Journal of Theoretical Biology 503 (2020) 110389

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

A simple stochastic model describing genomic evolution over time of GC content in microbial symbionts

Jon Bohlin [a,b,c,*], Brittany Rose [a,d], Ola Brynildsrud [a,c], Birgitte Freiesleben De Blasio [a,d]

[a] *Division of Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway*
[b] *Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway*
[c] *Department of Production Animals, Faculty of Veterinary Medicine, Norwegian University of Life Science, Oslo, Norway*
[d] *Department of Biostatistics, Oslo Centre for Biostatistics and Epidemiology, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway*

ARTICLE INFO

Article history:
Received 21 October 2019
Revised 21 April 2020
Accepted 24 June 2020
Available online 4 July 2020

Keywords:
Modeling of genomic evolution
Genome reduction
Müllers ratchet
Stochastic mutation rates
Base composition

ABSTRACT

An organism's genomic base composition is usually summarized by its AT or GC content due to Chargaff's parity laws. Variation in prokaryotic GC content can be substantial between taxa but is generally small within microbial genomes. This variation has been found to correlate with both phylogeny and environmental factors. Since novel single-nucleotide polymorphisms (SNPs) within genomes are at least partially linked to the environment through natural selection, SNP GC content can be considered a compound measure of an organism's environmental influences, lifestyle, phylogeny as well as other more or less random processes. While there are several models describing genomic GC content few, if any, consider AT/GC mutation rates subjected to random perturbations. We present a mathematical model that describes how GC content in microbial genomes evolves over time as a function of the AT → GC and GC → AT mutation rates with Gaussian white noise disturbances. The model, which is suited specifically to non-recombining vertically transmitted prokaryotic symbionts, suggests that small differences in the AT/GC mutation rates can lead to profound differences in outcome due to the ensuing stochastic process. In other words, the model indicates that time to extinction could be a consequence of the mutation rate trajectory on which the symbiont embarked early on in its evolutionary history.

1. Introduction

GC content varies considerably between prokaryotic species but is remarkably stable genome-wide, despite the fact that bacterial genomes are predominantly functional and expressed in some sense (Rocha and Feil, 2010). Bacteria can have an average genomic GC content of as low as 13.5% (*Candidatus* Zinderia insecticola) or of as high as 75% (*Anaeromyxobacter dehalogenans*) (Bohlin et al., 2018). While both large and small bacteria can be either GC-rich or AT-rich, there seems to be a tendency—at least in some phylogentic groups—for symbionts with smaller genomes to be more AT-rich, while soil-dwelling bacteria with large genomes tend to be more GC-rich (Bohlin et al., 2014; Agashe and Shankar, 2014).

The mechanisms responsible for GC richness in bacteria with large genomes are poorly understood; far more can be deduced from AT-rich bacteria with small genomes (see Agashe and Shankar, 2014 for a general review of GC content in prokaryotes). For instance, it was conjectured (Bentley and Parkhill, 2004) (before being later demonstrated (Hershberg and Petrov, 2010)) that mutations are generally AT-biased due to frequent methylation of cytosine that can subsequently change to thymine. Phylogenetic relatedness, on the other hand, exerts strong selective pressures against changes in GC content. This is due in large part to the significant role that protein coding genes play in bacteria and to the fact that mutations in the first two positions of a codon change the amino acid it codes for (Reichenberger et al., 2015). Phylogenetic influence on base composition in prokaryotes seems to be most prominent at the genus level and below (Bohlin et al., 2017).

Free-living bacteria can develop a sustained symbiotic relationship with a host, typically an insect, either through a facultative, horizontal exchange of nutrients or vertically, through cultivation within bacteriocytes (Wernegreen, 2017). Host-symbiont relationships typically start with horizontal exchange of beneficial nutrients (Fisher et al., 2017; Boscaro et al., 2017). According to some recent findings, a host that is dependent on a symbiont can exchange it with another as long as it receives the necessary nutrients (Moran and Bennett, 2014; Wernegreen, 2015; Hosokawa et al., 2016; Boscaro et al., 2017). Sometimes an established horizontal host-symbiont relationship will progress to a vertical relationship resulting in a strong host-

[*] Corresponding author at: Division of Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway.

https://doi.org/10.1016/j.jtbi.2020.110389
0022-5193/© 2020 Elsevier Ltd. All rights reserved.

symbiont dependecy (Fisher et al., 2017; Boscaro et al., 2017). This can, for instance, happen when the host is able to access particular environments, due to the symbiont, that it would not survive in otherwise (Fisher et al., 2017). The symbionts can, in turn, come to rely on provisions from the within–host environment. This environment will then drive the genomic evolution of the symbiont (Wernegreen, 2015). In a stable within–host environment the host may render several bacterial products redundant. That is, proteins, amino acids and nutrients that the symbiont expresses but are also available from the host could give a symbiont an advantage over another if lost (Batut et al., 2014). Ensuing mutations may accumulate in such genetic regions and form novel proteins that may eventually be beneficial to the host/symbiont relationship or not expressed at all and subsequently lost (Boscaro et al., 2017). Furthermore, maintaining a large genome requires energy and so genome size reduction may be advantageous (Lane and Martin, 2010). Indeed, genome reduction appears to start early in a microbial symbiont with an established vertical relationship with a host (Sabater-Muñoz et al., 2017; Bennett and Moran, 2015). If the host has a low effective population size (N_e), genetic drift may further influence the size and base composition of the symbiont's genome (Wernegreen, 2017; Lynch et al., 2016). The outside environment can thus also affect genomic base composition in symbionts (Foerstner et al., 2005).

There are several indicators that genome size reduction in microbial symbionts occur before genomic GC content drops (Wernegreen, 2017). Loss of DNA mismatch repair (MMR) genes and proofreading enzymes can nevertheless lead to a relatively quick decrease in genomic GC content (Lind and Andersson, 2008). An increase in genomic GC content, on the other hand, can result in increased fitness (Raghavan et al., 1450), and this is associated with stronger selection on base composition (Hildebrand et al., 2010; Bohlin et al., 2017; Bobay and Ochman, 2017). Abundance of nitrogen, as in soil, has been identified as a driver for increased genomic GC content (Seward and Kelly, 2016).

A recent study (Bohlin et al., 2018) found that single-nucleotide polymorphisms (SNPs) in microbial core genomes from different taxa were surprisingly GC-rich, except in cases where the genomes themselves were already among the most GC-rich. The study presented a mathematical model describing SNP GC content as a function of core genome GC content. The model indicated that GC → AT mutations occurred at roughly double the rate of AT → GC mutations, which suggests that most GC → AT mutations are lost prior to fixation (Bohlin et al., 2018).

In another recent study (Bohlin et al., 2019), it was shown that while GC → AT mutation rates are remarkably consistent across bacterial taxa, AT → GC mutation rates vary considerably. Since the environment exerts selective pressure on bacterial base composition (Foerstner et al., 2005; Reichenberger et al., 2015), it should, at least partly, be reflected in core genome SNPs, together with evolutionary history, lifestyle and taxon.

Even in stable environments, stochastic events impact the influence of the environment on genomic base composition in symbionts (Wernegreen, 2017). Inspired by Motoo Kimura's seminal paper (Kimura, 1980), we modify a previously described model (Bohlin et al., 2018) to investigate GC content evolution with respect to time; we extend the model with the assumption that changes to genomic GC content with respect to time, *i.e.* SNP GC content, can be described by parameters multiplied by genomic GC- and AT content, respectively, both randomly perturbed according to Gaussian white noise. We thus assume that SNP GC content is subject to Chargaff's parity rules (Elson and Chargaff, 1954). In practice, this means that SNP GC content is assumed to be computed from base pair substitutions that are selected for and not from random mutations that are purged before fixation in a generation or two. We employ Itô calculus to solve the stochastic differential equation (SDE) that accounts for the random perturbations in the AT → GC and GC → AT mutation rate parameters. The degree to which these random perturbations will affect the mutation rates can also be adjusted for. Finally, we discuss implications of the model and demonstrate that Muller's ratchet (Moran et al., 1996) can take on several different scenarios that may be unavoidable due to the mutation rates of the symbiont.

2. Methods

2.1. The mathematical model

The mathematical model presented here is an extension of the model presented in Bohlin et al. (2018). The original model, which describes the change in core genome SNP GC content with respect to core genome GC content, is

$$\frac{dF_{GC}(x)}{dx} = \alpha F_{GC}(x) + \beta(1 - F_{GC}(x)). \tag{1}$$

x represents core genome GC content, while $F_{GC}(x)$ represents SNP GC content. These terms are subject to the constraints $0 < x < 1$ and $0 < F_{GC}(x) < 1$. In Bohlin et al. (2018), the parameters α and β were estimated by fitting the model to empirical data using either non-linear least square regression (Bohlin et al., 2018) or Bayesian inference (Bohlin et al., 2019).

In the present study, we are concerned with genomic GC content with respect to time in a stochastic setting. That is, we are now interested in the relation

$$F_{t+\Delta t}(\omega) = F_t(\omega) + \alpha F_t(\omega)\Delta t + \beta(1 - F_t(\omega))\Delta t, \tag{2}$$

where $F_{t+\Delta t}(\omega) - F_t(\omega)$ represents change in GC content, or SNP GC content, at time $t + \Delta t$ for trajectory $\omega \in \Omega$. SNP GC content is thus modeled as a parameter α times $F_t(\omega)$ (GC content at time t) multiplied by time duration Δt plus a parameter β times $1 - F_t(\omega)$ (AT content at time t) times Δt. In other words, SNP GC content is assumed to be determined by the sum of parameter multiples, that represent base substitution rates, of genomic GC- and AT content, respectively. In classical calculus notation, we write

$$\frac{dF_t(\omega)}{dt} = \alpha F_t(\omega) + \beta(1 - F_t(\omega)), \tag{3}$$

Here, $\frac{dF_t(\omega)}{dt}$ represents SNP GC content and $F_t(\omega)$ genomic GC content at time t, and we let mutation rates $\alpha = a + W_t(\omega)$ and $\beta = b + W_t(\omega)$, where $a, b \in \mathbb{R}$ and $W_t(\omega)$ is a Gaussian white noise process with respect to trajectory $\omega \in \Omega$. Eq. (3) is subject to the probability space $(\Omega, \mathcal{F}_t, P)$ as well as the measure space $(\mathbb{R}^+, \mathcal{G}, dt)$. Ω is the space of all trajectories $\omega, \mathcal{F}_t$ is its filtration with respect to each time $t \in \mathbb{R}^+$ (*i.e.* $[0, \infty)$ of which $\mathcal{G}$ is the corresponding Borel algebra and dt Lebesgue measure), and P is a probability measure on Ω. We now have:

$$\begin{aligned}\tfrac{dF_t(\omega)}{dt} = \quad & (a + W_t(\omega))F_t(\omega) + (b + W_t(\omega))(1 - F_t(\omega)) \\ & = aF_t(\omega) + F_t(\omega)W_t(\omega) + \\ & +b(1 - F_t(\omega)) + W_t(\omega)(1 - F_t(\omega)) \\ & = aF_t(\omega) + b(1 - F_t(\omega)) + W_t(\omega).\end{aligned}$$

Hence,

$$\frac{dF_t(\omega)}{dt} = aF_t(\omega) + b(1 - F_t(\omega)) + W_t(\omega). \tag{4}$$

It is important to note that, in the present form, this derivative does not exist in the classical sense or in the Radon–Nikodym sense for $F_t(\omega)$. However, if we assume that $F_t(\omega)$ is a semimartingale (allowing for countable and bounded jumps), the Doob–Meyer decomposition theorem (pp. 129–133 of Protter (2005)) guarantees that $F_t(\omega) = F_0 + A(t) + X_t(\omega)$, where $A(t)$ is a function of bounded variation and $X_t(\omega)$ is a local martingale. Moreover, this decomposition is unique, and both $A(t)$ and $X_t(\omega)$ are adapted to $\mathcal{F}_t$. If we assume that $X_t(\omega)$ is a Brownian motion, then by chapter 3 of Øksendal (2005), (4) can be written as

$$dF_t(\omega) = (aF_t(\omega) + b(1 - F_t(\omega)))dt + dB_t(\omega). \tag{5}$$

Though the term $(aF_t(\omega) + b(1 - F_t(\omega)))dt$ resembles (1) it has a Brownian motion term $dB_t(\omega)$ and must therefore be handled in a non-classical way. We allow the Brownian motion to have scaled volatility c, as it is not unreasonable to expect variance differences across organisms and/or environments in addition to time t. It can be shown that a scaled Brownian motion is also a Brownian motion: Let U_t be a Brownian motion (see, for instance, ch. 2 of Øksendal (2005)). Then,

$$\mathbb{E}(U_t) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} ue^{-\frac{u^2}{2t}} du.$$

Letting $u = cz$ and $\frac{du}{dz} = c$, it follows that

$$\frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} ue^{-\frac{u^2}{2t}} du = \frac{1}{\sqrt{2\pi \frac{t}{c^2}}} \int_{\mathbb{R}} ze^{-\frac{c^2 z^2}{2t}} cdz$$

$$= \frac{1}{\sqrt{2\pi \frac{t}{c^2}}} \int_{\mathbb{R}} cze^{-\frac{c^2 z^2}{2t}} dz = \mathbb{E}\left(cZ_{\frac{t}{c^2}}\right).$$

$\mathbb{E}$ is the expectation operator with respect to probability measure $P(\omega)$, *i.e.* $\mathbb{E}(X) = \int_\Omega XdP$.

We do not presume that $F_t(\omega)$ can see into the future. We assume that $F_t(\omega)$ is adapted to the filtration $\mathcal{F}_t$ for each t, which motivates the use of the Itô integral instead of the Fisk–Stratonovich integral (Protter, 2005). It is therefore enough to assume that $F_t(\omega)$ is a cádlág process, *i.e.* $\lim_{s \to t^+} F_s(\omega) = F_t(\omega)$ (left-continuous with right limits; see ch. 2 of Protter (2005)), implying that $F_t(\omega)$ has a countable number of bounded jumps. We can then use the Itô formula (see ch. 4 of Øksendal (2005)) to solve (5). Furthermore, since we assume that $0 < F_t(\omega) < 1$ and that a, b are finite constants, it is guaranteed that (5) has a strong and unique solution (see ch. 5 of Øksendal (2005)).

First, we must identify an integrating factor that removes $F_t(\omega)$ from the right-hand side. Let

$$\begin{aligned} dF_t(\omega) = \quad & (aF_t(\omega) + b(1 - F_t(\omega)))dt + d\hat{B}_t(\omega) \\ & = (aF_t(\omega) - bF_t(\omega) + b)dt + d\hat{B}_t(\omega) \\ & = ((a - b)F_t(\omega) + b)dt + d\hat{B}_t(\omega), \end{aligned}$$

where $\hat{B}_t(\omega)$ is a c-scaled Brownian motion. Letting $g(t, x) = e^{(-(a-b)t)}x$, we get the integrating factor $g(t, F_t(\omega)) = Y_t(\omega) = e^{(-(a-b)t)}F_t(\omega)$. Applying Itô's formula (p. 44 of Øksendal (2005)), we see that

$$\begin{aligned} dY_t(\omega) = & \frac{\partial g}{\partial t}(t, F_t(\omega))dt + \frac{\partial g}{\partial t}(t, F_t(\omega))dF_t(\omega) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2} \\ & \times (t, F_t(\omega))(dF_t(\omega))^2. \end{aligned} \tag{6}$$

Because $\frac{\partial^2 g}{\partial x^2}(t, x) = 0$, the last term of (6) is equal to zero. As a result,

$$\begin{aligned} dY_t(\omega) = \quad & \tfrac{\partial g}{\partial t}(t, F_t(\omega))dt + \tfrac{\partial g}{\partial x}(t, F_t(\omega))dF_t(\omega) \\ & = -(a - b)e^{(-(a-b)t)}F_t(\omega)dt + e^{(-(a-b)t)}dF_t(\omega) \\ & = -(a - b)e^{(-(a-b)t)}F_t(\omega)dt + \\ & + e^{(-(a-b)t)}\left(((a - b)F_t(\omega) + b)dt + d\hat{B}_t\right) \\ & = be^{(-(a-b)t)}dt + e^{(-(a-b)t)}d\hat{B}_t. \end{aligned}$$

Thus, we have the differential

$$dY_t(\omega) = be^{(-(a-b)t)}dt + e^{(-(a-b)t)}d\hat{B}_t, \tag{7}$$

and so

$$dY_t(\omega) = d\left(e^{(-(a-b)t)}F_t(\omega)\right) = be^{(-(a-b)t)}dt + e^{(-(a-b)t)}d\hat{B}_t.$$

We can then find the formula for $F_t(\omega)$, by setting $s \in [0, t]$, and letting

$$d\left(e^{(-(a-b)t)}F_t(\omega)\right) = be^{(-(a-b)t)}dt + e^{(-(a-b)t)}d\hat{B}_t$$

which gives

$$e^{(-(a-b)t)}F_t(\omega) - F_0(\omega) = \int_0^t be^{(-(a-b)s)}ds + \int_0^t e^{(-(a-b)s)}d\hat{B}_s,$$

and

$$F_t(\omega) = F_0(\omega)e^{(a-b)t} + \int_0^t be^{(a-b)(t-s)}ds + \int_0^t e^{(a-b)(t-s)}d\hat{B}_s. \tag{8}$$

$F_t(\omega)$ is a semimartingale and $\int_0^t be^{(a-b)(t-s)}ds$ is of bounded variation. However,

$$\int_0^t e^{(a-b)(t-s)}d\hat{B}_s$$

is not a local martingale and therefore this is not the unique Doob-Meyer decomposition (see pp. 129–133 of Protter (2005)) described earlier. While the latter Brownian motion term must be solved numerically, the anti-derivative of the bounded variation term can be solved using the chain rule:

$$\int_0^t be^{(a-b)(t-s)}ds = c_0 + \frac{b}{(a - b)}\left(e^{(a-b)t} - 1\right).$$

We thus obtain the explicit equation for $F_t(\omega)$:

$$F_t(\omega) = F_0(\omega)e^{(a-b)t} + \frac{b}{(a - b)}\left(e^{(a-b)t} - 1\right) + \int_0^t e^{(a-b)(t-s)}d\hat{B}_s$$

that can be written as:

$$\begin{aligned} F_t(\omega) = & -\frac{b}{(a - b)} + \left(F_0(\omega) + \frac{b}{(a - b)}\right)e^{(a-b)t} \\ & + \int_0^t e^{(a-b)(t-s)}d\hat{B}_s \end{aligned} \tag{9}$$

which is subject to the constraints $t \in [0, \infty)$ and $0 < F_t(\omega) < 1$. The integration constant c_0 is just assumed included in F_0. It should be noted that for $F_0 = 0$,

$$\mathbb{E}(F_t(\omega)) = \frac{b}{(a - b)}\left(e^{(a-b)t} - 1\right). \tag{10}$$

Since the Brownian motion term vanishes (see p. 30 of Øksendal (2005)), we get the solution to (1) when $t = x$ (Bohlin et al., 2018). Furthermore, we do not need to bother with the Brownian motion term when estimating parameters a and b. The variance is given by $\mathrm{Var}(F_t(\omega)) = \mathbb{E}\left(\left(F_t(\omega) - \mathbb{E}(F_t(\omega))^2\right)\right.$, which we can solve by setting

$$A := F_0(\omega)e^{(a-b)t} + \frac{b}{(a - b)}\left(e^{(a-b)t} - 1\right)$$

and

$$B := \int_0^t e^{(a-b)(t-s)}d\hat{B}_s.$$

This gives:

$$\begin{aligned}\mathrm{Var}(F_t(\omega)) = &\ \mathbb{E}\Big(\Big(F_t(\omega) - \mathbb{E}(F_t(\omega))^2 = \mathbb{E}\Big((A+B)^2 - 2(A+B)A + A^2\Big)\\ &= \mathbb{E}\Big(A^2 + 2AB + B^2 - 2A^2 - 2AB + A^2\Big) = \mathbb{E}\Big(B^2\Big)\\ &= \mathbb{E}\left(\left(\int_0^t e^{(a-b)(t-s)} d\hat{B}_s\right)^2\right).\end{aligned}$$

The Itô isometry (see p. 26 of Øksendal (2005)) gives:

$$\mathbb{E}\left(\left(\int_0^t e^{(a-b)(t-s)} d\hat{B}_s\right)^2\right) = \mathbb{E}\left(\int_0^t \left(e^{(a-b)(t-s)}\right)^2 ds\right)$$

$$= \int_0^t e^{2(a-b)(t-s)} ds.$$

We can solve $\int_0^t e^{2(a-b)(t-s)} ds$ explicitly by calculating its antiderivative,

$$\int_0^t e^{2(a-b)(t-s)} ds = d_0 + \frac{1}{2(a-b)} \left(e^{2(a-b)t} - 1\right).$$

Hence, we recover the expectation for $F_t(\omega)$,

$$\mathbb{E}(F_t(\omega)) = F_0(\omega)e^{(a-b)t} + \frac{b}{(a-b)}\left(e^{(a-b)t} - 1\right), \tag{11}$$

and the corresponding variance (integration constant d_0 set to zero),

$$\mathrm{Var}(F_t(\omega)) = \frac{1}{2(a-b)}\left(e^{2(a-b)t} - 1\right). \tag{12}$$

2.2. The parameters a and b

We note that

$$0 < \mathbb{E}(F_t(\omega)) = F_0(\omega)e^{(a-b)t} + \frac{b}{(a-b)}\left(e^{(a-b)t} - 1\right) < 1. \tag{13}$$

For $t = 0$ we see from condition (13) that $0 < F_0(\omega) < 1$. For $(a-b) > 0, e^{(a-b)t}$ approaches infinity so this condition is not reasonable. We are therefore left with the condition $(a-b) \leq 0$. Since $0 < F_0 < 1$ we get

$$0 < F_0(\omega)e^{(a-b)t} + \frac{b}{(a-b)}\left(e^{(a-b)t} - 1\right) < 1$$

Letting $t \to \infty$ we see that

$$0 < \frac{b}{b-a} < 1$$

which implies that $b > 0$ and that $a < 0$. For $a = b$ the bounded variation term $A(t)$ in Eq. (9) collapses into a linear equation:

$$\begin{aligned}&A(t) = \tfrac{b}{a-b}\left(e^{(a-b)t} - 1\right)\\ &= \tfrac{b}{a-b}\left(1 + (a-b)t + \tfrac{(a-b)^2t^2}{2!} + \cdots + \tfrac{(a-b)^n t^n}{n!} + \cdots - 1\right)\\ &= b\left(\tfrac{1}{a-b} + t + \tfrac{(a-b)^1 t^2}{2!} + \cdots + \tfrac{(a-b)^{n-1}t^n}{n!} + \cdots - \tfrac{1}{a-b}\right)\\ &= b\left(t + \tfrac{(a-b)^1t^2}{2!} + \cdots + \tfrac{(a-b)^{n-1}t^n}{n!} + \cdots\right)\\ &= bt\end{aligned} \tag{14}$$

We will henceforth assume that $F_0 > 0$ and $(a-b) < 0$.

2.3. The Brownian motion term

We use Gaussian white noise to model perturbations in the AT → GC (a) and GC → AT (b) mutation rates. We also allow for scaling of $c > 0$, as mentioned above. The scale can be determined by factors such as species/strain, environment, host and presence of MMR genes. The Brownian motion term,

$$\int_0^t e^{(a-b)(t-s)} d\hat{B}_s, \tag{15}$$

depends on the parameters a and b as well as on the duration of the time period. Since we assume that $(a-b) < 0$, (15) approaches 0 as $t \to \infty$ and Brownian motion $\hat{B}_t(\omega)$ for $a = b$. For $(a-b) < 0$ it can be seen that (15) increases as $s \to t$.

We can reach the same conclusion by examining the variance of $F_t(\omega)$ (described in (12) above). The Brownian motion is assumed to have mean $\mu = 0$ and variance $\mathbb{E}\left(\hat{B}_t^2(\omega)\right) = t$. Thus, the variance of Brownian motion is in general expected to increase with time t. Since there is no simple way to calculate (15) analytically, we do so numerically:

$$\int_0^t e^{(a-b)(t-s)} d\hat{B}_s = \sum_{s_0}^{s_N} e^{(a-b)(t-s_i)}\left(\widehat{W}_{s_{i+1}}(\omega) - \widehat{W}_{s_i}(\omega)\right)\Delta s_i, \tag{16}$$

where $\hat{W}_s(\omega)$ is c-scaled white noise, $\Delta s_i = s_{i+1} - s_i$, and $s_0 = 0, \ldots, s_i = t_i, \ldots, s_N = t$.

2.4. The Girsanov transform

Eq. (7) can be written as

$$dF_t(\omega) = ((a-b)F_t(\omega) + b)dt + d\hat{B}_t(\omega).$$

Since we know from (9) that

$$F_t(\omega) = -\frac{b}{(a-b)} + \left(F_0(\omega) + \frac{b}{(a-b)}\right)e^{(a-b)t} + \int_0^t e^{(a-b)(t-s)} d\hat{B}_s$$

is a semimartingale, if we let

$$Z_t(\omega) = ((a-b)F_t(\omega) + b),$$

we can write

$$dF_t(\omega) = Z_t(\omega)dt + d\widehat{B}_t(\omega).$$

The Girsanov theorem allows us to compute the Radon-Nikodym derivative (see ch. 3, p. 143 of Protter (2005)) of a measure Q with respect to the probability measure P as follows:

$$\frac{dQ}{dP} = exp\left(-\int_0^t Z_s(\omega)d\widehat{B}_s - \frac{1}{2}\int_0^t Z_s^2(\omega)ds\right).$$

This means that $F_t(\omega)$ is a Brownian motion under the measure Q, since we assume that $(a-b) < 0$ which implies that Kazamaki's (and hence Novikov's condition) apply $\forall t$ (see chs. 4 and 8 of Øksendal (2005)).

2.5. Further generalizations

The model describing SNP GC content can be made more general if we assume that the parameters a and b are functions. It is important to note that if a and b are functions with respect to time, obtaining an analytical solution may be impossible. While up to this point we have assumed that variation in the model is described by a white noise process, a more complicated noise term X_t could also be used. For instance, if we let

$$\frac{dF_t(\omega)}{dt} = (a + X_t(\omega))F_t(\omega) + (b + X_t(\omega))(1 - F_t(\omega)),$$

we have

$$\frac{dF_t(\omega)}{dt} = aF_t(\omega) + X_t(\omega)F_t(\omega) + b + X_t(\omega) - (b + X_t(\omega))F_t(\omega).$$

This reduces to

$$\frac{dF_t(\omega)}{dt} = (a-b)F_t(\omega) + b + X_t(\omega),$$

where

$$X_t(\omega) = \theta(t,\omega) + \kappa(t,\omega)\hat{W}_t(\omega).$$

Thus,

$$\frac{dF_t(\omega)}{dt} = aF_t(\omega) + b(1-F_t(\omega)) + \Big(\theta(t,\omega) + \kappa(t,\omega)\hat{W}_t(\omega)\Big),$$

and after rearranging:

$$\begin{aligned} dF_t(\omega) &= ((a-b)F_t(\omega) + \theta(t,\omega) + b)dt + \kappa(t,\omega)dB_t(\omega). \\ &= be^{(-(a-b)t)}dt + e^{(-(a-b)t)}d\hat{B}_t \end{aligned} \tag{17}$$

We could, for instance, let $X_t(\omega)$ be a mean-reverting Ornstein–Uhlenbeck process, *i.e.*

$$\frac{dX_t(\omega)}{dt} = GC_0 - F_t(\omega) + \hat{W}_t(\omega).$$

Hence, we let $\theta(t,\omega) = GC_0 - F_t(\omega)$ and $\kappa(t,\omega) = 1$. Plugging these into (17), we see

$$dF_t(\omega) = ((a-b)F_t(\omega) + GC_0 + b)dt + d\hat{B}_t(\omega). \tag{18}$$

We can now use the integrating factor $g(t, F_t(\omega)) = Y_t(\omega) = e^{(-(a-b)t)}F_t(\omega)$ to solve (18) in a similar fashion to (5).

3. Results

3.1. Genomic evolution in microbial symbionts

Vertically transmitted symbionts tend to have reduced mismatch and repair (MMR) genes (Moran et al., 1996). This could be mediated by selection for AT nucleotides that are involved in ATP production as they are the least costly nucleotides in terms of energy (Chen et al., 2016; Batut et al., 2014). Purines A and G always bind to pyrimidines C and T and therefore A and T, as well as G and C, increase or decrease at similar rates, *i.e.* genomic % AT content = 100-% GC content. These relations are known as Chargaff's parity laws (Elson and Chargaff, 1954). In the present study, we are primarily concerned with modeling change in GC content with respect to time as fractions of genomic GC and AT content, respectively:

$$F_{t+\Delta t}(\omega) - F_t(\omega) = \alpha F_t(\omega)\Delta t + \beta(1-F_t(\omega))\Delta t \tag{19}$$

The change in $F_{t+\Delta t}(\omega) - F_t(\omega)$ can be considered as the SNP GC content in a colony during time Δt, which may be a very long time. Nevertheless, we let $\Delta t \to 0$ and write the equation out as a differential equation:

$$\frac{dF_t(\omega)}{dt} = \alpha F_t(\omega) + \beta(1-F_t(\omega)), \tag{20}$$

We allow the GC and AT mutation rates, $\alpha F_t(\omega)$ and $\beta(1-F_t(\omega))$ respectively, to be subjected to random perturbations, *i.e.* $\alpha = a + W_t(\omega)$ and $\beta = b + W_t(\omega)$, where $W_t(\omega)$ is Gaussian white noise with respect to every trajectory $\omega \in \Omega$. After some algebra (see Methods section) we get:

$$\frac{dF_t(\omega)}{dt} = aF_t(\omega) + b(1-F_t(\omega)) + W_t(\omega), \tag{21}$$

where $\frac{dF_t(\omega)}{dt}$ is the change in GC content (alternatively the SNP GC content) at any time during $\Delta t \to dt$. In practice this can be accomplished by carrying out metagenomic sequencing focusing on a particular strain of a specific symbiont species in a host, or from an environment, at a given time, subsequently performing SNP calling and, finally, assessing the GC content of the SNPs. This has been demonstrated recently from an mutation accumulation experiment, albeit for a shorter time span, in the facultative symbiont *Teredinibacter turnerae* (Senra et al., 2018). If metagenomic sampling and sequencing is carried out within shorter periods of time, as is the case in the previously mentioned study (Senra et al., 2018), many of the called SNPs will only be spurious mutations that will be purged by purifying selection (Castillo-Ramirez et al., 2011). In the present work, however, we are concerned with base substitutions (Bohlin et al., 2017) that are selected for and that comply with Chargaff's parity rules. Hence, the sampling frequency will typically be centuries and millennia.

Reduction or loss of the efficiency of the symbionts' MMR system will result in cumulated mutations since homologous recombination among intracellular, vertical symbionts is rare (McCutcheon and Moran, 2012). Since increased AT-content is associated with dysfunctional or loss of MMR genes genome reduction, certainly pseudogenization, may already be an ongoing process (Klasson, 2017) as the first genes lost are, unlike MMR genes, typically those least conserved within a species (Bolotin and Hershberg, 2016). Hence, it is only after a vertical symbiotic relationship has been established with the host that a dramatic drop in genomic GC content seems likely to occur (McCutcheon and Moran, 2012). If we assume that the host-symbiont relationship has become stable we can let a and b respectively represent the fraction of AT → GC and GC → AT base substitutions to be fixed parameters. In a recent study (Bohlin et al., 2018), we found that a similar model with fixed parameters (See Eq. (5) in Methods section), describing changes in core genome SNP GC content with respect to core genome GC content for 35 different species/core genomes (716 genomes in total), fitted empirical genomic data remarkably well regardless of the fact that most species were distantly related (See Fig. 2 in Bohlin et al. (2018)). Indeed, it has been shown that mutation rates may be quite similar amongst species with the same effective population size N_e (Lynch et al., 2016). Although the interpretation of the present model is somewhat different than the model described by Eq. (5) it is concerned with how the population of only one species evolve over time. We therefore believe that it is not unreasonable that the mutation parameters a and b are constant, although the model can describe the mutation rates a and b as respective functions $a(t)$ and $b(t)$ as it is argued in Section 2.5.

The effective population size N_e is typically small for microbial symbionts, reducing the effect of genome streamlining, leading to further cumulation of mutations (Lynch et al., 2016). Moreover, if the symbionts MMR system is deficient, or lost, the cumulated mutations will typically lead to increased protein evolution as well as biased genomic amino acid content due to the increasingly AT-biased base composition (Wernegreen, 2015). Indeed, genes related to protein folding, such as chaperones, are typically highly expressed in symbionts compared to other bacteria (McCutcheon and Moran, 2012). Fitness decreasing or lethal mutations must be purged by purifying selection otherwise Muller's ratchet sets in leading eventually to extinction (Moran et al., 1996). We mainly interpret an uncontrollable increase in the variance of the mutation rates, as can be observed in Fig. 1 and described by the Brownian motion term in the presented model (see Section 2.3 for more details)

$$\int_0^t e^{(a-b)(t-s)}d\hat{B}_s, \tag{22}$$

as the onset of Muller's rachet since this process is a consequence of cumulation of fitness decreasing and or deleterious mutations. However, it could also, potentially, represent a speciation event (Campbell et al., 2015). The mutation rate parameters a and b of a species influence the Brownian motion term. The scaling c that determines the variance of the Brownian motion term discussed in Section 2.1, not unlike the concept of 'quasi species' described

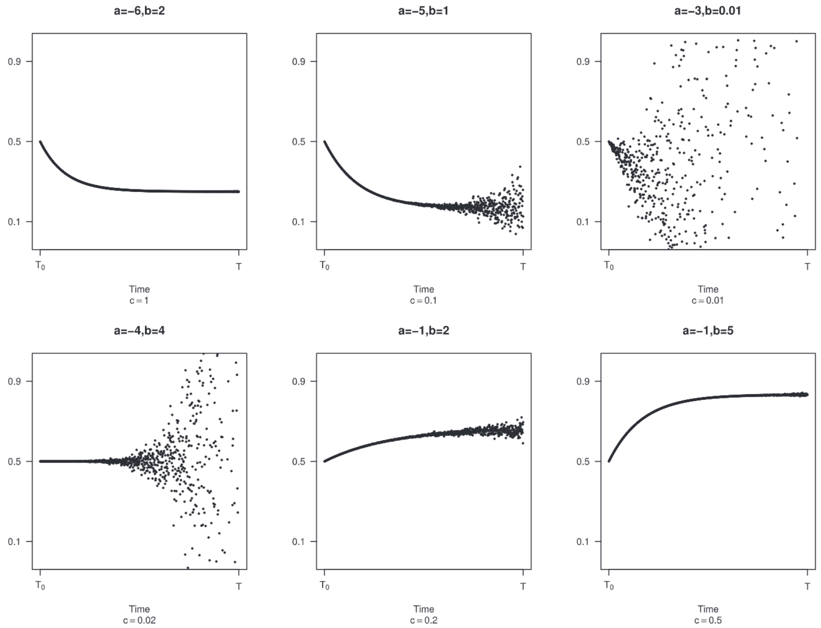

Fig. 1. The model (9) with different combinations of parameters a, b and Brownian scaling coefficient c, all starting at $F_0 = 0.5$. The vertical axis describes SNP GC content, while the horizontal axis describes time t from $T_0 = 0$ to $T = 1$.

for virus (Eigen, 1993), will likely also have an impact on the evolutionary history of the species. Thus, changes to mutation rate parameters a and/or b, as well as the variance/scaling c, could have dramatic implications for the evolutionary history of the species (see Fig. 1) something that has recently been shown for the spittlebug endosymbiont *Candidatus* Zinderia insecticola (*Zinderia*) (Koga et al., 2013). The symbiont has an extremely low GC content of only 13.5% and a genome size of 208 kilobases. There are indications that *Zinderia* is about to reach the drift-barrier (Lynch et al., 2016) however as it has been demonstrated that several of its hosts is replacing it with another symbiont (Koga et al., 2013).

While Gaussian white noise has been used to describe the random perturbations of the mutation rates it can be seen in Section 2.5 that more advanced models for instance, a mean-reverting Ornstein-Uhlenbech process to account for GC content 'intertia' in the symbiont or, alternatively, in the host, can be used instead but at the cost of more complicated calculations. Changing the scale of Brownian motion alters its variance but it is still a Brownian motion as can be seen in Section 2.1. In the Methods Section 2.4 it is shown that when the mutation rate parameters a and b are constants the solution to the stochastic differential Eq. (9) is essentially a Brownian motion, if looked at from a different perspective, *i.e.* the Girsanov transform.

4. Discussion

Randomness is widespread in genomics; it has been stated that mutations are the engine of evolution and mutations are, by most accounts, random (Hershberg, 2015). Eq. (9) describes a model for genomic GC content in prokaryotes which takes random variations of mutation rates into account. Using Itô calculus we see that the resulting model handles stochastic fluctuations in a natural manner; it's not merely a deterministic model with added 'noise' but a model that incorporates random events naturally. In the present study, the $AT \to GC$ and $GC \to AT$ mutation rates directly influence how the Brownian motion term evolves with time. In addition, the variance of the Brownian motion term can be set independently from the deterministic part of the model. From the previous Section 3, we can see (Fig. 1) that the Brownian motion term contains information not obtainable with traditional differential equation models, even if these models were to add random 'noise'. By analyzing the Brownian motion term we can se how random perturbations of the mutation rates increasingly influence the outcome over time eventually leading to dramatic consequences such as speciation or extinction (Campbell et al., 2015).

The model described here is based on a previous model describing core genome SNP GC content with respect to core genome GC content (Bohlin et al., 2018; Bohlin et al., 2019). However, the model has been altered to describe change in genomic GC content over time for a particular symbiont species and allowing for random perturbations of the AT/GC mutation rates. The solution to the differential Eq. (9) is a function describing genomic GC content in a symbiont species. The use of Itô calculus has provided us with novel insights into how genomic evolution in microbial symbionts may progress that would have been impossible using the previously described models based on ordinary differential equations.

One of the striking results from the Itô calculus based model is that a microbial symbiont may have been set for dramatic events, such as speciation or extinction, long before any signs of such events appear. Indeed, the path to extinction may already have been determined as a mathematical and statistical consequence at the time the symbiont entered a stable relationship with its host although the event would not occur for thousands or even million years before any signs of it. While it is difficult to present evidence of microbial symbionts going through all the phases of extinction described by our model there have recently been published several examples of different genome evolution in same species symbionts that certainly give some support to what our model indicates (Campbell et al., 2015; Bennett and Moran, 2015; Santos-Garcia et al., 2017). Another interesting consequence of our presented model that has some support in the scientific literature is that an event that alters the mutation rate parameters of the symbiont, for instance the introduction of an additional symbiont species to the host (Van Leuven et al., 2014), may also influence the fate of the first symbiont (Mao et al., 2018).

Intracellular pathogens do not appear to engage in symbiotic relationships with a host, most likely due to the increased constraints of a pathogen–host relationship (Weinert and Welch, 2017). Although these pathogens may undergo genome reduction, they do not seem to experience the same dramatic gene loss observed in some symbionts (Moran and Bennett, 2014; Wernegreen, 2015). It is not uncommon, however, for the genomic base composition of intracellular pathogens to be AT-biased but less so than what is observed for microbial symbionts (Weinert and Welch, 2017).

There appear to be some similarities between the evolutionary mechanisms of symbionts and those of free-living bacteria that undergo changes in environment even if not through attachment to a host (Batut et al., 2014; Klasson, 2017). There are only a few documented examples of free-living bacteria that experience genome reduction. One of these is the cyanobacterium *Prochlorococcus* spp. (Martínez-Cano et al., 2015; Batut et al., 2014), whose high-light ecotypes living close to the water surface are more AT-rich and have smaller genomes than the low-light ecotypes living at greater depths (Batut et al., 2014). Microbial organisms in the same environments often acquire the same nucleotide biases if enough time is allowed to pass (Reichenberger et al., 2015). Such environmental signatures become particularly evident in SNPs since, as discussed above and in Section 1, these polymorphisms arise as a consequence of natural selection regulated by the environment (Foerstner et al., 2005).

5. Conclusions

We have presented a mathematical model that describes the evolution of genomic GC content in a microbial symbiont over time. This was modeled using a stochastic differential equation where the difference in GC content with respect to time (alternatively, SNP GC content) was equal to parameter multiples a and b of genomic GC- and AT content, respectively. The model contains a stochastic term that indicates that minuscule, random changes in mutation rates early on can lead to abrupt, fluctuations in genomic GC content considerably later.

In the model, the variance of the mutation rates, as described by a Brownian motion term, must be kept low to avoid genomic base composition spiraling out of control, which becomes progressively harder as time passes. The model also indicates that differing mutation rate parameters a, b as well as the variance c in the Brownian motion term, could lead to severe consequences for the evolutionary history of microbial symbionts. Furthermore, the model demonstrates that a microbial symbiont may have been destined for speciation events and/or the onset of Muller's ratchet, long before any variation in the genomic base composition could have been detected, as a consequence of the AT/GC mutation rates.

Our model, based on the use of stochastic differential equations that allow for seamless integration of certain random processes, revealed that the evolution of genomic base composition in microbial symbionts could differ dramatically from what a non-stochastic model describes (Bohlin et al., 2019). Although conceptually simple, the model demonstrated here provides novel insight into stochastic evolutionary processes with mathematical rigor.

CRediT authorship contribution statement

Jon Bohlin: Conceptualization, Methodology, Formal analysis, Writing - review & editing. **Brittany Rose:** Writing - original draft. **Ola Brynildsrud:** Formal analysis. **Birgitte Freiesleben De Blasio:** Validation, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The figure was generated using the free statistical package R (Core Team, 2018). We thank Håkon Gjessing for helpful comments and suggestions. In particular, we extend our gratitude to one of the reviewers whose comments and suggestions improved the manuscript substantially.

References

Agashe, D., Shankar, N., 2014. The evolution of bacterial DNA base composition. J. Exp. Zool. Part B: Mol. Develop. Evol. 322 (7), 517–528.

Batut, B., Knibbe, C., Marais, G., Daubin, V., 2014. Reductive genome evolution at both ends of the bacterial population size spectrum. Nat. Rev. Microbiol. 12 (12), 841.

Bennett, G.M., Moran, N.A., 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. Proc. Nat. Acad. Sci. 112 (33), 10169–10176.

Bentley, S.D., Parkhill, J., 2004. Comparative genomic structure of prokaryotes. Annu. Rev. Genet. 38, 771–791.

Bobay, L.M., Ochman, H., 2017. Impact of recombination on the base composition of bacteria and archaea. Mol. Biol. Evol. 34 (10), 2627–2636.

Bohlin, J., Sekse, C., Skjerve, E., Brynildsrud, O., 2014. Positive correlations between genomic % AT and genome size within strains of bacterial species. Environ. Microbiol. Rep. 6 (3), 278–286.

Bohlin, J., Eldholm, V., Pettersson, J.H., Brynildsrud, O., Snipen, L., 2017. The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. BMC Genom. 18 (1), 151.

Bohlin, J., Eldholm, V., Brynildsrud, O., Petterson, J.H.O., Alfsnes, K., 2018. Modeling of the GC content of the substituted bases in bacterial core genomes. BMC Genom. 19 (1), 589.

Bohlin, J., Rose, B., Petterson, J.H.O., 2019. Estimation of AT and GC content distributions of nucleotide substitution rates in bacterial core genomes. Big Data Anal 4 (5).

Bolotin, E., Hershberg, R., 2016. Bacterial intra-species gene loss occurs in a largely clocklike manner mostly within a pool of less conserved and constrained genes. Sci. Rep. 6, 35168.

Boscaro, V., Kolisko, M., Felletti, M., Vannini, C., Lynn, D.H., Keeling, P.J., 2017. Parallel genome reduction in symbionts descended from closely related free-living bacteria. Nat. Ecol. Evol. 1 (8), 1160.

Campbell, M.A., Van Leuven, J.T., Meister, R.C., Carey, K.M., Simon, C., McCutcheon, J.P., 2015. Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont Hodgkinia. Proc. Nat. Acad. Sci. 112 (33), 10192–10199.

Castillo-Ramirez, S., Harris, S.R., Holden, M.T., He, M., Parkhill, J., Bentley, S.D., Feil, E.J., 2011. The impact of recombination on dN/dS within recently emerged bacterial clones. PLoS Pathogens 7 (7).

Chen, W.H., Lu, G., Bork, P., Hu, S., Lercher, M.J., 2016. Energy efficiency trade-offs drive nucleotide usage in transcribed regions. Nat. Commun. 7 (1), 1–10.

Eigen, M., 1993. Viral quasispecies. Sci. Am. 269 (1), 42–49.

Elson, D., Chargaff, E., 1954. Regularities in the composition of pentose nucleic acids. Nature 173 (4413), 1037–1038.

Fisher, R.M., Henry, L.M., Kiers, E.T., West, S.A., 2017. The evolution of host-symbiont dependence. Nat. Commun. 8, 15973.

Foerstner, K.U., Von Mering, C., Hooper, S.D., Bork, P., 2005. Environments shape the nucleotide composition of genomes. EMBO Rep. 6 (12), 1208–1213.

Hershberg, R., Petrov, D.A., 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 6, (9) e1001115.

Hershberg, R., 2015. Mutation–the engine of evolution: studying mutation and its role in the evolution of bacteria. Cold Spring Harbor perspectives in Biology 7, (9) a018077.

Hildebrand, F., Meyer, A., Eyre-Walker, A., 2010. Evidence of selection upon genomic GC-content in bacteria. PLoS Genet. 6 (9), e1001107.

Hosokawa, T., Ishii, Y., Nikoh, N., Fujie, M., Satoh, N., Fukatsu, T., 2016. Obligate bacterial mutualists evolving from environmental bacteria in natural insect populations. Nat. Microbiol. 1 (1), 15011.

Kimura, M.A., 1980. simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16 (2), 111–120.

Klasson, L., 2017. The unpredictable road to reduction. Nat. Ecol. Evol. 1 (8), 1062.

Koga, R., Bennett, G.M., Cryan, J.R., Moran, N.A., 2013. Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. Environ. Microbiol. 15 (7), 2073–2081.

Lane, N., Martin, W., 2010. The energetics of genome complexity. Nature 467 (7318), 929.

Lind, P.A., Andersson, D.I., 2008. Whole-genome mutational biases in bacteria. Proc. Nat. Acad. Sci. 105 (46), 17878–17883.

Lynch, M., Ackerman, M.S., Gout, J.F., Long, H., Sung, W., Thomas, W.K., Foster, P.L., 2016. Genetic drift, selection and the evolution of the mutation rate. Nat. Rev. Genet. 17 (11), 704.

Mao, M., Yang, X., Bennett, G.M., 2018. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. Proc. Nat. Acad. Sci. 115 (50), E11691–E11700.

Martínez-Cano, D.J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L.P., Latorre, A., Moya, A., Delaye, L., 2015. Evolution of small prokaryotic genomes. Front. Microbiol. 5, 742.

McCutcheon, J.P., Moran, N.A., 2012. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10 (1), 13.

Moran, N.A., 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Nat. Acad. Sci. 93 (7), 2873–2878.

Moran, N.A., Bennett, G.M., 2014. The tiniest tiny genomes. Annu. Rev. Microbiol. 68, 195–215.

Protter, P.E., 2005. Stochastic differential equations, Springer, Berlin, Heidelberg..

Core Team, R., 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raghavan, R., Kelkar, Y.D., Ochman, H., 1450. A selective force favoring increased G+C content in bacterial genes. Proc. Nat. Acad. Sci. 109 (36), 14504–14507.

Reichenberger, E.R., Rosen, G., Hershberg, U., Hershberg, R., 2015. Prokaryotic nucleotide composition is shaped by both phylogeny and the environment. Genome Biol. Evol. 7 (5), 1380–1389.

Rocha, E.P., Feil, E.J., 2010. Mutational patterns cannot explain genome composition: are there any neutral sites in the genomes of bacteria? PLoS Genet. 6, (9) e1001104.

Sabater-Muñoz, B., Toft, C., Alvarez-Ponce, D., Fares, M.A., 2017. Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. ISME J. 11 (6), 1291–1304.

Santos-Garcia, D., Silva, F.J., Morin, S., Dettner, K., Kuechler, S.M., 2017. The all-rounder Sodalis: a new bacteriome-associated endosymbiont of the lygaeoid bug Henestaris halophilus (Heteroptera: Henestarinae) and a critical examination of its evolution. Genome Biol. Evol. 9 (10), 2893–2910.

Senra, M.V., Sung, W., Ackerman, M., Miller, S.F., Lynch, M., Soares, C.A.G., 2018. An unbiased genome-wide view of the mutation rate and spectrum of the endosymbiotic bacterium Teredinibacter turnerae. Genome Biol. Evol. 10 (3), 723–730.

Seward, E.A., Kelly, S., 2016. Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms. Genome Biol. 17 (1), 226.

Van Leuven, J.T., Meister, R.C., Simon, C., McCutcheon, J.P., 2014. Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. Cell 158 (6), 1270–1280.

Weinert, L.A., Welch, J.J., 2017. Why might bacterial pathogens have small genomes? Trends Ecol. Evol. 32 (12), 936–947.

Wernegreen, J.J., 2015. Endosymbiont evolution: predictions from theory and surprises from genomes. Ann. N. Y. Acad. Sci. 1360 (1), 16–35.

Wernegreen, J.J., 2017. In it for the long haul: evolutionary consequences of persistent endosymbiosis. Curr. Opin. Genet. Develop. 47, 83–90.

Øksendal, B., 2005. Stochastic differential equations. Springer, Berlin, Heidelberg (2005).