

# An Unbiased Genome-Wide View of the Mutation Rate and Spectrum of the Endosymbiotic Bacterium *Teredinibacter turnerae*

Marcus V.X. Senra<sup>1,†</sup>, Way Sung<sup>2,†</sup>, Matthew Ackerman<sup>3</sup>, Samuel F. Miller<sup>3</sup>, Michael Lynch<sup>3</sup>, and Carlos Augusto G. Soares<sup>4,\*</sup>

<sup>1</sup>Departamento de Zoologia, Universidade Federal de Juiz de Fora, Brazil

<sup>2</sup>Department of Bioinformatics and Genomics, University of North Carolina, Charlotte

<sup>3</sup>Biodesign Center for Mechanisms of Evolution, Arizona State University

<sup>4</sup>Departamento de Genética, Universidade Federal do Rio de Janeiro, Brazil

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding author: E-mail: soares@biologia.ufrj.br.

Accepted: February 2, 2018

**Data deposition:** Illumina DNA sequences of all *Teredinibacter turnerae* MA lines generated in this work are deposited under the Bioproject PRJNA427401 at the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>).

## Abstract

Mutations contribute to genetic variation in all living systems. Thus, precise estimates of mutation rates and spectra across a diversity of organisms are required for a full comprehension of evolution. Here, a mutation-accumulation (MA) assay was carried out on the endosymbiotic bacterium *Teredinibacter turnerae*. After ~3,025 generations, base-pair substitutions (BPSs) and insertion–deletion (indel) events were characterized by whole-genome sequencing analysis of 47 independent MA lines, yielding a BPS rate of  $1.14 \times 10^{-9}$  per site per generation and indel rate of  $1.55 \times 10^{-10}$  events per site per generation, which are among the highest within free-living and facultative intracellular bacteria. As in other endosymbionts, a significant bias of BPSs toward A/T and an excess of deletion mutations over insertion mutations are observed for these MA lines. However, even with a deletion bias, the genome remains relatively large (~5.2 Mb) for an endosymbiotic bacterium. The estimate of the effective population size ( $N_e$ ) in *T. turnerae* is quite high and comparable to free-living bacteria ( $\sim 4.5 \times 10^7$ ), suggesting that the heavy bottlenecks associated with many endosymbiotic relationships is not prevalent during the life of this endosymbiont. The efficiency of selection scales with increasing  $N_e$  and such strong selection may have been operating against the deletion bias, preventing genome erosion. The observed mutation rate in this endosymbiont is of the same order of magnitude of those with similar  $N_e$ , consistent with the idea that population size is a primary determinant of mutation-rate evolution within endosymbionts, and that not all endosymbionts have low  $N_e$ .

**Key words:** mutation-accumulation (MA) assay, endosymbiosis, *Teredinibacter turnerae*, drift-barrier hypothesis.

## Introduction

Spontaneous mutations contribute largely to the input of genetic variation into living systems, and compose a major force in driving the evolutionary process. Accurate estimates of the spontaneous rate and spectrum of mutations across a large number of species are required to create a better comprehension of evolutionary patterns (Lynch et al. 2016). The least biased approach for mutation-rate estimation is the

mutation-accumulation (MA) assay (Denver 2009; Lee et al. 2012; Sung et al. 2012). Initially proposed by Mukai (1964), an MA experiment uses continuous individual passages of several lineages derived from a single ancestral colony. The reduction in effective population size,  $N_e$ , reduces the efficiency of selection on spontaneous mutations, allowing for the accumulation of all but the most deleterious mutations. Genome-wide sequencing of multiple MA lines has generated

unbiased estimates of the rate and spectrum of spontaneous mutations across eukaryotes and bacteria (Lynch et al. 2016). Mutation rates vary  $\sim 1,000$ -fold across species ( $\sim 10^{-7}$ – $10^{-10}$  mutations per site per generation) and the evolution of mutation rate has been linked to the population size of the organisms (Lynch 2010a; Sung et al. 2012, 2016). Specifically, the effective population size determines the power of random genetic drift, which defines the lower limit to which selection can promote reduction of deleterious mutation rates (Lynch 2008, 2010b; Sung et al. 2012; Lynch et al. 2016).

Although the number of MA genome-wide sequencing experiments has increased dramatically in the last few years, these experiments have been restricted to free-living model organisms. Endosymbiotic bacteria that are phylogenetically diverse (Moya et al. 2008), ecologically ubiquitous, and central to host evolution (Margulis and Bermudes 1985; Schink 1997; Minic and Hervé 2004; Zientz et al. 2004; Croft et al. 2005; Stewart et al. 2005; Kneip et al. 2007) have not been subject to such assays, mostly because of methodological reasons. As consequence of their long-term and intimate intracellular associations, endosymbionts are thought to display elevated mutation rates in comparison with their free-living close relatives (Moran 1996; Itoh et al. 2002; Woolfit and Bromham 2003) and eroded genomes (often  $< 1$  Mb coding for  $\sim 300$  genes), with the loss of essential cell functions forcing most of them to rely on host cells for survival (McCutcheon and Moran 2011). This pattern of evolution is thought to start just after the establishment of the host-restricted association (Moran 1996). Such a lifestyle can force the bacterium through constant population bottlenecks (between-generation host transmissions) and drive a reduction of the  $N_e$  of the symbiont (now limited to the host abundance). Under this view, the increased power of genetic drift (which is inversely related to  $N_e$ ) reduces the efficiency of purifying selection at removing slightly deleterious mutations, leading to a deleterious pattern of genome evolution (Lynch et al. 2016) through processes associated with Muller's ratchet (Moran 1996).

Here, we have performed an MA experiment on the endosymbiotic bacterium *Teredinibacter turnerae*. This cellulolytic and nitrogen-fixing  $\gamma$ -proteobacterium colonizes specialized structures called glands of Deshayes in both demibranchs of mollusks of the family Teredinidae (Bivalve: Pholadoidea). *Teredinibacter turnerae* supplies cellulolytic enzymes and nitrogen compounds (Carpenter and Culliney 1975; Trytek and Allen 1980; Gallager et al. 1981; Waterbury et al. 1983; Distel et al. 1991, 2002; Lechene et al. 2007) that allow their hosts to feed on wooden material during their juvenile and adulthood stages. The complete genome sequencing of this endosymbiont has revealed that  $\sim 7\%$  of its genome is devoted to proteins involved in the biosynthesis of secondary metabolites (Yang et al. 2009), including antibiotics, suggesting that this bacterium is highly involved in the synthesis of bioactive metabolites required for the host survival (Trindade-Silva et al. 2009; Han et al. 2013;

Elshahawi et al. 2013). Yet, despite its intracellular life-style, this bacterium can be cultivated independent of its host (Waterbury et al. 1983), allowing an evaluation of the mutation pattern of this unique intracellular bacterium by MA assays.

## Materials and Methods

### MA Assay

The *T. turnerae* strain used in this work (CS30) was isolated from *Neoteredo reynei* (Teredinidae) sampled in a Mangrove in Rio de Janeiro, Brazil (Trindade-Silva et al. 2009). Starting from a single cell,  $\sim 100$  independent MA lines were derived. Every 2 days, a single individual colony from each MA line was isolated and transferred to a fresh solid BMS (Trindade-Silva et al. 2009) and incubated at  $30^\circ\text{C}$  over the course of the experiment. The imposed bottlenecking process ensures that mutations accumulate in an effectively neutral fashion, as demonstrated by Kibota and Lynch (1996). Generation time was estimated monthly using an entire colony from 12 randomly selected MA lines, transferred to  $1 \times$  PBS saline buffer, vortexed, serially diluted, and replated for CFU counting. The generation time for each MA line was calculated using the harmonic mean of the cell divisions per transfer over the course of the experiment. After an average of 3,025 generations, the MA assay was concluded and the genomic DNA of the MA lines were extracted using the wizard DNA extraction kit (Promega) to Illumina library standards.

### Sequencing and Alignment

101-bp paired-end Illumina (Illumina Hi-Seq platform) sequencing was applied to randomly selected 47 *T. turnerae* MA lines. The coverage depth of each MA line was  $\sim 100\times$  and the average library fragment size (distance between paired end reads) was  $\sim 175$  bp. The paired-end reads from each MA line were individually mapped against the reference genome *T. turnerae* T7901 (Yang et al. 2009; available at <https://www.ncbi.nlm.nih.gov>) using two different mappers: BWA v0.7.4 (Li and Durbin 2009) and NOVOALIGN v2.08.02 (available at [www.novocraft.com](http://www.novocraft.com)). The generated pileup files were converted to SAM format using SAMTOOLS v0.1.18 (Li 2011). We used in-house perl scripts to parse the alignments and to produce forward and reverse mapping information at each site, resulting in a configuration of 8 numbers for each line (A, a, C, c, G, g, T, t) corresponding to the number of reads mapped at each genomic position in the reference sequence. A separate file was also generated to display sites that had insertion–deletion (indels) mutation calls from the two alignment algorithms.

### Mutation Calling

Base-pair substitutions (BPSs) identification was carried out as previously described (Sung et al. 2015) using a consensus approach to identify putative mutations by comparing each

individual MA line (focal line) against the consensus of all the other MA lines. Indels were identified as in Lynch (2010a) and briefly, followed the criteria: 1) At each position, a consensus indel requires 30% of the reads spanning a position in a line to indicate the same indel (size and motif); 2) Each consensus indel requires a minimum of two forward and two reverse reads spanning the indel; 3) At a single site, any identical indel event that is shared across >50% of the lines is considered either a progenitor indel that existed prior to the initiation of the MA line, or a genome assembly error, and not included in the final indel list; 4) BWA and NOVOALIGN alignment algorithms must both independently identify the site as a putative indel. The use of two separate alignment algorithms ensures that algorithm-specific alignment errors are not responsible for false-positive mutation calls. Short-read mapping algorithms have difficulties mapping indel events >10 bp, so we fed the BWA and NOVOALIGN alignment output into PINDEL (Ye et al. 2009), a short-read realignment algorithm used to identify indels through paired-end information. Our criteria for PINDEL for indels included the following: 1) Each indel requires a minimum of six forward and six reverse reads indicating the indel, with a minimum of 20 reads overall supporting the indel call; and 2) any indel shared across >50% of the lines is considered a progenitor indel that existed prior to the initiation of the MA line, and excluded from the final indel list.

### Mutation Rate Calculation

The base-substitution (BPS) mutation rate per site per generation ( $\mu_{bs}$ ) for each MA line is estimated as equation (1):

$$\mu_{bs} = \frac{m}{nT}, \quad (1)$$

where  $m$  is the number of observed base substitutions,  $n$  is the number of nucleotide sites analyzed, and  $T$  is the total number of generations in the MA assay. The standard error for each MA line was calculated using equation (2):

$$SE_x = \sqrt{\frac{\mu_{bs}}{nT}}. \quad (2)$$

The total standard error of the BPS mutation rate is estimated by equation (3):

$$SE_{pooled} = \frac{s}{\sqrt{N}}, \quad (3)$$

where  $s$  represents the standard deviation of the mutation rates across all lines, and  $N$  is the number of lines analyzed.

The same calculation was used to calculate indel mutation rate, with  $\mu_{bs}$  replaced with  $\mu_{indel}$ .

### Effective Population Size

Under the assumption of neutrality, *T. turnerae*'s  $N_e$  was indirectly derived from the average nucleotide heterozygosity at

silent-sites from complete and draft genomes of *T. turnerae* strains publicly available in the GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) as described (Sung et al. 2012).

### Multinucleotide Mutations Probability Estimation

If we assume mutations are random with respect to their position, the chance of occurrence of one or more adjacent mutations in any given genome can be estimated. The expected probability of occurrence of multinucleotide mutations (MNMs) within window sizes of 20, 50, or 100 nt in the genome of *T. turnerae* MA lines, is the number of observed BPSs (779) divided by the genome size ( $5.19 \times 10^6$ ), and by the number of MA lines (47), and multiplied the corresponding window sizes (20, 50, or 100 nt). The chance of occurrence of one MNM in at least one window in a single MA line is then (Schridder et al. 2011):

$$1 - \text{cdf}(1, \lambda)^n, \quad (4)$$

where cdf is the cumulative distribution function for a Poisson process,  $\lambda$  is the per-window rate, and  $n$  is the number of windows in the genome. The probability to observe even one MNM in the genome of these MA lines was <0.01 for all tested window lengths (only data for 50 nt shown, [supplementary table S4, Supplementary Material online](#)).

## Results

### MA Assay

The *T. turnerae* strain CS30 used in this work was originally isolated from a *N. reynei* (Teredinidae, Bivalve) host sampled in a mangrove in Rio de Janeiro, Brazil (Trindade-Silva et al. 2009). 47 MA lines of *T. turnerae* underwent an average of 3,025 generations of parallel single-individual passages from the ancestral CS30 strain on solid basal medium with sucrose (BMS) at 30 °C (Trindade-Silva et al. 2009). By the end of this process, a number of MA lines were displaying smaller colony sizes, consistent with reduced fitness from the accumulation of deleterious load (Kibota and Lynch 1996). The *T. turnerae* MA lines were subjected to 101-bp paired-end high-throughput whole-genomic sequencing (Illumina Hi-Seq platform); the resulting reads were mapped against the *T. turnerae* T7901 (NC\_012997) reference genome; and mutations were identified as previously described (Sung et al. 2015). Spontaneous-mutation data are summarized in [table 1](#) (with further details in [supplementary tables S1–S3, Supplementary Material online](#)). As in other MA experiments with other organisms (Denver et al. 2012; Lee et al. 2012; Sung et al. 2012), the number of BPSs in coding regions ( $\chi^2 = 5.748$ ,  $df = 1$ ,  $P = 0.016$ ) and the expected ratio of non-synonymous to synonymous mutations ( $\chi^2 = 3.220$ ,  $df = 1$ ,  $P = 0.070$ ) are not significantly different from random expectations. Because a large proportion of nonsynonymous

**Table 1**

Summarized Mutation-Accumulation (MA) Assay Data

	MA Lines (n)	Gen.	BPSs	Indels	Total Mutation Events	BPSs Rate ( $\times 10^{-10}$ )/ Site/Generation	BPSs Rate ( $\times 10^{-3}$ )/ Genome Replication	Ts/Tv
Total	47	3,025	779	106	885	—	—	—
Ave.	—	—	16.57	2.26	18.83	11.4	5.59	3.48
Max	—	—	28	6	34	19.8	9.26	17
SEM	—	—	0.81	0.24	1.05	0.55	1.61	0.47

NOTE.—Ave., average; Max, maximum; SEM, standard error of the mean; Gen., number of generations; indels, insertion/deletion events; Ts/Tv, transition/transversion ratio.

mutations are expected to be removed by natural selection if effective, a random distribution of nonsynonymous and synonymous mutations indicates that selection played a minimal role on the mutation process during this MA experiment.

### Mutation Rate and Distribution of Mutations

A total of 885 spontaneous mutations accumulated within the genomes of the 47 tested *T. turnerae* MA lines, of which 779 (88.1%) are BPSs and 106 (11.9%) are indels (fig. 1A, table 1, supplementary tables S1–S3, Supplementary Material online). The genome-wide BPSs rate of  $1.14 \times 10^{-9}$  (standard error of the mean [SEM] = 0.55) per site per generation, or  $5.59 \times 10^{-3}$  (SEM = 1.61) mutations per genome per generation, and the indel rate  $1.55 \times 10^{-10}$  (SEM = 0.16) events per site per generation are among the highest within free living and facultative intracellular bacteria (fig. 1B). Although the BPSs are distributed randomly with respect to function, the distribution of mutations with respect to chromosomal position is not homogeneous. Similar to *Escherichia coli* (Foster et al. 2013), a peak of mutations is observed near the replication terminus located at chromosome position  $\sim 2.5$  Mb (Yang et al. 2009). As shown in supplementary figure S1, Supplementary Material online, a 50-kb bin containing the replication terminus of *T. turnerae* accumulated significantly more BPSs than the other bins elsewhere in the genome of this bacterium (one sample-*t* test = 36.50, df = 103,  $P = 3.15 \times 10^{-37}$ ; supplementary fig. S1, Supplementary Material online). Moreover, we noticed an excess of adjacent mutations (within single MA lines) corresponding to 21 BPSs and 16 indels (4.2% of the total de novo mutations) arising within <50 nucleotides from each other (supplementary table S4, Supplementary Material online). Under the assumption that mutations are randomly distributed across the genome, the probability of the occurrence of MNMs (Schridder et al. 2011), defined here as two or more mutations within windows of 50 nucleotides is expected to be extremely low ( $P = 1.96 \times 10^{-4}$ ). Yet, MNMs are observed in over a third (34.0%) of the MA lines. Most clustered mutations are found as pairs, with a single instance of three closely neighboring mutations in MA line 29 (supplementary table S4, Supplementary Material online). Our data do not allow us to determine whether these MNMs arose as a single event, but, if not, they had at least occurred in a short period of time

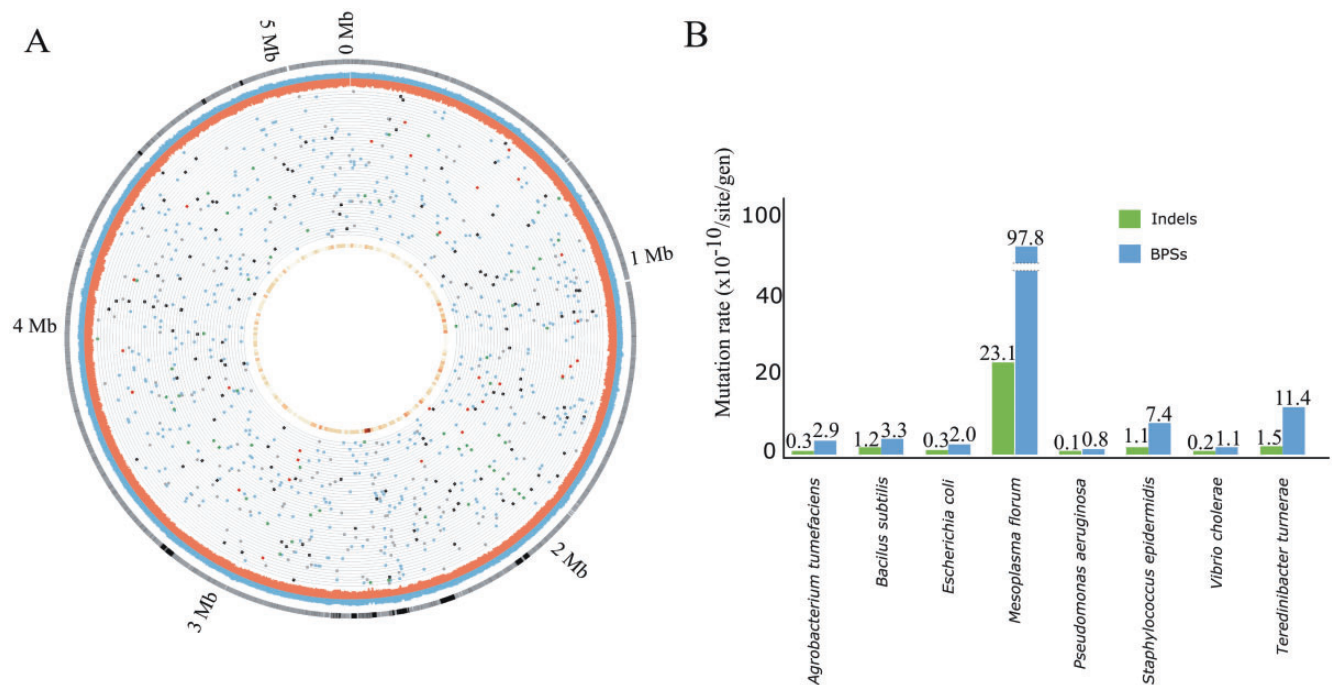
(during the course of this MA assay). These data are consistent with the idea that spontaneous mutations may not be independent with respect to position (Amos 2010; Schridder et al. 2011; Sung et al. 2015).

### Mutation Spectrum

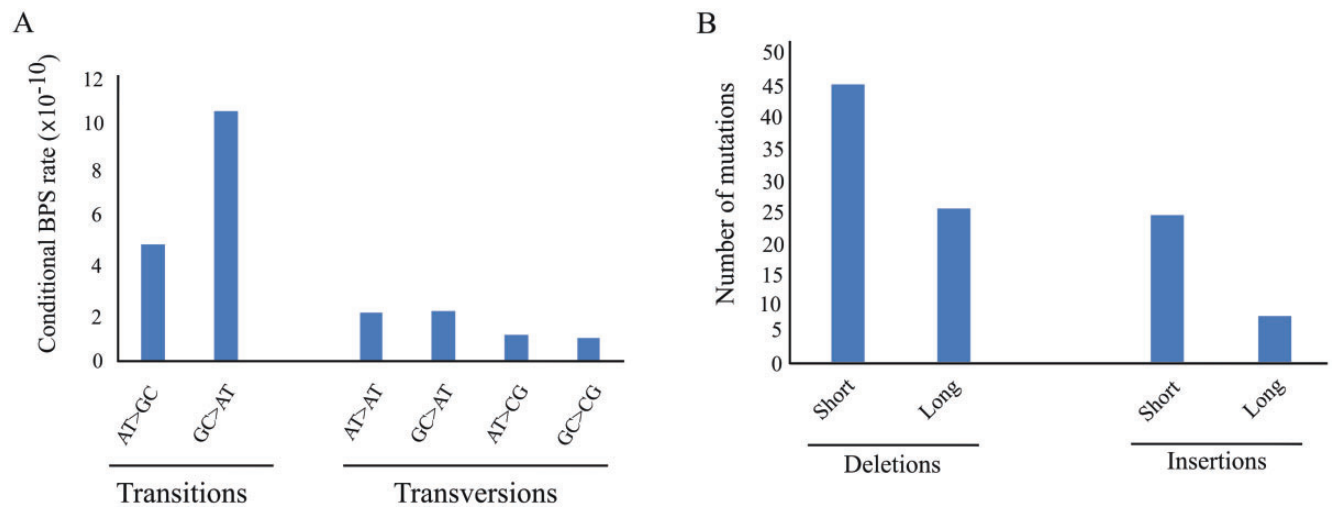
Of the total 779 BPSs, 71.6% (558/779) are transitions (supplementary table S1, Supplementary Material online), yielding a Transition/Transversion (Ts/Tv) BPS ratio of 3.48 (SEM 0.47; table 1). The observed transition bias is consistent with that observed in other MA experiments (Ochman 2003; Lynch 2010b). BPSs toward AVT (fig. 2A) account for 68.4% (533/779) of all BPSs, which is also commonly observed in other bacterial systems (Lynch et al. 2008; Denver 2009; Keightley et al. 2009; Hershberg and Petrov 2010; Hildebrand et al. 2010; Sung et al. 2012), and thought to be linked to the high observed rate of spontaneous deamination of cytosine and the conversion of guanine to 8-oxo-guanine (Duncan and Miller 1980). Considering deletion mutation events, 64% (47/73) are short (1–3 bp) and 36% (26/73) long (4 bp or more; fig. 2B), including a 12,581-bp long deletion in MA line 56. Short insertions account for 75% (25/33) of all insertions (fig. 2B). Overall, there is a significant excess of deletions (68.8%, 73/106) over insertions in this MA experiment ( $\chi^2 = 15.09$ , df = 1,  $P = 0.10 \times 10^{-3}$ ), with deletion accounting for a loss of  $9.89 \times 10^{-2}$  nucleotides per generation per MA line (175.93-fold higher than the gain by insertions). These data are consistent with a general deletion bias observed in bacteria (Mira et al. 2001; Kuo and Ochman 2009; Koskiniemi et al. 2012).

### Effective Population Size ( $N_e$ ) Estimates

The effective population size ( $N_e$ ) is inversely proportional to the power of random genetic drift, which has been suggested to have an influence on the evolution of mutation rate (Lynch 2010a; Sung et al. 2012; Lynch et al. 2016). According to the drift-barrier hypothesis (DBH), the genome-wide mutation rate affecting fitness (estimated by the total protein coding nucleotides) is expected to be inversely related to  $N_e$ . In haploid organisms and under the assumption of neutrality,  $N_e$  can be estimated using the formula  $\pi_S = 2N_e\mu_{bs}$ , where  $\pi_S$  is equal to the variation at silent sites among natural isolates, and  $\mu_{bs}$  is



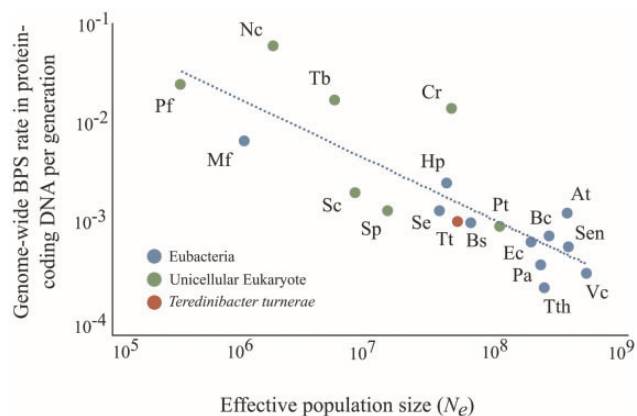
**Fig. 1.**—(A) Distribution of mutations, BPSs, and indels mapped in the 47 *Teredinibacter turnerae* CS30 MA lines. From the outer ring to inner ring scaled to total genome size: gene density (grey), significantly elevated (1 kb blocks >2 standard deviations from the mean) G/C (blue) or A/T content (red), position of base substitutions in each line (black—intergenic substitution; grey—synonymous substitution; blue—nonsynonymous substitution; red—insertions; green—deletions), and base-substitution density (25 kb blocks, red > orange > yellow). Circos plot (Krzywinski et al. 2009) was used to create this figure. Please access the online version for color information on this figure. (B) Mutation rates (BPSs and indels per site per generation) across different bacterial species. The data for this analysis were extracted from Sung et al. (2016).



**Fig. 2.**—Mutation spectrum of *Teredinibacter turnerae* MA lines. (A) Conditional BPSs rate normalized to the number of AT or GC base pairs in the genome. (B) Indels. Short indels are defined here as indels ranging from 1 to 3 bp, whereas long indels are ≥ 4 bp.

the BPSs rate per site per generation. *Teredinibacter turnerae*'s  $\pi_S$  was measured by comparing silent-site diversity in draft *T. turnerae* genomes (deposited in GenBank) according to Sung et al. (2012). By substituting the estimated  $\mu_{bs}$  ( $1.14 \times 10^{-9}$  per site per generation) in the formula, we estimate  $N_e$  to be

$\sim 4.5 \times 10^7$ . As shown in figure 3, this  $N_e$  is slightly lower than comparable measures in most free-living bacteria (e.g., *E. coli*  $N_e$  is estimated to be on the order of  $10^8$  [Sung et al. 2016]). Moreover, the mutation rate in this endosymbiont is quite similar to rates in other species with similar  $N_e$  (fig. 3),



**Fig. 3.**—Average total genome-wide BPSs rate in protein coding DNA per generation as a function of  $N_e$ . Trend line  $F(x) = 64.35 \times x^{-0.59}$  ( $r^2 = 0.693$ ). The data for this analysis were extracted from Lynch et al. (2016). Labels: Eubacteria: At, *Agrobacterium tumefaciens*; Bs, *Bacillus subtilis*; Bc, *Burkholderia cenocepacia*; Ec, *Escherichia coli*; Hp, *Helicobacter pylori*; Mf, *Mesoplasma florum*; Pa, *Pseudomonas aeruginosa*; Sen, *Salmonella enterica*; Se, *Staphylococcus epidermidis*; Tth, *Thermus thermophila*; Tt, *Teredinibacter turnerae*; and Vc, *Vibrio cholerae*. Unicellular Eukaryotes: Cr, *Chlamydomonas reinhardtii*; Nc, *Neurospora crassa*; Pf, *Plasmodium falciparum*; Pt, *Paramecium tetraurelia*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Tb, *Trypanosoma brucei*.

reinforcing the idea that  $N_e$  (Sung et al. 2012) has a large effect in determining the mutation rate evolution in endosymbionts as in other species (McCutcheon and Moran 2011).

## Discussion

Endosymbiotic bacteria have a unique lifestyle of coexistence with their hosts, and the forces driving their evolution differ from those of free-living bacteria and have a profound impact on their patterns of genome evolution (Wernegreen 2015). From evolutionary theory, one may predict that endosymbionts are subject to relaxed purifying selection on metabolic pathways that are redundant in host's genomes (Moran 2003), resulting in a reductive genome evolution, and increased fixation of slightly deleterious mutations by genetic drift due to  $N_e$  reduction caused by population bottlenecks intrinsic to the host-restricted lifestyle (for more exhaustive review on this topic, see Wernegreen 2015).

Confirming theoretical predictions on the role of the endosymbiotic lifestyle in genome evolution, prior mutation studies using phylogenetic comparisons and recent genomic data from a number of insect endosymbiotic species have brought to light that these organisms have rapid sequence evolution in comparison to closely related free-living bacteria (Moran 1996; Itoh et al. 2002). In addition, they often display severely eroded genomes (Kuo and Ochman 2009), where all but essential genes for the maintenance of the host association are lost. An extreme example of this process is

"*Candidatus Tremblaya princeps*" that has a complete genome length of  $\sim 139$  kb coding only for 110 genes (López-Madriral et al. 2011; McCutcheon and Moran 2011).

In MA lines of the endosymbiont *T. turnerae*, we found a rate of BPSs ( $1.14 \times 10^{-9}$  per-site per generation) higher than those observed in most free-living and facultative intracellular pathogenic bacteria and of the same order of magnitude of other nonfree-living bacteria such as *Buchnera aphidicola* ( $\sim 4 \times 10^{-9}$  per site per generation; Moran et al. 2009). The observed indel rate in the present study ( $1.55 \times 10^{-10}$  events per site per generation) is also one of the highest among bacteria (with established indel rates based on whole-genome sequencing data generated after mutation accumulation assay; fig. 1B). Furthermore, we observed a strong deletion bias ( $\chi^2 = 15.09$ ,  $df = 1$ ,  $P = 0.10 \times 10^{-3}$ ), with  $2.2 \times$  more spontaneous deletion events than insertion events, resulting in a loss of  $9.83 \times 10^{-2}$  nucleotides per generation per MA line. This observation suggests that genome erosion might be strong and have a role in *T. turnerae* evolution.

Although *T. turnerae* exhibits mutational properties that are similar to other endosymbionts (higher mutation rate and strong deletion bias), *T. turnerae* retains a relatively large genome ( $\sim 5.2$  Mb), comparable in size to many free-living bacteria such as *E. coli* (Blattner 1997), *Pseudomonas aeruginosa* (Stover et al. 2000), and *Bacillus subtilis* (Kunst et al. 1997). Because genomic erosion is a time dependent process, it may be suggested that this symbiotic relationship has arisen recently. Yet, fossil records reveal that this is not a recent association, with the family Teredinidae dating back to lower Cretaceous (100–145 million years [Myr]; Lopes et al. 2000), similar to the 150 Myr proposed for the symbiotic association between aphids and *B. aphidicola* that has a genome size of 0.6 Mb (Bennett and Moran 2014). However, population data show that the measured effective population size of *T. turnerae* ( $\sim 4.5 \times 10^7$ ) lies between the  $N_e$  measured in the free-living bacteria *B. subtilis* ( $10^7$ ) and *E. coli* ( $10^8$ ) (Sung et al. 2012), suggesting that *T. turnerae* may not be subject to heavy bottlenecks events such as those theoretically predicted to occur along the life history of host-restricted endosymbionts (Wernegreen 2015). Thus, although *T. turnerae* exhibits a statistically significant deletion bias, it has not experienced the large-scale genome erosion usually found in endosymbionts, presumably because of more effective selection in removing deletions from population. Our findings are consistent with the expectation that effective population size has a large influence on the mutation rates (both indel and BPSs; Sung et al. 2012, 2016) and genome size (Moran 2003; Kuo and Ochman 2009; McCutcheon and Moran 2011; Wernegreen 2015).

*Teredinibacter turnerae* is an intracellular bacterium that can be cultivated under in vitro conditions. Moreover, based on morphological data and the high sequence identity of the 16S rDNA from *T. turnerae* isolated from many Teredinidae species from unrelated geographical regions, this bacterium

might be subject to horizontal transfer across hosts (Waterbury et al. 1983; Distel et al. 1991). In fact, many endosymbiotic bacteria can be horizontally transmitted between hosts (Sandström et al. 2001; Haselkorn et al. 2009; Russell et al. 2009; Duron et al. 2010), and some intracellular opportunistic pathogenic bacteria such as *P. aeruginosa* (LaBauve and Wargo 2012) are capable of surviving in the environment as free-living organisms. Through similar mechanisms, these endosymbionts might be able to maintain large  $N_e$  during their life cycles. Thus, we propose that the endosymbiont *T. turnerae* is maintained in large enough populations when transferred from one generation to another, with selection acting efficiently enough to counteract a deletion bias and to maintain mutation rates at levels that lie between those of free-living bacteria and endosymbiotic bacteria that undergo frequent bottlenecks.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

This work was supported by Instituto Nacional de Ciência e Tecnologia (INCT) para o Controle do Câncer; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) (484005/2013-8); the Marie Curie Actions, European Commission FP7-PEOPLE-2009-IRSES project CINAR PATHOBACTER (project 247658); and by the Multidisciplinary University Research Initiative Award W911NF-09-1-0444 from the US Army Research Office to M.L., H. Tang, and S. Finkel. This work was also supported by NIH (R01-GM036827 and R35-GM122566) and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) (Post-Doctoral fellowship, “PNPD” to M.S. project 006995/2011-51).

## Literature Cited

- Amos W. 2010. Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence? *Proc Roy Soc Lond B Biol.* 277(1686):1443–1449.
- Bennett G, Moran NA. 2014. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci U S A.* 112(33):10169–10176.
- Blattner FR. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453–1462.
- Carpenter EJ, Culliney JL. 1975. Nitrogen fixation in marine shipworms. *Science* 187(4176):551–552.
- Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. 2005. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438(7064):90–93.
- Denver DR. 2009. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc Natl Acad Sci U S A.* 106(38):16310–16314.
- Denver DR, et al. 2012. Variation in base-substitution mutation in experimental and natural lineages of *Caenorhabditis* nematodes. *Genome Biol Evol.* 4(4):513–522.
- Distel DL, DeLong EF, Waterbury JB. 1991. Phylogenetic characterization and *in situ* localization of the bacterial symbiont of shipworms (Teredinidae: bivalvia) by using 16S rRNA sequence analysis and oligodeoxynucleotide probe hybridization. *Appl Environ Microbiol.* 57(8):2376–2382.
- Distel DL, Morrill W, MacLaren-Toussaint N, Franks D, Waterbury J. 2002. *Teredinibacter Turnerae* gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: teredinidae). *Int J Syst Evol Microbiol.* 52(Pt 6):2261–2269.
- Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* 287(5782):560–561.
- Duron O, Wilkes TE, Hurst GD. 2010. Interspecific transmission of a male-killing bacterium on an ecological timescale. *Ecol Lett.* 13(9):1139–1148.
- Elshahawi SI, et al. 2013. Boronated tartronic antibiotic produced by symbiotic cellulose-degrading bacteria in shipworm gills. *Proc Natl Acad Sci U S A.* 110(4):E295–E304.
- Foster PL, Hanson AJ, Lee H, Popodi EM, Tang H. 2013. On the mutational topology of the bacterial genome. *G3 (Bethesda)* 3(3):399–407.
- Gallager SM, Turner RD, Berg CJ. 1981. Physiological aspects of wood consumption, growth, and reproduction in the shipworm *Lyrodus pedicellatus* Quatrefages (Bivalvia: teredinidae). *J Exp Mar Bio Ecol.* 52(1):63–77.
- Han AW, et al. 2013. Turnerbactin, a novel triscatecholate siderophore from the shipworm endosymbiont *Teredinibacter turnerae* T7901. *PLoS ONE.* 8(10):e76151.
- Haselkorn TS, Markow TA, Moran NA. 2009. Multiple introductions of the *Spiroplasma* bacterial endosymbiont into *Drosophila*. *Mol Ecol.* 18(6):1294–1305.
- Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in Bacteria. *PLoS Genet.* 6(9):e1001115.
- Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content in Bacteria. *PLoS Genet.* 6(9):e1001107.
- Itoh T, Martin W, Nei M. 2002. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci U S A.* 99(20):12944–12948.
- Keightley PD, et al. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation Lines. *Genome Res.* 19(7):1195–1201.
- Kibota TT, Lynch M. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* 381(6584):694–696.
- Kneip C, Lockhart P, Voß C, Maier U-G. 2007. Nitrogen fixation in Eukaryotes – new models for symbiosis. *BMC Evol Biol.* 7(1):55.
- Koskiniemi S, Sun S, Berg OG, Andersson DI. 2012. Selection-driven gene loss in Bacteria. *PLoS Genet.* 8:1–7.
- Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19(9):1639–1645.
- Kunst F, et al. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390(6657):249–256.
- Kuo C-H, Ochman H. 2009. Deletional bias across the three domains of life. *Genome Biol Evol.* 1:145–152.
- LaBauve AE, Wargo MJ. 2012. Growth and Laboratory Maintenance of *Pseudomonas aeruginosa*. *Curr Protoc Microbiol.* Unit-6E.1.
- Lechene CP, Luyten Y, McMahon G, Distel DL. 2007. Quantitative imaging of nitrogen fixation by individual Bacteria within animal cells. *Science* 317(5844):1563–1566.
- Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A.* 109(41):E2774–E2783.

- Li H, Durbin R. 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987–2993.
- Lopes SGBC, Domaneschi O, De Moraes DT, Morita M, Meserani GDLC. 2000. Functional anatomy of the digestive system of *Neoteredo reynei* (Bartsch, 1920) and *Psiloteredo healdi* (Bartsch, 1931) (Bivalvia: tereidinidae). *Geol Soc SP*. 177(1):257–271.
- López-Madrugal S, Latorre A, Porcar M, Moya A, Gil R. 2011. Complete genome sequence of ‘*Candidatus Tremblaya princeps*’ strain PCVAL, an intriguing translational machine below the living-cell status. *J Bacteriol*. 193:5587–5588.
- Lynch M, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A*. 105(27):9272–9277.
- Lynch M. 2010a. Evolution of the mutation rate. *Trends Genet*. 26(8):345–352.
- Lynch M. 2010b. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A*. 107(3):961–968.
- Lynch M, et al. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet*. 17(11):704–714.
- Margulis L, Bermudes D. 1985. Symbiosis as a mechanism of evolution: status of cell symbiosis theory. *Symbiosis* 1:101–124.
- McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic Bacteria. *Nat Rev Microbiol*. 10(1):13–26.
- Minic Z, Hervé G. 2004. Biochemical and enzymological aspects of the symbiosis between the deep-sea tubeworm *Riftia pachyptila* and its bacterial endosymbiont. *Eur J Biochem*. 271(15):3093–3102.
- Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet*. 17(10):589–596.
- Moran NA. 1996. Accelerated evolution and Muller’s ratchet in endosymbiotic Bacteria. *Proc Natl Acad Sci U S A*. 93(7):2873–2878.
- Moran NA. 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin Microbiol*. 6(5):512–518.
- Moran NA, McLaughlin HJ, Sorek R. 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic Bacteria. *Science* 323(5912):379–382.
- Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into Prokaryote–animal symbioses. *Nat Rev Genet*. 9(3):218–229.
- Mukai T. 1964. The genetic structure of natural populations of *Drosophila melanogaster*. I. spontaneous mutation rate of polygenes controlling viability. *Genetics* 50(500):1–19.
- Ochman H. 2003. Neutral mutations and neutral substitutions in bacterial genomes. *Mol Biol Evol*. 20(12):2091–2096.
- Russell JA, et al. 2009. Specialization and geographic isolation among *Wolbachia* symbionts from ants and lycaenid butterflies. *Evolution* 63(3):624–640.
- Sandström JP, Russell JA, White JP, Moran NA. 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Mol Ecol*. 10(1):217–228.
- Schink B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol Rev*. 61(2):262–280.
- Schrider DR, Hourmozdi JN, Hahn MW. 2011. Pervasive multinucleotide mutational events in Eukaryotes. *Curr Biol*. 21(12):1051–1054.
- Stewart FJ, Newton ILG, Cavanaugh CM. 2005. Chemosynthetic endosymbioses: adaptations to oxic–anoxic interfaces. *Trends Microbiol*. 13(9):439–448.
- Stover CK, et al. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406(6799):959–964.
- Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci U S A*. 109(45):18488–18492.
- Sung W, et al. 2015. Asymmetric context-dependent mutation patterns revealed through mutation–accumulation experiments. *Mol Biol Evol*. 32(7):1672–1683.
- Sung W, et al. 2016. Evolution of the insertion–deletion mutation rate across the tree of life. *G3 (Bethesda)* 6(8):2583–2591.
- Trindade-Silva AE, et al. 2009. Physiological traits of the symbiotic bacterium *Teredinibacter turnerae* isolated from the mangrove shipworm *Neoteredo reynei*. *Genet Mol Biol*. 32(3):572–581.
- Trytek RE, Allen WV. 1980. Synthesis of essential amino acids by bacterial symbionts in the gills of the shipworm *Bankia setacea* (Tryon). *Comp Biochem Physiol A Physiol*. 67(3):419–427.
- Waterbury JB, Calloway CB, Turner RD. 1983. A cellulolytic nitrogen-fixing bacterium cultured from the gland of *Deshayes* in shipworms (Bivalvia: tereidinidae). *Science* 221(4618):1401–1403.
- Wernegreen JJ. 2015. Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann NY Acad Sci*. 1360(1):16–35.
- Woollfit M, Bromham L. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol Biol Evol*. 20(9):1545–1555.
- Yang JC, et al. 2009. The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS ONE*. 4(7):e6085.
- Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25(21):2865–2871.
- Zientz E, Dandekar T, Gross R. 2004. Metabolic interdependence of obligate intracellular Bacteria and their insect hosts. *Microbiol Mol Biol Rev*. 68(4):745–770.

Associate editor: Howard Ochman