Research Article　　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Estimation of Genome Evolution Time by Mutation Rate

Ahmed M El-Shehawi[1,2]* and Mona M Elseehy[2]

[1]Department of Biotechnology, Faculty of Science, Taif University, Taif, Saudi Arabia
[2]Department of Genetics, Faculty of Agriculture, University of Alexandria, Alexandria, Egypt

Abstract

Mutations, transposable elements, and recombination are the main mechanisms for genome size evolution. The quantitative impact of mutations, excluding polyploidy, on genome size is well studied in some genomes while the impact of other factors has not been investigated. Mutation rate was used to estimate the evolution time of origin genome to form a higher size genome and test if the estimated age of earth fits for these evolution events. Results indicated that the evolution time of the smallest detected genome through mutation rate to the largest detected genome is much higher than the estimated age of earth. The cumulative evolution time of the studied origin genomes was estimated at 5300 folds of earth's age and the average evolution time is 2.7×10^{12} years per genome. The relationship among genome size, mutation rate, and evolution time indicated that evolution time is positively correlated with genome size suggesting that larger genomes take longer time to evolve in size. Estimation of evolution time would lead to establishment of genome evolution timeline to replace or support the fossil evolution timeline.

Keywords: Mutation rate; Evolution time; Genome size; Cumulative evolution time

Introduction

Genome is the unique string of nucleotide sequence tailored and organized in a very unique architecture to reproduce unique specific features of a species. It is determined by the amount of DNA in the cell "C-value" distributed on a certain number of chromosomes. When mutated to a critical level, this string will not be able to define the species' unique characteristics. The detailed unique organization of a genome is very significant because two distinct species could have the same genome size. For example, *Homo sapiens* (human) [1] and the grass *Festuca tatrae* [2] have the same genome size (3.5 pg) and they have completely different structural features and development. Charles Darwin introduced his evolution theory of common ancestor [3] based on series of continuous phenotypic and structural similarities in the absence of principles of genetics and genomics, yet he was not able to explain how the original cell had formed. In the view of evolution theory and the uniqueness of genomes and species, the common ancestor genome had to contain a specific amount of DNA (genome). Based on our current knowledge in genomics, this common ancestor genome (original genome) had to gain more DNA and arrange it in a new unique format to generate new genome and then new species with distinct features. This way the original genome can change and evolve to other genomes to define new species.

Mechanisms of genome evolution

Various mechanisms contribute to genome evolution (change in genome size) including recombination [4,5], transposition [6,7], and mutations [8,9]. Recombination affects the genome architecture and evolutionary rate. Its effect on genome evolution is not well understood because its impact on genomes requires whole genome sequencing and global recombination rate [4]. Previous studies indicated that high recombination rate is negatively correlated with genome size [10], positively correlated with Long Terminal Repeats (LTR) content [11], and positively correlated with GC% content and CpG density [5].

Transposable elements (TEs) are highly represented in nearly all genomes. They make up about 45% of the human genome and most of plant genomes [6]. For example, about 85% of maize genome is TEs [12]. Their integration in host genomes leads to various types of rearrangements including insertions, deletions, duplications, inversions, and loss or change in gene expression [6].

Polyploidy is one type of mutations that involves the most drastic change in genome size, yet it is limited to the polyploid genomes. After polyplodization, rapid genome rearrangements and gene silencing occur [13] indicating that polyploidy is not a dead end in genome evolution. Several studies also reported that polyplodization occurred early during plant evolution [14].

The distinct role of recombination, TEs, and mutation is not clear because their impact on genome size evolution is extensively intermingled. It is believed that TEs increase genome reconstruction in polyploids [15] and they are suspected to be involved in evolution of gene silencing mechanisms including methylation and heterochromatin formation in eukaryotes [16]. Also, recombination is positively correlated with LTR content because removal of LTRs involves recombination processes [11]. These overlapping roles make it extremely difficult to estimate their distinct impact on genome size evolution in a quantitative way. On the other hand, there have been clear estimates of mutation rates, excluding polyploidy, in many different genomes [9,17-22].

Mutation rate (μ_g)

Two main approaches have been used to estimate mutation rates in living organisms quantitatively. The first is based on using function analysis (FA) of one locus [17,18] and the second is the recently introduced whole genome sequencing (WGS) for accurate estimation of mutation rate after high number of generations [9,19-22]. Many types of mutations contribute to genome evolution including insertion, deletion (indel), and base substitution. Various studies give estimation of total mutation rate because sometimes it is not practically possible to give separate estimates for different types of mutations, but generally, excluding polyploids, the change in genome size via mutations comes from the net difference between insertion and deletion mutations [9,17-22].

The estimated value of mutation rate differs depending on the estimation approach and the organism (Table 1). For example, the average mutation rate of *E. coli* determined by FA was 0.0025 base per genome per generation (bgg) [17,18] (Table 1), whereas it was 8.9×10^{-11} base per site per generation (bsg) (0.00045 bgg) when it was estimated using WGS [20]. This is much lower than the previous reported rate (0.0025 bgg) or the average microbial mutation rate (0.0034 bgg) [18]. Generally, the WGS approach gave more accurate but lower estimates of spontaneous mutations than the FA (Table 1). Base substitution represents the major percentage of mutations, whereas indel mutations are rare and sometimes are not estimated distinctly. This was reported in yeast [9], *C. reinhardtii* [21], and *D. melanogaster* [22-24] raising the question about the effectiveness of mutation to change genomes size enough to the level of genome evolution. In this study, we assumed that total mutation rate is insertion mutations for simplified calculations.

Previous studies have focused on getting estimations of mutation rates (Table 1), whereas the impact of mutations on genome size has not been investigated quantitatively. In this study, mutation rate, excluding polyploidy, of different genomes was used to estimate the evolution time of origin diploid genomes to higher size target diploid genomes and test the possibility of genome evolution via mutation rate during the estimated age of earth. The impact of recombination, transposition, and polyploidy mutations were excluded because their accurate quantitative rates have been not determined.

Materials and Methods

Estimation of genome evolution time

Estimated mutation rates in Table 1 were used to calculate evolution time (Et) from nine well studied origin genomes to target genomes of higher size. The smallest detected bacterial genome of *Buchnera sp.* (25) was used to represent the anonymous controversial common ancestral genome. Because its mutation rate has not been detected the average microbial mutation rate (0.0034 bgg, Table 1) was used. The human (*H. sapiens*) and the lung fish *Protopterus aethiopicus* (*P. aethiopicus*) genomes were used as target genomes because human is the most recent organism on earth [25,26] and *P. aethiopicus* has the largest detected genome (130000 Mb). When there is more than one estimate for mutation rate for an origin genome (Table 1), the highest rate was used in calculations of evolution time. Mutation rate (bsg, μ_s) was multiplied by genome size in base pairs (Gb) to give mutation rate per genome per generation (bgg, μ_g). This was multiplied by the number of generation per year to give mutation rate per year (bgy, μ_y). To calculate the Et of an origin genome to a target genome the genome difference in bp was divided by μ_y. The estimated age of earth is 4.5 billion years [27] and the estimated age of life on earth is 4 billion years [28]. The calculated Et was divided by the age of earth (A) to give the Et in folds of the earth's age (NA), Et/A. Calculation details are summarized in Supplementary Material S1.

Cumulative evolution time

The cumulative genome evolution time for the nine origin genomes was calculated to estimate the total time needed for the smallest genome (*Buchnera sp.*) to reach the largest genome of *P. aethiopicus* passing by the other 8 genomes. The calculation details of evolution time of origin genomes to target genomes are summarized in Supplementary Material S1. The cumulative evolution time was divided by the number of genomes [9] to give the average evolution time per genome. This was used to predict the total evolution time for the current characterized or predicted number of genomes on earth.

Organism, locus	EM[#]	G_b	μ_b	μ_g	Reference
E. coli, lacI	FA	4.6×10^6	6.93×10^{-10}	0.0033	[17]
E. coli, lacI	FA	4.6×10^6	4.08×10^{-10}	0.0019	[17]
E. coli, hisGDCBHAFE	FA	4.6×10^6	5.06×10^{-10}	0.0024	[17]
E. coli, average	FA	$\mathbf{4.6 \times 10^6}$	-	**0.0025**	**[18]**
E. coli, B REL606	WGS	4.6×10^6	8.9×10^{-11}	0.00041	[20]
E. coli	WGS	4.6×10^6	-	0.001	[23]
E. coli, average	FA	4.6×10^6	-	0.0016	-
S. cerevisiae, URA3	FA	12.1×10^6	2.76×10^{-10}	0.00381	[17]
S. cerevisiae, CAN1	FA	12.1×10^6	1.73×10^{-10}	0.00238	[17]
S. cerevisiae, average	FA	12.1×10^6	2.2×10^{-10}	0.0027	**[18]**
S. cerevisiae	WGS	$\mathbf{12.1 \times 10^6}$	-	**0.004**	**[24]**
S. cerevisiae	WGS	12.1×10^6	1.67×10^{-10}	0.002	[9]
N. crassa, ad-3AB	FA	43×10^6	4.47×10^{-11}	0.00187	[17]
N. crassa, mtr	FA	43×10^6	9.96×10^{-11}	0.00417	[17]
N. crassa, average	FA	43×10^6	-	0.003	**[18]**
Microbes, average	FA	**-**	**-**	**0.0034**	**[18]**
C. reinhardi	WGS	$\mathbf{111 \times 10^6}$	$\mathbf{3.23 \times 10^{-10}}$	**0.036**	**[21]**
A. thaliana	WGS	$\mathbf{157 \times 10^6}$	$\mathbf{7 \times 10^{-9}}$	**1***	**[19]**
C. elegans	FA	$\mathbf{97 \times 10^6}$	-	**0.036***	**[18]**
D. melanogaster	FA	130×10^6	-	0.14*	**[18]**
D. melanogaster	WGS	$\mathbf{130 \times 10^9}$	$\mathbf{2.8 \times 10^{-9}}$	**0.57***	**[22]**
M. musculus	FA	$\mathbf{2.7 \times 10^9}$	-	**0.9***	**[18]**
H. sapiens	FA	$\mathbf{3.2 \times 10^9}$	-	**1.6***	**[18]**

[#]Estimation Method. FA: Functional Analysis, WGS: Whole Genome Sequencing.
*Mutation rate per sexual generation. Estimates in bold face font were used in the estimation of evolution time (Table 2).

Table 1: Mutation rate of some studied genomes.

Results

Genome evolution time

Estimated evolution time (Et) from the nine genomes (Table 1) is summarized in Table 2. Estimation of Et from the eight genomes to human genome or nine genomes to *P. aethiopicus* genome revealed some interesting features. Et from the 8 studied genomes to the human genome ranged from 3.58×10^7 to 3.6×10^8 years (0.008-0.08 NA), whereas it ranged from 9.9×10^8 to 2.4×10^{13} years (0.22-5300 NA) for the 9 genomes to the *P. aethiopicus* genome (Table 2). The smallest detected genome *Buchnera sp* would take 3.58×10^7 years (0.008 NA) to reach human genome or 1.45×10^9 years (0.32 NA) to reach *P. aethiopicus* genome. The Et differed among the 9 origin genomes used in this study depending on the mutation rate of the origin genome and the difference (bp) between the origin and the target genome. For example, human genome would take 2.4×10^{13} years (5.3×10^3 NA) to reach *P. aethiopicus* genome (Table 2). Also, in one evolution leap, *S. cerevisiae* genome would take as twice as the earth's age and *C. elegans* genome would take as 9 folds as the earth's age to reach *P. aethiopicus* genome (Table 2). Total evolution time for the 8 origin genomes to reach human genome is 1.88×10^9 years (0.42 NA) and for the 9 genomes to reach *P. aethiopicus* genome is 2.4×10^{13} years (5300 NA) (Table 2).

Cumulative evolution time

To estimate the average genome evolution time of the 8 genomes used in this study based on their mutation rates, the cumulative Et was calculated. This gives indication about the time during which the smallest detected genome took to reach the largest detected genome passing by other origin genomes. This was done by estimation of Et of the smallest detected genome to the next genome in size then estimation of the second genome to the third and so on. These independent Ets of origin

| Origin genome | G_m | μ_g | μ_y | Target genome | | | |
| | | | | Homo sapiens | | P. aethiopicus | |
				Et	NA	Et	NA
Buchnera sp.	0.449	0.0034	89.35	3.58×10^7	0.008	1.45×10^9	0.32
E. coli	4.6	0.0025	65.7	48.64×10^6	0.01	1.98×10^9	0.44
S. cerevisiae	12.1	0.004	14.6	2.18×10^8	0.048	8.9×10^9	1.98
C. elegans	97	0.036	3.2	9.7×10^8	0.22	4×10^{10}	8.9
C. reinhardtii	111	0.036	131	2.36×10^7	0.005	9.9×10^8	0.22
D. melanogaster	130	0.57	24	1.3×10^8	0.03	5.4×10^9	1.2
A. thaliana	157	1*	8.5	3.6×10^8	0.08	1.5×10^{10}	3.3
M. musculus	2700	0.9*	5.4	9.3×10^7	0.02	2.4×10^{10}	5.3
H. sapiens	3200	1.6*	5.3×10^{-3}	-	-	2.4×10^{13}	5.3×10^3
Total	-	-	-	1.88×10^9	0.42	2.4×10^{13}	5.3×10^3

*Base per genome per sexual generation.

Table 2: Estimated evolution time of origin genomes to target genomes.

genomes were added to give the cumulative Et for the 9 genomes. The cumulative Et for the nine origin genomes used in this study to develop *P. aethiopicus* genome was 2.4×10^{13} years (5.3×10^3 NA). Based on this estimation, the average Et of the 9 genomes can be estimated as 2.7×10^{13} years per genome (5.9×10^2 NA) (Table 3). Although the average Et per genome comes from a quiet few number of genomes compared to the number of characterized genomes (1.2×10^6)[30] it could give an approximate estimation because the studied genomes represent a wide range of genome size. They include the smallest detected genome of *Buchnera sp.* (0.449 Mb), the largest detected genome of *P. aethiopicus* (130000 Mb), and other small and average size genomes (Table 2). Therefore, it could represent the average evolution time based on mutation rate without recombination, TEs, and polyploidy.

Mutation rate, genome size, and evolution time

Using genome size (G_m), mutation rate per year (μ_y), and estimated evolution time (Et) of the 9 genomes (Table 2), a relationship was drawn to investigate how these three genomic factors are interrelated. Mutation rate per generation (bgg) of microorganisms is lower than that of higher organisms [18]. On the other hand, mutation rate per year (μ_y) is higher in microorganisms because of the high number of generation per year (Table 2). Data obtained showed that Et was found to be positively correlated with genome size, whereas μ_y was found to be negatively correlated with G_m (Figure 1). This introduces a very interesting conclusion that larger genomes would take longer time to evolve to higher size genomes because of their low mutation rate per year.

Discussion

The WGS approach gives more accurate estimates of spontaneous mutations but generally it showed lower estimates of mutation rate (Table 1). This could be due to the correction of mutations by repair systems during high number of generations or successive mutations in the same site [10,29]. The high number of generation before calculation of mutation rate in the WGS approach gives the repair system to correct mutations so that they are not detectable. Also, in some origin genomes of this study the deletion mutation rate was higher than the insertion mutation rate. For example, in *C. reinhardtii*, spontaneous mutation rate was estimated using WGS after mutation accumulation experiment for 350 generations. Only 7% of detected mutations were insertions, whereas 29% were deletions [21]. These observations suggest that the estimated impact of mutations on genome evolution might be slower than expected. This is supported by the long estimations of Et in Table 2.

Based on the principle of evolution from common ancestor, all living organisms should have been evolved from one common ancestor genome in a number of evolution steps equal to the number of living species. Currently, there are 1.2×10^6 characterized and 8.7×10^6 predicted species [30]. Therefore, the smallest genome had to take 1.2 million evolution leaps to form the largest known genome among the characterized species or 8.7 million evolution leaps to develop the expected number of genomes. Moreover, each evolution leap includes unlimited processes of genome evolution mechanisms (mutations, transposition, recombination) to make up the new genome. When the average Et and number of characterized or predicted genomes are put together in this context some estimations can be laid down. The average Et of genome is estimated as 2.7×10^{12} years per genome (Table 3). Therefore, the cumulative Et of characterized species would be 3.2×10^{18} years (7.2×10^8 NA). Similarly, the cumulative Et of predicted species would be 2.3×10^{19} years (5.2×10^9 NA) (Table 3). These estimates give indication that it is impractical for the characterized or predicted number of genomes to have been evolved from common ancestral genome through mutations during the estimated age of earth [27].

In addition, the devolution is a continuous process working against the evolution mechanisms to repair and tune the genome and keep it

Origin genome	Et
Buchnera sp.	4.65×10^4
E. coli	1.1×10^5
S. cerevisiae	5.8×10^6
C. elegans	4.4×10^6
C. reinhardi	1.45×10^5
D. melanogaster	1.13×10^6
A. thaliana	3×10^8
M. musculus	9.3×10^7
H. sapiens	2.4×10^{13}
P. aethiopicus	-
Total	2.4×10^{13}
NA	5.3×10^3
Average of Et	2.7×10^{12}
Average of NA	5.9×10^2
Et for characterized species	3.2×10^{18}
NA for characterized species	7.2×10^8
Et for predicted species	2.3×10^{19}
NA for predicted species	5.2×10^9

Table 3: Cumulative Et of origin genomes used in this study. Value opposite to a genome indicates its Et to the following genome in the Table. For calculation details see Supplementary materials S1.
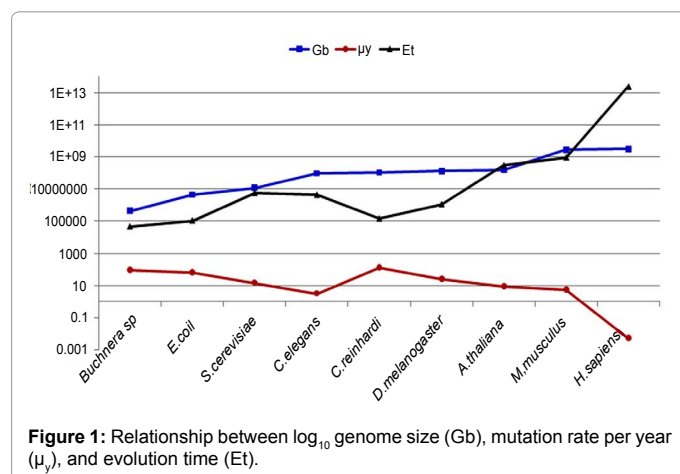

Figure 1: Relationship between $\log_{10}$ genome size (Gb), mutation rate per year (μ_y), and evolution time (Et).

functional. Recombination is corrected by the repair systems and it is involved in the efficient elimination of transposons after their integration in the genomes. Also, Long Terminal Repeats retrotransposons (LTR-RTs), one major type of TEs, can be eliminated from the genome via different mechanisms including deletions and recombinations [7]. It was suggested that genome size is maintained by retrotranspositon (addition of DNA) and elimination of TE sequences through deletions and recombination [7]. In addition, after polyploidization, genome rearrangement and gene silencing can lead to diploidization which can remove the polyploid feature of the genome. This phenomenon was found in maize, sorghum, polyploid species of sugarcane, and *Brassica sp.* [13]. This gives evidence that, the impact of genome evolution mechanisms is counteracted by devolution processes which slow down genome evolution process than have been estimated in this study.

This study represents the first report about the quantitative impact of mutation rate on genome size evolution. This study investigated the impact of mutations only on genome size evolution time while other genome evolution mechanisms (recombination and transposition) were excluded because their independent impact on genome size has not been determined quantitatively. It shows the need for more extensive studies to estimate the impact of recombination and transposition. Results also show that the estimated genome evolution time can be used in establishing genome evolution timeline based on changes in genome size which provide another novel quantitative tool for evolution of living species. This can be used in parallel or alternative to the fossil evolution timeline, the principle of Darwinian evolution. Genome evolution timeline is expected to be more accurate and reliable than fossil evolution timeline because the first is based on changes in genome size, common to all living species, not on gradual phenotypic and anatomical similarities.

Conflict of Interest

Authors have nothing to disclose.

References

1. Krishan A, Dandekar P, Nathan N, Hamelik R, Miller, et al. (2005) DNA index, genome size, and electronic nuclear volume of vertebrates from the Miami Metro Zoo. Cytometry 65: 26-34.

2. Smarda P, Bures P, Horova L, Foggi B, Rossi G (2008) Genome size and GC content evolution of Festuca: Ancestral expansion and subsequent reduction. Annals of Botany 101: 421-433.

3. Darwin C (1859) On the Origin of Species. London: John Murray.

4. Tiley GP, Burleigh G (2015) The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. BMC Evol Biol 15: 194.

5. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, et al. (2015) Comparative Recombination Rates in the Rat, Mouse and Human Genomes. Genome Res 14: 528-538.

6. Alzohairy AM, Gyulai G, Jansen RK, Bahieldin A (2013) Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid 69: 1-15.

7. El Baidouri M, Panaud O (2013) Comparative Genomic Paleontology across Plant Kingdom Reveals the Dynamics of TE-Driven Genome Evolution. Genome Biol Evol 5: 954-965.

8. Lynch M (2010) Evolution of the mutation rate. Trends Genet 6: 345-352.

9. Zhu YO, Siegal ML, Halld DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. Proc Natl Acad Sci USA 111: E2310-E2318.

10. Lynch M (2006) The origins of eukaryotic gene structure. Mol Biol Evol 23: 450-468.

11. Gaut BS, Wright SI, Rizzon C, Dvorak, Anderson LK, et al. (2007) Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet 8: 77-84.

12. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

13. Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. TREE 14: 348-352.

14. Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 16: 805-814.

15. Wendel JF, Small RL, Cronn RC, Brubaker CL (1999) Genes, jeans, and genomes: Reconstructing the history of cotton. In: Plant evolution in man-made habitats. Proceedings of the VIIth international symposium of the international organization of plant biosystematists (L.W.D. van Raamsdonk and J. C. M. den Nijs (eds.) Hugo de Vries Laboratory, Amsterdam, The Netherlands pp: 133-161.

16. McDonald JF (1998) Transposable elements, gene silencing, and macroevolution. Trends Ecol Evol 13: 94-95.

17. Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA 88: 7160-7164.

18. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148: 1667-1686.

19. Ossowski S, Schneeberger K, Lucas-Lledó JL, Warthmann N, Clark RM, et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327: 92-94.

20. Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, et al. (2011) Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With *Escherichia coli*. G3 1: 183-186.

21. Ness RW, Morgan AD, Colegrave N, Keightley PD (2012) Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*. Genetics 192: 1447-1454.

22. Keightley PD, Ness RW, Halligan DL, Haddrill PR (2014) Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. Genetics. 196: 313-320.

23. Lee H, Popodib E, Tanga H, Fosterb PL (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. PNAS 109: E2774-E2783.

24. Lynch L, Sung W, Morris K, Coffey N, Landry CR, et al. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. PNAS 105: 9272-9277.

25. Gil R, Sabater-Munoz B, Latorre A, Silva FJ, Moya A, et al. (2002) Extreme genome reduction in *Buchnera spp.*: toward the minimal genome needed for symbiotic life. Proc Natl Acad Sci USA 99: 4454-4458

26. Elhaik E, Tatarinova TV, Klyosov AA, Graur D (2014) The extremely ancient chromosome that isn't: a forensic bioinformatic investigation of Albert Perry''s X-degenerate portion of the Y chromosome. EJHG 22: 1111-1116.

27. Dalrymple GB (2001) The age of the Earth in the twentieth century: a problem (mostly) solved. Special Publications, Geological Society of London 190: 205-221.

28. Schopf JW, Kudryavtsev AB, Czaja AD, Tripathi AB (2007) Evidence of Archean life: Stromatolites and microfossils. Precambrian Research 158:141-155.

29. Williams AB (2014) Spontaneous mutation rates come into focus in *Escherichia coli.* DNA Repair 24: 73-79.

30. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? PLoS Biol 9: e1001127.