwant J. *Programming Perl*. China: O'Reilly; 2000.

78. Wang F, Hu S, Liu W, et al. Deep–sequencing analysis of the mouse transcriptome response to infection with Brucella eliteness strains of differing virulence. *PloS One*. 2011;6(12):e28485.