Research Article

Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes

Amit N. Khachane, Kenneth N. Timmis, and Vítor A.P. Martins dos Santos*

Department of Environmental Microbiology, Helmholtz Center for Infection Research, Braunschweig, Germany.

*** Corresponding author**:
Vitor A.P. Martins dos Santos, Helmholtz Center for Infection Research,
Inhoffenstrasse 7, D-38124 Braunschweig, Germany.
Tel: +49(0) 531 6181 4008. Fax: +49(0) 531 6181 4199.
e-mail: vds@helmholtz-hzi.de

Key words:
GC content, genome size, mitochondria, obligate intracellular microbes, exponential decay

Running head:
Dynamics of reductive genome evolution

Abbreviations:
GC: guanine plus cytosine content; 16S rDNA: 16S ribosomal DNA gene; SSU rDNA: small subunit ribosomal DNA gene; MB: Megabases; Myr: million years; Gyr: billion years.

Abstract

Reductive evolution in mitochondria and obligate intracellular microbes has led to a significant reduction in their genome size and GC content. We show that genome shrinkage during reductive evolution in prokaryotes follows an exponential decay pattern and provide a method to predict the extent of this decay on an evolutionary time-scale. We validated predictions by comparison with estimated extents of genome reduction known to have occurred in mitochondria and *Buchnera aphidicola*, through comparative genomics and by drawing on available fossil evidence. The model shows how the mitochondrial ancestor would have quickly shed most of its genome, shortly after its incorporation into the proto-eukaryotic cell, and prior to co-divergence subsequent to the split of eukaryotic lineages. It also predicts that the primary rickettsial parasitic event would have occurred between 140 and 240 million years (Myr) ago, an event of relatively recent evolutionary origin considering the fact that *Rickettsia* and mitochondria evolved from a common Alphaproteobacterial ancestor. This suggests that the symbiotic events of *Rickettsia* and mitochondria originated at different time points. Moreover, our model results predicts that the ancestor of *Wigglesworthia glossinidia brevipalpis*, dated around the time of origin of its symbiotic association with the tsetse fly (50 -100 Myr ago), was likely to have been an endosymbiont itself, thus supporting an earlier proposition that *Wigglesworthia*, which is currently a maternally inherited primary endosymbiont, evolved from a secondary endosymbiont.

Introduction

It is widely believed that present day mitochondria have originated from a symbiotic fusion event (Margulis 1985) that occurred ~ 2.0 billion years (2 Gyr) ago between an amitochondriate pro-eukaryote and a free-living Alphaproteobacterial progenitor (Dyall, Brown and Johnson 2004; Embley and Martin 2006). Subsequent to the symbiotic event, the mitochondrial ancestor (the Alphaproteobacterial progenitor) in a host-restricted intracellular environment, underwent a massive reduction in its genome size until its current size of 0.005 - 0.16 megabases (MB), a process similar to that ongoing in obligate intracellular parasites and endosymbionts (Andersson and Kurland 1998; Andersson and Andersson 1999a; Moran and Wernegreen 2000; Gil et al. 2002). Lack of selection for biosynthetic pathway genes that perform functions redundant to that of the host (Andersson and Kurland 1998) and severe population bottlenecks resulting in an increased fixation of deleterious mutations leading to inactivation of a gene and its subsequent deletion (Moran 1996; Andersson et al. 1998; Ochman and Moran 2001) are some of the factors responsible for genome reduction in obligate intracellular microbes. Many reports relying on cues from phylogenetic analyses suggest a relatively fast genome decay in early stages of the reductive process (Andersson and Andersson 1999b; Wernegreen et al. 2000; Moran and Mira 2001). However, the actual dynamics of such a reductive process has never been directly demonstrated.

Fossil record of hosts are very helpful in dating the origin of endosymbiotic events (Moran et al.1993; Ochman, Elwyn and Moran 1999). But because of the lack of sufficient fossils that represent the various stages of genome reduction, elucidating the

dynamics of genome shrinkage has remained a major challenge. A step towards understanding the decay process was achieved by studying the decay rate of a pseudogene (Gómez-Valero, Latorre and Silva 2004). However, pseudogenization mechanism only partly accounts for the actual ongoing genome decay process, as during the early stages of genome reduction, genes can be shed in chunks, for example, resulting from chromosome rearrangement events (Moran and Mira 2001). Supporting the latter scenario, a recent experimental evidence directly revealed that extensive genome reduction can occur within a very short evolutionary time-span (Nilsson et al. 2005). According to another proposition, the process of genome reduction begins with a gradual gene-by-gene pseudogenization, which at some point of time renders a crucial gene in a pathway nonfunctional, thereby triggering a mass deletion of the dependent genes in the pathway (Dagan, Blekhman and Graur 2006). These findings indicate that a combination of various mechanisms are responsible for genome reduction, and that the study of the dynamics of such a process is rather complex.

In some cases, comparative genomics approaches enable a fair assessment of the magnitude of genome decay that an obligate intracellular microorganism has undergone, for example in *Buchnera aphidicola* (Delmotte et al. 2006; Toft and Fares 2006). However, a pre-requisite while implementing such an approach is availability of sequenced genomes of a relatively large number of close relatives of an obligate intracellular microbe that have diverged at various time points. Here, in this report, we show by a phylogeny-independent approach (one that does not rely specifically on comparisons of phylogenetically-related organisms) that prokaryotic genomes, in general, decay exponentially during the reductive evolutionary process, and provide a

quantitative framework to predict the extent of this decay along the evolutionary time-scale.

Materials and Methods

Data curation and analysis: SSU rDNA sequences from 230 prokaryotes and 67 mitochondria (refer supplementary data, Table S1) along with their genome size information were obtained from the NCBI website (www.ncbi.nlm.nih.gov) and Genomes OnLine database (www.genomesonline.org). Only one strain per species was included in the study to avoid statistical bias. Statistical analyses were done in SigmaPlot 2000 (version 9.0, SPSS Inc.) and MS-EXCEL (Microsoft Corp.)

Results and Discussion

Rate of genome shrinkage in obligate intracellular microbes and mitochondria

We previously have found that the nucleotide composition of 16S ribosomal small subunit RNA genes (SSU rDNA) of obligate intracellular microbes is biased towards high T/AT content (Khachane, Timmis and Martins dos Santos 2005). Since these organisms are known to have small genome sizes and low genomic GC contents (Moran 2002; Wernegreen 2002), we checked whether GC content of 16S sequence as well co-varies with the genome size of prokaryotes like the genome GC content. Here, by plotting the %GC content of the SSU rDNA of mesophiles and mitochondria versus their genome size (n = 298), we observed an exponential relationship (r^2 = 0.737, Fig. 1) of the form: size=p * e $^{q * (GC)}$, where 'size' represents the genome size in megabases, 'GC' the percentage GC content of the SSU rDNA, 'p' = 0.00006 (95%

confidence interval: 0.000121 to 0.000031), 'q' = 0.1971 (95% confidence interval: 0.184 to 0.21). [Note: coefficient of determination (r^2) = 0.737 and *P-value* < 0.0001 for a linear correlation between %SSU rDNA GC contents and the natural logarithm of the genome sizes. See Supplementary Fig. S1]. Thus, the relationship between the SSU rDNA %GC content of mesophiles and mitochondria and their genome size can be written as:

$$size = 0.00006 * e^{0.1971 * (GC)} \quad \text{(Eq. 1)}$$

It is clear from Fig. 1 that the SSU rDNA sequences of mitochondria and obligate intracellular microbes are more AT-rich than those of the free-living bacteria, which clearly suggests that reductive evolution in obligate intracellular microbes and mitochondria is accompanied by a reduction in GC content of the SSU rDNA sequence. Fig. 1 also shows that at higher GC content values in SSU rDNA, a small range of GC content correlates with a wide range of genome size, whereas at lower GC content values, a large range of GC content is associated with a smaller range of genome size. Thus, genomes seem to shrink much more rapidly than the rate of reduction in SSU rDNA GC content.

Model formulation:

Since 16S rRNA has been widely used as a molecular clock to time various aspects of evolutionary events, on similar lines we tested whether we could use 16S rRNA to study the evolutionary dynamics of reductive genome evolution. To this end, we drew on the model developed by Lawrence and Ochman (1997) to estimate the rate at which the GC content of a horizontally-acquired gene adjusts to that of the

background genome, and which Andersson and Andersson (2001) have applied to estimate the rate at which the GC content of an ancestral gene evolved during its reductive evolution. Here, we adapted the model to predict the dynamics of change in GC content of SSU rDNA gene of endosymbionts and protomitochondrion, viz:

$$\Delta GC \text{ per unit time} = m * \frac{(IV \text{ ratio} + ½)}{(IV \text{ ratio} + 1)} * (GC_f - GC_c) \quad \text{(Eq. 2)}$$

where ΔGC is the change in the GC content of a given gene, 'm' is the mutation rate of a gene, 'IV' is the transition to transversion ratio taken as 2:1 from a previous study (Lawrence and Ochman 1997), 'GC_c' is the GC content before every simulation time-step, and 'GC_f' is the expected final SSU rDNA GC content.

Combining equations (1) and (2), yields an equation to predict the drop in the genome size over time as a function of SSU rDNA gene mutation rate:

$$size_{(t)} = [0.00006^{m*k}] * size_{(t-1)}{}^{(1- m*k)} * e^{0.1971*m*k*GC_f} \quad \text{(Eq. 3)}$$

where $size_{(t)}$ is the genome size in MB after time 't' (in Myr), '$size_{(t-1)}$ ' is the genome size in MB at previous time-step, 'm' represents the mutation rate of the SSU rDNA gene (in percentage per time 't'), and k= (5/6) is obtained from the ratio (IV ratio + ½)/(IV ratio + 1).

Simulation parameters: The following simulation parameters were considered for 16S (SSU) rDNA sequences of intracellular bacteria: 1) a constant mutation rate (m) of 4% per 100 Myr, which is roughly an average value of the range that is characteristic

of *Buchnera* and *Carsonella* species (Douglas and Raven 2003; Clark, Moran and Baumann 1999), 2) a transition to transversion ratio 'IV' of 2:1 (Lawrence and Ochman 1997), and 3) a final %GC content (GC_f) content value of zero [Note: the lowest small subunit rRNA GC values found in nature is 12%, that of *Aleurodicus dugesii* mitochondrion. Although theoretically the (GC_f) is considered to be zero, one may not see such a low value due to earlier extinction of the genome. Thus, with mitochondria as model systems for studying the evolutionary fate of intracellular bacteria, a similar fate can be expected for genomic properties of current intracellular bacteria which also share same habitat].

Simulation procedure:

The total simulation time is divided into smaller time intervals, say 1 Myr or 100 Myr. Next, by using the above listed simulation parameter values in Eq. 3 [$size_{(t)} = (0.00006^{m*k}) * size_{(t-1)}{}^{(1-m*k)}$], the genome size (reduced state) at the end of each time interval is predicted. The process is repeated for the intended study period while considering the final genome size estimated in the previous time interval to be the ancestor genome size for the next time interval. 'm' represents the SSU rDNA mutation rate, which can be varied in different time intervals. The difference between the initial genome size and the final genome size gives an estimate of the extent of genome decay that is expected for the studied time-period.

Model assumptions

i) Genome reduction as a regular process

It is believed that during the process of reductive genome evolution, non-functional sequences / pseudogenes are formed as intermediates before complete disintegration

of the coding regions. For example, in the sequenced genome of *Rickettsia prowazekii*, nearly a quarter of the genome was found to be composed of non-coding sequences and these had GC contents significantly lower than that of the coding regions (Andersson et al. 1998). This suggested that these non-functional DNA sequences are in the process of being purged out of the genome (Andersson et al. 1998). Nevertheless, another equally possible scenario is that chunks of coding sequences can be lost abruptly without awaiting inactivation, *i.e.*, without taking degenerative steps (Andersson et al. 1998). Deletion of large contiguous genomic regions has also been demonstrated (Moran and Mira 2001). According to a two-step 'domino effect' model (Dagan, Blekhman and Graur 2006) genome reduction begins with gradual gene-by-gene nonfunctionalization. Consequently, a crucial gene in a pathway is rendered nonfunctional triggering a mass deletion of the dependent genes in the pathway. Furthermore, transfer of genomic fragment from a *Wolbachia* endosymbiont to the insect host nucleus (X chromosome) has also been reported (Kondo et al. 2002). Since genome reduction occurs by a combination of the above discussed mechanisms, the model proposed here reflects a net, average genome decay process with time as a function of the initial genome size and mutation rate of the SSU 16S RNA content and does not describe individual mechanism *per se*.

ii) Obligate intracellular organisms on an evolutionary trajectory towards extinction

The model assumes that the genomes of obligate intracellular microbes would undergo continual gene loss that ultimately may lead to their extinction (or of negligible genome size). A recent study showed that, despite an apparent conserved genomic architecture for the past 50 Myr (Tamas et al. 2002), the genomes of

Buchnera species are still shrinking (Gil et al. 2002, Latorre et al. 2005) as evidenced by lineages with further genome reduction, and that they are possibly on an evolutionary trajectory towards extinction (Latorre et al. 2005). This trend is evident in mitochondria, since certain eukaryotes have lost previously-acquired mitochondrial genomes (Palmer 1997; Knight 2004). In Rickettsial species as well the genome decay process is ongoing (Andersson and Andersson 1999a). Endosymbionts retain genes (or few relevant pathways) that are necessary for producing essential metabolites needed by the host. The rest of the genome is expected to be lost over the time, including the most conserved pathways in free-living bacteria such as glycolysis and TCA cycle. This is evident in insect endosymbionts, *Blochmannia*, *Buchnera*, *Wigglesworthia*: none have a complete TCA cycle. The input metabolites are taken from the host. Thus, only a small number of genes will be retained. Eventually, the genome will be lost and replaced by other secondary endosymbionts for complementing the host physiology (Latorre et al. 2005, Pérez-Brocal et al., 2006). For example, mitochondrion of *P.falciparum* has retained just 3 genes, indicating that the genome is near extinction. A minimum set of genes is essential for an organism to lead a free-living life-style, however, since endosymbionts are dependent on their host for their nutritional support, the concept of minimum genes set is probably not applicable to them. Indeed, the recent sequencing of the genome of the smallest known endosymbiont (0.16 MB, 182 ORFs), *Carsonella ruddii*, suggested that it may be evolving into an organelle (Nakabachi et al. 2006). Interestingly as well, is that the genome size and 16rDNA GC content of *Carsonella ruddi* clearly fits into the area exclusively "populated" by mitochondria, an observation that underscores our model assumptions (see Fig. 1).

Verification of the model

In view of the moderate degree of correlation between SSU rDNA %GC content and genome size (Fig 1), we suggest that the model describes an average genome decay curve for prokaryotes. Using this model, we ask, in general, what is the average extent of drop in the genome size of an intracellular microbe for a given period of time. We used equation three to predict the average extent and speed of drop in the genome size, that would be expected for the duration of reductive evolution mitochondria and *Buchnera* have undergone, and compared it with the estimated degree of genome shrinkage they have experienced as determined by comparative genomics approaches.

1) Genome reduction in mitochondria

It has been proposed that mitochondria originated from a symbiotic associative event that occurred some 2Gyr ago triggered by a rise in the atmospheric concentration of highly toxic and reactive oxygen radicals (Andersson and Kurland 1999; Dyall, Brown and Johnson 2004; Embley and Martin 2006). Phylogenomic reconstructions indicate that present-day mitochondria have evolved from a free-living universal ancestor of Alphaproteobacteria that had a genome containing between 3000 and 5000 ORFs (Boussau et al. 2004). This corresponds to an initial genome size of about 3 – 5 MB based on a linear correlation between the number of ORFs in a genome and genome size (Konstantinidis and Tiedje 2004). Mutation rates of SSU rDNA sequence of obligate intracellular microbes, viz. *Buchnera* and *Carsonella* species, range between 1.9 and 6.0% per 100 Myr (Clark, Moran and Baumann 1999; Douglas and Raven 2003). Thus, assuming an average SSU rDNA mutation rate of 4% per 100 Myr for the mitochondrial ancestor (since it shared the same intracellular habitat as that by these obligate intracellular microbes) or even allowing for higher rates of >

4% per 100 Myr, we predicted with Eq.3 current mitochondrial genome sizes to be between 0 and 0.02 MB in all cases (albeit at different times; Fig. 2 and supplementary Table S2). These values clearly fall within the range of genome size values observed in extant mitochondria (mostly between 0.005 to 0.1 MB), and are consistent with the fact that certain eukaryotes have completely lost previously-acquired mitochondrial genomes (Knight 2004). The figure also shows that most of the genome shrinkage had occurred before the divergence of eukaryotic lineages (~1200 Myr ago, Douzery et al. 2004) which is in agreement with the existing notion. Varying the mutation rates in different time intervals during the course of reductive evolution did not alter the outcome of the predictions (supplementary Fig. S2). These results imply that a major part of the genome is exponentially lost within a relatively short interval of evolutionary time, a finding that had been only hypothesized thus far. A recent experimental study showing that a microbial genome could shed as much as 1MB in a very short evolutionary period of ~ 50000 years (Nilsson et al. 2005) supports these conclusions, although direct comparisons need to be of course made with caution.

Mitochondrion evolution following eukaryotic divergence

Fig. 1 shows that the large majority of mitochondria have lower SSU rDNA GC content and smaller genome size than extant obligate intracellular microorganisms. Also, it is clear that their SSU rDNA GC contents vary widely, whereas they have relatively similar genome sizes (Supplementary Fig. S3). For example, the GC content of mitochondrial SSU rDNAs of metazoans vary from 12% to 54%, whereas their genome sizes are rather similar at around ~ 0.015 MB. This may be explained by a two-tier evolutionary scenario (Fig. 3) in which the universal common mitochondrial

ancestor (proto-mitochondrion) would have first shed a major portion of its genome shortly after making the transition from the free-living form to the intracellular environment 2000 Myr ago (Dyall, Brown and Johnson 2004; Embley and Martin 2006) but prior to the subsequent divergence of the eukaryotic lineages, estimated to have occurred around 1200 Myr ago (Embley and Martin 2006; Douzery et al. 2004; Javaux, Knoll and Walter 2001). At that point, the reduced proto-mitochondrion within the eukaryotic ancestor would have retained less than 8 % of its original genome, having thus lost most of what it could shed. After the eukaryotic split, the greatly reduced mitochondrial genomes would have decayed slowly (as predicted by an exponentially decaying curve) while undergoing major changes in SSU rDNA GC content that were determined by disparate mutation rates in different eukaryotic hosts (Fig. 3). For example, plant mitochondria (phylum *Streptophyta*), in general have a higher SSU rDNA GC content in comparison to their counterparts. These changes in the SSU RNA GC content thus reflect the adaptative responses of the distinct mitochondria to their eukaryotic hosts undergoing themselves the (on-going) accelerated evolution process that resulted ultimately in the sheer diversity of past and present eukaryotic organisms. It has been suggested that mitochondrial endosymbiosis may have triggered (or contributed to trigger) an "eukaryotic big bang" (Philippe and Adoutte 1998). Whether this was indeed so, is uncertain but the two-tiered evolutionary scenario proposed here suggests that the bulk of the eukaryotic split as currently acknowledged took place when mitochondrion had lost already most its genome. Whether this was a pre-condition for evolutionary divergence of eukaryotes or simply an on-going, parallel process, remains to be elucidated. Whatever the case, these findings underscore the importance of endosymbiosis in eukaryotic evolution.

The model developed provides valuable insights and sets a plausible, quantitative framework for the study of the evolutionary history of mitochondria.

2) Genome reduction in Buchnera aphidicola

It has been suggested that the symbiotic association between *Buchnera* with its aphid host originated about 250 Myr ago (Moran et al.1993; Ochman, Elwyn and Moran 1999; Moran and Wernegreen 2000), thus in Eq. 3, 't'=250 Myr. Phylogenetic comparisons of gene orthologs amongst the free-living relatives of *E.coli* and *Buchnera*, and subsequent phylogenomic reconstructions, indicate that a free-living *Buchnera* ancestor would have had a genome containing between 1818 ORFs (Silva, Latorre and Moya 2001) and 2425 ORFs (Moran and Mira 2001). Assuming a linear correlation between the number of ORFs in a genome and genome size (r^2 = 0.98, Konstantinidis and Tiedje 2004), this corresponds to a *Buchnera* common ancestral genome size between ~1.85 MB and ~2.5 MB. According to Eq. 3, for a period of 250 Myr of intracellular lifestyle and a constant SSU rDNA sequence mutation rate of 4% per 100 Myr, the extant *Buchnera* genome size should range between 1.06 MB and 0.80 MB, and for a mutation rate of 5% per 100 Myr, between 0.87 and 0.66 MB, which agrees reasonably well with the actual genome size range of 0.67 - 0.42 MB that was experimentally determined for various *Buchnera* species by Gil et al. (2002) and (Pérez-Brocal et al., 2006), see Fig. 4. In addition, the genome sizes predicted at two different time points *viz*., 70 Myr ago and midway between 70 and 160 Myr ago (Fig. 4), were reasonably close to the sizes estimated by comparative genomics (Delmotte et al. 2006).

These predictitions were derived assuming a constant SSU rDNA rate. Nevertheless, we certainly cannot ignore that the SSU rDNA mutation rate, or AT-biased mutations,

is likely to be relatively higher immediately following a change in the habitat (from a free-living to a host-restricted environment) than that is at present, as has been indicated for protein coding genes in endosymbionts (Clark, Moran and Baumann 1999). To test how would this possibly affect the outcome of the model, we varied the SSU rDNA mutation rates (m) at different time intervals during the course of reductive evolution, and by varying the coefficient parameter of equation 1 by implementing upper and lower-end values of the 95% confidence interval as well as a five-fold change. Globally, we found that this did not influence significantly the predictions outcome (supplementary Fig. S4 and Fig. S5). An exponential decay such as the one we propose here is intrinsically consistent as the loss of genes or chunks thereof of a genome implies that the room for further reduction becomes more limited as the genome shrinks (*i.e.*, the more it sheds the less it can loose further). In other words, a genome may appear to be in stasis towards the later stages of reductive evolution but in fact is still undergoing slow shrinkage, as evidenced by the sequencing of the 0.42-MB genome of *Buchnera aphidocola* BCc, which, remarkably, appear to have lost most of its metabolic functions (Pérez-Brocal et al. 2006).

In the above examples (mitochondria and *Buchnera*), the predictions agreed reasonably to the estimated degree of genome decay as determined by comparative genomics approaches for their respective time-periods of reductive evolution. This shows that the method can be a useful tool for a rough approximation of the extent of prokaryotic genome decay over time. The model also enables prediction of the time of extinction, and directly supports the hypothesis that, in the absence of counter-selective pressures, obligate intracellular microbes may become extinct after sufficiently long period of intracellular residence.

Implications of our model

1) Origin of non-organelle primary endosymbiosis

Although *Rickettsia* and mitochondria both evolved from common Alphaproteobacteria ancestors (Andersson et al. 1998; Gray, Burger and Lang 2001; Boussau et al. 2004), and reside in intracellular environments, they appear to be at different evolutionary stages, since extant *Rickettsia* have genome sizes around 2 MB which are significantly larger than those of mitochondria. Our model predicts that the time needed for shrinkage of an initial common Alphaproteobacteria ancestoral genome, 3-5 MB in size, to 2 MB, to be in the range of 140-240 Myr (Fig. 5). Thus, eukaryotic parasitism by *Rickettsia* is likely to be of recent origin.

2) Genome size of the ancestor of Wigglesworthia glossinidia brevipalpis

The *Wigglesworthia*-tsetse fly symbiotic association originated 50-100 Myr ago (Ochman, Elwyn and Moran 1999; Moran et al. 1993), so our genome decay model predicts that the genome size of the ancestor *Wigglesworthia glossinidia brevipalpis* would have been 0.83 - 0.97 MB (Fig. 6). This seems low for a free-living ancestor, given that the smallest free-living microbe known has a genome size of ~1.3 MB (*Pelagibacter ubique* HTCC1062). This may suggest that 100 Myr ago, the *Wigglesworthia* ancestor was already an endosymbiont. This conclusion is consistent with that of a phylogeny-based study, which proposed that *Wigglesworthia*, a maternally-inherited primary endosymbiont, may have evolved from a secondary endosymbiont (Herbeck, Degnan and Wernegreen 2005).

Conclusion

In summary, we propose a mathematical framework to study the evolutionary dynamics of genome reduction in endosymbionts and obligate intracellular parasites and show that their genomes decay exponentially. In combination with comparative genomics and phylogenetic studies, the evolutionary model described here can be a useful predictor of the extent of genome reduction in prokaryotes that are under reductive evolutionary pressure.

Supplementary material is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgements: A.N.K and V.A.P.M.d.S. gratefully acknowledge financial support from the BMBF (project Intergenomics). K.N.T. thanks the Fonds der Chemischen Industrie for generous support.

Literature cited

Andersson JO, Andersson SG. 1999a. Genome degradation is an ongoing process in *Rickettsia*. Mol. Biol. Evol. 16:1178-1191.

Andersson JO, Andersson SG. 1999b. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. 9:664-671.

Andersson JO, Andersson SG. 2001. Pseudogenes, junk DNA, and the dynamics of *Rickettsia* genomes. Mol. Biol. Evol. 8:829-839.

Andersson SG, Kurland CG. 1998. Reductive evolution of resident genomes. Trends Microbiol. 6:263-268.

Andersson SG, Kurland CG. 1999. Origins of mitochondria and hydrogenosomes. Curr. Opin. Microbiol. 2:535-541.

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396:133-140.

Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SG. 2004. Computational inference of scenarios for α-Proteobacterial genome evolution. Proc. Natl. Acad. Sci. U.S.A. 101:9722-9727.

Clark MA, Moran NA, Baumann P. 1999. Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol. Biol. Evol. 16:1586-1598.

Dagan T, Blekhman R, Graur D. 2006. The "domino theory" of gene death: gradual and mass gene extinction events in three lineages of obligate symbiotic bacterial pathogens. Mol. Biol. Evol. 23:310-316.

Delmotte F, Rispe C, Schaber J, Silva FJ, Moya A. 2006. Tempo and mode of early gene loss in endosymbiotic bacteria from insects. BMC Evol. Biol. 6:56.

Douglas AE, Raven JA. 2003. Genomes at the interface between bacteria and organelles. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 358:5-17.

Douzery EJ, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. U.S.A. 101:15386-15391.

Dyall SD, Brown MT, Johnson PJ. 2004. Ancient invasions: from endosymbionts to organelles. Science 304:253-257.

Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. Nature 440:623-630.

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

Gómez-Valero L, Latorre A, Silva FJ. 2004. The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. Mol. Biol. Evol. 21:2172-2181.

Gray MW, Burger G, Lang BF. 2001. The origin and early evolution of mitochondria. Genome Biol. 2.

Herbeck JT, Degnan PH, Wernegreenn JJ. 2005. Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the enterobacteriales (gamma-Proteobacteria). Mol. Biol. Evol. 22:520-532.

Javaux EJ, Knoll AH, Walter MR. 2001. Morphological and ecological complexity in early eukaryotic ecosystems. Nature 41:66-69.

Khachane AN, Timmis KN, Martins dos Santos VA. 2005. Uracil content of 16S rRNA of thermophilic and psychrophilic prokaryotes correlates inversely with their optimal growth temperatures. Nucleic Acids Res. 33:4016-4022.

Knight J. 2004. Giardia: not so special, after all? Nature 429:236-237.

Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T. 2002. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. Proc. Nat. Acad. Sci. U.S.A. 99:14280–14285.

Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proc. Natl. Acad. Sci. U.S.A. 101:3160-3165.

Latorre A, Gil R, Silva FJ, Moya A. 2005. Chromosomal stasis versus plasmid plasticity in aphid endosymbiont *Buchnera aphidicola*. Heredity 95:339-347.

Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. 44:383-397.

Margulis L, Bermudes D. 1985. Symbiosis as a mechanism of evolution: status of cell symbiosis theory. Symbiosis 1:101-124.

Moran NA, Munson MA, Baumann P, Ishikawa H. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. London Ser. B. 253:167–171.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U.S.A. 93:2873-2878.

Moran NA, Wernegreen JJ. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol. Evol. 15:321-326.

Moran NA, Mira A. 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome Biol. 2.

Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108:583-586. Review.

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science 314:267.

Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JC, Andersson DI. 2005. Bacterial genome size reduction by experimental evolution. Proc. Natl. Acad. Sci. U.S.A. 102:12112-12116.

Ochman H, Elwyn S, Moran NA. 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. U.S.A. 96:12638-12643.

Ochman H, Moran NA. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292:1096-1099. Review.

Palmer JD. 1997. Organelle genomes: going, going, gone! Science 275:790-791. Review.

Philippe H, Adoutte A. 1998. The molecular phylogeny of Eukaryota: solid facts and uncertainties. In: Evolutionary Relationships Among Protozoa (Coombs GH, Vickerman K, Sleigh MA, Warren A, editors. Chapman & Hall, London: p. 25–56.

Silva FJ, Latorre A, Moya A. 2001. Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. Trends Genet. 17:615-618.

Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SG. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science 296:2376-2379.

Toft C, Fares MA. 2006. GRAST: a new way of genome reduction analysis using comparative genomics. Bioinformatics 22:1551-1561.

Wernegreen JJ, Ochman H, Jones IB, Moran NA. 2000. Decoupling of genome size and sequence divergence in a symbiotic bacterium. J. Bacteriol. 182:3867-3869.

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. 3:850-861. Review.

Figure legends

Figure 1:

Scatter plot of GC content of SSU rDNA versus genome size of free-living and intracellular prokaryotes and mitochondria. The majority of the GC contents of mitochondrial SSU rDNAs are lower than those of intracellular prokaryotes.

Figure 2:

Prediction of the genome size of extant mitochondria. The three simulations are based on assumed genome sizes of the proto-mitochondrion of 3, 4 and 5 MB (Boussau et al. 2004), and for SSU rDNA mutation rate scenarios of 4% (a), 7% (b) and 10% (c) per 100 Myr.

Figure 3:

Proposed evolutionary model of mitochondria depicting the period in which the genome reduction and GC content variation in the SSU rDNA may have occurred.

Figure 4:

Prediction of the genome size of extant *Buchnera aphidicola* species. The two simulations are based on ancestral genome sizes of 1.8 MB (Silva, Latorre and Moya 2001; pink line) and 2.5 MB (Moran and Mira 2001; blue line), and a 16S mutation rate of 4% per 100 Myr (a), and for 5% per 100 Myr (b). The zero time point represents the transition point from the free-living to the intracellular states, and the 250 Myr point represents the present. The curly bracket represents the extant genome size range of various *Buchnera* species identified by Gil et al. (2002). Filled circles

represent predictions of ancestoral genomes sizes of *Buchnera aphidicola*, using comparative genomics approach, at different time points (Delmotte et al. 2006). [Note: time points denoted by 'A' (250 Myr ago - origin of endosymbiosis), 'B' (midpoint of the range 70 to 160 Myr ago, for the common ancestor of *Buchnera aphidicola* Sg, Bp and Ap species) and 'C' (70 Myr ago - common ancestor of *Buchnera aphidicola* Sg and Ap), have been derived based on aphid host fossil records (Moran et al. 1993).]

Figure 5:

Estimated evolutionary age of parasitism of an eukaryotic cell by *Rickettsia*. The simulation shows the genome decay kinetics of the common Alphaproteobacterial ancestor, 3-5 MB in size (Boussau et al. 2004), to that of current *Rickettsia* species, 2.1 MB in size (filled circle).

Figure 6:

Prediction of the genome size of the ancestor of *Wigglesworthia glossinidia brevipalpis*. The current genome size of *Wigglesworthia glossinidia brevipalpis* is ca. 0.7 MB (filled circle). The ancestor of *W. glossinidia* has been dated at 50-100 Myr ago (Ochman, Elwyn and Moran 1999; Moran et al. 1993). The simulation predicts a genome size of this ancestor of 0.83-0.97 MB.

Figure 1:

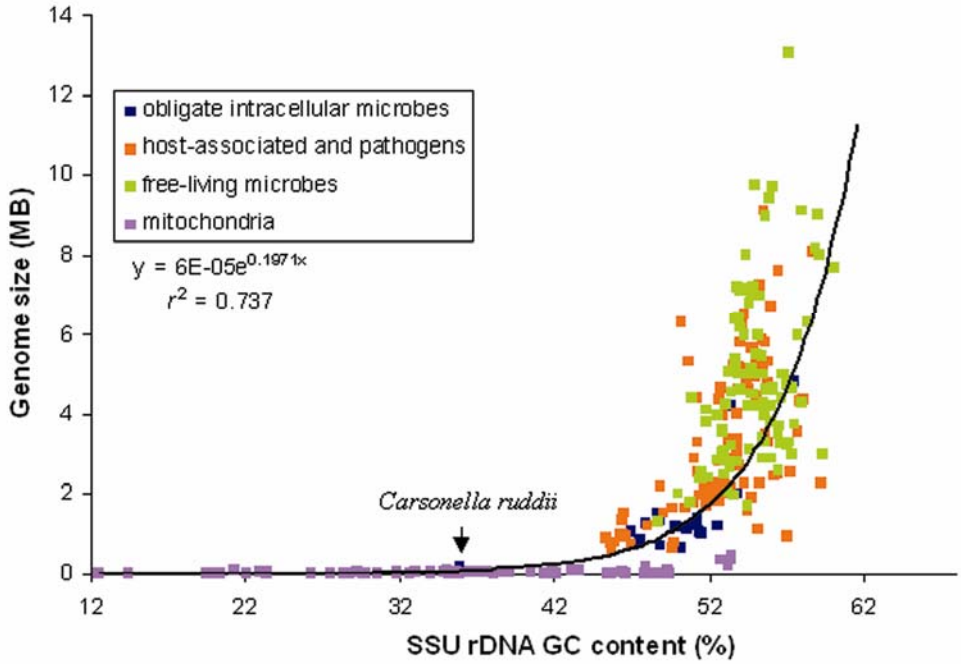

obligate intracellular microbes
host-associated and pathogens
free-living microbes
mitochondria
y = 6E-05e^0.1971x
r^2 = 0.737
Carsonella ruddii
Genome size (MB)
SSU rDNA GC content (%)
0
2
4
6
8
10
12
14
12
22
32
42
52
62

Figure 2:

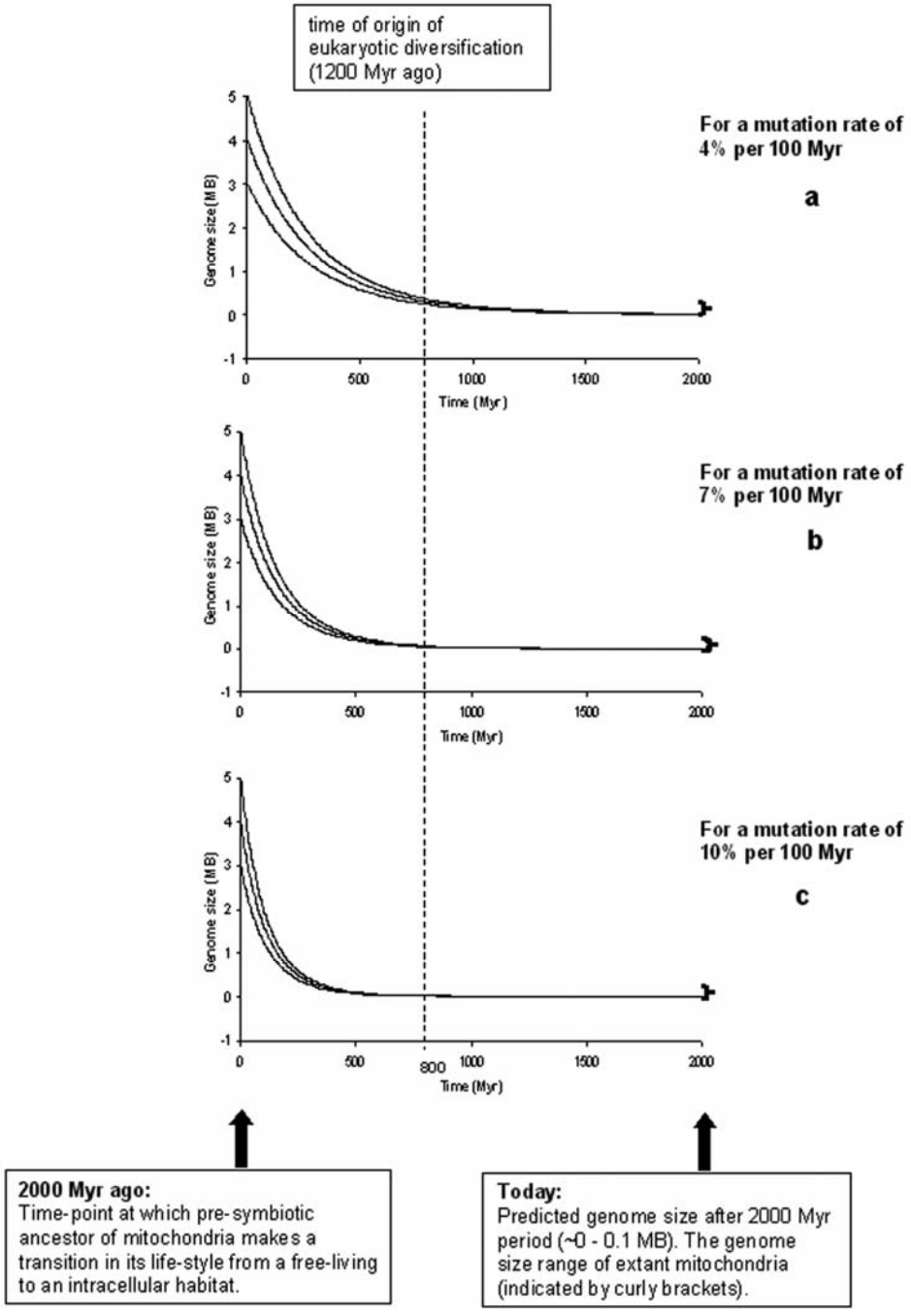

time of origin of eukaryotic diversification (1200 Myr ago)
For a mutation rate of 4% per 100 Myr
a
For a mutation rate of 7% per 100 Myr
b
For a mutation rate of 10% per 100 Myr
c
Genome size (MB)
Time (Myr)
0
500
800
1000
1500
2000
2000 Myr ago:
Time-point at which pre-symbiotic ancestor of mitochondria makes a transition in its life-style from a free-living to an intracellular habitat.
Today:
Predicted genome size after 2000 Myr period (~0 - 0.1 MB). The genome size range of extant mitochondria (indicated by curly brackets).

Figure 3:

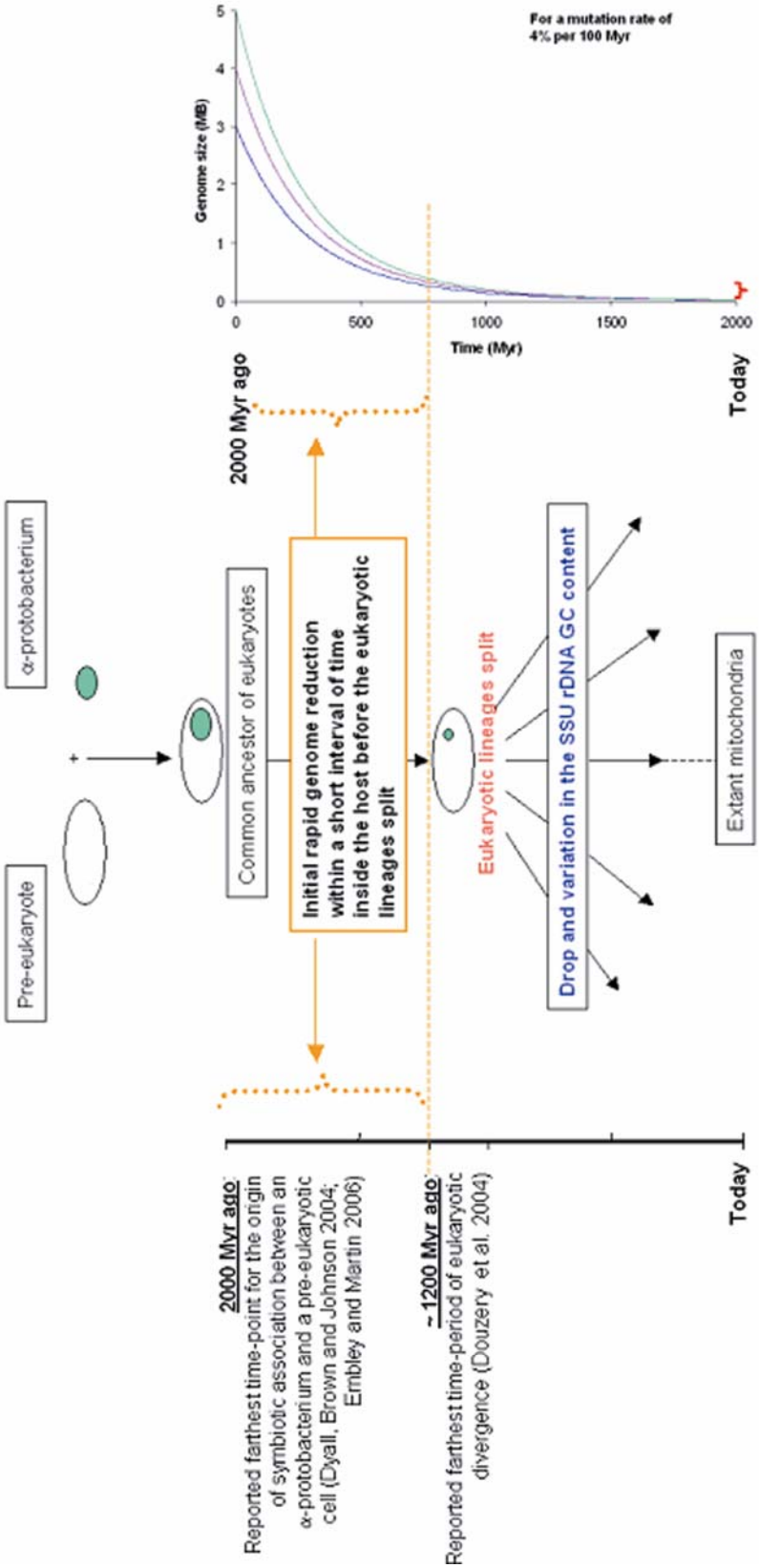
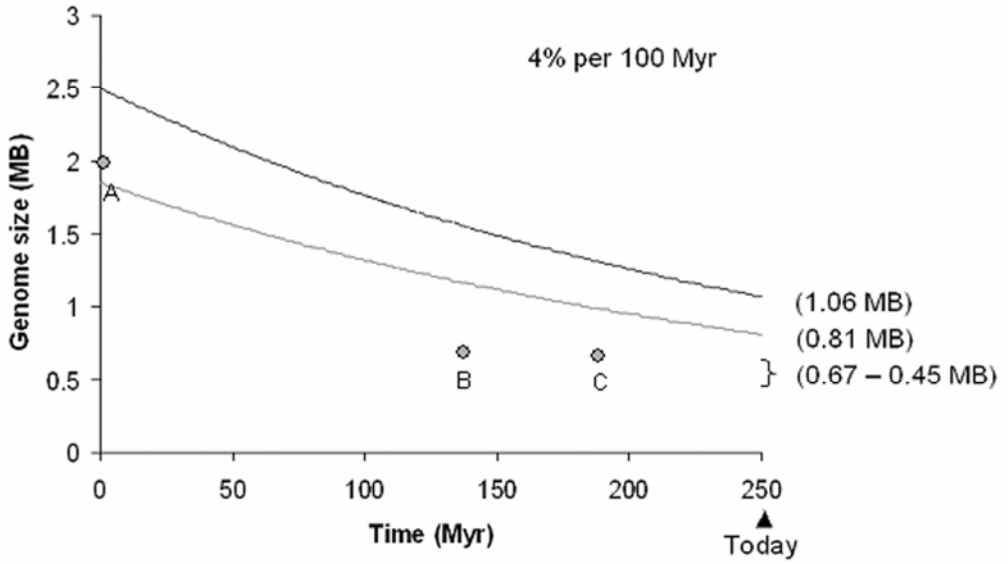

For a mutation rate of
4% per 100 Myr
Genome size (MB)
Time (Myr)
2000 Myr ago
Today
α-protobacterium
Pre-eukaryote
Common ancestor of eukaryotes
Initial rapid genome reduction within a short interval of time inside the host before the eukaryotic lineages split
Eukaryotic lineages split
Drop and variation in the SSU rDNA GC content
Extant mitochondria
2000 Myr ago:
Reported farthest time-point for the origin of symbiotic association between an α-protobacterium and a pre-eukaryotic cell (Dyall, Brown and Johnson 2004; Embley and Martin 2006)
~1200 Myr ago:
Reported farthest time-period of eukaryotic divergence (Douzery et al. 2004)
Today

Figure 4:

(a)


4% per 100 Myr
Genome size (MB)
3
2.5
2
1.5
1
0.5
0
A
B
C
(1.06 MB)
(0.81 MB)
(0.67 – 0.45 MB)
0
50
100
150
200
250
Time (Myr)
Today

(b)

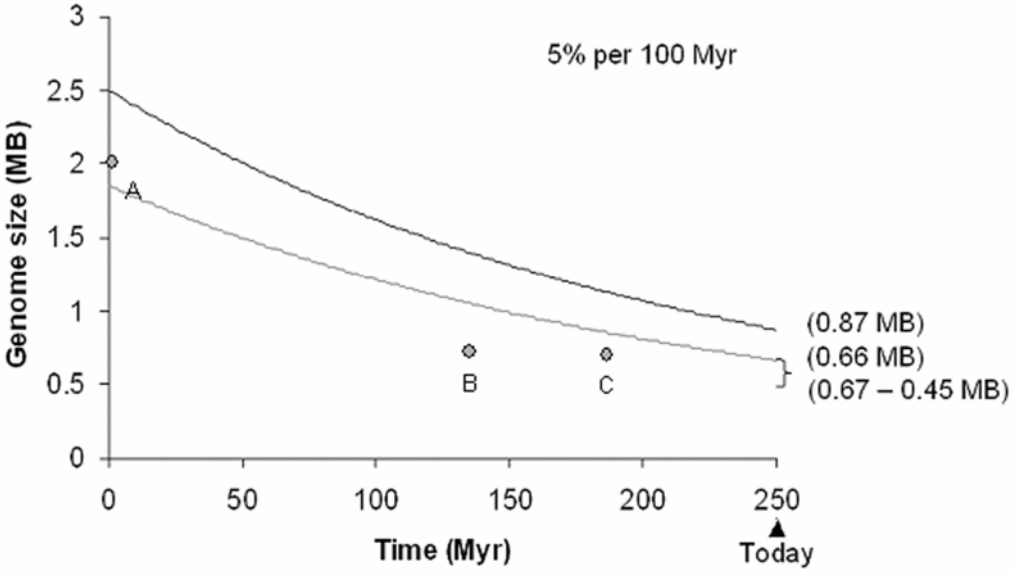

5% per 100 Myr
Genome size (MB)
3
2.5
2
1.5
1
0.5
0
A
B
C
(0.87 MB)
(0.66 MB)
(0.67 – 0.45 MB)
0
50
100
150
200
250
Time (Myr)
Today

Figure 5:

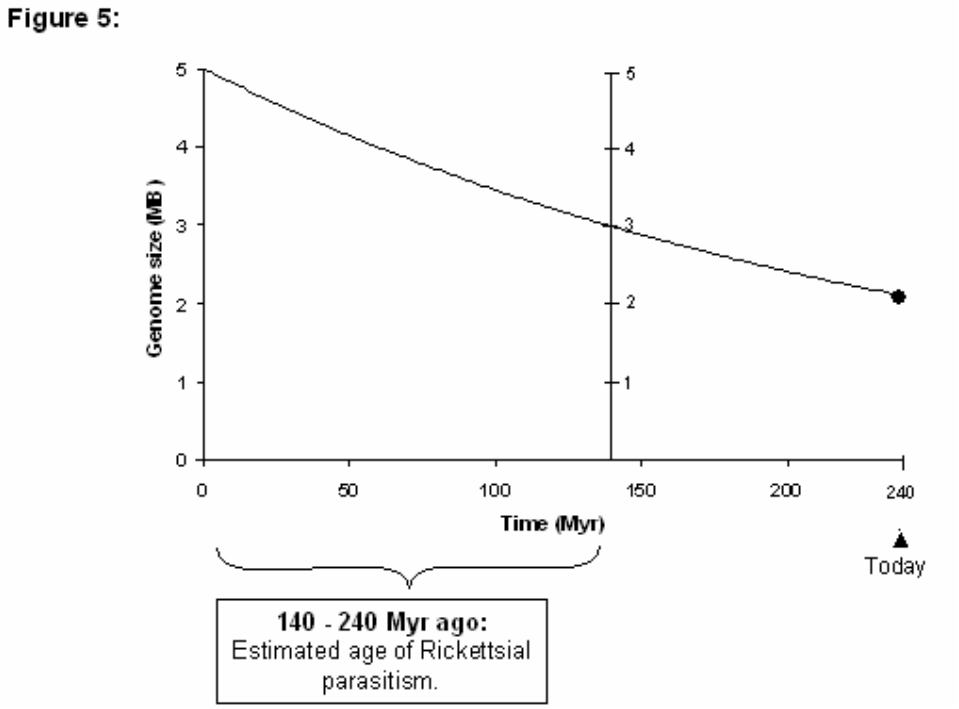
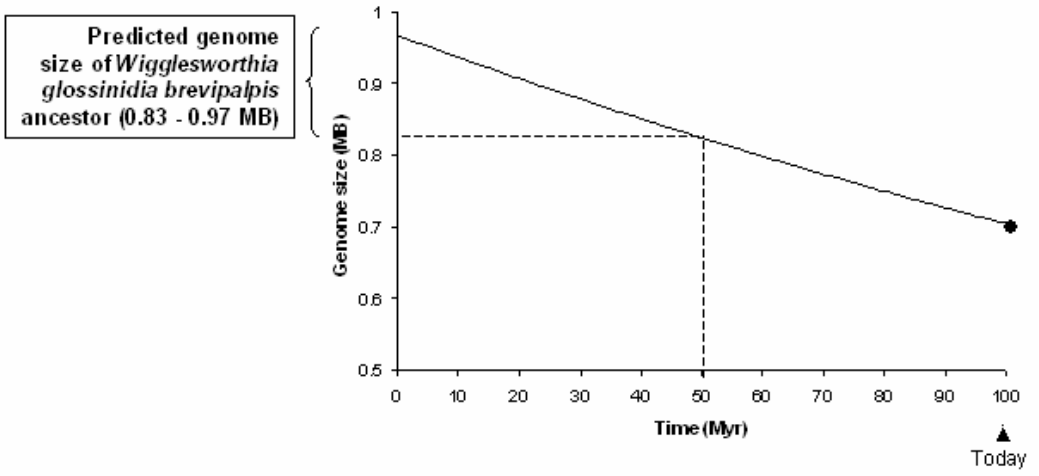

Genome size (MB)
0
1
2
3
4
5
0
50
100
150
200
240
Time (Myr)
Today
140 - 240 Myr ago:
Estimated age of Rickettsial parasitism.

Figure 6:


Predicted genome size of *Wigglesworthia glossinidia brevipalpis* ancestor (0.83 - 0.97 MB)
Genome size (MB)
1
0.9
0.8
0.7
0.6
0.5
0
10
20
30
40
50
60
70
80
90
100
Time (Myr)
Today