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

E-mail: vd@helmholtz-hzi.de.

doi:10.1093/molbev/msl174
Advance Access publication November 15, 2006

© The Author 2006. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

---

**Dynamics of Reductive Genome Evolution in Mitochondria and Obligate Intracellular Microbes**

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

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

Reductive evolution in mitochondria and obligate intracellular microbes has led to a significant reduction in their genome size and guanine plus cytosine content (GC). 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 timescale.

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 evidences. The model shows how the mitochondrial ancestor would have quickly shed most of its genome, shortly after its incorporation into the proteokaryotic cell and prior to codivergence subsequent to the split of eukaryotic lineages. It also predicts that the primary rickettsial parasitic event would have occurred between 180 and 425 million years ago (MYA), 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 predict that the ancestor of *Wigglesworthia glossinidia brevipalpis*, dated around the time of origin of its symbiotic association with the tsetse fly (50–100 MYA), 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 and Bermudes 1985) that occurred ~2.0 billion years (2 Gyr) ago between an amitochondriate proeukaryote and a free-living alphaproteobacterial progenitor (Dyall et al. 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 MB, a process similar to that ongoing in obligate intracellular parasites and endosymbionts (Andersson and Kurland 1998; Andersson JO and Andersson SG 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 JO and Andersson SG 1999b; Wernegreen et al. 2000; Moran and Mira 2001). However, the actual dynamics of such a reductive process has never been directly demonstrated.

Fossil records of hosts are very helpful in dating the origin of endosymbiotic events (Moran et al. 1993; Ochman et al. 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 toward understanding the decay process was achieved by studying the decay rate of a pseudogene (Gomez-Valero et al. 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 (Dagan et al. 2006). These findings indicate that a combination of various mechanisms is 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 prerequisite 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 phylogenetically 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 timescale.

**Materials and Methods**

**Data Curation and Analysis**

Small subunit ribosomal DNA gene (SSU rDNA) sequences from 230 prokaryotes and 67 mitochondria (refer supplementary data, table S1, Supplementary Material online) along with their genome size information were
observed an exponential relationship (intracellular prokaryotes. of the GC contents of mitochondrial SSU rDNAs are lower than those 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. obtained from the National Center for Biotechnology Information Web site (http://www.ncbi.nlm.nih.gov) and Genomes online database (http://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, Systat Software Inc., Richmond, CA) and MS‐EXCEL (Microsoft Corp., Redmond, WA).

Results and Discussion

Rate of Genome Shrinkage in Obligate Intracellular Microbes and Mitochondria

We previously have found that the nucleotide composition of small subunit ribosomal DNA sequences (SSU rDNA) of obligate intracellular microbes is biased toward high T/AT content (Khachane et al. 2005). Because these organisms are known to have small genome sizes and low genomic guanine plus cytosine contents (GC) (Moran 2002; Wernegreen 2002), we checked whether GC content of 16S sequence as well covaries with the genome size of prokaryotes, like the genome size of mitochondria versus their genome size (Khachane et al. 2005). Because these small subunit rRNA GC values found in nature is 12%, that suggests that reductive evolution in obligate intracellular microbes and mitochondria is accompanied by a reduction in GC content of the SSU rDNA sequence. Figure 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

Because 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 J0 and Andersson SG (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 promitochondrion, namely,

\[ \Delta GC \text{ per unit time} = m \frac{(IV \text{ ratio} + \frac{1}{2})^{GCf}(GC_f - GC_c)}{(IV \text{ ratio} + 1)} \]

where \( \Delta GC \) 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

\[ \text{size}_{(t)} = \left[ 0.00006 \exp \left[ \frac{1.971}{GC} \right] \right] \left[ \text{size}_{(t-1)} \right] \left[ 0.00006 \left( \frac{1}{IV \text{ ratio} + 1} \right) \right] \]

where \( \text{size}_{(t)} \) is the genome size in MB after time \( t \) (in Myr), \( \text{size}_{(t-1)} \) 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 \((IV \text{ ratio} + \frac{1}{2})/(IV \text{ ratio} + 1)\).

Simulation Parameters. The following simulation parameters were considered for 16S (SSU) ribosomal DNA gene (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 (Clark et al. 1999; Douglas and Raven 2003), 2) a transition to transversion ratio of 2:1 (Lawrence and Ochman 1997), and 3) a final GC content \( GC_f \) of zero (Note: the lowest small subunit rRNA GC values found in nature is 12%, that of Aleurodicus dugesii mitochondrion. Although theoretically the GCf 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 or 100 Myr. Next, by using the above-listed simulation parameter values in equation (3) \[\text{size}_{(t)} = (0.00006^{m^n}) \times \text{size}_{(t-1)}\], the genome size (reduced state) at the end of each time interval is predicted. The process is repeated for the intended study period, whereas 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 vary 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**

**Genome Reduction As a Regular Process.** It is believed that during the process of reductive genome evolution, nonfunctional 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 noncoding sequences and these had GC contents significantly lower than that of the coding regions (Andersson et al. 1998). This suggested that these nonfunctional 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, that is, 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 2-step “domino effect” model (Dagan et al. 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). Because 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 16S rDNA and does not describe individual mechanism per se.

**Obligate Intracellular Organisms on an Evolutionary Trajectory toward 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 toward extinction (Latorre et al. 2005). This trend is evident in mitochondria because certain eukaryotes have lost previously acquired mitochondrial genomes (Palmer 1997; Knight 2004). In *Rickettsia* species as well, the genome decay process is ongoing (Andersson JO and Andersson SG 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*, and *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 *Plasmodium 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 lifestyle, however, because 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 open reading frames [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 16 rDNA GC content of *C. ruddii* 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 (3) 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.

**Genome Reduction in Mitochondria.** It has been proposed that mitochondria originated from a symbiotic associative event that occurred some 2 Gyr ago, triggered by a rise in the atmospheric concentration of highly toxic and reactive oxygen radicals (Andersson and Kurland 1999; Dyall et al. 2004; Embley and Martin 2006). Phylogenetic reconstructions indicate that present-day mitochondria have evolved from a free-living universal ancestor of Alphaproteobacteria that had a genome containing between 3,000 and 5,000 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 et al. 1999; Douglas and Raven 2003). Thus, assuming an average SSU rDNA
mutation rate of 4% per 100 Myr for the mitochondrial ancestor (because 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 equation (3), current mitochondrial genome sizes to be between 0 and 0.02 MB in all cases (albeit at different times, fig. 2). These values clearly fall within the range of genome-size values observed in extant mitochondria (mostly between 0.005 and 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 MYA, 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, Supplementary Material online). 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 1 MB in a very-short evolutionary period of ~50,000 years (Nilsson et al. 2005) supports these conclusions, although direct comparisons need to be made with caution.

Mitochondrion Evolution following Eukaryotic Divergence. Figure 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, Supplementary Material online). 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 2-tier evolutionary scenario (fig. 3) in which the universal common mitochondrial ancestor (protomitochondrion) would have first shed a major portion of its genome shortly after making the transition from the free-living form to the intracellular environment 2,000 Myr ago (Dyall et al. 2004; Embley and Martin 2006), but prior to the subsequent divergence of the eukaryotic lineages, estimated to have occurred around 1,200 Myr ago (Javaux et al. 2001; Douzery et al. 2004; Embley and Martin 2006). At that point, the reduced protomitochondrion within the eukaryotic ancestor would have 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 adaptive responses of the distinct mitochondria to their eukaryotic hosts undergoing themselves the (ongoing) accelerated evolution process that resulted ultimately in the sheer diversity of past and present eukaryotic organisms.

Genome Reduction in Β. aphidicola. It has been suggested that the symbiotic association between Buchnera with its aphid host originated about 250 MYA (Moran et al. 1993; Ochman et al. 1999; Moran and Wernegreen 2000), thus in equation (3), \( t = 250 \text{ Myr} \). Phylogenetic comparisons of gene orthologs amongst the free-living relatives of Escherichia coli and Buchnera, and subsequent phylogenomic reconstructions, indicate that a free-living Buchnera ancestor would have had a genome containing between 1,818 ORFs (Silva et al. 2001) and 2,425 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 \( 1.85 \) and \( 2.5 \) MB. According to equation (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

---

**Fig. 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.**
1.06 and 0.80 MB, and for a mutation rate of 5% per 100 Myr, between 0.87 and 0.66 MB, respectively, 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 figure 4. In addition, the genome sizes predicted at 2 different time points viz. 70 MYA and midway between 70 and 160 MYA (fig. 4) were reasonably close to the sizes estimated by comparative genomics (Delmotte et al. 2006). These predictions were derived assuming a constant SSU rDNA rate. Nevertheless, we certainly cannot ignore that the SSU rDNA mutation rate, or AT-biased mutation, is likely to be relatively higher immediately following a change in the habitat (from a free-living to a host-restricted environment) than at present, as has been indicated for protein-coding genes in endosymbionts (Clark et al. 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 5-fold change. Globally, we found that this did not influence significantly the prediction’s outcome (supplementary figs. S4 and S5).

**Implications of Our Model**

**Origin of Nonorganelle Primary Endosymbiosis.** Although Rickettsia and mitochondria both evolved from common Alphaproteobacteria ancestors (Andersson et al. 1998; Gray et al. 2001; Boussau et al. 2004) and reside in intracellular environments, they appear to be at different evolutionary stages because extant Rickettsia have genome sizes around 1.35 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 1.1 to 1.6 MB, to be in the range of 180–425 Myr (fig. 5). Thus, eukaryotic parasitism by Rickettsia is likely to be of recent origin.
Genome Size of the Ancestor of Wigglesworthia glossinidia brevipalpis. The Wigglesworthia–tsetse fly symbiotic association originated 50–100 MYA (Moran et al. 1993; Ochman et al. 1999), so our genome decay model predicts that the genome size of the ancestor W. 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 MYA, 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 et al. 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

The supplementary table S1 and figures S1–S5 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

A.N.K. and V.A.P.M.d.S. gratefully acknowledge financial support from the Bundesministerium für Bildung und Forschung (project Intergenomics) and the European Union (New and Emerging Science and Technology Project Programmable Bacterial Catalysts, Contract Nr. 029104). K.N.T. thanks the Fonds der Chemischen Industrie for generous support.

Literature Cited


William Martin, Associate Editor

Accepted November 6, 2006