Article - New Results

The rate and molecular spectrum of spontaneous mutations in the GC-rich multi-chromosome genome of *Burkholderia cenocepacia*

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

[a]Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, 105 Main Street, Durham, NH, 03824

[b]Department of Biology, Indiana University, Bloomington, IN, 47405

Corresponding Author:
Vaughn Cooper
46 College Road, 212 Rudman Hall
Durham, NH 03824
Phone: (603)-862-3422
Email: vaughn.cooper@unh.edu

Keywords: Mutation rate; Mutation Spectra; genome evolution; replication timing

Author Contributions: M.D., W.S., M.L., and V.C. designed the research; M.D., and W.S. performed the research; M.D., and W.S. analyzed the data; and M.D., W.S., M.L., and V.C. wrote the paper.

ABSTRACT

Spontaneous mutations are ultimately essential for evolutionary change and are also the root cause of nearly all disease. However, until recently, both biological and technical barriers have prevented detailed analyses of mutation profiles, constraining our understanding of the mutation process to a few model organisms and leaving major gaps in our understanding of the role of genome content and structure on mutation. Here, we present a genome-wide view of the molecular mutation spectrum in *Burkholderia cenocepacia*, a clinically relevant pathogen with high %GC content and multiple chromosomes. We find that *B. cenocepacia* has low genome-wide mutation rates with insertion-deletion mutations biased towards deletions, consistent with the idea that deletion pressure reduces prokaryotic genome sizes. Unlike previously assayed organisms, *B. cenocepacia* exhibits a GC-mutation bias, which suggests that at least some genomes with high GC content may be driven to this point by unusual base-substitution mutation pressure. Notably, we also observed variation in both the rates and spectra of mutations among chromosomes, and a significant elevation of G:C>T:A transversions in late-replicating regions. Thus, although some patterns of mutation appear to be highly conserved across cellular life, others vary between species and even between chromosomes of the same species, potentially influencing the evolution of nucleotide composition and genome architecture.

INTRODUCTION

As the ultimate source of genetic variation, mutation is implicit in every aspect of genetics and evolution. However, as a result of the genetic burden imposed by deleterious mutations, remarkably low mutation rates have evolved across all of life, making detection of these rare events technologically challenging and accurate measures of mutation rates and spectra exceedingly difficult (Kibota and Lynch 1996; Lynch and Walsh 1998; Sniegowski et al. 2000; Lynch 2011; Fijalkowska et al. 2012; Zhu et al. 2014). Consequently, most estimates of mutational properties have been derived indirectly using comparative genomics at putatively neutral sites (Graur and Li 2000; Wielgoss et al. 2011) or by extrapolation from small reporter-construct studies (Drake 1991). Both of these methods are subject to potentially significant biases, as many putatively neutral sites are subject to selection and mutation rates can vary substantially among different genomic regions (Lynch 2007).

To avoid the potential biases of these earlier methods, pairing classic mutation accumulation (MA) with whole-genome sequencing (WGS) has become the preferred method for obtaining direct measures of mutation rates and spectra (Lynch et al. 2008; Denver et al. 2009; Ossowski et al. 2010; Lee et al. 2012; Sung, Ackerman, et al. 2012; Sung, Tucker, et al. 2012; Heilbron et al. 2014). Using this strategy, a single clonal ancestor is used to initiate several replicate lineages that are passaged through repeated single-cell bottlenecks for several thousand generations. The complete genomes of each evolved lineage are then sequenced and compared with the other lines to identify *de novo* mutations that occurred over the course of the experiment. The bottlenecking regime minimizes the ability of natural selection to eliminate deleterious

mutations, and the parallel sequencing provides a large enough body of information to yield a nearly unbiased picture of the natural mutation spectrum of the study organism (Lynch et al. 2008).

The MA-WGS method has now been used to examine mutational processes in several model eukaryotic and prokaryotic species, yielding a number of apparently generalizable conclusions about mutation rates and spectra. For example, a negative scaling between base-substitution mutation rates and both effective population size (N_e) and the amount of coding DNA supports the hypothesis that the refinement of replication fidelity that can be achieved by selection is determined by the power of random genetic drift among phylogenetic lineages (Lynch 2011; Sung, Ackerman, et al. 2012). This “drift-barrier hypothesis” therefore predicts that organisms with very large population sizes such as some bacteria should have evolved very low mutation rates (Lee et al. 2012; Sung, Ackerman, et al. 2012; Foster et al. 2013).

Universal transition and G:C>A:T biases have also been observed in all MA studies to date (Lind and Andersson 2008; Lynch et al. 2008; Denver et al. 2009; Ossowski et al. 2010; Lee et al. 2012; Sung, Ackerman, et al. 2012; Sung, Tucker, et al. 2012), corroborating previous findings using indirect methods (Hershberg and Petrov 2010; Hildebrand et al. 2010). However, several additional characteristics of mutation spectra vary among species (Lynch et al. 2008; Denver et al. 2009; Ossowski et al. 2010; Lee et al. 2012; Sung, Tucker, et al. 2012; Sung, Ackerman, et al. 2012), and examining the role of genome architecture, size, and lifestyle in producing these idiosyncrasies will require a considerably larger number of detailed MA studies. Among bacterial species that have been subjected to mutational studies, genomes with high

GC content are particularly sparse and no studies have been conducted on bacteria with multiple chromosomes, a genome architecture of many important bacterial species (e.g *Vibrio*, *Brucella*, *Burkholderia*).

The *Burkholderia cepacia* complex is a diverse group of bacteria with important clinical implications for patients with cystic fibrosis (CF), in whom it can form persistent lung infections and highly resistant biofilms (Coenye et al. 2004; Mahenthiralingam et al. 2005; Traverse et al. 2013). *Burkholderia cenocepacia* is the most threatening pathogenic member of this complex in CF patients, and is renowned for its rapid diversification following infection (Coenye et al. 2004; Zlosnik et al. 2011). The core genome of *B. cenocepacia* HI2424 has a high %GC content (66.8%) and harbors three chromosomes, each containing rDNA operons (LiPuma et al. 2002), although the third chromosome can be eliminated under certain conditions (Agnoli et al. 2012). The primary chromosome (Chr1) is ~3.48 Mb and contains 3253 genes; the secondary chromosome (Chr2) is ~3.00 Mb and contains 2709 genes; and the tertiary chromosome (Chr3) is ~1.06 Mb and contains 929 genes. In addition, *B. cenocepacia* HI2424 contains a 1.64 Kb plasmid, which contains 159 genes and lower GC content than the core genome (62.0%). Although the GC content is consistent across the three core chromosomes, the proportion of coding DNA declines from Chr1 to Chr3, while the evolutionary rate of genes increases (Cooper et al. 2010; Morrow and Cooper 2012). Whether this variation in evolutionary rate is driven by variation in non-adaptive processes like mutation bias or variation in the relative strength of purifying selection remains a largely unanswered question in the evolution of bacteria with multiple chromosomes.

Here, we applied whole-genome sequencing to 47 MA lineages derived from *B. cenocepacia* HI2424 that were evolved in the near absence of natural selection for over 5550 generations each. We identified a total of 282 mutations spanning all three replicons and the plasmid, enabling a unique perspective on inter-chromosomal variation in both mutation rate and spectra, in a bacterium with the highest %GC content studied with MA-WGS to date.

RESULTS

A classic mutation-accumulation experiment was carried out for 217 days with 75 independent lineages all derived from the same ancestral colony of *B. cenocepacia* HI2424 (LiPuma et al. 2002) using a daily serial transfer regime in which a single colony from each line was re-streaked onto a fresh plate. Measurements of generations incurred each day were taken monthly and varied from 26.2 ± 0.12 to 24.9 ± 0.14 (mean ± 95% CI of highest and lowest measurements, respectively) (Figure S1), resulting in an average of 5554 generations per line over the course of the MA experiment. Thus, across the 47 lines whose complete genomes were sequenced, we were able to visualize the natural mutation spectrum of *B. cenocepacia* HI2424 over 261,038 generations of mutation accumulation.

Whole-genome sequencing was performed using the 151-bp paired-end Illumina HiSeq platform to an average depth of ~50x. Mutations were identified using a consensus approach that leverages the parallel sequencing of nearly isogenic lineages to verify the ancestral consensus base at each site in the reference genome, then compare that to base calls of individual lineages. This approach allows us to minimize

false-positive identifications while missing few true mutations, as evidenced by previous studies that have verified mutations called by this method through conventional sequencing (Sung, Ackerman, et al. 2012; Sung, Tucker, et al. 2012). From the comparative sequence data, we identified 245 base-substitutional (bps) changes, 33 short-insertion/deletion (indel) mutations (with sizes in the range of 1 to 145 base pairs), and four plasmid-loss events spanning the entire genome (Figure 1, Table S1, S2). With means of 5.21 bps and 0.70 indel mutations per line, the distribution of bps and indels across individual lines did not differ significantly from a Poisson distribution (bps: χ^2 = 1.81, p = 0.99; indels: χ^2 = 0.48, p = 0.92), indicating that mutation rates did not vary over the course of the MA experiment.

Mutation-accumulation experiments rely on the basic principle that when the effective population size (N_e) is sufficiently reduced, the efficiency of selection is minimized to the point at which all mutations become fixed by genetic drift with equal probability (Kibota and Lynch 1996). N_e in this mutation accumulation study was calculated to be ~12.86, using the harmonic mean of the population size over 24 hours of colony growth (Hall et al. 2008). Thus, only mutations conferring effects of s > 0.078 will be subject to the biases of natural selection (Lynch et al. 2008), which is expected to be a very small fraction of mutations (Kibota and Lynch 1996; Elena et al. 1998; Zeyl and DeVisser 2001).

We tested for selection in our observed mutation spectra using the ratio of synonymous substitutions per synonymous site to non-synonymous substitutions per non-synonymous site. Given the codon usage and conditional mutation rates of *B. cenocepacia* HI2424, 27.8% of coding substitutions are expected to be synonymous.

The observed percentage of synonymous substitutions (25.5%) did not differ significantly from this null-expectation (χ^2 = 0.54, df = 1, p = 0.46). Although both base-substitutions (χ^2 = 4.20, df = 1, p = 0.04) and indels (χ^2 = 21.3, df = 1, p < 0.0001) were biased to non-coding DNA, evidence exists that mismatch repair preferentially repairs damage in coding regions, which can create artificial signatures of selection in MA experiments (Lee et al. 2012). Thus, our overall observations are consistent with the fact that MA experiments induce limited selection on the mutation spectra, at least as far as base substitutions are concerned.

Low base-substitution and indel mutation rates. The preceding results imply that base-substitution and indel mutation rates for *B. cenocepacia* are 1.33 (0.008) × 10^{-10} /bp/generation and 1.68 (0.003) × 10^{-11}/bp/generation (SEM), respectively. Based on the 7.70 Mb genome size, these per-base mutation rates correspond to a genome-wide base-substitution mutation rate of only 0.0010/genome/generation, and an indel mutation rate of only 0.00013/genome/generation. Although the ~1:3 ratio of synonymous to non-synonymous substitutions is consistent with negligible influence of selection on base-substitution mutations in this study, too few indels occurred to evaluate a signature of selection, although their scarcity could reflect some selective loss of genotypes with loss-of-function mutations (Heilbron et al. 2014; Zhu et al. 2014). Moreover, although we applied PINDEL to identify indels of any size based on the aberrant mapping of paired-end reads, intermediate and large indels cannot be identified with our multi-aligner consensus method using short-read aligners, making them more difficult to accurately assign than base-substitutions or short indels. Thus,

because some indel mutations may have been either purged by selection or overlooked by our analysis, our estimate of the indel rate should be considered a lower limit.

Base-substitution mutations are not GC>AT biased. One of the central motivations for studying the molecular mutation spectrum of *B. cenocepacia* was its high GC content (66.8%). A universal mutation bias in the direction of AT has been observed in all other wild-type species studied by MA, and has also been inferred in comparative analyses of several bacterial species, including *Burkholderia pseudomallei* (Lynch et al. 2008; Denver et al. 2009; Hershberg and Petrov 2010; Hildebrand et al. 2010; Ossowski et al. 2010; Lee et al. 2012; Sung, Tucker, et al. 2012). If this G:C>A:T mutation bias extends to *B. cenocepacia,* biased gene conversion or selection in the direction of GC content would have to occur (Lynch et al. 2008; Duret and Galtier 2009; Raghavan et al. 2012; Zhu et al. 2014).

In comparing the relative mutation rates of G:C>A:T transitions and G:C>T:A transversions with those of A:T>G:C transitions and A:T>C:G transversions, corrected for the ratio of G:C to A:T sites analyzed in this study, we found no mutational bias in the A:T direction. Rather, substitutions in the G:C direction were 17% more frequent than mutations in the A:T direction per base pair, although the rates were not significantly different (χ^2 = 0.91, df = 1, p = 0.33). The lack of mutational bias in the A:T direction can largely be attributed to A:T>C:G transversions occurring at significantly higher rates than any other transversion type, most notably the G:C>T:A transversions (χ^2 = 8.68, df = 1, p = 0.0032). However, A:T>G:C transitions also occurred at nearly the same rate as G:C>A:T transitions, the latter of which have been the most commonly

observed substitution in other studies, putatively due to deamination of cytosine or 5-methyl-cytosine (Figure 2) (Lee et al. 2012; Sung, Tucker, et al. 2012; Zhu et al. 2014).

Using the ratio of the conditional rate of mutation in the G:C direction to that in the A:T direction (x), the expected GC content under mutation-drift equilibrium is x/(1+x) = 0.539 ± 0.043 (SEM). Therefore, it is clear that the observed mutation bias is not sufficient to drive the overall GC content of 66.8%. Either the *B. cenocepacia* genome is still moving towards mutation-drift equilibrium, or GC-biased gene conversion and/or natural selection are responsible for the observed %GC content (Lynch et al. 2008; Duret and Galtier 2009; Raghavan et al. 2012; Zhu et al. 2014).

Deletion bias favors genome-size reduction and AT composition. Although our lower bound estimates of the insertion and deletion mutation rates are both ~15-fold lower than the base-substitution mutation rate, many indels affect more than one base. Specifically, the 17 deletions observed in this study result in the deletion of a total of 376 bases, while the 16 insertions result in a gain of 121 bases. Therefore, the number of bases that are impacted by indels in this study is still more than twice the number impacted by bps, indicating that indels may still play a central role in the genome evolution of *B. cenocepacia* if they are not purged by natural selection.

As noted above, of the 33 short indels observed in this study, 17 were deletions and 16 were insertions, suggesting that small-scale insertions and deletions occur with similar probability in *B. cenocepacia*. However, the average size of deletions was higher than the average size of insertions, leading to an experiment-wide deletion and insertion rates of 1.97 (0.86) × 10^{-10} and 6.11 (1.90) × 10^{-11}/bp/generation (SEM). Thus, there is a

net deletion rate of 1.36 (5.95) × 10^{-10}/bp/generation (Table 1). Although no indels >150 bp were observed in this study, examining the depth of coverage of the *B. cenocepacia* HI2424 plasmid relative to the rest of the genome revealed that the plasmid was lost at a rate of 1.53 × 10^{-5} per cell division, while gains in plasmid copy number were not observed (Table 1).

The base composition of deletions was also biased, with GC bases being deleted significantly more than expected based on the genome content (χ^2 = 30.4, df = 1, $p < 0.0001$). In contrast, no detectable bias was observed towards insertions of GC over AT bases (χ^2 = 1.20, df = 1, p = 0.27) (Table 1). Thus, indels in *B. cenocepacia* are expected to reduce genome wide GC content, further supporting the implied need for other population-genetic processes favoring GC content (Lynch et al. 2008; Duret and Galtier 2009; Raghavan et al. 2012; Zhu et al. 2014). Overall, the greater number of bases that were deleted than inserted in this study suggests that the natural indel spectrum of *B. cenocepacia* causes both genome-size reduction and increased AT content.

Non-uniform chromosomal distribution of mutations. Another major goal of this study was to investigate whether mutation rates and spectra vary among chromosomes and chromosomal regions. The three core chromosomes of *B. cenocepacia* vary in size and content but are sufficiently large to have each accumulated a considerable number of mutations in this study (Morrow and Cooper 2012). Chromosome 1 (chr1) is the largest chromosome (both in size and in gene count), with more essential and highly expressed genes than either chromosome 2 (chr2) or 3 (chr3) (Figure S4). Expression

and number of essential genes are second highest on chr2 and lowest on chr3 (Cooper et al. 2010; Morrow and Cooper 2012). In contrast, average non-synonymous and synonymous variation among orthologs shared by multiple strains of *B. cenocepacia,* as well as fixed variation among *Burkholderia* species (dN and dS), are highest on chr3 and lowest on chr1 (Figure S4) (Cooper et al. 2010; Morrow and Cooper 2012).

The base-substitution mutation rates of the three core chromosomes differ significantly based on a chi-square proportions test, where the null expectation was that the number of substitutions would be proportional to the number of sites covered on each chromosome (χ^2 = 6.77, df = 2, p = 0.034) (Figure 3A). Specifically, base-substitution mutation rates are highest on chr1, and lowest on chr2, which is the opposite of observed evolutionary rates on these chromosomes (Figure S4) (Cooper et al. 2010). There was moderate variation in the ratio of GC to AT base-pairs covered on each chromosome, and because AT bases experience slightly higher mutation rates overall than GC bases in *B. cenocepacia* (Figure 2), we set up a second chi-squared test to test whether the inter-chromosomal variation in substitution rates could be due to variation in nucleotide content. Here, the null expectation for the frequency of base-substitutions expected on each chromosome was calculated by taking the product of the number of GC bases covered across all lines, the number of generations incurred per line, and the overall GC substitution rate across the genome. The resultant product was then added to the product of the same calculation for AT substitutions to obtain the total expected number of substitutions on each chromosome, given both their size and nucleotide content. The differences in the base-substitution mutation rates of the three core chromosomes remained significant when this test was performed (χ^2 = 6.88, df = 2,

p = 0.032), indicating that the intra-chromosomal heterogeneity in base-substitution mutation rates cannot be explained by variation in nucleotide content.

The conditional base-substitution mutation spectra were also significantly different in all pairwise chi-squared proportions tests between chromosomes (chr1/chr2: χ^2=14.3, df=5, p=0.014; chr1/chr3: χ^2=17.0, df=5, p=0.004; chr2/chr3: χ^2=13.4, df=5, p=0.020) (Figure 3C). These comparisons further illustrate that the significant variation in conditional base-substitution mutation rates is mostly driven by a few types of substitutions that occur at higher conditional rates on particular chromosomes. Specifically, although their individual differences were not quite statistically significant, G:C>T:A transversions seem to occur at the highest rate on chr3 (χ^2 = 5.94, df = 2, p = 0.051) and A:T>C:G transversions occur at the highest rate on chr1 (χ^2 = 5.67, df = 2, p = 0.059) (Figure 3B; Figure 4A).

Unlike base-substitution mutation rates, neither the deletion or insertion mutation rate varied significantly among chromosomes (Deletions: χ^2=3.81, df=2, p=0.15; Insertions: χ^2=0.64, df=2, p=0.73), (Figure 3B; Figure 4B). No indels were observed on the 0.16 Mb plasmid, but as noted above, four plasmid loss events were observed. The latter events involve the loss of 157 genes, and are expected to have phenotypic consequences. The relative rarity of indels observed in this study limits our ability to analyze their intra-chromosomal biases in great detail, but the repeated occurrence of indels within microsatellites (57.6% of all indels) suggests that replication slippage is a common cause of indels in the *B. cenocepacia* genome (Figure 4B).

DISCUSSION

Despite their relevance to both evolutionary theory and human health, the extent to which generalizations about mutation rates and spectra are conserved across organisms remains unclear. Because of their diverse genome content, bacterial genomes are particularly amenable to studying these issues (Lynch 2007). In measuring the rate and molecular spectrum of mutations in the high-GC, multi-replicon genome of *B. cenocepacia*, we have corroborated some prior findings of MA studies in model organisms, but also demonstrated idiosyncrasies in the *B. cenocepacia* spectrum that may extend to other organisms with high %GC content and/or with multiple chromosomes. Specifically, *B. cenocepacia* has a low mutation rate and is consistent with a universal deletion bias in prokaryotes (Mira et al. 2001). However, the lack of G:C>A:T bias is inconsistent with all previous findings in mismatch-repair proficient organisms (Lynch et al. 2008; Denver et al. 2009; Hershberg and Petrov 2010; Hildebrand et al. 2010; Ossowski et al. 2010; Lee et al. 2012; Sung, Tucker, et al. 2012).

Bacterial genomes are also advantageous study subjects for their relatively ordered patterns of replication initiated at only one origin per chromosome. In genomes with multiple chromosomes, the origins apparently fire at different times to maintain termination synchrony, causing smaller chromosomes to be replicated later (Rasmussen et al. 2007; Cooper et al. 2010). With this model in mind, it becomes noteworthy that both mutation rates and spectra differed significantly among chromosomes in this multi-replicon genome and in a manner suggesting greater oxidative damage or more inefficient repair in late replicated regions.

As a member of a species complex with broad ecological and clinical significance, *B. cenocepacia* is a taxon with rich genomic resources that enable comparisons between the *de novo* mutations reported here and extant sequence diversity. With 7050 genes, *B. cenocepacia* HI2424 has a large amount of coding DNA (G_E) (6.8 x 10^6 base pairs), and a high average nucleotide heterozygosity at silent-sites (π_s) (6.57 x 10^{-2}) relative to other strains (Watterson 1975; Mahenthiralingam et al. 2005). By combining this π_s measurement and the base-substitution rate from this study, we estimate that the N_e of *B. cenocepacia* is approximately 247 x 10^6, which is in the upper echelon among species whose N_e has been derived in this manner (Figure S5).

Under the drift-barrier hypothesis, high target size for functional DNA and high N_e increase the ability of natural selection to reduce mutation rates (Lynch 2010; Lynch 2011; Sung, Ackerman, et al. 2012). Thus, given the large proteome and N_e of *B. cenocepacia*, it is unsurprising that *B. cenocepacia* has relatively low base-substitution and indel mutation rates when compared to other organisms (Sung, Ackerman, et al. 2012). However, the low substitution and indel mutation rates observed in this study need not imply limited genetic diversity among species of the *Burkholderia cepacia* complex. Rather, because of their high N_e and evidently frequent lateral genetic transfer, species of the *Burkholderia cepacia* complex are remarkably diverse (Baldwin et al. 2005; Pearson et al. 2009), demonstrating that low mutation rates need not imply low levels of genetic diversity.

Because mutations provide the raw material for evolutionary processes, the mutational spectrum of *B. cenocepacia* has important implications for its genome

evolution, which possibly extend to other GC-rich or multi-replicon genomes. Despite similar rates of insertion and deletion events, deletions were larger than insertions, and plasmids were lost relatively frequently, which together support the model that bacterial genomes are subject to a deletion bias (Mira et al. 2001; Kuo and Ochman 2009). Ultimately, this dynamic has the potential to drive the irreversible loss of previously essential genes during prolonged colonization of a host and may enable host dependence to form more rapidly in prokaryotic organisms than in eukaryotes, which do not have a strong deletion bias (Denver et al. 2004; Kuo and Ochman 2009; Dyall et al. 2014).

The lack of GC mutation bias observed in *B. cenocepacia* has not been seen previously in non-mutator MA lineages of any kind (Lind and Andersson 2008; Lynch et al. 2008; Denver et al. 2009; Ossowski et al. 2010; Lee et al. 2012; Sung, Ackerman, et al. 2012; Sung, Tucker, et al. 2012). Interestingly, the lack of GC mutation bias appears to be primarily caused by a substantial elevation of the A:T>C:G mutation rate relative to all other transversion types on chromosome 1 (Figure 2; Figure 3C). A decreased ratio of G:C>A:T to A:T>G:C transition mutations relative to that seen in other bacteria was also observed (Lee et al. 2012; Sung, Tucker, et al. 2012). In principle, a decreased rate of G:C>A:T transition mutation could be achieved by an increased abundance of uracil-DNA-glycosylases, which remove uracils from DNA following cytosine deamination (Pearl 2000), or by a lack of cytosine methyltransferases, which methylate the C-5 carbon of cytosines and expose them to increased rates of cytosine deamination (Kahramanoglou et al. 2012). However, *B. cenocepacia* HI2424 does not appear to have an exceptionally high number of UDGs, and it does contain an obvious cytosine

methyltransferase homolog, suggesting that active methylation of cytosines does occur in *B. cenocepacia*. Extending these methods to more genomes with high GC content will be required to determine whether a lack of AT mutation bias is a common feature of GC-rich genomes.

Perhaps the most important finding from this study is that both mutation rates and spectra vary significantly among the three autonomously replicating chromosomes that make up the *B. cenocepacia* genome (Figure 3). The possibility that mutation rates vary among genome regions has been demonstrated several times using reporter genes and comparative methods (Hudson et al. 2002; Mira and Ochman 2002; Hawk et al. 2005; Cooper et al. 2010; Lang and Murray 2011; Agier and Fischer 2012; Morrow and Cooper 2012), and also directly in a more recent study that used similar methods to those described here (Foster et al. 2013). Although comparative evidence demonstrates that evolutionary rates in multi-chromosome bacteria increase on secondary chromosomes (Cooper et al. 2010; Morrow and Cooper 2012), differences among taxa can be a consequence of biases at the level of mutation and/or selection. Our data demonstrate that base-substitution mutation rates vary significantly among chromosomes, but not in the direction predicted by comparative studies (Cooper et al. 2010). Specifically, we find that base-substitution mutation rates are highest on the primary chromosome (Figure 3A,B), where evolutionary rates are lowest. Thus, purifying selection must be substantially stronger on the primary chromosome to offset the effect of an elevated mutation rate.

The spectra of base-substitutions also differed significantly among chromosomes, with two types of transversions occurring much more frequently on only

one of the three replicons. While A:T>C:G transversions are more than twice as likely to occur on the primary chromosome as elsewhere, G:C>T:A transversions are more than twice as likely to occur on the third chromosome (Figure 3C). The G:C>T:A transversions are a particularly interesting class of substitutions because they can arise through oxidative damage (Michaels et al. 1992; Lee et al. 2012) and may be elevated late in the cell cycle when intracellular levels of reactive oxygen species are high. Models of replication timing in another multi-chromosome bacterium, *Vibrio cholerae*, have demonstrated that smaller secondary chromosomes initiate replication later in the cell cycle (Rasmussen et al. 2007). While not all mutations arise during replication, late replicating regions have been associated with higher transversion rates in prokaryotes and multicellular eukaryotes (Mira and Ochman 2002; Stamatoyannopoulos et al. 2009; Chen et al. 2010), and specifically with G:C>T:A transversions in several species comparisons (Mira and Ochman 2002). Thus, because late-replicated regions on the larger primary and secondary chromosomes are expected to replicate concordantly with those on the tertiary chromosome (Rasmussen et al. 2007), we would not only expect elevated rates of these mutations on the tertiary chromosome, but also on the later replicated regions of the primary and secondary chromosomes.

We tested this prediction by measuring the overall rates of G:C>T:A transversions in the early replicated regions on chr1 and chr2 (prior to chr3 initiation), and comparing them to the late replicated regions on chr1 and chr2 (following chr3 initiation), as well as the rates on chr3. Although the low number of total G:C>T:A transversions observed in this study prevents us from statistically distinguishing conditional G:C>T:A transversion rates between late and early replicated regions of

chr1 and chr2, the conditional G:C>T:A transversion rate is higher in late than early replicated regions of chr1 and chr2 (Figure S6), which is remarkable considering that early replicated genes on chr1 and chr2 are expressed more, which has been shown to induce G:C>T:A transversions independent of replication (Klapacz and Bhagwat 2002; Kim and Jinks-Robertson 2012; Alexander et al. 2013). Thus, we suggest that late replicating DNA, particularly in divided genomes, is inherently predisposed to increased rates of G:C>T:A transversions, possibly due to increased exposure to oxidative damage or variation in DNA-repair mechanisms, although the transversion type responsible for these increases may vary between species (Mira and Ochman 2002).

A mechanism of an increased A:T>C:G transversion mutation rate on the primary chromosome is less clear, but a decreased rate of A:T>C:G transversions in a late replicating reporter relative to that on an intermediate replicating reporter has been demonstrated previously in *Salmonella enterica* (Hudson et al. 2002). Thus, it is possible that this form of transversion is reduced in late replicating DNA, or that it is primarily caused by other forms mutagenesis (Klapacz and Bhagwat 2002), although transcriptional mutagenesis is unlikely as A:T>C:G transversions occur relatively frequently in non-coding DNA relative to other substitution types (Figure S7).

In summary, this study has demonstrated that the GC-rich genome of *B. cenocepacia* has a relatively low mutation rate, with a mutation spectrum biased toward deletion and G:C production. Moreover, both the rate and types of base-substitution mutations that occur most frequently vary by chromosome, likely related to replication dynamics, the cell cycle, and transcription (Klapacz and Bhagwat 2002; Cooper et al. 2010; Merrikh et al. 2012). Although this study represents an essential first step in

broadening our understanding of mutation rates and spectra beyond that of model organisms, whether the observed mutational traits are common to all GC-rich genomes with multiple replicons, or are merely species-specific idiosyncrasies will require a more thorough investigation across a more diverse collection of GC-rich and multi-replicon bacterial genomes. Ultimately, by better understanding the core mutational processes that generate the raw variation on which evolution acts, we can aspire to develop true species-specific null-hypotheses for molecular evolution, and by extension, enable more accurate analyses of the role of all evolutionary forces in driving genome evolution.

MATERIALS AND METHODS

Mutation accumulation. Seventy-five independent lineages were founded by single cells derived from a single colony of *Burkholderia cenocepacia* HI2424, a soil isolate that had only previously been passaged in the laboratory during isolation (Coenye and LiPuma 2003). Independent lineages were then serially propagated every 24 hours onto fresh high nutrient Tryptic Soy Agar (TSA) plates (30 g/L Tryptic Soy Broth (TSB) Powder, 15 g/L Agar). Two lineages were maintained on each plate at 37°C, and the isolated colony closest to the base of each plate half was chosen for daily re-streaking. Backups were maintained in a 4°C walk-in refrigerator in case of line extinction or experimental error, but were never used in any of the sequenced lineages. Following 217-days of MA, frozen stocks of all lineages were prepared by growing a final colony per isolate in 5 ml TSB (30 g/L TSB) overnight at 37°C, and freezing in 8% DMSO at -80°C.

Daily generation times were estimated each month by placing a single representative colony from each line in 2 ml of Phosphate Buffer Saline (80 g/L NaCl, 2 g/L KCl, 14.4 g/L $Na_2HPO_4 \bullet 2H_2O$, 2.4 g/L KH_2PO_4), serially diluting to 10^{-3} and spread plating 100 ul on TSA. By counting the colonies on the resultant TSA plate, we calculated the number of viable cells in a single colony and thus the number of generations between each transfer. The average generation time across all lines was then calculated and used as the daily generation time for that month. These generation-time measurements were used to evaluate potential effects of declining colony size over the course of the MA experiment as a result of mutational load, a phenotype that was observed (Figure S1). Final generation numbers per line were estimated as the sum of monthly generation estimates, which were derived by multiplying the number of generations per day in that month by the number of days between measurements (Figure S1).

DNA extraction and sequencing. Genomic DNA was extracted from 1 ml of overnight culture inoculated from 47 frozen derivatives of MA lines using the Wizard Genomic DNA Purification Kit (Promega Inc.). Concentration and purity were analyzed using a Thermo Scientific Nanodrop 2000c (Thermo Scientific Inc.) and a 1% Agarose gel with a Quick-Load 2-log ladder (New England BioLabs Inc.). Following library preparation, sequencing was performed using the 151-bp paired-end Illumina HiSeq platform at the University of New Hampshire Hubbard Center for Genomic Studies with an average fragment size between paired-end reads of ~386 bp. Sequenced lineages were then individually mapped to the reference genome of *Burkholderia cenocepacia* HI2424

(LiPuma et al. 2002), with both the Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009) and Novoalign (www.novocraft.com), producing an average sequence depth of ~50x.

Molecular analysis and mutation identification. To identify base-substitution mutations, the sam alignment files that were produced by each reference aligner were first converted to mpileup format using samtools (Li et al. 2009). Forward and reverse read alignments were then produced for each position in each line using in-house perl scripts. Next, a three-step process was used to detect polymorphisms.

First, a base for each individual line was called if a site was covered by at least two forward and two reverse reads, and at least 80% of those reads identified the same base. Second, an ancestral consensus was called as the base with the highest support among reads across all lines, as long as there were at least three lines with sufficient coverage to identify a base. Lastly, at sites where both an individual line base and ancestral consensus were identified, individual line bases were compared to the ancestral base, and if they were different, a putative base-substitution mutation was identified. Putative base-substitution mutations were identified as true substitutions if both aligners independently identified the mutation.

Although the above criteria for identifying individual line bases and overall consensus bases are relatively lenient given our coverage of ~50x for individual lines and ~2200x across all lines, both the coverage and support for all substitutions that were called dramatically exceeded those criteria, demonstrating that we were not simply obtaining false positives in regions of lower coverage (Table S1). Furthermore, these

same methods have been used to identify base-substitution mutations in both *Escherichia coli* and *Bacillus subtilis* MA lines, where 19 of 19 and 69 of 69 base-substitution mutations called were confirmed by conventional sequencing, respectively (Lee et al. 2012; Sung, Ackerman, et al. 2012). Thus, these criteria are unlikely to result in false positives, while allowing us to cover the majority of the *B. cenocepacia* genome and reduce false negatives.

For insertion-deletion mutations (indels), inherent difficulties with gaps and repeat elements can reduce agreement in the alignment of single reads using short-read alignment algorithms, even in the case of true indels. Thus, putative indels were first extracted from both BWA and Novoalign at all sites where at least two forward and two reverse reads covered an indel, and 30% of those reads identified the exact same indel (size and motif). Next, the alignment output was additionally passaged through the pattern-growth algorithm PINDEL to verify putative indels from the alignment and identify larger indels using paired-end information (Ye et al. 2009). Here, a total of twenty reads, including at least six forward and six reverse reads were required to extract a putative indel. Putative indels were only kept as true indels for further analysis if: a) they were independently identified by both alignment algorithms and PINDEL, and at least 50% of the full-coverage reads (>25 bases on both sides of the indel) from the initial alignment identified the mutation; b) they were identified only by BWA and Novoalign, and at least 80% of the good-coverage reads from the initial alignment identified the mutation; or c) they were larger indels that were only identified by the more strict requirements of PINDEL.

Unlike base-substitutions mutations, many reads that cover an indel mutation may fail to identify the mutation because they lack sufficient coverage on both sides of the mutation to anchor the read to the reference genome, particular when they occur at simple sequence repeats. Therefore, applying the initially lenient filter to extract putative indels is justified to identify all potential indels. By then focusing only on the good-coverage reads and applying an independent paired-end indel identifier (PINDEL), we can filter out indels that are more likely to be false positives, while keeping only the high concordance indels supported by multiple algorithms. Although there remains more uncertainty with indel calls than with base-substitutions mutations, we are confident that we have obtained an accurate picture of the naturally occurring indels from this study because of the high concordance across algorithms and reads (Table S2; Figure S2), and the fact that no indels were called independently by more than 2 lines (Figure S3). A complete list of the indels identified in this study, along with the algorithms that identified them, their coverage, and concordance across well-covered reads can be found is Table S2.

Mutation-rate Analysis. Once a complete set of mutations had been identified in each lineage, we calculated the substitution and indel mutation rates for each line using the equation $\mu = m/nT$, where μ represents the mutation rate (μ_{bs} for bps, μ_{indel} for indels), m represents the number of mutations observed, n represents the number of sites that had sufficient depth and consensus to analyze, and T represents the total generations over the course of the MA study for an individual line. The standard error of the mutation rate for each line was measured as described previously with the equation $SE_x = \sqrt{\mu/nT}$ (Denver et al. 2004; Denver et al. 2009).

The final μ_{bs} and μ_{indel} for *B. cenocepacia* were calculated by taking the average μ of all sequenced lineages, and the total standard error was calculated as the standard deviation of the mutation rates across all lines (*s*) divided by the square root of the number of lines analyzed (*N*): $SE_{pooled} = s/\sqrt{N}$. Specific base-substitution mutation rates were further divided into conditional rates for each substitution type using the equation $\mu_{bs} = m/nT$, where *m* is the number of substitutions of a particular type, and *n* is the number of ancestral bases that can lead to each substitution with sufficient depth and consensus to analyze.

Calculation of G_E, π_s, and N_E. Effective genome size (G_E) was determined as the total coding bases in the *B. cenocepacia* genome. Silent site diversity (π_s) was derived using a survey of 200 *B. cenocepacia* strains across 7 loci (*atpD*, *gltB*, *gyrB*, *recA*, *lepA*, *phaC*, *trpB*), which were concatenated and aligned using BIGSdb (Jolley and Maiden 2010), and analyzed using DNAsp (Librado and Rozas 2009). Using the value of μ_{bs} obtained in this study, N_e was estimated by dividing the value of π_s by $2\mu_{bs}$ ($\pi_s = 2N_e\mu_{bs}$) (Kimura 1983).

ACKNOWLEDGMENTS

We thank Kenny Flynn for helpful discussion and Brian VanDam for technical support. 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); and the National Science Foundation Career Award (DEB-0845851 to VSC).

REFERENCES

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.

Alexander MP, Begins KJ, Crall WC, Holmes MP, Lippert MJ. 2013. High levels of transcription stimulate transversions at GC base pairs in yeast. Environ. Mol. Mutagen. 54:44–53.

Baldwin A, Mahenthiralingam E, Thickett KM, Honeybourne D, Maiden MCJ, Govan JR, Speert DP, Lipuma JJ, Vandamme P, Dowson CG. 2005. Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. J. Clin. Microbiol. 43:4665–4673.

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.

Coenye T, LiPuma JJ. 2003. Population structure analysis of *Burkholderia cepacia* genomovar III: varying degrees of genetic recombination characterize major clonal complexes. Microbiology-Sgm 149:77–88.

Coenye T, Spilker T, Van Schoor A, LiPuma JJ, Vandamme P. 2004. Recovery of *Burkholderia cenocepacia* strain PHDC from cystic fibrosis patients in Europe. Thorax 59:952–954.

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

Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. Proc. Natl. Acad. Sci. U. S. A. 106:16310–16314.

Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature 430:679–682.

Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. U. S. A. 88:7160–7164.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu. Rev. Genomics Hum. Genet. 10:285–311.

Dyall SD, Brown MT, Johnson PJ. 2014. Ancient invasions : from endosymbionts to organelles. Science 304:253–257.

Elena SF, Ekunwe L, Hajela N, Oden SA, Lenski RE. 1998. Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. Genetica 102:349–358.

Fijalkowska IJ, Schaaper RM, Jonczyk P. 2012. DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair. Fems Microbiol. Rev. 36:1105–1121.

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.

Graur D, Li W-H. 2000. Fundamentals of molecular evolution. Sunderland, Mass.: Sinauer Associates

Hall DW, Mahmoudizad R, Hurd AW, Joseph SB. 2008. Spontaneous mutations in diploid *Saccharomyces cerevisiae*: another thousand cell generations. Genet. Res. (Camb). 90:229–241.

Hawk JD, Stefanovic L, Boyer JC, Petes TD, Farber RA. 2005. Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. Proc. Natl. Acad. Sci. U. S. A. 102:8639–8643.

Heilbron K, Toll-Riera M, Kojadinovic M, Maclean RC. 2014. Fitness Is strongly influenced by rare mutations of large effect in a microbial mutation accumulation experiment. Genetics 197:981–990.

Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 6:e1001115.

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

Hudson RE, Bergthorsson U, Roth JR, Ochman H. 2002. Effect of chromosome location on bacterial mutation rates. Mol. Biol. Evol. 19:85–92.

Jolley KA, Maiden MCJ. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595–606.

Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, Fraser GM, Luscombe NM, Seshasayee ASN. 2012. Genomics of DNA cytosine methylation in

Escherichia coli reveals its role in stationary phase transcription. Nat. Commun. 3:886.

Kibota TT, Lynch M. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature 381:694–696.

Kim N, Jinks-Robertson S. 2012. Transcription as a source of genome instability. Nat. Rev. Genet. 13:204–214.

Kimura M. 1983. The Neutral Theory of Molecular Evolution. Cambridge, New York: Cambridge University Press

Klapacz J, Bhagwat AS. 2002. Transcription-dependent increase in multiple classes of base substitution mutations in *Escherichia coli*. J. Bacteriol. 184:6866–6872.

Kuo C-H, Ochman H. 2009. Deletional bias across the three domains of life. Genome Biol. Evol. 1:145–152.

Lang GI, Murray AW. 2011. Mutation rates across budding yeast chromosome VI are correlated with replication timing. Genome Biol. Evol. 3:799–811.

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.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria. Proc. Natl. Acad. Sci. U. S. A. 105:17878–17883.

LiPuma JJ, Spilker T, Coenye T, Gonzalez CF. 2002. An epidemic *Burkholderia cepacia* complex strain identified in soil. Lancet 359:2002–2003.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc. Natl. Acad. Sci. U. S. A. 105:9272–9277.

Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits. Sunderland (MA): Sinauer Associates

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates

Lynch M. 2010. Evolution of the mutation rate. Trends Genet. 26:345–352.

Lynch M. 2011. The lower bound to the evolution of mutation rates. Genome Biol. Evol. 3:1107–1118.

Mahenthiralingam E, Urban TA, Goldberg JB. 2005. The multifarious, multireplicon *Burkholderia cepacia* complex. Nat. Rev. Microbiol. 3:144–156.

Merrikh H, Zhang Y, Grossman AD, Wang JD. 2012. Replication-transcription conflicts in bacteria. Nat. Rev. Microbiol. 10:449–458.

Michaels ML, Cruz C, Grollman AP, Miller JH. 1992. Evidence that mutY and mutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. Proc. Natl. Acad. Sci. U. S. A. 89:7022–7025.

Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet. 17:589–596.

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.

Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327:92–94.

Pearl LH. 2000. Structure and function in the uracil-DNA glycosylase superfamily. Mutat. Res. - DNA Repair 460:165–181.

Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A, Price EP, Glass MB, Leadem B, Beckstrom-Sternberg JS, et al. 2009. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. BMC Biol. 7:78–92.

Raghavan R, Kelkar YD, Ochman H. 2012. A selective force favoring increased G plus C content in bacterial genes. Proc. Natl. Acad. Sci. U. S. A. 109:14504–14507.

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.

Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. 2000. The evolution of mutation rates: separating causes from consequences. Bioessays 22:1057–1066.

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.

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. 109:18488–18492.

Sung W, Tucker AE, Doak TG, Choi E, Thomas WK, Lynch M. 2012. Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. Proc. Natl. Acad. Sci. U. S. A. 109:19339–19344.

Traverse CC, Mayo-Smith LM, Poltak SR, Cooper VS. 2013. Tangled bank of experimentally evolved *Burkholderia* biofilms reflects selection during chronic infections. Proc. Natl. Acad. Sci. U. S. A. 110:E250–E259.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7:256–276.

Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Medigue C, Lenski RE, Schneider D. 2011. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. G3-Genes Genomes Genet. 1:183–186.

Ye K, Schulz MH, Long Q, Apweiler R, Ning ZM. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25:2865–2871.

Zeyl C, DeVisser JA. 2001. Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. Genetics 157:53–61.

Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. Proc. Natl. Acad. Sci. U. S. A. 111:E2310–E2318.

Zlosnik JEA, Costa PS, Brant R, Mori PYB, Hird TJ, Fraenkel MC, Wilcox PG, Davidson AGF, Speert DP. 2011. Mucoid and nonmucoid *Burkholderia cepacia* complex bacteria in cystic fibrosis infections. Am. J. Respir. Crit. Care Med. 183:67–72.

FIGURE LEGENDS

Figure 1: Distribution of observed mutations in the 47 sequenced lineages derived from *B. cenocepacia* HI2424 following an average of 5554 generations of mutation accumulation per line. Minimal variation exists in the number of sites analyzed per line, making differences in mutation number across lineages representative of the random variance in mutation rate measurements across lineages.

Figure 2: Conditional base-substitution mutation rates of *B. cenocepacia* mutation accumulation (MA) lines across all three chromosomes. Conditional base-substitution rates are estimated by scaling the base-substitution mutation rates to the analyzed nucleotide content of the *B. cenocepacia* genome, whereby only covered sites capable of producing a given substitution are used in the denominator of each calculation. Error bars represent one standard error of the mean.

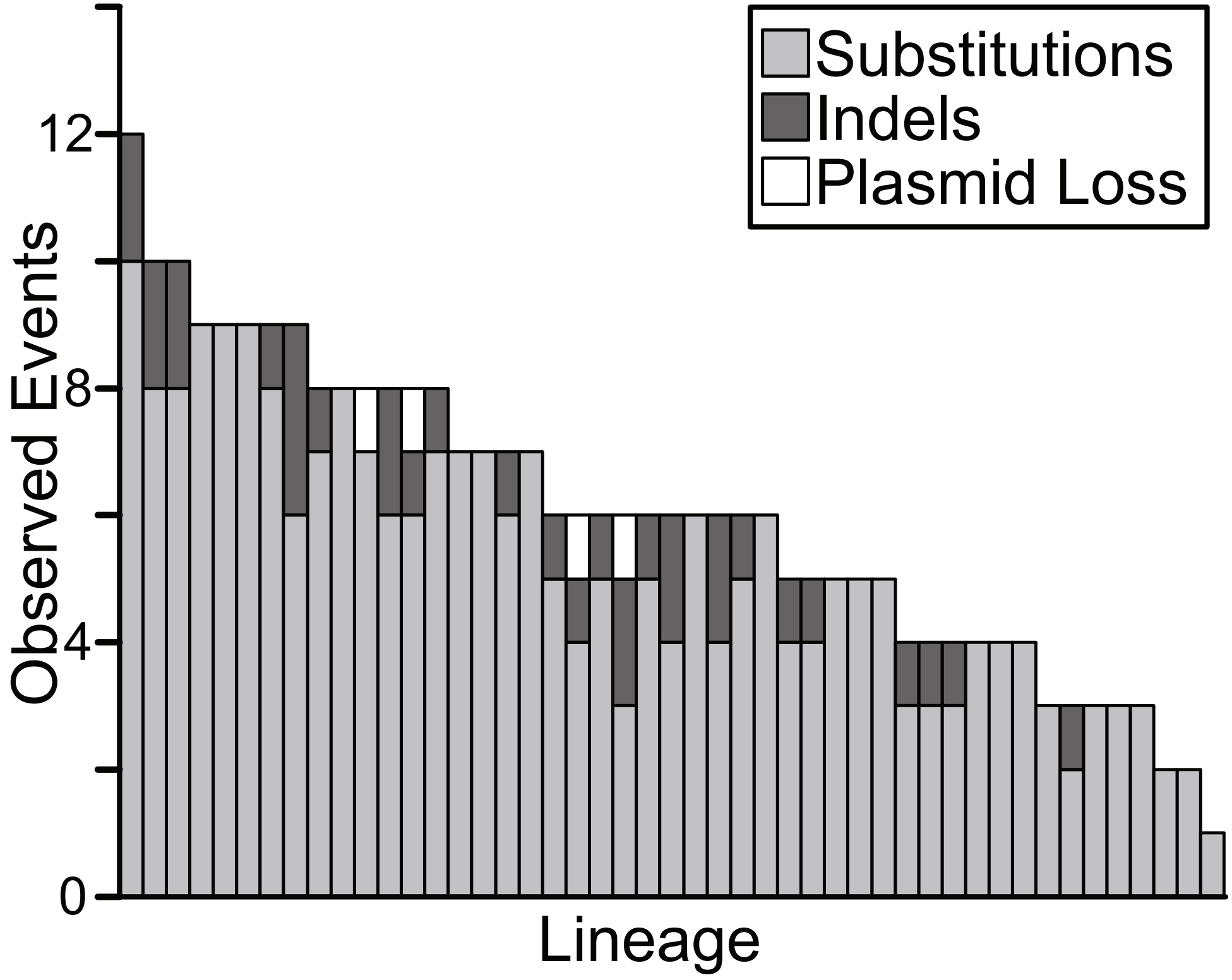
Figure 3: Substitution and indel mutation rates for the three chromosomes of *B. cenocepacia*; error bars indicate one standard error of the mean. A, B) Overall base-substitution and indel mutation rates. C) Conditional base-substitution mutation rates for each chromosome of *B. cenocepacia* estimated as described in Figure 2, based on the analyzed nucleotide content of each chromosome.

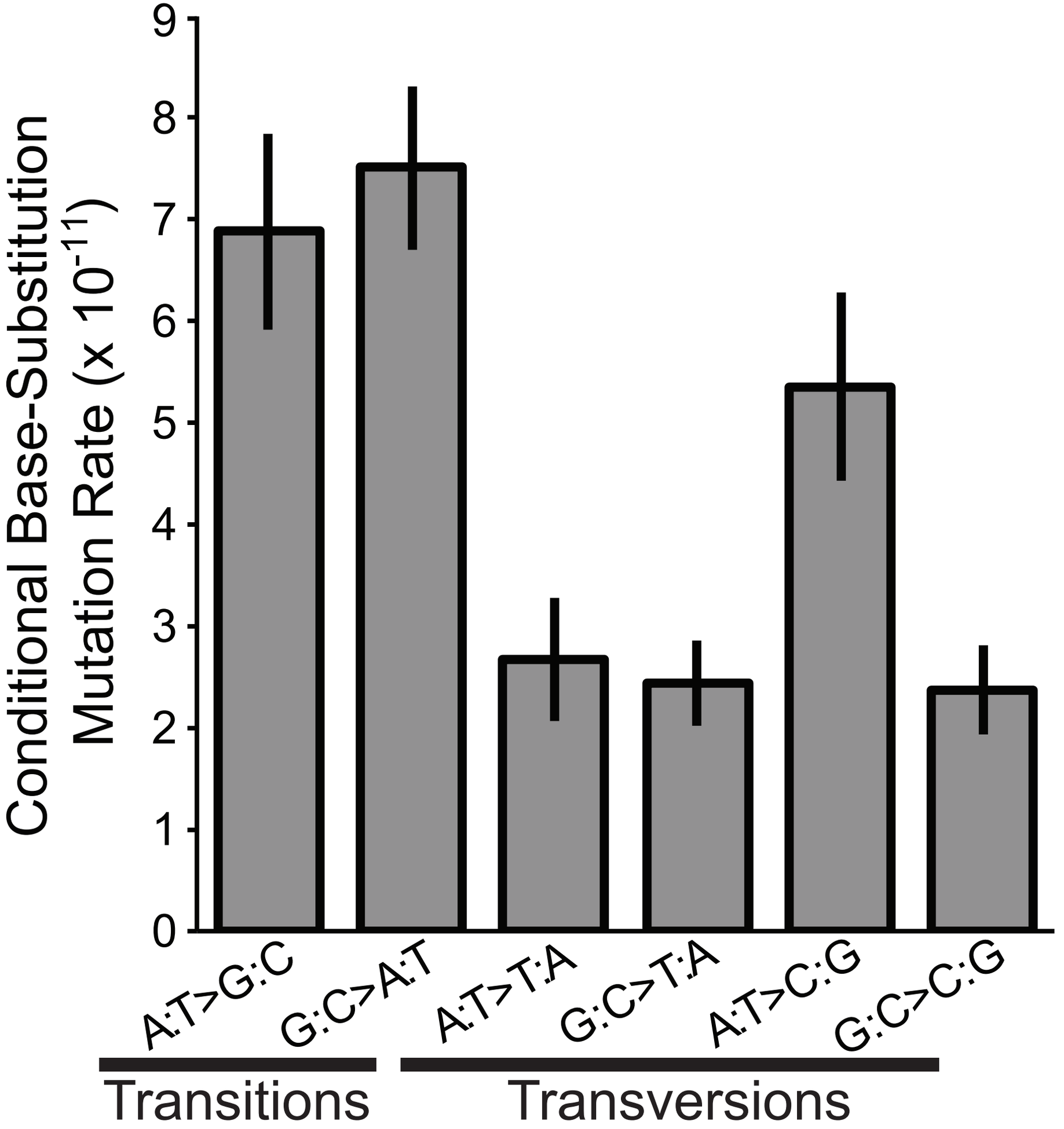
Figure 4: Overall base-substitution (A) and indel mutation rates (B) in 100 kb (outer), 25 kb (middle), and 5 kb (inner) bins extending clockwise from the origin of replication (oriC). Mutation rates were analyzed independently for each bin size, so color shades in smaller bins don't directly compare to the same color shades in larger bins. The 1.64 Kb plasmid is not to scale.

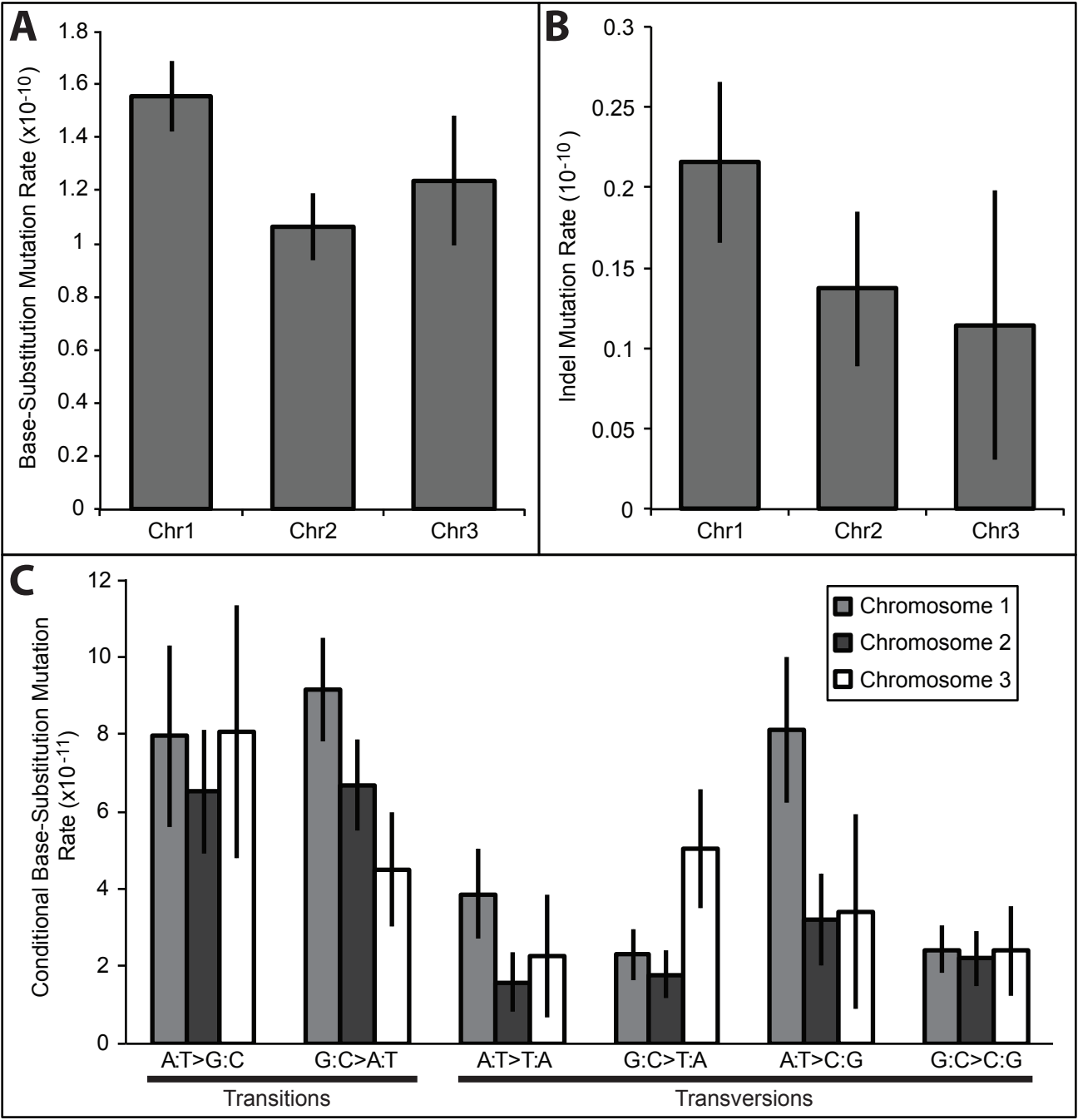
TABLES

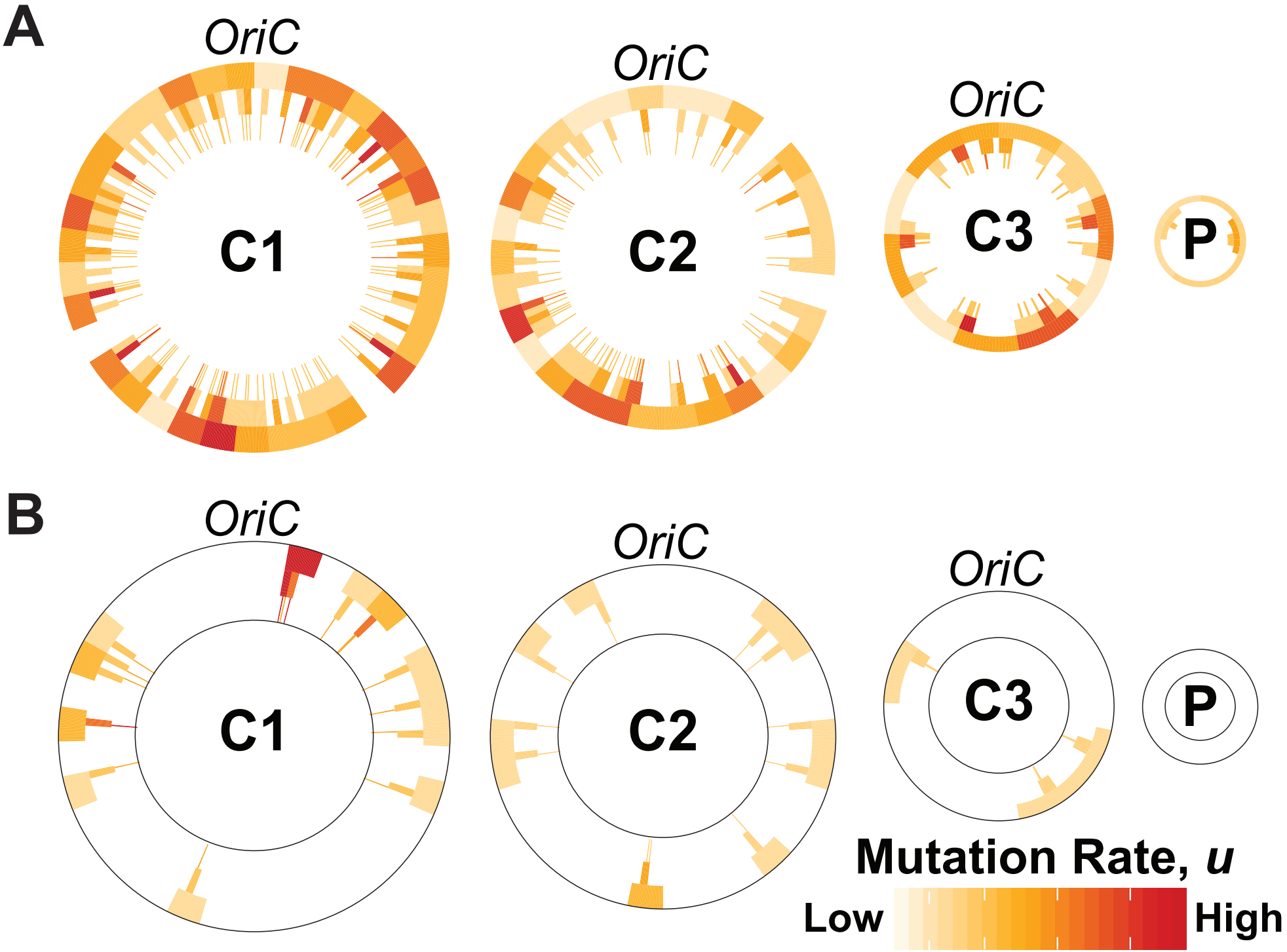
Table 1: Parameters of insertion and deletion mutations during 261,038 generations of spontaneous mutation accumulation in *B. cenocepacia*.

Parameter	Deletions	Insertions
Events Observed	17	16
Total Nucleotides Affected	376	121
Total GC Bases Affected	302	87
Total AT Bases Affected	74	34
Proportion of GC Bases Affected	0.80	0.72
Plasmid Copy Number Loss/Gain	4	0


Substitutions
Indels
Plasmid Loss
Observed Events
0
4
8
12
Lineage


Conditional Base-Substitution Mutation Rate (x 10^-11)
0
1
2
3
4
5
6
7
8
9
A:T>G:C
G:C>A:T
A:T>T:A
G:C>T:A
A:T>C:G
G:C>C:G
Transitions
Transversions


A
Base-Substitution Mutation Rate (x10-10)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Chr1
Chr2
Chr3
B
Indel Mutation Rate (10-10)
0.3
0.25
0.2
0.15
0.1
0.05
0
Chr1
Chr2
Chr3
C
Conditional Base-Substitution Mutation Rate (x10-11)
12
10
8
6
4
2
0
Chromosome 1
Chromosome 2
Chromosome 3
A:T>G:C
G:C>A:T
A:T>T:A
G:C>T:A
A:T>C:G
G:C>C:G
Transitions
Transversions


A
OriC
C1
OriC
C2
OriC
C3
P
B
OriC
C1
OriC
C2
OriC
C3
P
Mutation Rate, u
Low
High