ECRONICON OPEN ACCESS

EC CLINICAL AND MEDICAL CASE REPORTS
Research Article

Inconsistent Phylogenetic Trees from Nucleotide or Amino Acid Sequences from Mammalian Mitochondrial Genomes

Behnjemyn JK Loh[1,2], Katheresan S Sooriya Kannan[1,2], Tanmay Patil[1,2], Rohit Vij[1,2] and Maurice HT Ling[1,2,3,4]*

[1]School of Life Sciences, Management Development Institute of Singapore, Singapore

[2]School of Applied Sciences, Northumbria University, United Kingdom

[3]School of Data Sciences, Perdana University, Malaysia

[4]HOHY PTE LTD, Singapore

***Corresponding Author:** Maurice HT Ling, School of Life Sciences, Management Development Institute of Singapore, Singapore.

Received: June 06, 2022; **Published:** June 27, 2022

Abstract

Phylogenetic trees using orthologs are commonly used to analyze such evolutionary histories of organisms. A recent study suggests that phylogenetic analysis from single orthologous genes or multiple single orthologous genes are not likely to reflect actual evolutionary history and core genome is required. However, it is not clear whether this finding can be extrapolated to orthologous peptide sequences. In this study, we compare the generated phylogenetic trees constructed using nucleotide and peptide sequences from mitochondrial genomes of fourteen mammals. Our results confirmed that different orthologous nucleotide sequences may result in different phylogenetic trees with 20.5% of pairwise comparisons being significantly different (p-value < 0.05). This result is extrapolatable to orthologous peptide sequences with 33.3% of the pairwise comparisons being significantly different. In addition, the phylogenetic tree constructed using core genome is significantly different (paired t-test p-value = 5.52E-7) from the phylogenetic tree constructed using core proteome.

Keywords: *Phylogenetic Tree; Orthologs; Core Genome; Core Proteome*

Introduction

Darwinian Theory of Evolution [1] has been shown to be applicable from multicellular organism [2] to proteins [3] to bacteria and viruses [4], which has been integrated with George Mendel's work on genetics into Neo-Darwinism [5]. If two more recent organisms share a common ancestor, then there has to be one common ancestor at the root of all organisms today, which is known as the Last Common Universal Ancestor [6] - a mesophilic DNA and protein-based entity [7].

Inter-species and intra-species evolutionary relationships are commonly deduced using phylogenetic analyses [8], using nucleotide and amino acid sequences of a single protein to the full genome sequence of different organisms [9]. Phylogeny is the key to comprehend the evolution of genes, genomes, and species [10]. However, the question as to whether to use one or more orthologous sequences to even core genome [11-13] or core proteome [14] sequences remained unanswered. A recent study by Wang., *et al*. [15] on thirteen mitochondrial genomes suggests that phylogenetic trees constructed using sets of single orthologous nucleotide sequences can vary substantially;

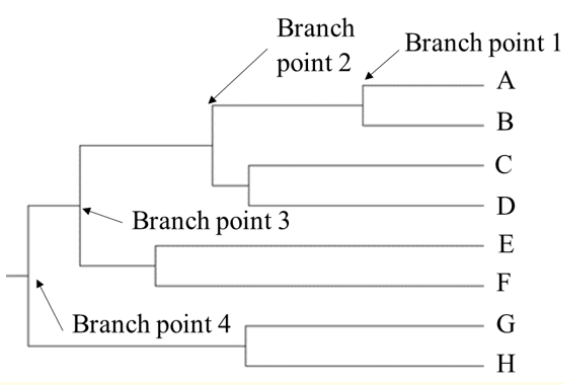
thus, proposing the use of core genomic sequences for phylogenetic tree construction. However, it is not clear whether the same finding holds for amino acid sequences.

In this study, we repeat the work of Wang., *et al*. [15] using another set of mitochondria genomes and extend it to its corresponding amino acid sequences. The origins of the mitochondria can be traced back to the Last Eukaryote Common Ancestor [16]. Our results suggest that different orthologous peptide sequences may result in different phylogenetic trees and the phylogenetic tree from core genome may differ from that of core proteome.

Materials and Methods

Mitochondrial genome sequence mining: Complete coding nucleotide and protein sequences of mitochondrial genomes of fourteen mammals were acquired from the GenBank. The fourteen mammals chosen for the analysis are namely (accession numbers are in brackets): *Bos taurus* (NC_006853.1), *Canis lupus familiaris* (NC_002008.4), *Felis catus* (NC_001700.1), *Gorilla gorilla* (NC_001645.1), *Homo sapiens* (NC_012920.1), *Lepus oiostolus* (NC_050983.1), *Macropus giganteus* (NC_027424.1), *Moschus moschiferus* (NC_013753.1), *Mus musculus* (NC_005089.1), *Ochotona thibetana sacraria* (MH345726.1), *Ornithorhynchus anatinus* (NC_000891.1), *Rattus norvegicus* (NC_001665.2), *Tachyglossus aculeatus* (NC_003321.1) and *Tarsipes rostratus* (NC_006518.1).

Phylogenetic tree construction, scoring, and analysis: Multiple sequence alignments were constructed using ClustalW [17] in MEGA X [18]. Default settings were used except those recommended by Hall [19] for amino acid sequences - (a) multiple alignment gap opening penalty was set to 3 and (b) multiple alignment gap extension penalty was set to 1.8. Phylogenetic trees were constructed using the Maximum Likelihood method [20] with default settings. Each phylogenetic tree was scored based on the method described by Wang., *et al*. [15] and is illustrated in figure 1. The average scores of the phylogenetic tree branch between two trees were analyzed using paired t-test.



	A	B	C	D	E	F	G	H
A								
B	1							
C	2	2						
D	2	2	1					
E	3	3	3	3				
F	3	3	3	3	1			
G	4	4	4	4	3	3		
H	4	4	4	4	3	3	1	

Figure 1: *An example of the phylogenetic tree branch scoring system. Branch point 1 is the common ancestor for A and B. Branch point 2 is the common ancestor for A, B, C and D. Branch point 3 is the common ancestor for A, B, C, D, E and F. Branch point 4 is the common ancestor for A, B, C, D, E, F, G and H. Table of ancestry showing the scoring of phylogenetic branch tree points of organisms A, B, C, D, E, F, G and H. We can conclude from the diagram that organisms A and B, C and D, E and F, as well as organisms G and H were the most closely related based on the branch point score of 1. Similarly, a branch point score of 2 revealed that organisms A and G, A and H, B and G, B and H, C and G, C and H, D and G, as well as organisms D and H were the least closely related.*

Results and Discussion

Similar genetic framework in the mitochondrial genomes: An identical gene order was observed in the mitochondrial genomes of the fourteen mammals which consisted thirteen protein-coding genes and aligned in the following chronology: (a) NADH dehydrogenase subunit 1 (ND1; 955 to 957 base pairs), (b) NADH dehydrogenase subunit 2 (ND2; 1038 to 1047 base pairs), (c) cytochrome c oxidase subunit 1 (COX1; 1542 to 1551 base pairs), (d) cytochrome c subunit 2 (COX2; 682 to 693 base pairs), (e) ATP synthase Fo subunit 8 (ATP8; 201 to 210 base pairs), (f) ATP synthase Fo subunit 6 (ATP6; 679 to 681 base pairs), (g) cytochrome c oxidase subunit 3 (COX3; 781 to 785 base pairs), (h) NADH dehydrogenase subunit 3 (ND3; 346 to 351 base pairs), (i) NADH dehydrogenase subunit 4L (ND4L; 297 base pairs), (j) NADH dehydrogenase subunit 4 (ND4; 1378 base pairs), (k) NADH dehydrogenase subunit 5 (ND5; 1809 to 1,830 base pairs), (l) NADH dehydrogenase subunit 6 (ND6; 501 to 528 base pairs), and (m) cytochrome b (CYTB; 1135 to 1,146 base pairs). This suggests that the length of each gene was similar across the fourteen mammals.

Confirmation of Wang., *et al.*'s key finding that different orthologous nucleotide sequences may result in different phylogenetic trees: A recent study by Wang., *et al*. [15] on thirteen mitochondrial genomes from the same order suggests that phylogenetic trees constructed using sets of single orthologous nucleotide sequences can vary substantially. Our results show there are significant differences (paired t-test p-value of 0.013) in phylogenetic trees for nucleotide sequences of ND1 and CYTB (Figure 2). Of the 156 pairwise comparisons of phylogenetic trees (Figure 3), 32 (20.5%) are significantly different (p-value < 0.05). This confirmed the findings of by Wang., *et al*. [15].

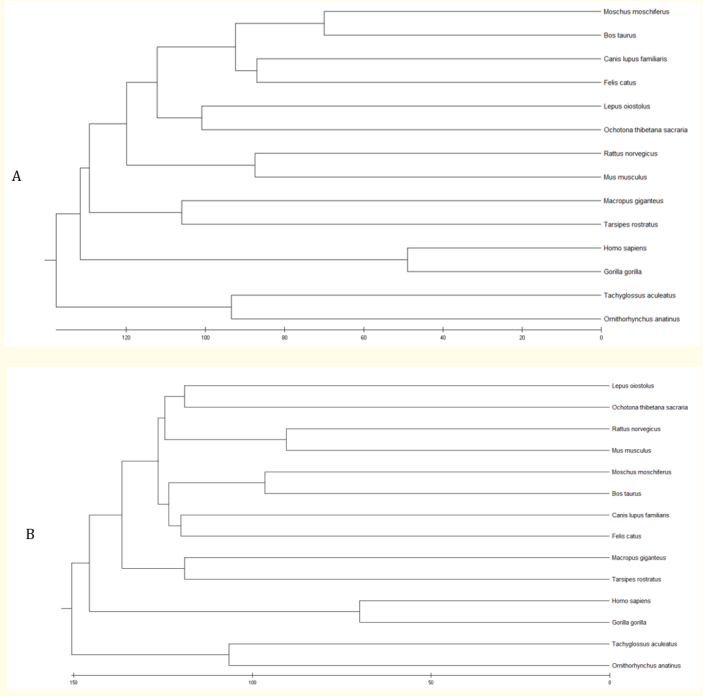


Figure 2: *Different phylogenetic trees using different orthologous nucleotide sequences. (A) Phylogenetic tree based on the nucleotide sequence of ND1. (B) Phylogenetic tree based on the nucleotide sequence of CYTB.*

	ND1	ND2	COX1	COX2	ATP8	ATP6	COX3	ND3	ND4L	ND4	ND5	ND6	CYTB
ND1		0.875	0.383	**0.021**	0.195	0.511	**0.013**	0.613	0.604	0.823	0.875	0.420	**0.013**
ND2	0.875		0.398	0.065	0.291	0.370	0.116	0.118	0.224	0.511	1.000	0.417	0.116
COX1	0.383	0.398		**0.010**	0.148	0.728	**0.021**	0.645	0.717	0.525	0.398	0.228	**0.021**
COX2	**0.021**	0.065	**0.010**		0.171	**0.035**	0.610	**0.011**	**0.014**	**0.017**	0.065	0.126	0.610
ATP8	0.195	0.291	0.148	0.171		0.129	0.259	0.094	0.102	0.182	0.291	0.640	0.259
ATP6	0.511	0.370	0.728	**0.035**	0.129		**0.036**	0.886	1.000	0.714	0.370	0.140	**0.036**
COX3	**0.013**	0.116	**0.021**	0.610	0.259	**0.036**		**0.047**	**0.047**	0.070	0.116	0.117	1.000
ND3	0.613	0.118	0.645	**0.011**	0.094	0.886	**0.047**		0.765	0.657	0.118	0.147	**0.047**
ND4L	0.604	0.224	0.717	**0.014**	0.102	1.000	**0.047**	0.765		0.610	0.224	0.140	**0.047**
ND4	0.823	0.511	0.525	**0.017**	0.182	0.714	0.070	0.657	0.610		0.511	0.357	0.070
ND5	0.875	1.000	0.398	0.065	0.291	0.370	0.116	0.118	0.224	0.511		0.417	0.116
ND6	0.420	0.417	0.228	0.126	0.640	0.140	0.117	0.147	0.140	0.357	0.417		0.117
CYTB	**0.013**	0.116	**0.021**	0.610	0.259	**0.036**	1.000	**0.047**	**0.047**	0.070	0.116	0.117	

Figure 3: *P-values from paired t-test of average scores between phylogenetic trees of orthologous nucleotide sequences.*

Different orthologous peptide sequences may result in different phylogenetic trees: Although both nucleotide and peptide sequences can be used for phylogenetic tree construction [21], peptide sequences are generally considered to be more stable than nucleotide sequences [22]. Hence, nucleotide sequences are considered more useful for comparing from closely related organisms while peptide sequences are considered appropriate for distantly related organisms [23]. Therefore, it is plausible to consider whether orthologous peptide sequences will yield the same findings after confirming Wang., *et al*. [15]'s findings. Our results show there are significant differences (paired t-test p-value of 0.012) in phylogenetic trees for peptide sequences of ND1 and COX2 (Figure 4). Of the 156 pairwise comparisons of phylogenetic trees (Figure 5), 52 (33.3%) are significantly different (p-value < 0.05). This suggests that sets of sequences from either single orthologous genes or single orthologous peptides may result in different phylogenetic trees; thereby, cautioning the use of sequence sets from single orthologous genes or peptides for phylogenetic tree construction for the purpose of studying evolution of organisms.

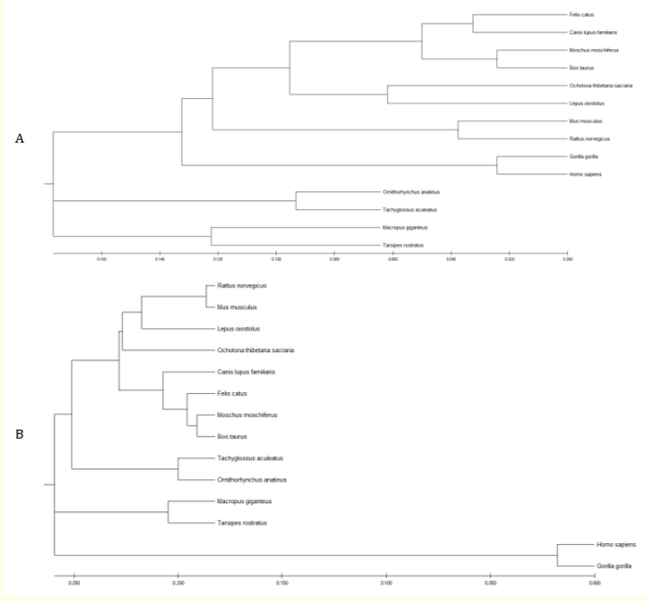


Figure 4: *Different phylogenetic trees using different orthologous peptide sequences. (A) Phylogenetic tree based on the amino acid sequence of ND1. (B) Phylogenetic tree based on the amino acid sequence of COX2.*

	ND1	ND2	COX1	COX2	ATP8	ATP6	COX3	ND3	ND4L	ND4	ND5	ND6	CYTB
ND1		0.634	**0.012**	0.283	**0.000**	0.877	1.000	**0.014**	1.000	**0.014**	1.000	0.192	0.118
ND2	0.634		0.077	0.482	**0.002**	0.822	0.634	**0.007**	0.664	**0.005**	0.181	**0.039**	0.372
COX1	**0.012**	0.077		0.489	0.348	**0.033**	**0.012**	0.793	0.113	0.819	**0.047**	**0.020**	0.262
COX2	0.283	0.482	0.489		0.077	0.420	0.283	0.602	0.385	0.602	0.373	0.093	0.938
ATP8	**0.000**	**0.002**	0.348	0.077		**0.001**	**0.000**	0.203	**0.001**	0.198	**0.001**	**0.000**	**0.039**
ATP6	0.877	0.822	**0.033**	0.420	**0.001**		0.877	0.097	0.939	0.097	0.912	0.235	0.327
COX3	1.000	0.634	**0.012**	0.283	**0.000**	0.877		**0.014**	1.000	**0.014**	1.000	0.192	0.118
ND3	**0.014**	**0.007**	0.793	0.602	0.203	0.097	**0.014**		**0.027**	1.000	**0.001**	**0.004**	0.420
ND4L	1.000	0.664	0.113	0.385	**0.001**	0.939	1.000	**0.027**		**0.027**	1.000	0.096	0.367
ND4	**0.014**	**0.005**	0.819	0.602	0.198	0.097	**0.014**	1.000	**0.027**		**0.001**	**0.004**	0.420
ND5	1.000	0.181	**0.047**	0.373	**0.001**	0.912	1.000	**0.001**	1.000	**0.001**		0.070	0.262
ND6	0.192	**0.039**	**0.020**	0.093	**0.000**	0.235	0.192	**0.004**	0.096	**0.004**	0.070		0.066
CYTB	0.118	0.372	0.262	0.938	**0.039**	0.327	0.118	0.420	0.367	0.420	0.262	0.066	

Figure 5: *P-values from paired t-test of average scores between phylogenetic trees of orthologous peptide sequences.*

Phylogenetic tree from genome sequence is different from phylogenetic tree from proteome sequence: Given that different orthologous nucleotide sequences and different orthologous peptide sequences may result in distinctly different phylogenetic trees, it may be plausible to consider the use of core genome or core proteome for phylogenetic tree construction. The use of core genome [15,24-27] or core proteome [28-30] in phylogenetic tree construction to study evolution of organisms has been advocated. Mitochondrial proteomes of each organism were constructed by gene ordered concatenation of peptide sequences. Our results show there are significant differences (paired t-test p-value of 5.52E-7) in between phylogenetic trees from mitochondrial genomes and mitochondrial proteomes (Figure 6). This suggests that phylogenetic tree from core genomes may differ from that of core proteomes.

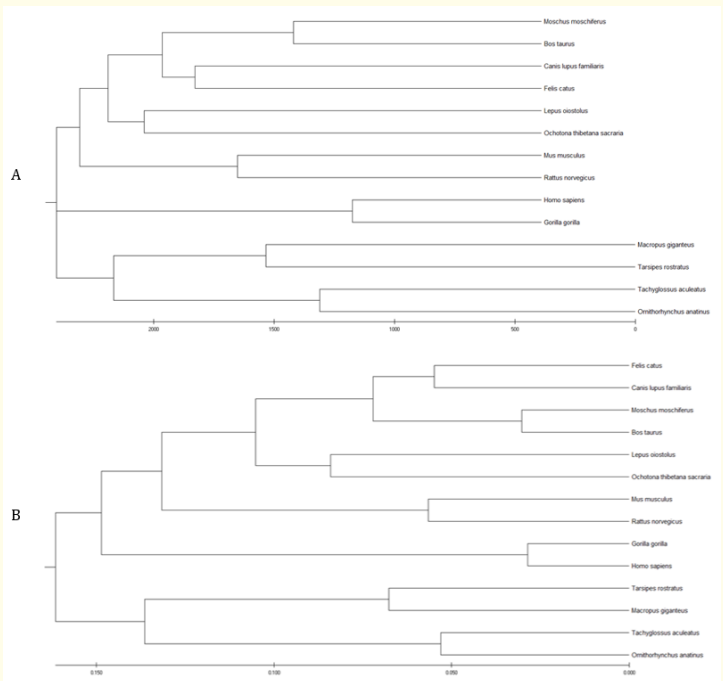


Figure 6: *Phylogenetic tree from genome vs phylogenetic tree from proteome. (A) Phylogenetic tree based on the complete mitochondrial genome. (B) Phylogenetic tree based on the complete mitochondrial proteome.*

Conclusion

Using nucleotide and peptide sequences from 14 mammalian mitochondria, we show that (a) different orthologous nucleotide sequences may result in different phylogenetic trees, (b) different orthologous peptide sequences may result in different phylogenetic trees, and (c) the phylogenetic tree constructed using core genome may be different from the phylogenetic tree constructed using core proteome.

Supplementary Materials

Supplementary materials for this study can be downloaded at https://bit.ly/DNA_AA_Phylogeny.

Conflict of Interest

The authors declare no conflict of interest.

Bibliography

1. Darwin C. On the Origin of Species (Harvard University Press) (1964).
2. Soltis PS., *et al*. "Polyploidy and Genome Evolution in Plants". *Current Opinion in Genetics and Development* 35 (2015): 119-125.
3. Phillips JC. "Reply to Koonin et al.: Evolution of Proteins is Darwinian". *Proceedings of the National Academy of Sciences of the United States of America* 117.33 (2020): 19641-19642.
4. Koonin EV and Wolf YI. "Evolution of Microbes and Viruses: A Paradigm Shift in Evolutionary Biology?" *Frontiers in Cellular and Infection Microbiology* 2 (2012): 119.
5. Hancock ZB., *et al*. "Neo-Darwinism Still Haunts Evolutionary Theory: A Modern Perspective on Charlesworth, Lande, and Slatkin (1982)". *Evolution; International Journal of Organic Evolution* 75.6 (1982): 1244-1255.
6. Weiss MC., *et al*. "The Physiology and Habitat of the Last Universal Common Ancestor". *Nature Microbiology* 1.9 (2016): 16116.
7. Penny D and Poole A. "The Nature of the Last Universal Common Ancestor". *Current Opinion in Genetics and Development* 9.6 (1999): 672-677.
8. Young AD and Gillung JP. "Phylogenomics - Principles, Opportunities and Pitfalls of Big-Data Phylogenetics". *Systematic Entomology* 45.2 (2020): 225-247.
9. Gabaldón T. "Evolution of Proteins and Proteomes: A Phylogenetics Approach". *Evolutionary Bioinformatics Online* 1 (2007): 51-61.
10. Higgins D. "Alignment Problem". Encyclopedia of Genetics, editions Brenner S, Miller J (Academic Press, New York) (2001): 9-35.
11. Tan XT., *et al*. "Core Pseudomonas Genome from 10 Pseudomonas Species". *MOJ Proteomics and Bioinformatics* 9.3 (2020): 68-71.
12. Kuan ZJ and Ling MH. "Core Genome of Poales, An Economically Important Order of Monocotyledons". *EC Agriculture* 7.2 (2021): 24-29.
13. Aggelen H van., *et al*. "A Core Genome Approach that Enables Prospective and Dynamic Monitoring of Infectious Outbreaks". *Scientific Reports* 9.1 (2019): 7808.

14. Yang L., *et al*. "Systems Biology Definition of the Core Proteome of Metabolism and Expression is Consistent with High-Throughput Data". *Proceedings of the National Academy of Sciences of the United States of America* 112.34 (2015): 10810-10815.

15. Wang VC., *et al*. "A Case Study Using Mitochondrial Genomes of the Order Diprotodontia (Australasian Marsupials) Suggests that Single Ortholog is Not Sufficient for Phylogeny". *EC Clinical and Medical Case Reports* 3.9 (2020): 93-114.

16. O'Malley MA., *et al*. "Concepts of the Last Eukaryotic Common Ancestor". *Nature Ecology and Evolution* 3.3 (2019): 338-344.

17. Thompson JD., *et al*. "CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice". *Nucleic Acids Research* 22.22 (1994): 4673-4680.

18. Kumar S., *et al*. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms". *Molecular Biology and Evolution* 35.6 (2018): 1547-1549.

19. Hall BG. "Building Phylogenetic Trees from Molecular Data with MEGA". *Molecular Biology and Evolution* 30.5 (2013): 1229-1235.

20. Truszkowski J and Goldman N. "Maximum Likelihood Phylogenetic Inference is Consistent on Multiple Sequence Alignments, with or without Gaps". *Systematic Biology* 65.2 (2016): 328-333.

21. Foth BJ. "Phylogenetic Analysis to Uncover Organellar Origins of Nuclear-Encoded Genes". *Methods in Molecular Biology* 390 (2007): 467-488.

22. Zhang D., *et al*. "Using Phylogenetic Analysis to Investigate Eukaryotic Gene Origin". *Journal of Visualized Experiments* 138 (2018): 56684.

23. Baldauf SL. "Phylogeny for the Faint of Heart: A Tutorial". *Trends in Genetics* 19.6 (2003): 345-351.

24. Stott CM and Bobay L-M. "Impact of Homologous Recombination on Core Genome Phylogenies". *BMC Genomics* 21.1 (2020): 829.

25. Shakya M., *et al*. "Standardized Phylogenetic and Molecular Evolutionary Analysis Applied to Species Across the Microbial Tree of Life". *Scientific Reports* 10.1 (2020): 1723.

26. Sakoparnig T., *et al*. "Whole Genome Phylogenies Reflect the Distributions of Recombination Rates for Many Bacterial Species". *eLife* 10 (2021): e65366.

27. Abdel-Glil MY., *et al*. "Phylogenetic Relatedness and Genome Structure of Yersinia ruckeri Revealed by Whole Genome Sequencing and a Comparative Analysis". *Frontiers in Microbiology* 12 (2021): 782415.

28. Artuso I., *et al*. "Phylogenomic Analysis and Characterization of Carbon Monoxide Utilization Genes in the Family Phyllobacteriaceae with Reclassification of Aminobacter carboxidus (Meyer et al. 1993, Hördt et al. 2020) as Aminobacter lissarensis comb. nov. (McDonald et al. 2005)". *Systematic and Applied Microbiology* 44.3 (2021): 126199.

29. Callister SJ., *et al*. "Comparative Bacterial Proteomics: Analysis of the Core Genome Concept". *PLoS ONE* 3.2 (2008): e1542.

30. Crapitto AJ., *et al*. "A Consensus View of the Proteome of the Last Universal Common Ancestor". *Ecology and Evolution* 12.6 (2022): e8930.

Volume 5 Issue 7 July 2022