Periodic variation of mutation rates in bacterial genomes associated with replication timing

Marcus M. Dillon[a,b], Way Sung[c,d], Michael Lynch[d], and Vaughn S. Cooper[b,e]

[a]Department of Cell and Systems Biology, University of Toronto, Toronto, ON, CAN

[b]Graduate Program in Microbiology, University of New Hampshire, Durham, NH, USA

[c]Department of Bioinformatics and Genomics, University of North Carolina Charlotte, Charlotte, NC, USA

[d]Department of Biology, Indiana University, Bloomington, IN, USA

[e]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Corresponding Author:
Vaughn Cooper
425 Bridgeside Point II, 450 Technology Drive
Pittsburgh, PA 15219
Phone: (412)-624-1265
Email: vaughn.cooper@pitt.edu

Running title: Periodic variation in bacterial mutation rates.

Keywords: Mutation accumulation, mutation rate, genome structure, periodicity

ABSTRACT

Spontaneous mutation rates may vary within genome regions across the tree of life, but the causes and consequences of this spatiotemporal variation are uncertain because studies of genome-wide mutation rates with sufficient resolution are rare. Here we examine relationships between local mutation rates and replication timing throughout the genomes of three bacterial species using results from five mutation accumulation (MA) experiments analyzed by whole genome sequencing (WGS). Three MA experiments were carried out with wild-type strains of the species *Vibrio fischeri*, *Vibrio cholerae*, and *Burkholderia cenocepacia*, each of which have multiple chromosomes, while two additional experiments were carried out using mismatch repair-deficient strains of *V. fischeri* and *V. cholerae*. We find that in the absence of mismatch repair, base-substitution mutation rates vary in a mirrored wave-like pattern on opposing replichores of the large chromosome of *V. fischeri* and *V. cholerae*, where concurrently replicated regions experience similar base-substitution mutation rates. The base-substitution mutation rates on the small chromosome are less variable in both species, but tend to occur at similar rates as the concurrently replicated regions of the large chromosome. We lack a sufficient quantity of base-substitution mutations in the MA-WGS experiments with intact mismatch repair to rigorously evaluate regional variation in base-substitution mutation rates, although it is possible that a larger collection of base-substitution mutations would reveal rate periodicity in these experiments. These results support the notion that base-substitution mutation rates are likely to vary systematically across many bacterial genomes, which exposes certain genes to elevated deleterious mutational load.

INTRODUCTION

Although genome architecture varies markedly across the tree of life, some level of spatiotemporal organization of the genome is essential for all organisms because DNA must be compact but also available for gene expression, DNA replication, and chromosome segregation (Lynch 2007; Dorman 2013; Herrick 2011). Consistent with this requirement, relative distance from the origin of replication and replication terminus is one of the most conserved properties of orthologous genes, suggesting that genomes are under relatively strict organizational constraints (Sobetzko et al. 2012). These constraints have historically been attributed mostly to gene expression, where intragenic variation in the binding of nucleotide associated proteins (NAPs) and compaction of the nucleoid induce selection on gene order and location for optimal expression (Dorman 2013; Sobetzko et al. 2012; Dame et al. 2011; Warnecke et al. 2012). However, variation also exists in the propensity of different genome regions to incur mutations as a result of NAP binding (Sobetzko et al. 2012), error prone polymerases (Courcelle 2009), error prone repair pathways (Lang and Murray 2011), and inconsistent nucleotide pools (Zhang and Mathews 1995). This variation in mutation rates has important implications for the functional evolution of genes and may impose increased mutation pressure on specific genes as a result of their genome location, yet the patterns of spatiotemporal variation in mutation rates remain poorly understood.

A series of comparative studies in multicellular eukaryotes (Chen et al. 2010; Stamatoyannopoulos et al. 2009; Herrick 2011; Mugal et al. 2010), unicellular eukaryotes (Agier and Fischer 2012; Herrick 2011), archaea (Flynn et al. 2010), and bacteria (Cooper et al. 2010; Mira and Ochman 2002) have shown that synonymous substitution rates vary across the genome and generally increase with replication timing. However, this correlation could result from higher base-substitution mutation (bpsm) rates or weaker purifying selection in late replicating regions (Ochman 2003; Cooper et al. 2010; Lynch et al. 2016). A more direct test of intragenomic variation in mutation rates has come from mutation accumulation (MA) experiments paired with whole-genome sequencing (WGS), in which several replicate lineages are passaged through hundreds of single cell bottlenecks in the near absence of natural selection and all mutations incurred are identified. A number of MA-WGS studies have now been conducted in diverse bacterial species but they do not generally support the positive correlation between replication timing and

substitution rates. Specifically, MA experiments with mismatch repair (MMR)-proficient organisms do not reveal any significant variation in base-substitution mutation (bpsm) rates among regions of bacterial genomes (Lee et al. 2012; Dillon et al. 2015, 2017). MA studies with MMR-deficient organisms that generate much larger collections of mutations do show that polymerase error rates are non-uniform across genomes, but show no significant positive correlation between bpsm rates and replication timing (Foster et al. 2013; Long et al. 2014; Sung et al. 2015; Dettman et al. 2016; Dillon et al. 2017). Rather, these MMR-deficient MA studies of *Escherichia coli*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, *Vibrio cholerae*, and *Vibrio fischeri* reveal mirrored bpsm rates across the two replichores of the primary circular chromosome. This suggests a shared mechanism by which the replication cycle influences mutation rates in diverse bacterial genomes and indicates that common evolutionary pressures may be operating on genes residing at shared distances from the origin of replication.

Although most published bacterial MA-WGS studies have focused on well-studied bacterial genomes with a single circular chromosome, our studies focus on bacteria with multiple circular chromosomes (Dillon et al. 2015, 2017), an underappreciated, but not uncommon bacterial genome architecture (Egan et al. 2005; Ochman 2002; Cooper et al. 2010; Val et al. 2014). Setting aside the distinction between chromosomes and megaplasmids (Agnoli et al. 2012), the *Vibrio cholerae* and *Vibrio fischeri* genomes are composed of two chromosomes, while the *Burkholderia cenocepacia* genome is composed of three. In all three of these species, the first chromosome (chr1) is largest, harbors the most essential genes and is expressed at the highest levels (Cooper et al. 2010; Morrow and Cooper 2012). Secondary chromosomes (chr2, chr3) are similar in structure, as they are also circular, initiate replication from a single origin, and are replicated bi-directionally on two replichores (Val et al. 2014; Egan and Waldor 2003; Rasmussen et al. 2007). However, while secondary chromosomes are replicated at the same rate as the first chromosome, their origins of replication (*oriCII*) have distinct initiation requirements from those of chr1 origins (*oriCI*) (Egan et al. 2005; Duigou et al. 2006). Importantly, chr2 (or chr3) replication is delayed relative to chr1, which ensures that replication of all chromosomes will terminate synchronously (Rasmussen et al. 2007; Val et al. 2014; Baek and Chattoraj 2014). The delayed initiation of chr2 replication means that the genome region near the origin of chr1 is always replicated prior to the replication of secondary chromosomes, while late replicated regions

of chr1 are replicated concurrently with chr2. This general replication timing pattern in bacteria with multiple circular chromosomes begs the question of whether secondary chromosomes experience similar mutation rates and periodicity to concurrently late replicated regions of primary chromosomes, and more broadly, whether replication timing is a good predictor of mutation rates. Shared fluctuations in mutation rates across chromosomes would imply that the causative factors act not just spatially but also temporally as a function of the cell cycle.

Here, we performed detailed analyses of the genome-wide distribution of spontaneous bpsms generated by MA-WGS experiments with MMR-deficient strains of *V. fischeri* (4313 bpsms) and *V. cholerae* (1022 bpsms), and spontaneous bpsms generated by MA-WGS experiments with MMR-proficient strains of *V. fischeri* (219 bpsms), *V. cholerae* (138 bpsms), and *B. cenocepacia* (245 bpsms) (Dillon et al. 2015, 2017). The bpsm rates in MMR-deficient MA-WGS experiments reveal that the patterns of bpsm rates on chr1, particularly G:C>A:T transitions, vary in a mirrored wave-like pattern that is robust to changing interval sizes, consistent with MMR-deficient MA-WGS experiments in some single-chromosome bacterial genomes. Specifically, bpsm rates are low near the *oriC*, reach intermediate peaks approximately 600 Kb from the *oriC* on each replichore, then decline into valleys on each replichore before rising and falling again as they approach the replication terminus. Although the distributions of bpsms on chr2 do not show significant periodicity, regional patterns of bpsm rates on chr2 are similar to those of late replicated regions of chr1, suggesting that concurrently replicated regions on different chromosomes experience similar bpsm rates.

RESULTS

Two MMR-deficient (mutator) and three MMR-proficient (wild-type) MA-WGS experiments were founded by five different ancestral strains: a) *V. fischeri* ES114 *ΔmutS* (*Vf*-mut), b) *V. cholerae* 2740-80 *ΔmutS* (*Vc*-mut), c) *V. fischeri* ES114 wild-type (*Vf*-wt), d) *V. cholerae* 2740-80 wild-type (*Vc*-wt), and e) *B. cenocepacia* HI2424 wild-type (*Bc*-wt). Forty-eight independent MA lineages were propagated for 43 days in the two mutator experiments and seventy-five MA lineages were propagated for 217 days in the three wild-type experiments. In total, we performed successful WGS on evolved clones of 19 *Vf*-mut lineages, 22 *Vc*-mut

lineages, 48 *Vf*-wt lineages, 49 *Vc*-wt lineages, and 47 *Bc*-wt lineages. However, despite the fact that the mutator experiments were shorter and involved fewer lineages, the vast majority of bpsms were generated in the *Vf*-mut and *Vc*-mut lineages, as their bpsm rates are 317-fold and 85-fold greater than those of their wild-type counterparts, respectively. Consequently, we can study the effects of genomic position on bpsm rates in much greater detail in the mutator lineages, where we observe a considerable number of bpsms distributed across the genome at intervals as low as 10 Kb in length (Table 1), the approximate length of bacterial microdomains (Dorman 2013).

Table 1. Number of base-substitution mutations (bpsms) in each mutation accumulation experiment, and the associated average number of bpsms in intervals of variable sizes.

MA	No. of	500 Kb		250 Kb		100 Kb		50 Kb		25 Kb		10 Kb	
Lines	bpsm	Avg.	SEM	Avg.	SEM	Avg.	SEM	Avg.	SEM	Avg.	SEM	Avg.	SEM
Vf-mut	4313	499.00	35.82	253.50	13.77	101.05	3.12	50.53	1.26	25.26	0.51	10.08	0.18
Vc-mut	1022	141.50	21.14	65.33	6.23	25.47	1.53	12.51	0.59	6.22	0.25	2.50	0.09
Vf-wt	219	22.25	3.28	12.25	1.39	4.95	0.40	2.48	0.20	1.24	0.09	0.50	0.03
Vc-wt	138	18.00	3.09	8.75	0.95	3.42	0.30	1.72	0.14	0.83	0.07	0.34	0.03
Bc-wt	245	15.90	1.43	7.58	0.57	3.27	0.22	1.62	0.11	0.81	0.05	0.32	0.02

We analyzed bpsm rates in 10 Kb, 25 Kb, 50 Kb, 100 Kb, 250 Kb, and 500 Kb intervals that extended bi-directionally from the *oriCI* as the replication forks proceed during replication to determine the effects of genomic position on bpsm rates. For secondary chromosomes, we analyzed bpsm rates using the same interval sizes as chr1 but these chromosomes were analyzed relative to the inferred replication timing of *oriCI* (Figure 1A). This enables direct comparisons between intervals on secondary chromosomes with late replicating intervals on chr1, allowing us to ask whether concurrently replicated regions on different chromosomes experience similar bpsm rates. To illustrate how this analysis works, we have plotted the patterns of bpsm rates from a recent *E. coli* mutator MA experiment where mutation rates were demonstrated to vary in a wave-like pattern that is mirrored on the two replichores of its singular circular chromosome (Foster et al. 2013; Lee et al. 2012) (Figure 1B). If replication timing is responsible for this pattern, a hypothetical secondary chromosome in *E. coli* would be expected to mirror concurrently replicated (late replicating) regions on the primary chromosome (Figure 1B).

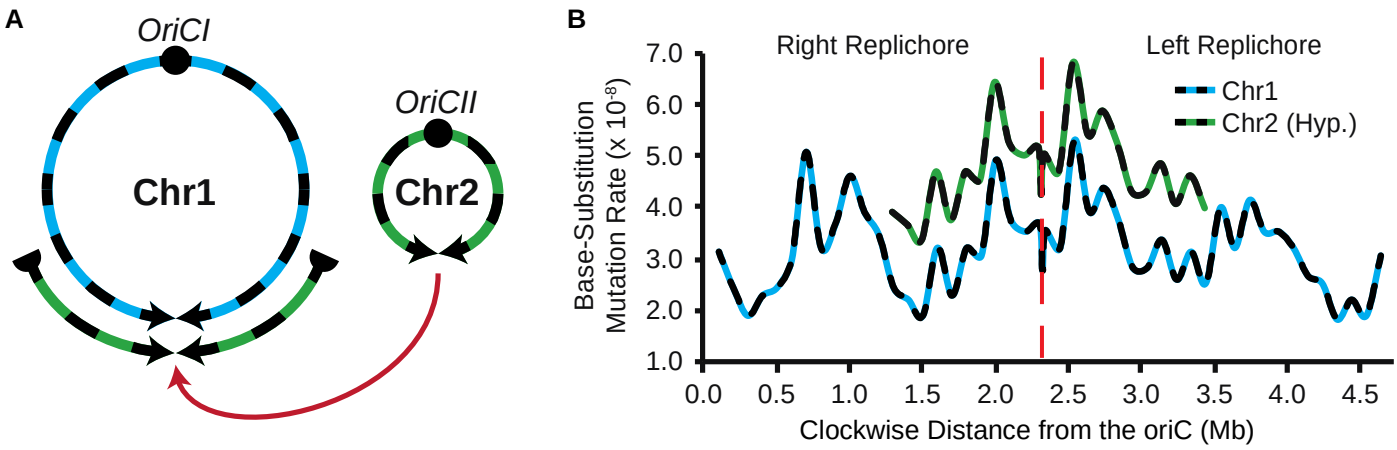

Figure 1. Design of the interval analysis used in this study to enable direct comparisons of base-substitution mutation (bpsm) rates of concurrently replicated regions on chromosome 1 (chr1) and chromosome 2 (chr2). A) For all multi-chromosome species analyzed in this study, secondary chromosomes are split at their origin of replication (*oriCII*), and mapped directly to concurrently replicated intervals in late replicating regions of chr1. All intervals on both chromosomes are thus relative to the initiation of replication of *oriCI*, and the boundaries of the intervals are consistent with their replication timing. B) Patterns of bpsm rates on the single chromosome of *Escherichia coli* MG1655 rph+ Δ*mutL*, derived from (Lee et al. 2012), show a wave-like mirrored pattern of bpsm rates on the two opposing replichores. If replication timing governs this pattern, a hypothetical secondary chromosome would be expected to mirror patterns of bpsm rates of late replicated regions on the primary chromosome.

Patterns of mutator base-substitution mutation rate variation within single chromosomes. We can reject the null hypothesis that bpsms were uniformly distributed across 500 Kb, 250 Kb, 100 Kb, 50 Kb, and 10 Kb intervals on chr1 in both the *Vf*-mut and *Vc*-mut MA experiments (Table S1), but not for any interval size on chr2 for these MA experiments, suggesting that regional variation in bpsm rates on chr1 exceeds that of chr2 (Table S1). The patterns of bpsm rate variation on chr1 for both the *Vf*-mut and *Vc*-mut MA experiments are wave-like and mirrored on both replichores as they extend bi-directionally from the origin of replication (Figure 2A, B). This mirrored pattern is evident at multiple interval sizes and is consistent with some other bacterial genomes that have a single chromosome, like *E. coli* (Foster et al. 2013), although the length of the wave periods that we observe here are shorter than those observed in *E. coli* (Figure 1B). Specifically, bpsm rates are low at the *oriCI*, increase to bpsm rate peaks approximately 600 Kb from the *oriCI* on both replichores, then decline into another valley before rising and falling again as they approach the replication terminus. This sequence generates two distinct waves on each replichore of chr1 (Figure 2A, B), but these same patterns were not observed on chr2 (Figure S1). We chose to focus our most detailed analyses on patterns of variation in bpsm rates at the intermediate interval length of 100 Kb because it is the maximum interval length with which all of the features described above are pronounced for both

species. Thus, the 100 Kb interval analysis maximizes the number of bpsms per interval while still retaining two mirrored bpsm rate waves on each replichore.

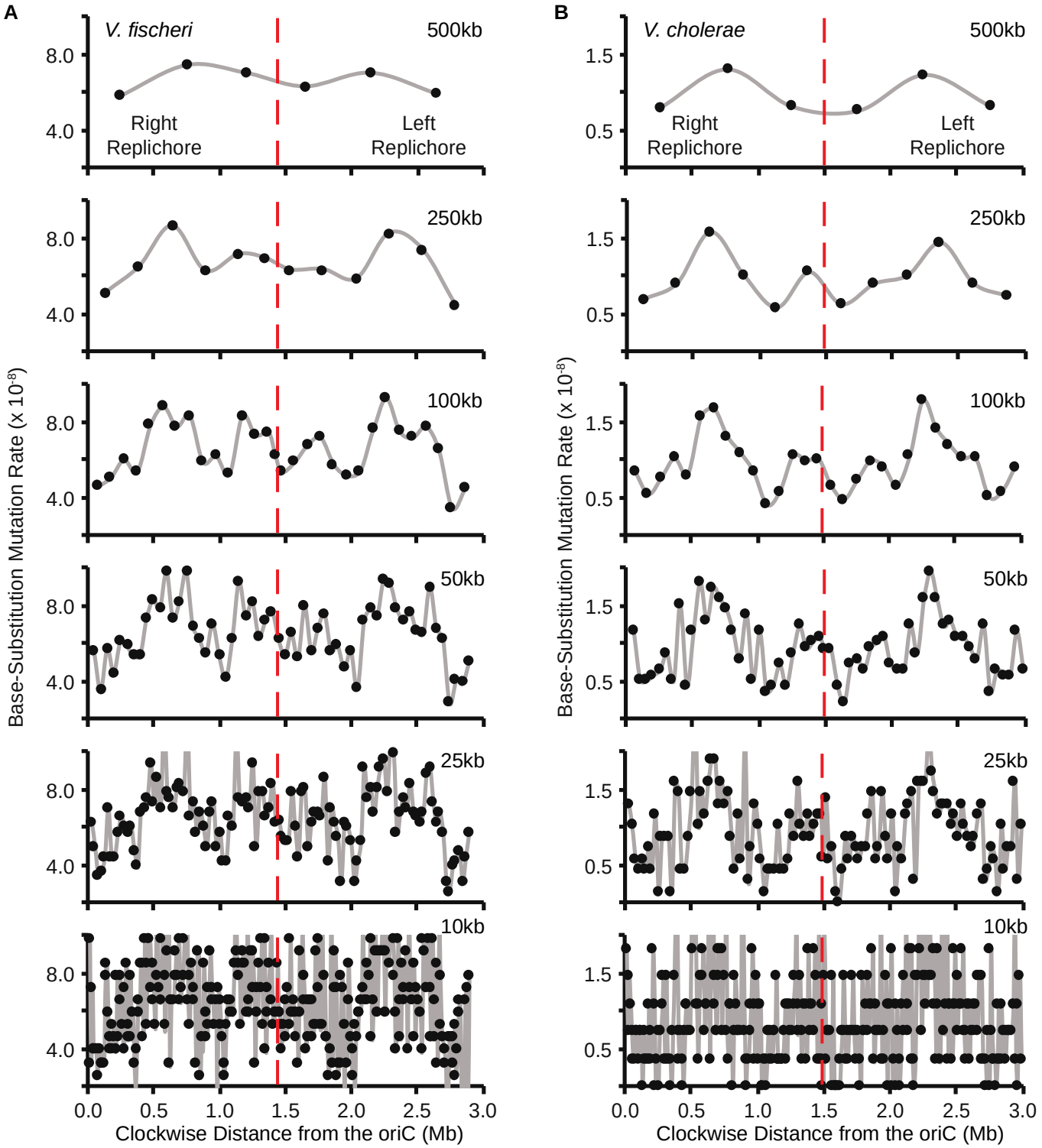

Figure 2. Patterns of base-substitution mutation (bpsm) rates at various size intervals extending clockwise from the origin of replication (*oriC*) in MMR-deficient mutation accumulation lineages of *Vibrio fischeri* (A) and *Vibrio cholerae* (B) on chromosome 1. Bpsm rates are calculated as the number of mutations observed within each interval, divided by the product of the total number of sites analyzed within that interval across all lines and the number of generations of mutation accumulation. The two intervals that meet at the

terminus of replication (dotted red line) on each replichore are shorter than the interval length for that analysis, because the size of chromosome 1 is never exactly divisible by the interval length.

As expected from the visual similarity of bpsm rates on the opposing replichores of chr1 (Figure 2A, B), the bpsm rates of concurrently replicated intervals across replichores are significantly correlated in both the *Vf*-mut and *Vc*-mut experiments (Figure 3A, B) (linear regression; *Vf*-mut: F = 10.98, df = 13, p = 0.0060, r^2 = 0.46; *Vc*-mut: F = 6.76, df = 13, p = 0.0221, r^2 = 0.34). This relationship is also positive in analyses of all other interval lengths and is significantly positive for most comparisons (Table S2; Table S3). In contrast, none of these comparisons between replichores of chr2 revealed significant correlations (Table S2; Table S3) owing to considerably less variation in bpsm rates on these chromosomes (Figure S1). These data suggest that when there is sufficient variation in bpsm rates within chromosomes, as is the case in chr1 of the *Vf*-mut and *Vc*-mut MA experiments, replication timing can guide mutation rate estimates among concurrently replicated regions.

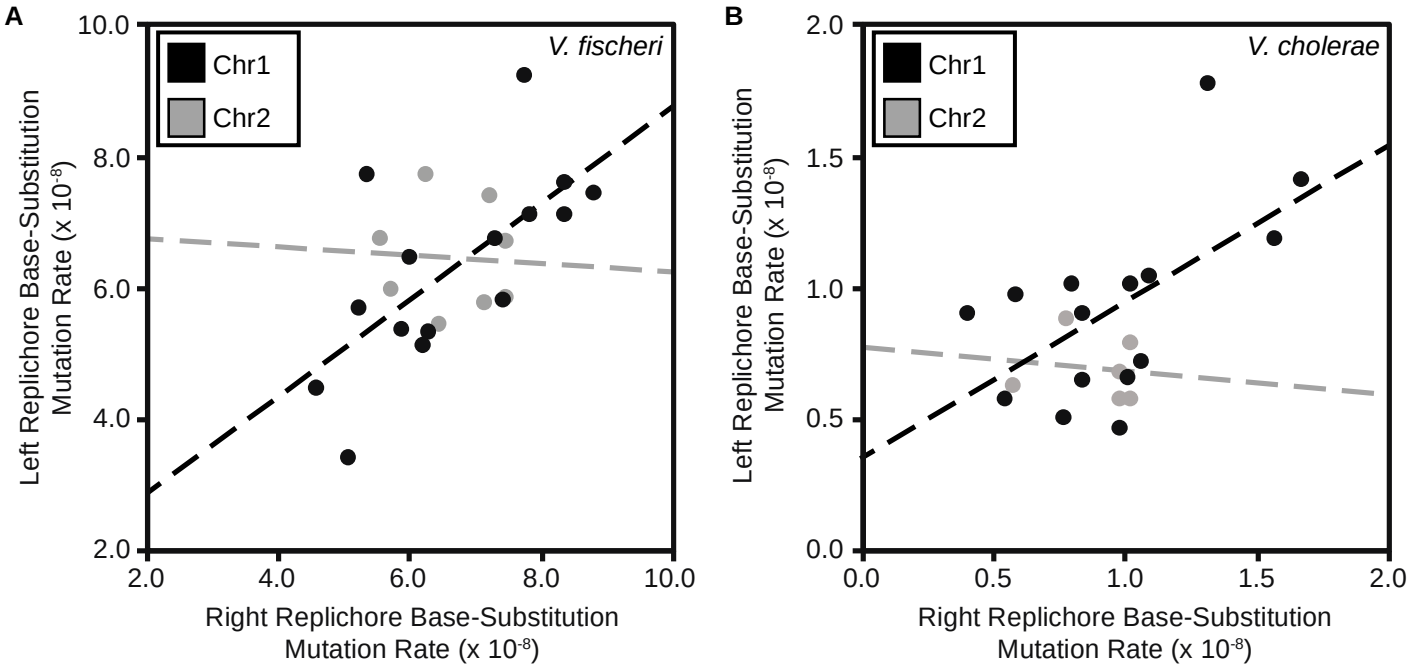

Figure 3. Relationship between base-substitution mutation (bpsm) rates in 100 Kb intervals on the right replichore with concurrently replicated 100 Kb intervals on the left replichore in MMR-deficient *Vibrio fischeri* (A) and *Vibrio cholerae* (B). Bpsm rates were calculated as the number of mutations observed within each interval, divided by the product of the total number of sites analyzed within that interval across all lines and the number of generations of mutation accumulation. The range of bpsm rates between intervals on chromosome 1 is greater for both *V. fischeri* and *V. cholerae* and both linear regressions are significant (*V. fischeri*: F = 10.98, df = 13, p = 0.0060, r^2 = 0.46; *V. cholerae*: F = 6.76, df = 13, p = 0.0221, r^2 = 0.34), while neither linear regressions on chromosome 2 are significant (*V. fischeri*: F = 0.02, df = 6, p = 0.8910, r^2 = 0.03×10^{-1}; *V. cholerae*: F = 0.06, df = 4, p = 0.8140, r^2 = 0.02).

Patterns of mutator base-substitution mutation rates of concurrently replicated regions on different chromosomes. Late replicated regions of chr1 may experience similar bpsm rates to chr2 because like the opposing replichores of chr1, they are replicated concurrently. To study this relationship, we mapped the patterns of bpsm rates in 100 Kb intervals on chr2 to those of late replicated 100 Kb intervals on chr1 for both *Vf*-mut and *Vc*-mut (Figure 1). The patterns of bpsm rates on chr2 are visually similar to the bpsm rates of the late replicated regions on chr1 in both species (Figure 4A, B), but the overall variation in bpsm rates between different intervals is limited in late replicated regions. Consequently, the bpsm rates of concurrently replicated 100 Kb intervals (or any other interval lengths, see Table S4) on chr1 and chr2 are not significantly correlated (Linear regression; *Vf*-mut: $F = 0.62$, $df = 14$, $p = 0.4442$, $r^2 = 0.04$; *Vc*-mut: $F = 0.07$, $df = 10$, $p = 0.7941$, $r^2 = 0.01$), despite the fact that chr2 mutation bpsm rates appear to be more similar to late replicating regions on chr1 than they are to any other regions on the primary chromosome.

To further explore the extent to which bpsm rates on chr2 are similar to late replicating regions on chr1, we evaluated whether regional bpsm rates on chr2 were more similar to the concurrently replicated regions on chr1 than they were to any other chr1 regions. Specifically, we mapped chr2 bpsm rates to all possible interval combinations on the right and left replichores of chr1, starting with the concurrently replicated terminal intervals and shifting back one interval per replichore until we reached the origin of replication of chr1. We then calculated the sum of the residuals in each analysis to identify the best fit for chr2 bpsm rates. For *Vf*-mut, the lowest sum of the residuals (14.01×10^{-8}) occurs when the chr2 intervals were mapped to the concurrently late replicated intervals on chr1 (Figure 4A; Table S5). Similarly, the lowest sum of the residuals in *Vc*-mut (2.53×10^{-8}) occurs when the chr2 intervals are mapped to their concurrently replicated intervals on chr1 (Figure 4B; Table S5). Thus, despite the fact that there is no significant linear correlation in mutation rate periodicity between chr1 and chr2, bpsm rates on chr2 most closely resemble the bpsm rates of concurrently replicated regions on chr1 in both *V. cholerae* and *V. fischeri.* Furthermore, because of its smaller size and delayed replication, chr2 narrowly avoids the high bpsm rate peaks on the right and left replichores of chr1 in both *Vf*-mut and *Vc*-mut (Figure 4A, B), which suggests that genes on chr2 will be subjected to less deleterious load than many of the genes on chr1, particularly in *V. cholerae*.

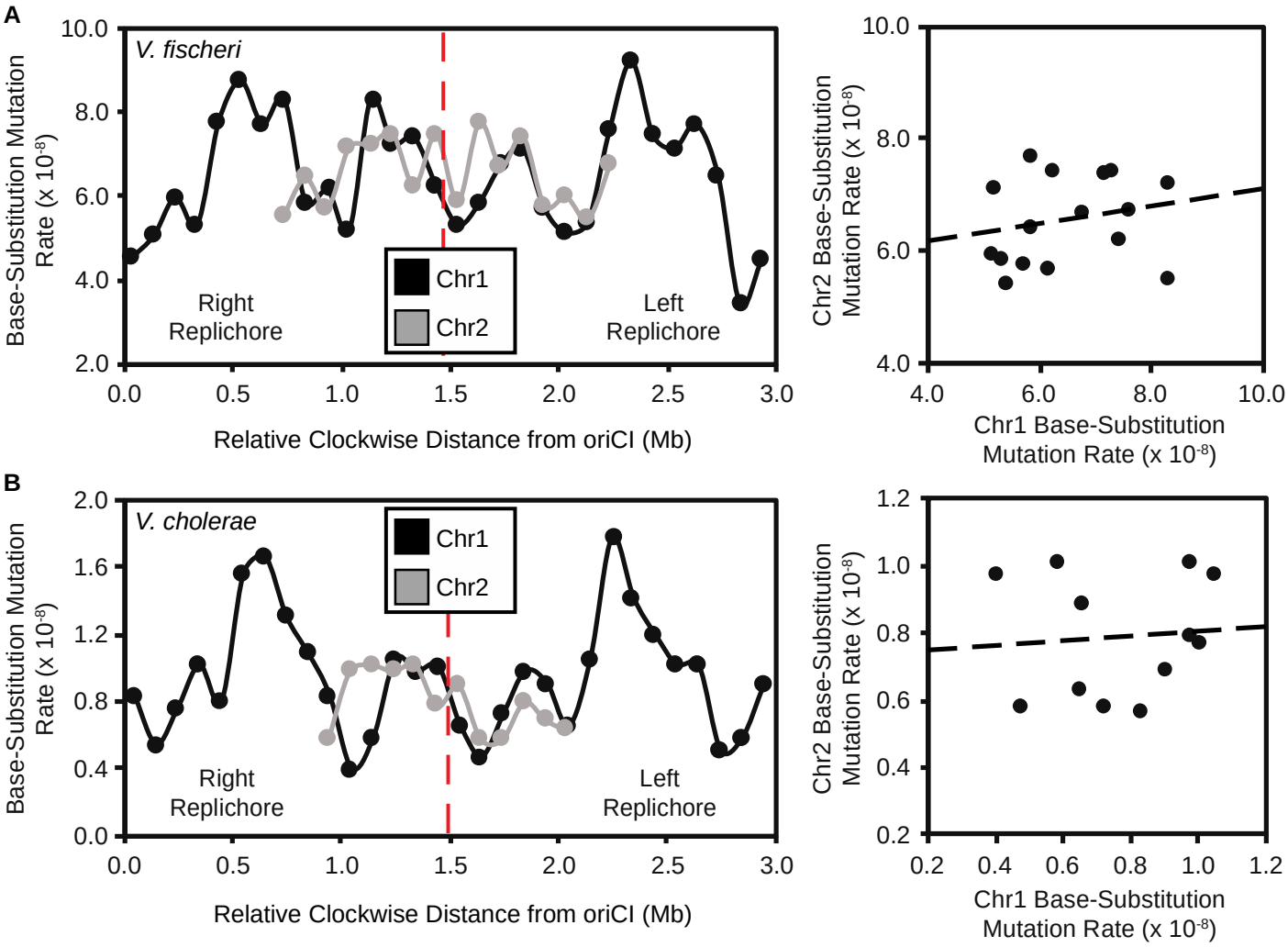

Figure 4. Patterns of base-substitution mutation (bpsm) rates in 100 Kb intervals extending clockwise from the origin of replication (*oriCI*) on chromosome 1 (chr1) and patterns of bpsm of concurrently replicated 100 Kb intervals on chromosome 2 (chr2) for MMR-deficient *Vibrio fischeri* (A) and *Vibrio cholerae* (B). Bpsm rates were calculated as the number of mutations observed within each interval, divided by product of the total number of sites analyzed within that interval across all lines and the number of generations of mutation accumulation. Patterns of bpsm rates on chr2 appear to map to those of concurrently replicated regions on chr1 in both species, but the variance in bpsm rate between intervals is not sufficient to produce significant linear regressions between concurrently replicated intervals on chr1 and chr2 in either *V. fischeri* or *V. cholerae* (*V. fischeri*: F = 0.62, df = 14, p = 0.4442, r^2 = 0.04; *V. cholerae*: F = 0.07, df = 10, p = 0.7941, r^2 = 0.01).

Wavelet transformations capture periodicities in base-substitution mutation rates. Recognizing that regional or cyclic variation in mutation rates may not be captured by linear models, we used wavelet transformations to identify periodicities in the mirrored wave-like patterns in bpsm rates. Specifically, the frequency structure of bpsm rates on each chromosome in the *Vf*-mut and *Vc*-mut studies was modeled as a univariate series using the Morlet wavelet (Roesch and Schmidbauer 2014). Extending clockwise from the *oriCI* region in the *Vf*-mut lineages, we find significant wave periods of approximately 1.6 Mb and 0.8 Mb

(Figure 5A). The 1.6 Mb period is significant across the entire chromosome and depicts a single long period wave on each replichore. The ~0.8 Mb period is also significant across most of chr1 but its inferred length varies across the chromosome, between 0.6 and 1.0 Mb. This period reflects the wave-like patterns depicted in Figure 2, with two synchronous periods per replichore or four periods in total around the chromosome. Similarly, we found two wave periods at approximately 1.6 Mb and 0.8 Mb on chr1 in the *Vc*-mut lineages (Figure 5E) that are consistently significant across the chromosome. Using only periods with high power across multiple intervals, we were also able to effectively reproduce the mirrored wave-like patterns in bpsm rates seen on chr1 at intervals of 100 Kb in both the *Vf*-mut and *Vc*-mut lineages (Figure 5B, F). No continuously significant wave periods originating from the *oriCII* region of chr2 for either the *Vf*-mut or the *Vc*-mut experiment were identified (Figure 5C, G); consequently, we had limited success at reconstructing the observed bpsm rates on Chr2 in both studies (Figure 5D, H).

To further evaluate whether waves on opposing replichores of Chr1 were synchronous in the *Vf*-mut and *Vc*-mut experiments, we employed a cross-wavelet transformation to identify shared periodicities (Roesch and Schmidbauer 2014). Two waves with periods of approximately 1.6 Mb and 0.8 Mb were synchronous across the genome in both the *Vf*-mut and *Vc*-mut experiments (Figure S2A, B). This analysis suggests that significant wave periods in bpsm rates on the right replichore can predict significant wave periods on the left replichore, and vice-versa. It is also noteworthy that the statistically synchronous waves become smaller near the replication terminus, particularly in the *Vf-mut* lineages, (Figure S2A, B), as can be seen in the raw data presented in Figure 2. Together, these analyses suggest that bpsm rates in MMR-deficient *V. fischeri* and *V. cholerae* vary in a periodic function modeled by two synchronous waves extending bi-directionally from the origin of replication on Chr1. The first is a long period wave function where each replichore represents a single period and two total periods exist around the chromosome, while the second is a shorter period wave function with two waves on each replichore that occur synchronously as the chromosome is replicated bi-directionally.

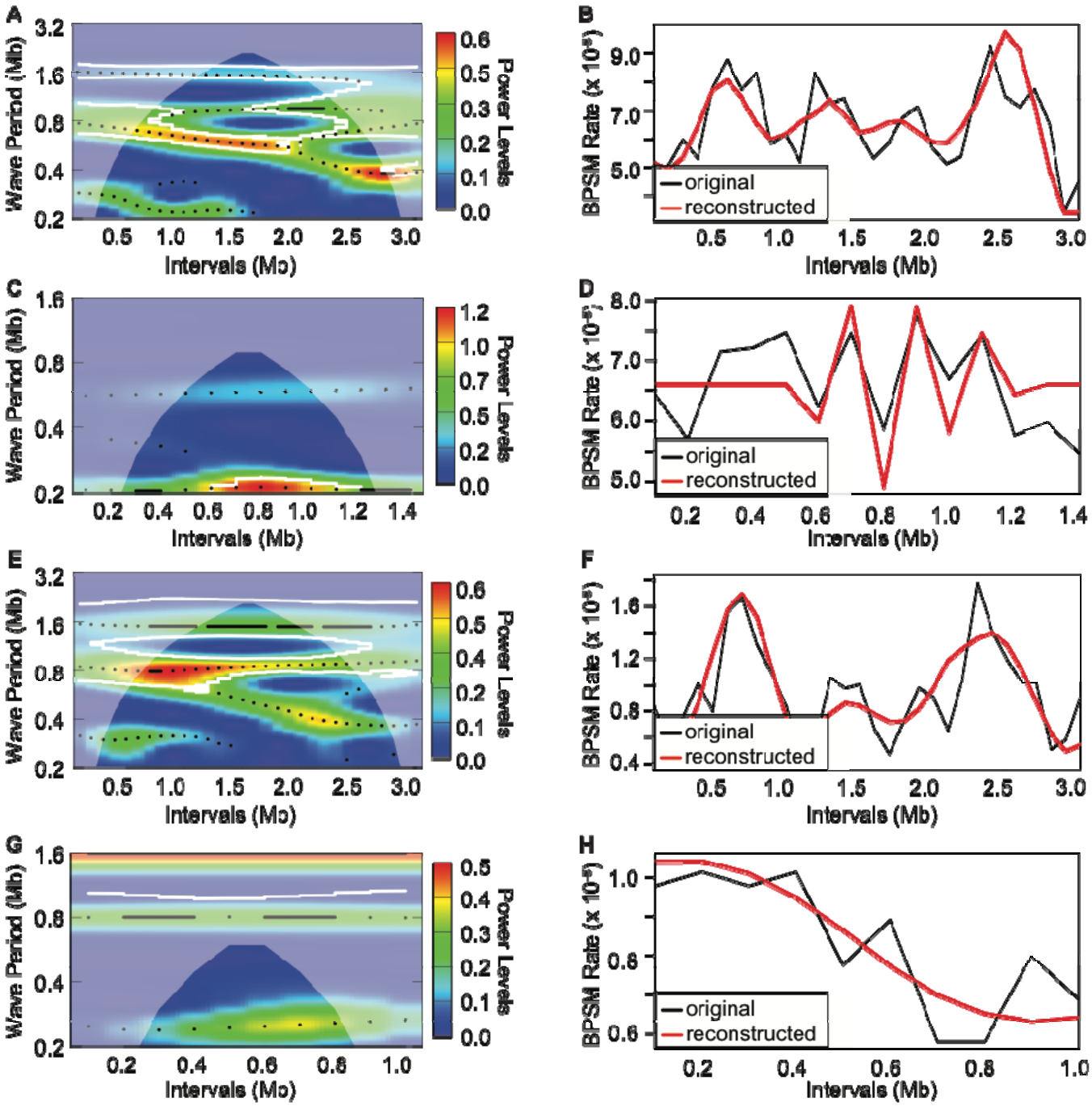

Figure 5. Wavelet power spectrum and resultant reconstruction of the patterns of base-substitution mutation (bpsm) rates in 100 Kb intervals extending clockwise from the *oriCI* region of chromosome 1 (A, B: *V. fischeri*; E, F: *V. cholerae*) and the *oriCII* region of chromosome 2 (C,D: *V. fischeri*; G,H: *V. cholerae*) using the MMR-deficient mutation accumulation lineages. All plots were generated using the WaveletComp package for Computational Wavelet Analysis in R. Wavelet power spectra were analyzed using the "white.noise" method, with a significance cutoff of 0.1 for contour, a period range of 0.2 to 3.2 Mb, and 100 simulations. The results of the wavelet power analyses were plotted using an interval color key (A, C, E, G). Reconstructed series were generated using only the periods whose average power was significant over the entire interval (B, D, F, H).

Replication-associated periodicity results from specific forms of base-substitution mutations. Nucleotide content varies across chromosomes and could conceivably underlie variation in bpsm rates among 100kb intervals. To address this question, we focus independently on A:T>G:C and G:C>A:T transitions in both the *Vf*-mut and *Vc*-mut studies, as these two forms of bpsm represent 97.93% and 98.34% of all of the bpsms that we observed in the *Vf*-mut and *Vc*-mut studies, respectively (Dillon et al. 2017). Although nucleotide composition *per se* did not vary significantly, the spectra of bpsms corrected for nucleotide content varied significantly among intervals on chr1 in both the *Vf*-mut and *Vc*-mut MA experiments (Chi-square test; A:T>G:C: *Vf*-mut - χ^2 = 62.26, df = 29, p = 0.0003, *Vc*-mut - χ^2 = 49.04, df = 29, p = 0.0110; G:C>A:T: *Vf*-mut - χ^2 = 120.69, df = 29, p < 0.0001, *Vc*-mut - χ^2 = 111.19, df = 29, p < 0.0001). In contrast, we cannot reject a uniform distribution of content corrected bpsms on chr2 for either experiment, with the exception of G:C>A:T substitutions in the *Vf*-mut study (Chi-square test; A:T>G:C: *Vf*-mut - χ^2 = 5.91, df = 15, p = 0.9810, *Vc*-mut - χ^2 = 7.93, df = 11, p = 0.7193; G:C>A:T: *Vf*-mut - χ^2 = 26.81, df = 15, p = 0.0300, *Vc*-mut - χ^2 = 13.73, df = 11, p = 0.2484), as was the case when we pooled all bpsms. Altogether, these observations suggest that after correcting for nucleotide content, neither A:T>G:C and G:C>A:T bpsm rates are uniformly distributed across the genome, which confirms that differences in nucleotide content between intervals cannot explain the systematic variation in bpsm rates.

We further explored the genome-wide rates of A:T>G:C and G:C>A:T transitions to examine whether the same mirrored patterns of bpsms on chr1 were observed for both individual bpsm types. Interestingly, we find that there is considerably more variation in G:C>A:T mutation rates on chr1 in both the *Vf*-mut and *Vc*-mut studies, and that positive correlations in bpsm rates on opposing replichores are driven by G:C>A:T mutation rates, not A:T>G:C mutation rates (Figure S3; Figure S4) (linear regression; A:T>G:C: *Vf*-mut - F = 1.77, df = 13, p = 0.2067, r^2 = 0.12, *Vc*-mut: F = 0.24, df = 13, p = 0.6313, r^2 = 0.02; G:C>A:T: *Vf*-mut - F = 13.32, df = 13, p = 0.0029, r^2 = 0.51, *Vc*-mut: F = 28.99, df = 13, p = 0.0001, r^2 = 0.69). However, as expected, neither A:T>G:C or G:C>A:T mutation rates on opposing replichores of chr2 are significantly correlated in *V. fischeri* or *V. cholerae* (Figure S3: Figure S4) (linear regression; A:T>G:C: *Vf*-mut - F = 3.26, df = 6, p = 0.1209, r^2 = 0.35, *Vc*-mut: F = 1.74, df = 4, p = 0.2574, r^2 = 0.30; G:C>A:T: *Vf*-mut - F = 0.17, df = 6, p = 0.6947, r^2 = 0.03, *Vc*-mut: F = 0.15, df = 4, p = 0.7209, r^2 = 0.04). Finally, despite visual similarity

between the A:T>G:C and G:C>A:T mutation rates of concurrently replicated regions between chr1 and chr2 (Figure S5), linear correlations were not significant (linear regression; A:T>G:C: *Vf*-mut - F = 0.03, df = 14, p = 0.8638, $r^2 = 0.02 \times 10^{-1}$, *Vc*-mut: F = 0.15, df = 10, p = 0.7061, $r^2 = 0.01$; G:C>A:T: *Vf*-mut - F = 2.43, df = 14, p = 0.1413, $r^2 = 0.15$, *Vc*-mut: F = 1.56, df = 10, p = 0.2409, $r^2 = 0.13$). These findings reaffirm that differences in nucleotide content between intervals cannot explain the wave-like periodicity in bpsm rates in either *Vf*-mut or *Vc*-mut and further suggest that these patterns are predominantly caused by G:C>A:T transitions.

Base-substitution mutation rates in wild-type MA lineages. Despite conducting longer MA experiments and sequencing more lineages in the wild-type MA experiments that we conducted in *V. fischeri*, *V. cholerae*, and *B. cenocepacia*, considerably fewer bpsms accumulated in these studies. Consequently, we cannot reject the null hypothesis that bpsms are uniformly distributed across chr1, chr2, and chr3 (for *Bc*) in the *Vf*-wt, *Vc*-wt, or *Bc*-wt MA experiments (Table S1), and although the patterns of bpsm rates in the *Vc*-wt experiment are positively correlated with the corresponding mutator experiment, that is not the case for the *Vf*-wt experiment (linear regression, 100 Kb intervals; *Vf*-wt – *Vf*-mut: F = 0.21, df = 42, p = 0.6491, $r^2 = 0.01$; *Vc*-wt – *Vc*-mut: F = 5.07, df = 38, p = 0.0303, $r^2 = 0.12$). The lack of significant variation in bpsm rates within chromosomes in wild-type MA lineages may have also limited detection of correlations between replichores on chr1 (Figure 6A, B, C). There is no significant relationship between the bpsm rates on the right replichore and concurrently replicated left replichore intervals of chr1 for *Vf*-wt, *Vc*-wt, or *Bc*-wt (linear regression, 100 Kb intervals; *Vf*-wt: F = 0.82, df = 13, p = 0.3810, $r^2 = 0.06$; *Vc*-wt: F = 0.11, df = 13, p = 0.7480, $r^2 = 0.01$; *Bc*-wt: F = 0.86, df = 16, p = 0.3670, $r^2 = 0.05$). Furthermore, there was no significant relationship between the patterns of bpsm rates on chr2 and those of late replicating regions on chr1 in any of these studies (Figure 6) (linear regression; *Vf*-wt: F = 0.16, df = 14, p = 0.7001, $r^2 = 0.01$; *Vc*-wt: F = 2.72, df = 10, p = 0.1300, $r^2 = 0.21$; *Bc*-wt: F = 0.32, df = 30, p = 0.5760, $r^2 = 0.01$).

It is nevertheless important to recognize that even at large interval lengths of 100 Kb, we observe an average of only 4.65 (0.38), 3.29 (0.28), and 3.08 (0.22) (SEM) bpsms per interval for the *Vf*-wt, *Vc*-wt, and *Bc*-wt MA lineages, respectively. Therefore, the lack of any significant intragenomic variation in bpsm rates

in the wild-type lineages may imply that the patterns we observe in the mismatch repair deficient lines are uniquely associated with strains that lack mismatch repair, or simply be a reflection of the reduced sample sizes in the wild-type lineages. We estimate effect sizes (*w*) of 0.2138 for chr1 in the *Vf*-mut experiments and 0.3631 for chr1 in the *Vc*-mut experiments, where mutation rates were non-uniformly distributed. If we assume that the wild-type lineages should experience similar variation in their bpsm rates as the corresponding mutator lineages, a chi-squared power analysis reveals that the 132 mutations that we observed on chr1 in the *Vf*-wt experiment would only have yielded a significantly non-uniform distribution of bpsms in only 19.46% of cases. Similarly, the 97 mutations that we observed on chr1 in the *Vc*-wt experiment would only have yielded a significantly non-uniform distribution of bpsms in 43.95% of cases. Although we lack a corresponding mutator experiment for *B. cenocepacia*, even if we assume an effect size as large as we observed in the *Vc*-mut lineages, we would still only have observed a significant non-uniform distribution of bpsms in 55.16% of cases based on the 130 bpsms that we collected on chr1. Collectively, these analyses suggest that longer or broader experiments with wild-type genomes are needed to capture more mutations and determine whether the periodicity in mutation rates seen in mutator lines also occurs in wild-type genomes.

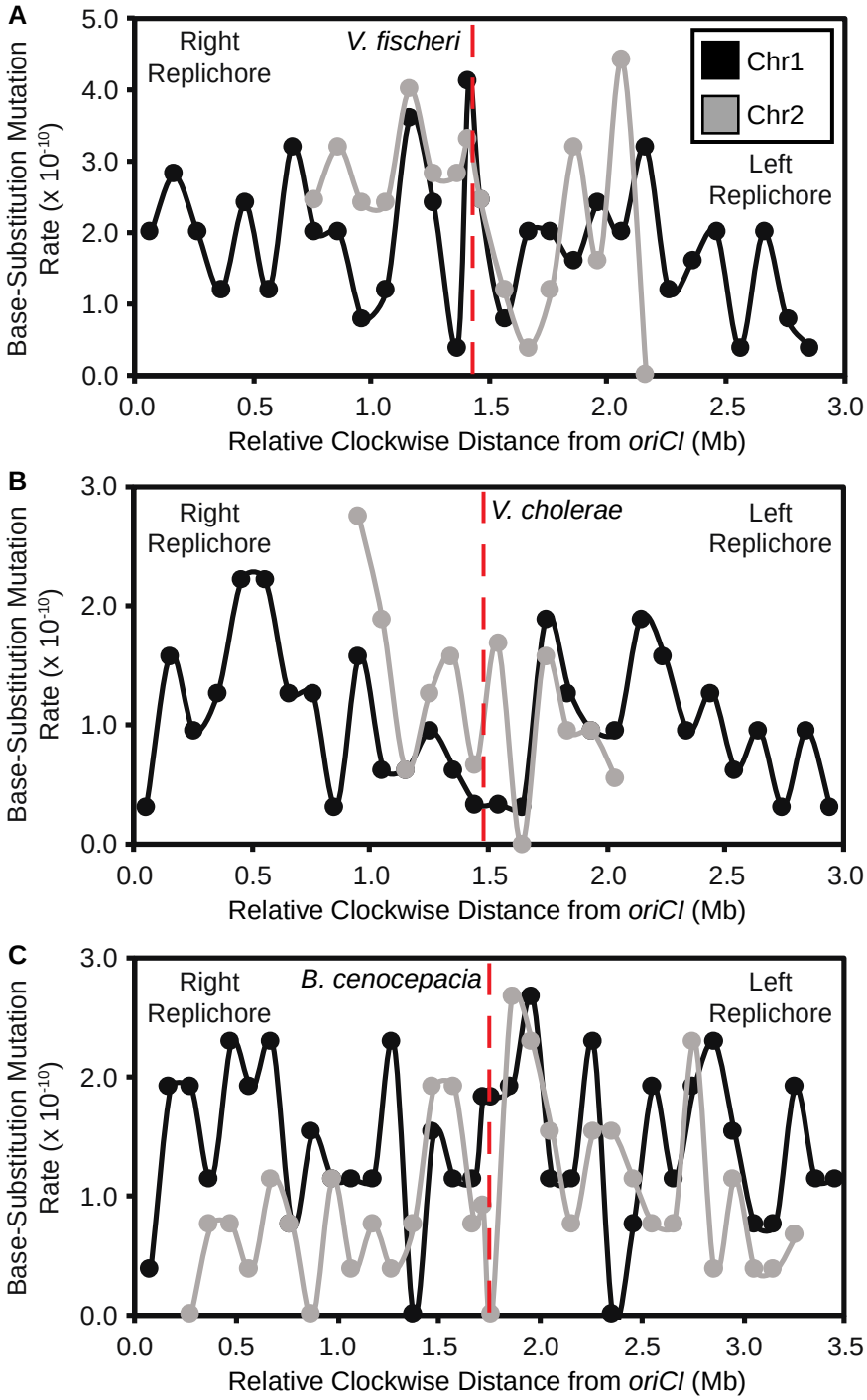

Figure 6. Patterns of base-substitution mutation (bpsm) rates in 100 Kb intervals extending clockwise from the origin of replication (*oriC*) on chromosome 1 (chr1) and concurrently replicated intervals of chromosome 2 (chr2) for MMR-proficient *Vibrio fischeri* (A), *Vibrio cholerae* (B), and *Burkholderia cenocepacia* (C). Bpsm rates were calculated as the number of mutations observed within each interval, divided by product of the total number of sites analyzed within that interval across all lines and the number of generations of mutation accumulation. *B. cenocepacia* also has a third chromosome, which is not shown.

DISCUSSION

Variation in mutation rates among genome regions can have important implications for genome evolution and diseases, including most cancers (Donley and Thayer 2013; Schuster-Böckler and Lehner 2012; Lawrence et al. 2013; Liu et al. 2013; Dame et al. 2011; Sobetzko et al. 2012; Dorman 2013). Recent evidence suggests that one of the most conserved properties of genome organization is the relative distance of genes from the origin of replication (Sobetzko et al. 2012), which is expected to result in the long-term conservation of traits like expression and mutation rates for genes harbored in divergent genomes. Consequently, molecular modifications that change genome wide-patterns of replication timing, expression, and mutation rates could increase the probability of acquiring defective alleles in typically conserved regions, leading to disease. Indeed, alteration of the replication timing program appears to be a very early step in carcinogenesis and a number of other somatic disease states (Donley and Thayer 2013). However, given the remarkable diversity in genome architecture across the tree of life, we still have much to learn about the nature of regional patterns of variation in bpsm rates and the genomic features and molecular processes that govern them.

Periodic variation in bpsm rates that is mirrored on the two replichores of bacterial chromosomes has been observed in genomes of some single-chromosome bacteria that are MMR-deficient (Foster et al. 2013; Dettman et al. 2016), yet not all MMR-deficient MA studies affirm these observations (Long et al. 2014), and the underlying causes of periodic variation in bacterial bpsm rates are unknown. Here we demonstrate that MMR-deficient bacterial genomes with multiple chromosomes do display mirrored, wavelike patterns of bpsm rates on chr1 (Figure 2A, B), and although we cannot reject the null hypothesis that bpsm rates are uniform on chr2, the patterns of bpsm rates on chr2 best match those of concurrently replicating regions on chr1 (Figure 4A, B). Furthermore, much of the genome-wide variation in bpsm rates that we observe appears to be generated by G:C>A:T transitions in both the *Vf*-mut and *Vc*-mut studies. Three MA experiments with MMR-proficient genomes may also experience regional variation in bpsm rates, but these studies were insufficiently powered to reject the null hypothesis of uniformity. Nonetheless, shared periodicities in mutation rates between replichores and coarse similarities across chromosome regions that are coordinately replicated suggests strongly that some processes that influence replication fidelity act throughout the genome across different active replication forks, which in our organisms causes bpsm to

begin at a minimum level near the replication origin, rise to roughly 2-4 times these rates, then decline and repeat this cycle before replication termination. More broadly, the finding that physically separate genome regions share common mutation rates because of their shared replication timing demonstrates the potential for shared forces of molecular evolution to operate in disparate genome regions with a common hand.

In principle, concurrently replicated genome regions within and between chromosomes may experience similar bpsm rates for several reasons. First, nucleotide context can generate heterogeneous bpsm rates because certain nucleotides or nucleotide sequences are more prone to incur bpsms than others (Sung et al. 2015; Dettman et al. 2014; Baer et al. 2007; Long et al. 2014). However, nucleotide composition varied little among the large genomic intervals analyzed in this study and most mutations were transitions (occurring equally in both directions) in both the *Vf*-mut and *Vc*-mut experiments (Dillon et al. 2017). Further, we show that when independently considering A:T>G:C and G:C>A:T transitions, which are corrected for nucleotide composition, there significant variation across chr1 in both the *Vf*-mut and *Vc*-mut studies. Rates of G:C>A:T transitions in particular vary widely among genome regions and exhibited highly significant correlations between concurrently replicated regions on chr1 in both *V. fischeri* and *V. cholerae*. Second, the replication machinery itself may generate heterogeneous bpsm rates because of biased usage of error prone polymerases (Courcelle 2009) or error prone repair pathways (Lang and Murray 2011) in certain genome regions. Both mechanisms have all been invoked to explain why substitution rates scale positively with replication timing (Chen et al. 2010; Stamatoyannopoulos et al. 2009; Herrick 2011; Mugal et al. 2010; Agier and Fischer 2012; Flynn et al. 2010; Cooper et al. 2010; Mira and Ochman 2002; Lang and Murray 2011), but it is difficult to imagine how they might create the mirrored wave-like patterns of bpsm rates observed in bacterial chromosomes across long intervals. Finally, a number of genomic features that also vary systematically with replication timing, like compaction of the bacterial nucleoid, binding of NAPs, and transcription could also generate the observed heterogeneous bpsm rates (Dorman 2013; Sobetzko et al. 2012; Dame et al. 2011; Schmidt et al. 2006; Warnecke et al. 2012). Indeed, negative DNA superhelicity does correlate positively with the mirrored wave-like patterns of bpsm rates on opposing replichores of *E. coli* (Foster et al. 2013), and patterns of extant sequence variation are impacted by NAPs in a growth phase-specific manner (Warnecke et al. 2012). Furthermore, sigma factors, DNA gyrase, and a number of

NAPs have mirrored patterns of activity on the right and left replichores of the single chromosome in *E. coli* (Sobetzko et al. 2012), possibly resulting from their concurrent replication. Transcription has also been speculated to impact bpsm rates through gene expression and replication-transcription conflicts (Paul et al. 2013), but oscillations in expression patterns and gene density are not consistent with concurrently replicated regions experiencing similar expression levels (Allen et al. 2006), and expression has not been significantly correlated with the patterns of bpsm rates in *E. coli* and other species (Foster et al. 2013; Lynch et al. 2016).

Although the relationship between some of these genomic and physiological features with mutation rates could be experimentally tested in future studies, perhaps the simplest and experimentally tractable dynamic that could produce the observed periodicity in bpsm rates is variation in levels of deoxyribonucleotides (dNTPs). Synthesis of dNTPs is governed primarily by the amount of active ribonucleotide reductase (RNR) enzyme, whose production is coordinated with the rate of DNA synthesis but reaches its maximum following a burst of NrdA and NrdB synthesis at the onset of DNA replication to meet demand (Sun and Fuchs 1992; Gon et al. 2006). High levels of dNTPs are known to be mutagenic in many organisms because of increased probability of misincorporation (Gon et al. 2006). In slow-growing bacteria whose division rates exceed the time required for chromosome replication, dNTP availability should peak after the start of replication and transiently increase the mutation rate; both nucleotides and mutation rates should then decline to steady state as replication proceeds and dNTP concentration is reduced. However, when bacterial generation times are faster than the time required for chromosome replication, which is commonplace for fast-growing *Vibrio* species that can double in <20 minutes, new rounds of replication are initiated and proceed before the first round concludes (Stokke et al. 2011). Further, in multi-chromosome genomes like those of *Vibrio* species, replication of secondary chromosomes requires an additional firing of an origin that increases dNTP demand and synthesis. Consequently, fast-growing bacteria with multi-chromosome genomes may experience multiple pulses of elevated dNTP pools, as origins repeatedly fire and stimulate bursts of RNR activity (Sun and Fuchs 1992), but the mutational effects of each successive pulse of dNTP synthesis should be diluted across a growing number of active replication forks. This dynamic could simply generate a wavelike pattern, whereby bpsm rates increase at the onset of replication

as RNR increases dNTP levels, and then decline steadily until the next origin fires (Figure S6). With each origin firing, this wave then repeats itself, though its amplitude may be diminished given the dilution of high dNTPs across multiple replication forks. It could also explain why not all bacteria appear to experience such a pattern (Long et al. 2014), because such pattern is not expected in single chromosome species that do not experience overlapping cycles of replication.

This hypothetical model of mutation rate periodicity brought on by varied replication processivity (Figure S6) could fit two major aspects of the chr1 bpsm waves shown in this study (Figure 2). First, the peaks in mutation rate seen after ~600kb and ~1200kb of replication in both *Vibrio* species can be explained by delayed maximal RNR activity that was activated by the onset of replication of Chr1 and Chr2, respectively. Second, a RNR activity peak associated with the onset of replication of Chr2 is predicted to be lower because dNTP levels per fork are halved, which is consistent with the lower second peak in bpsm rates observed in *V. cholerae* (Figure 4B) and the lower overall variance in mutation rates among late-replicated regions (Figure 4A, B). Importantly, single replication does not substantively change the overall dynamics predicted by our model, but may explain why some fast-growing single chromosome species like *E. coli* also experience mirrored wave-like patterns in bpsm rates around their single circular chromosomes. We acknowledge that this model is speculative and requires additional biochemical and genetic study, although the literature focused on relationships between RNR and mutation rates is vast (Mathews 2015). At a minimum, this simple model relating RNR activity with mutation-rate periodicity is empirically testable and may be revised or rejected by further studies using MA-WGS with defined mutants and altered growth conditions.

Yet another mechanism may more specifically explain the increased rate of G:C->A:T substitutions that underlies much of the periodicity observed here (Figure S5). It is conceivable that the plate growth conditions in these MA experiments promote the accumulation of reactive oxygen species (ROS), which leads to increased frequency of oxidized bases such as 8-oxo-G (Fuchs and Fujii 2013). This base modification commonly mispairs with A, which results in a G:C->A:T mutations. Such mutations are known to be corrected by MMR and thus are expected to be more common in MMR deficient MA lines. It remains

493 unclear how either the origin or failed repair of ROS-induced mutations would be periodic with respect to
494 replication timing, except that early-replicated nucleotides on Chr1 might be repaired more frequently by
495 alternative pathways like translesion synthesis (Fuchs and Fujii 2013) and access to these repair complexes
496 might also be diluted with each new round of replication. This hypothesis is also testable in future MA-WGS
497 experiments under conditions that alter the likelihood of ROS or with mutants deficient in alternative repair
498 pathways (Foster et al. 2015).
499
500 The conserved patterns of bpsm rates across concurrently replicated regions of MMR-deficient *Vf*-mut and
501 *Vc*-mut genomes further begs the question of whether these patterns contribute to extant sequence
502 variation in these genomes. If natural bpsm rates are in fact generally non-uniform and do reflect the
503 patterns of bpsm revealed in the MMR-deficient MA lineages, we would expect a positive correlation
504 between genetic variation among strains within these species and the bpsm rates in the 100 Kb intervals
505 used in this study. This correlation would be expected to be strongest on chr1, where the variation in bpsm
506 rates between 100 Kb intervals is significant. To test for a correlation between extant sequence variation
507 and bpsm rate, we calculated the average pairwise synonymous (dS) and non-synonymous (dN)
508 substitution rates in 100 Kb intervals *V. fischeri* ES114 and *V. cholerae* 2740-80 (see Methods). Indeed, we
509 find a significant positive correlation between dS, often used as a proxy to measure mutation rates, and the
510 bpsm rates in 100 Kb intervals on chr1 from the *Vf*-mut MA experiment ($F = 8.317$, $df = 28$, $p = 0.008$, $r^2 =$
511 0.229), but not on chr2 ($F = 0.558$, $df = 14$, $p = 0.468$, $r^2 = 0.038$) (Figure S7A). However, there is no
512 significant correlation between dN, often used as a proxy to measure evolutionary rates, and bpsm rates on
513 either chromosome (Figure S7). Furthermore, we find no significant correlations between dS or dN and
514 bpsm rates in 100 Kb intervals on either chromosome of *V. cholerae* (Figure S8). The lack of significant
515 correlations between dN and dS with bpsm rates in most cases here may simply reflect that selection
516 operating on both synonymous and non-synonymous sites is quite strong in bacteria (Lynch 2007).
517 Alternatively, it may suggest the natural patterns of bpsm rates in *V. fischeri* and *V. cholerae* are not
518 consistent with those observed in this study. Indeed, only 33.33% of the bpsms in the wild-type *V. fischeri*
519 MA experiment and 47.10% of bpsms in the wild-type *V. cholerae* MA experiment were G:C>A:T transitions
520 (Dillon et al. 2017), which we show are predominantly responsible for the wave-like patterns in bpsm rates

observed in this study. Assuming that most extant sequence variation was generated by wild-type mutation spectra, we would expect the wave-like patterns in bpsm rates to be dampened by other forms of spontaneous bpsms, particularly those in the A:T>G:C direction. Differentiating between the influence of selection and other bpsm types on patterns of extant sequence variation will require that additional mutation accumulation studies be carried out using different strains and under different environmental conditions for both species.

In summary, we have shown that in MMR-deficient lineages of bacteria with multiple chromosomes, bpsm rates are non-uniformly distributed on chr1, varying in a mirrored wave-like pattern extending bi-directionally from the origin of replication. In contrast, bpsm rates on chr2 are more constant, like those of late replicating regions on chr1. These observations suggest that concurrently replicated regions of bacterial genomes experience similar bpsm rates prior to MMR, which could be governed by a number of temporally regulated cellular processes, including variation in dNTP pools and the availability of replication machinery with secondary rounds of replication. Quantifying how different cellular processes influence bpsm rates (see for example (Strauss et al. 2017)) and identifying whether these patterns also apply to wild-type genomes in natural populations will undoubtedly shed new light on the nature of intra-genome variation in bpsm rates. Bacterial genomes are highly organized entities, and evidence now exists that one of the most conserved properties of this genome organization is their relative distance from the origin of replication and replication terminus (Sobetzko et al. 2012). Gene regulation by various NAPs has been invoked to explain this spatial organization, but this may ultimately expose certain genes to elevated deleterious mutational load. By revealing the underlying properties of variation in bpsm rates for a larger collection of bacteria, and determining the cellular processes that regulate them, we may ultimately be able to predict relative bpsm rates based on their proximity to the origin of replication and better understand how mutation risk is distributed across bacterial genomes.

METHODS

Bacterial strains and culture conditions. The founding strains of the five MA experiments conducted in this study were *V. fischeri* ES114 *ΔmutS* (Vf-mut), *V. cholerae* 2740-80 *ΔmutS* (*Vc*-mut), *V. fischeri* ES114

wild-type (*Vf*-wt), *V. cholerae* 2740-80 wild-type (*Vc*-wt), and *B. cenocepacia* HI2424 wild-type (*Bc*-wt). The mutator ancestors were generated by replacing the *mutS* gene in *V. fischeri* ES114 and *V. cholerae* 2740-80 with an erythromycin resistance cassette, as described previously (Val et al. 2012; Heckman and Pease 2007; Datsenko and Wanner 2000; Stabb and Ruby 2002). Completed genomes for *V. fischeri* ES114 and *B. cenocepacia* 2740-80 were downloaded from NCBI (Ruby et al. 2005; LiPuma et al. 2002) and the location of *oriCI*, *oriCII*, and *oriCIII* (if applicable) were downloaded from the dOriC 5.0 database (Gao et al. 2013). However, because the *V. cholerae* 2740-80 genome was only in draft form, we sequenced our *V. cholerae* 2740-80 ancestor using a long-insert library on single SMRT cell of a Pacific Biosciences RSII sequencer at the Icahn School of Medicine at Mount Sinai (Beaulaurier et al. 2015), then assembled the genome into two contigs using HGAP3 and Quiver, and identified the *oriCI* and *oriCII* regions of the resultant *V. cholerae* 2740-80 assembly using Ori-finder (Gao and Zhang 2008; Gao et al. 2013; Dillon et al. 2017) .

MA experiments with *Vf*-mut and *Vf*-wt were carried out on tryptic soy agar plates supplemented with NaCl (TSAN) (30 g/liter tryptic soy broth powder, 20 g/liter NaCl, 15 g/liter agar) and were incubated at 28°. MA experiments with *Vc*-mut, *Vc*-wt, and *Bc*-wt were carried out on tryptic soy agar plates (TSA) (30 g/liter tryptic soy broth powder, 15 g/liter agar) and were incubated at 37°. Frozen stocks of each MA lineage were prepared at the end of the experiment by growing a single colony overnight in 5 ml of tryptic soy broth supplemented with NaCl (TSBN) (30 g/liter tryptic soy broth powder, 20 g/liter NaCl) at 28° for *V. fischeri*, and in 5ml of tryptic soy broth (TSB) (30 g/liter tryptic soy broth powder) at 37° for *V. cholerae* and *B. cenocepacia*. Final isolates from all MA lineages in each experiment were stored in 8% DMSO at -80°.

MA-WGS Process. For each of the mutator MA experiments, forty-eight independent lineages were founded from single colonies of *V. fischeri* ES114 *ΔmutS* and *V. cholerae* 2740-80 *ΔmutS*. These lineages were independently propagated every 24 hours onto fresh media for 43 days and daily generations were estimated bi-monthly. For each of the three wild-type MA experiments, seventy-five independent lineages were founded from single colonies of *V. fischeri* ES114, *V. cholerae* 2740-80, and *B. cenocepacia* HI2424. These lineages were independently propagated every 24 hours onto fresh media for 217 days and daily

generations were estimated monthly. To estimate daily generation times, ten representative colonies following 24 hours of growth were placed in 2 ml of phosphate buffer saline, serially diluted, and spread plated on the appropriate media (see above) to calculate the number of viable cells in each colony. The number of generations elapsed over 24 hours of growth was then calculated, and the average number of generations across the ten representative colonies was used as the experiment-wide daily generations for each lineage at that time-point. The total generations elapsed between each measurement was calculated as the product of the average daily generations and the number of days before the next measurement, and the total of number of generations elapsed during the entire MA experiment, per lineage, was calculated as the sum of these totals. At the conclusion of each MA experiment, all lineages were stored at -80° in 8% DMSO and later revived for WGS.

Genomic DNA was extracted from 1 ml of overnight culture (TSBN at 28° for *V. fischeri*; TSB at 37° for *V. cholerae* and *B. cenocepacia*) in 50 representative lineages from the three wild-type experiments and 48 lineages from the two mutator experiments using the Wizard Genomic DNA Purification Kit (Promega). All libraries were prepared using a modified Illumina Nextera protocol designed for inexpensive library preparation of microbial genomes (Baym et al. 2015). Sequencing of the *Vf*-mut, *Vc*-mut, and *Bc*-wt lineages was performed using the 151-bp paired-end Illumina HISeq platform at the University of New Hampshire Hubbard Center for Genomic Studies, while sequencing for the *Vf*-wt and *Vc*-wt lineages was performed using the 101-bp paired-end Illumina HiSeq platform at the Beijing Genome Institute (BGI). In sum, we analyzed 19 *Vf*-mut lineages, 22 *Vc*-mut lineages, 48 *Vf*-wt lineages, 49 *Vc*-wt lineages, and 47 *Bc*-wt lineages, as fastQC revealed that the depth of coverage for the other sequenced lineages was insufficient for accurate detection of polymorphism (Andrews 2010). The reads from each of these lineages were mapped to their respective reference genomes with the Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009) and Novoalign (www.novocraft.com). The average depth of coverage was 124x for *Vf*-mut, 92x for *Vc*-mut, 100x for *Vf*-wt, 96x for *Vc*-wt, and 50x for *Bc*-wt.

Base-substitution mutation identification. For each MA experiment, bpsms were identified as described previously (Dillon et al. 2015, 2017). Briefly, we used SAMtools to convert the SAM alignment files from

each lineage to mpileup format (Li et al. 2009), then in-house perl scripts to produce the forward and reverse read alignments for each position in each line. A three-step process was then used to detect putative bpsms. First, pooled reads across all lines were used to generate an ancestral consensus base at each site in the reference genome. This allows us to correct for any differences that may exist between the reference genomes and the ancestral colony of each our MA experiments. Second, a lineage specific consensus base was generated at each site in the reference genome for each individual MA lineage using only the reads from that line. Here, a lineage specific consensus base was only called if the site was covered by at least two forward and two reverse reads and at least 80% of the reads identified the same base. Otherwise, the site was not analyzed. Third, each lineage specific consensus base that was called was compared to the overall ancestral consensus of the MA experiment and a putative bpsm was identified if they differed. This analysis was carried out independently with the alignments generated by BWA and Novoalign, and putative bpsms were considered genuine only if both pipelines independently identified the bpsm and they were only identified in a single lineage. All genuine bpsms analyzed in this study are summarized in Supplementary Dataset S1, which shows that nearly all bpsms were identified with high-confidence and were not clustered at the lower limits of detection.

Base-substitution mutation rate analysis at different interval lengths. Genomes were divided into intervals of 10 Kb, 25 Kb, 50 Kb, 100 Kb, 250 Kb, and 500 Kb, and bpsms were categorized into those intervals based on their location. On chr1, these intervals started at the origin of replication (*oriCI*), and extended bi-directionally to the replication terminus to mimic the progression of the two replication forks. Therefore, in each analysis, the final interval on each replichore of chr1 was smaller than the rest, since each half of chr1 is not exactly divisible by any of the intervals lengths. For example, chr1 of *V. fischeri* is 2,897,536 bps in length, so there are 1,448,768 bps on each replichore. Therefore, when we analyze bpsm rates in 100 Kb intervals from *oriCI*, the first fourteen intervals on each replichore are 100 Kb in length, while the final interval on each replichore is only 48.768 Kb in length.

On secondary chromosomes, bpsm rates were analyzed with the same interval sizes as chr1, but intervals were measured relative to the initiation of replication of *oriCI* rather than their own origin of replication

(*oriCII*) (Figure 1). This ensured that we could make direct comparisons between concurrently replicated intervals on chr1 and chr2 where the limits of each interval relative to *oriCI* initiation were identical based on the established models of secondary chromosome replication timing in *V. cholerae* (Rasmussen et al. 2007; Baek and Chattoraj 2014; Val et al. 2014). We achieved this by starting intervals bi-directionally from the replication terminus in the opposite direction of the replication forks, starting with the smaller intervals used in the corresponding interval analysis on chr1. All intervals were thus exactly the same length as those of the concurrently replicated intervals on chr1, with the exception of the first two intervals on each replichore, which extended into the concurrently replicated interval of chr1 until they reached their own origin of replication (*oriCII*). In the example presented above for *V. fischeri*, chr2 is 1,330,333 bps in length. Therefore, for 100 Kb intervals, the final interval on each replichore of chr2 was 48.768 Kb in length. These intervals were preceded by six 100 Kb intervals extending away from the terminus on each replichore, and finally, a single 16.398 Kb interval on each replichore, where the two replichores meet at *oriCII*. Bpsm rates in each interval were always calculated as the number of mutations observed in each interval, divided by the product of the total number of sites analyzed in that interval across all lines and the total number of generations of mutation accumulation, so even the bpsm rates in the truncated intervals could be directly compared to the full-size intervals. For independent analyses of A:T>G:C and G:C>A:T mutations, bpsm rates were calculated as the number of mutations observed in each interval, divided by the product of the total number of sites in that interval that could lead to the bpsm being analyzed (A+T sites for A:T>G:C; G+C sites for G:C>A:T) and the total number of generations of mutation accumulation.

Wavelet Transformations. We used WaveletComp, an open-source R package for wavelet analysis of univariate and bivariate series, to study the periodicity of the wave-like patterns in bpsm rates around the chromosomes of *V. fischeri* and *V. cholerae* and to ask whether waves on opposing replichores were synchronous (Roesch and Schmidbauer 2014)**.** We analyzed the periodicity of the variation in bpsm rates on each chromosome of the *Vf*-mut and *Vc*-mut lineages at an interval length of 100 Kb, treating each chromosome as a univariate series starting at the origin of replication and extending clockwise around the chromosome. Here, WaveletComp uses the Morlet wavelet to transform the series of mutation rates in successive 100 Kb intervals, then tests the null hypothesis of no periodicity for all combinations of intervals

and periods (Roesch and Schmidbauer 2014). We performed this analysis using the “white.noise” method, with no smoothing, and a period range of 0.2 Mb to the entire length of the respective chromosomes (0.2 to 3.2 Mb for Chr1, 0.2 to 1.6 Mb for Chr2). Default settings were used for all other parameters. To plot the results of this analysis on a wavelet power spectrum, we specified 250 color levels and used an interval color key.

To test whether opposing replichores on Chr1 where synchronous, we used a cross-wavelet transformation to compare the patterns of variation in bpsm rates extending clockwise from *oriCI* to those extending counter clockwise from *oriCI* (Roesch and Schmidbauer 2014). Here, we are testing the null-hypothesis of joint periodicity (synchronicity) between the two series as they traverse the circular primary chromosome in opposite directions. As with the univariate analysis, we used default settings, but turned off smoothing and specified a period range of 0.2 Mb to the entire length of Chr1 in both *V. fischeri* and *V. cholerae* (3.2 Mb). To plot the results of this analysis on a cross-wavelet power spectrum, we specified 250 color levels, used an interval color key, and applied cutoffs of $p < 0.1$ for contour and $p < 0.05$ for arrows.

Evolutionary rate analysis at different interval lengths. As we did with bpsm rates, we estimated extant sequence variation in the *V. fischeri* ES114 and *V. cholerae* 2740-80 genomes using the average synonymous (dS) and non-synonymous (dN) substitution rates of genes in each 100 Kb interval extending bi-directionally from the two origins of replication. Specifically, for *V. fischeri*, we measured dS and dN of all one-to-one orthologs shared between *V. fischeri* ES114 and *V. fischeri* MJ11, which have 83.66% symmetric identity. For *V. cholerae*, we measured dS and dN of all one-to-one orthologs shared between *V. cholerae* 2740-80 and *V. cholerae* HE-16, which have 84.54% symmetric identity. The genomes of *V. fischeri* MJ11 and *V. cholerae* HE-16 were chosen because they were sufficiently diverse relative to *V. fischeri* ES114 and *V. cholerae* 2740-80 to produce reliable evolutionary rate measurements, while also maintaining the synteny of the majority of the one-to-one orthologs.

To calculate evolutionary rates for all one-to-one orthologs, we first identified all orthologs using orthoMCL with default settings (Li et al. 2003). We then aligned all one-to-one orthologs using the Needleman-Wunsch

algorithm with a gap open penalty of 10 and a gap extend penalty of 0.5 (Rice et al. 2000), and converted these amino-acid alignments to nucleotide alignments using RevTrans (Wernersson and Pedersen 2003). Pairwise evolutionary rates were then computed in paml (Yang 2007), using the Yang and Nielson method for pairwise substitution rate estimates (Yang and Nielsen 2000). Lastly, any one-to-one orthologs with dS or dN values exceeding 2.0 were discarded because saturation can render these values unreliable. In sum, we used evolutionary rate estimates of 3,259 one-to-one orthologs shared by *V. fischeri* ES114 and *V. fischeri* MJ11, and 2,849 one-to-one orthologs shared by *V. cholerae* 2740-80 and *V. cholerae* HE-16. Estimates of dS and dN in 100 Kb intervals equivalent to those described above for bpsm rates were calculated as the average evolutionary rates of all one-to-one orthologs in each interval. In the rare cases where orthologs overlapped across two intervals, the ortholog was considered to be part of the interval in which its start codon resides.

Other Statistical analyses. All statistical analyses were performed in R Studio Version 1.1.383 using the Stats analysis package (R Development Core Team 2017).

DATA ACCESS

All of the whole-genome sequencing data produced by this study can be accessed on NCBI using the following Bioproject accession numbers: PRJNA256340 for *V. fischeri*, PRJNA256339 for *V. cholerae*, and PRJNA326274 for *B. cenocepacia*.

ACKNOWLEDGMENTS

This work was supported by the Multidisciplinary University Research Initiative Award from the US Army Research Office (W911NF-09-1-0444 to ML, P. Foster, H. Tang, and S. Finkel); a National Science Foundation CAREER Award (DEB-0845851 to VSC), and NASA Astrobiology Institute CAN-7 NNA15BB04A to VSC. We thank Cheryl Whistler, Randi Foxall, and Brian VanDam for technical support and Kevin Culligan, Greg Lang, and Jeffrey Lawrence for helpful discussion. M.D., W.S., M.L., and V.C. designed the research; M.D., and W.S. performed the research; M.D. analyzed the data; and M.D., and V.C. wrote the paper.

DISCLOSURE DECLARATION

The authors declare no conflict of interest.

REFRENCES

Agier N, Fischer G. 2012. The mutational profile of the yeast genome is shaped by replication. *Mol Biol Evol* **29**: 905–913.

Agnoli K, Schwager S, Uehlinger S, Vergunst A, Viteri DF, Nguyen DT, Sokol PA, Carlier A, Eberl L. 2012. Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Mol Microbiol* **83**: 362–378.

Allen TE, Price ND, Joyce AR, Palsson B. 2006. Long-range periodic patterns in microbial genomes indicate significant multi-scale chromosomal organization. *PLoS Comput Biol* **2**: e2.

Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data.

Baek JH, Chattoraj DK. 2014. Chromosome I controls chromosome II replication in Vibrio cholerae. *PLoS Genet* **10**: e1004184.

Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat Rev Genet* **8**: 619–631.

Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* **10**: e0128036.

Beaulaurier J, Zhang X, Zhu S, Sebra R, Rosenbluh C, Deikus G, Shen N, Munera D, Waldor MK, Chess A, et al. 2015. Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. *Nat Commun* **6**: 7438.

Chen C-L, Rappailles A, Duquenne L, Huvet M, Guilbaud G, Farinelli L, Audit B, D'Aubenton-Carafa Y, Arneodo A, Hyrien O, et al. 2010. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res* **20**: 447–457.

Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. 2010. Why genes evolve faster on secondary chromosomes in bacteria. *Plos Comput Biol* **6**: e1000732.

Courcelle J. 2009. Shifting replication between IInd, IIIrd, and IVth gears. *Proc Natl Acad Sci U S A* **106**: 6027–6028.

Dame RT, Kalmykowa OJ, Grainger DC. 2011. Chromosomal macrodomains and associated proteins: Implications for DNA organization and replication in gram negative bacteria. *PLoS Genet* **7**.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640–6645.

Dettman JR, Rodrigue N, Kassen R. 2014. Genome-wide patterns of recombination in the opportunistic human pathogen *Pseudomonas aeruginosa*. *Genome Biol Evol* **7**: 18–34.

Dettman JR, Sztepanacz JL, Kassen R. 2016. The properties of spontaneous mutations in the opportunistic pathogen *Pseudomonas aeruginosa*. *BMC Genomics* **17**: 27–41.

Dillon MM, Sung W, Lynch M, Cooper VS. 2015. The rate and molecular spectrum of spontaneous mutations in the GC-rich multichromosome genome of *Burkholderia cenocepacia*. *Genetics* **200**: 935–946.

Dillon MM, Sung W, Lynch M, S. C V. 2017. Genome-wide biases in the rate and molecular spectrum of spontaneous mutations in *Vibrio cholerae* and *Vibrio fischeri*. *Mol Biol Evol* **34**: 93–109.

Donley N, Thayer MJ. 2013. DNA replication timing, genome stability and cancer late and/or delayed DNA replication timing is associated with increased genomic instability. *Semin Cancer Biol* **23**: 80–89.

Dorman CJ. 2013. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat Rev Microbiol* **11**: 349–355.

Duigou S, Knudsen KG, Skovgaard O, Egan ES, Lobner-Olesen A, Waldor MK. 2006. Independent control of replication initiation of the two *Vibrio cholerae* chromosomes by DnaA and RctB. *J Bacteriol* **188**: 6419–6424.

Egan ES, Fogel MA, Waldor MK. 2005. Divided genomes: Negotiating the cell cycle in

prokaryotes with multiple chromosomes. *Mol Microbiol* **56**: 1129–1138.
Egan ES, Waldor MK. 2003. Distinct replication requirements for the two *Vibrio cholerae* chromosomes. *Cell* **114**: 521–530.
Flynn KM, Vohr SH, Hatcher PJ, Cooper VS. 2010. Evolutionary rates and gene dispensability associate with replication timing in the archaeon *Sulfolobus islandicus*. *Genome Biol Evol* **2**: 859–869.
Foster PL, Hanson AJ, Lee H, Popodi EM, Tang HX. 2013. On the mutational topology of the bacterial genome. *G3-Genes Genomes Genet* **3**: 399–407.
Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. *Proc Natl Acad Sci* **112**: E5990–E5999.
Fuchs RP, Fujii S. 2013. Translesion DNA synthesis and mutagenesis in prokaryotes. *Cold Spring Harb Perspect Biol* **5**: 1–22.
Gao F, Luo H, Zhang CT. 2013. DoriC 5.0: an updated database of *oriC* regions in both bacterial and archaeal genomes. *Nucleic Acids Res* **41**: D90–D93.
Gao F, Zhang C-T. 2008. Ori-Finder: a web-based system for finding *oriCs* in unannotated bacterial genomes. *BMC Bioinformatics* **9**: 79–85.
Gon S, Camara JE, Klungsøyr HK, Crooke E, Skarstad K, Beckwith J. 2006. A novel regulatory mechanism couples deoxyribonucleotide synthesis and DNA replication in *Escherichia coli*. *EMBO J* **25**: 1137–1147.
Heckman KL, Pease LR. 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2**: 924–932.
Herrick J. 2011. Genetic variation and DNA replication timing, or why is there late replicating DNA? *Evolution (N Y)* **65**: 3031–3047.
Lang GI, Murray AW. 2011. Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol Evol* **3**: 799–811.
Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts S a, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**: 214–8.
Lee H, Popodi E, Tang HX, 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**: E2774–E2783.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
Li L, Stoeckert CJJ, Roos DS. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**: 2178–2189.
LiPuma JJ, Spilker T, Coenye T, Gonzalez CF. 2002. An epidemic *Burkholderia cepacia* complex strain identified in soil. *Lancet* **359**: 2002–2003.
Liu L, De S, Michor F. 2013. DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. *Nat Commun* **4**: 1–9.
Long H, Sung W, Miller SF, Ackerman MS, Doak TG, Lynch M. 2014. Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair (MMR) deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol* **7**: 262–271.
Lynch M. 2007. *The origins of genome architecture*. Sinauer Associates, Sunderland (MA).
Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* **17**: 704–714.
Mathews CK. 2015. Deoxyribonucleotide metabolism, mutagenesis and cancer. *Nat Rev Cancer* **15**: 528–539.
Mira A, Ochman H. 2002. Gene location and bacterial sequence divergence. *Mol Biol Evol* **19**:

1350–1358.
Morrow JD, Cooper VS. 2012. Evolutionary effects of translocations in bacterial genomes. *Genome Biol Evol* **4**: 1256–1262.
Mugal CF, Wolf JBW, Von Grünberg HH, Ellegren H. 2010. Conservation of neutral substitution rate and substitutional asymmetries in mammalian genes. *Genome Biol Evol* **2**: 19–28.
Ochman H. 2002. Bacterial evolution: Chromosome arithmetic and geometry. *Curr Biol* **12**: R427–R428.
Ochman H. 2003. Neutral mutations and neutral substitutions in bacterial genomes. *Mol Biol Evol* **20**: 2091–2096.
Paul S, Million-Weaver S, Chattopadhyay S, Sokurenko E, Merrikh H. 2013. Accelerated gene evolution through replication-transcription conflicts. *Nature* **495**: 512–515.
R Development Core Team. 2017. R: A Language and Environment for Statistical Computing.
Rasmussen T, Jensen RB, Skovgaard O. 2007. The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle. *Embo J* **26**: 3124–3131.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276–277.
Roesch A, Schmidbauer H. 2014. WaveletComp: Computational Wavelet Analysis. R package version 1.0. *https://CRANR-project.org/package=WaveletComp*.
Ruby EG, Urbanowski M, Campbell J, Dunn A, Faini M, Gunsalus R, Lostroh P, Lupp C, McCann J, Millikan D, et al. 2005. Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. *Proc Natl Acad Sci U S A* **102**: 3004–3009.
Schmidt KH, Reimers JM, Wright BE. 2006. The effect of promoter strength, supercoiling and secondary structure on mutation rates in *Escherichia coli*. *Mol Microbiol* **60**: 1251–1261.
Schuster-Böckler B, Lehner B. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**: 504–507.
Sobetzko P, Travers A, Muskhelishvili G. 2012. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci U S A* **109**: E42–E50.
Stabb E V., Ruby EG. 2002. RP4-based plasmids for conjugation between *Escherichia* coli and members of the *Vibrionaceae*. *Methods Enzymol* **358**: 413–426.
Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov G V, Mirkin SM, Sunyaev SR. 2009. Human mutation rate associated with DNA replication timing. *Nat Genet* **41**: 393–395.
Stokke C, Waldminghaus T, Skarstad K. 2011. Replication patterns and organization of replication forks in *Vibrio cholerae*. *Microbiology* **157**: 695–708.
Strauss C, Long H, Patterson CE, Te R, Lynch M. 2017. Genome-wide mutation rate response to pH change in the coral reef. *MBio* **8**: e01021-17.
Sun L, Fuchs JA. 1992. *Escherichia coli* Rribonucleotide reductase expression is cell cycle regulated. *Mol Biol Cell* **3**: 1095–1105.
Sung W, Ackerman MS, Gout J-F, Miller SF, Williams E, Foster PL, Lynch M. 2015. Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. *Mol Biol Evol* **32**: 1672–1683.
Val M-E, Soler-Bistué A, Bland MJ, Mazel D. 2014. Management of multipartite genomes: the Vibrio cholerae model. *Curr Opin Microbiol* **22**: 120–126.
Val ME, Skovgaard O, Ducos-Galand M, Bland MJ, Mazel D. 2012. Genome engineering in *Vibrio cholerae*: A feasible approach to address biological issues. *Plos Genet* **8**: e1002472.
Warnecke T, Supek F, Lehner B. 2012. Nucleoid-associated proteins affect mutation dynamics in *E. coli* in a growth phase-specific manner. *PLoS Comput Biol* **8**: e1002846.
Wernersson R, Pedersen AG. 2003. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res* **31**: 3537–3539.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* **17**: 32–43.
Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.
Zhang XL, Mathews CK. 1995. Natural DNA precursor pool asymmetry and base sequence context as determinants of replication fidelity. *J Biol Chem* **270**: 8401–8404.

79