The Rate and Character of Spontaneous Mutation in *Thermus thermophilus*

Reena R. Mackwan,*,† Geraldine T. Carver,‡ Grace E. Kissling,§ John W. Drake† and Dennis W. Grogan*,‡

*Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio 45221-0006 and Laboratory of Molecular Genetics and Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709-2233

†These two authors contributed equally.

‡Corresponding author: Department of Biological Sciences, 614 Rieveschl Hall (ML 0006), Clifton Court, Cincinnati, Ohio 45221-0006. E-mail: gograndw@email.uc.edu
Running title: Spontaneous mutation in *Thermus thermophilus*

Key words: Base-pair substitutions, Indel mutations, Mutation rate, Mutation spectrum, *Thermus thermophilus*

Corresponding author: Dennis W. Grogan

Department of Biological Sciences
614 Rieveschl Hall, ML 0006
Clifton Court
Cincinnati, Ohio 45221-0006
513-556-9748
grogandw@email.uc.edu
ABSTRACT

Selection of spontaneous, loss-of-function mutations at two chromosomal loci (pyrF and pyrE) enabled the first molecular-level analysis of replication fidelity in the extremely thermophilic bacterium *Thermus thermophilus*. Two different methods yielded similar mutation rates, and mutational spectra determined by sequencing of independent mutants revealed a variety of replication errors distributed throughout the target genes. The genomic mutation rate estimated from these targets, 0.00097 ± 0.00052 per replication, was lower than corresponding estimates from mesophilic micro-organisms, primarily because of a low rate of base substitution. However, both the rate and spectrum of spontaneous mutations in *T. thermophilus* resembled those of the thermoacidophilic archaeon *Sulfolobus acidocaldarius*, despite important molecular differences between these two thermophiles and their genomes.

Experimental study of genetic stability in prokaryotes has focused historically on mesophilic bacteria that allow facile genetic manipulations. However, diverse prokaryotes adapted to extreme environments may also provide strategic insights into mechanisms of genetic stability, particularly with respect to environmental stresses exerted on DNA structure, replication, and repair. For instance, *Deinococcus radiodurans*, which belongs to a basal lineage of bacteria, exhibits extreme resistance to ionizing radiation and desiccation and has been used as a model system to understand the molecular basis of highly efficient repair of DNA double-strand breaks (Cox and Battista 2005).

Species of the related genus *Thermus* show a corresponding potential for the study of cellular and molecular adaptations of bacteria to extremely high temperature. This potential is reinforced by progress in the genetic analysis of *T. thermophilus* supported, in turn, by convenient growth and transformation conditions, cloning vectors, selectable markers, and complete genome sequences for two strains (Henne et al. 2004; Brüggemann and Chen 2006). One of these *T. thermophilus* strains (HB27) lacks a functional reverse gyrase, which is otherwise highly
conserved among extremely thermophilic bacteria and archaea (HENNE et al. 2004). In addition, both strains appear to lack family-Y DNA polymerases (HENNE et al. 2004; BRÜGGEMANN and CHEN 2006), which generate many of the mutations induced by DNA damage and may also contribute to spontaneous mutation (GOODMAN 2002). The apparent lack of these enzymes, combined with the ability to grow above 80°C, raises questions about the accuracy of genome replication in T. thermophilus. At the same time, genetic manipulations made possible by various T. thermophilus vectors and selections suggest ways to investigate the mechanisms affecting replication accuracy in vivo. Both considerations provide incentives to examine the rate and molecular nature of replication errors.

Errors that occur in the course of normal DNA replication and escape correction can be detected in descendant genomes as spontaneous mutations, provided the mutations do not impair survival. The errors include unforced misincorporations by replicative polymerases as well as the inaccurate insertion of bases opposite template damage, often by specialized family-Y polymerases (GOODMAN 2002). Despite the presumably chaotic nature of such replication errors, distinct patterns emerge from analysis of the resulting mutations among diverse organisms. For example, micro-organisms with DNA genomes (including DNA viruses, prokaryotes and eukaryotes) display very similar genomic mutation rates that cluster around 0.003 mutations per genome per replication despite tremendous differences in genome size and error rates per base pair (DRAKE et al., 1998). Furthermore, base-pair substitutions (BPSs) tend to outnumber insertion and deletion mutations (indels) in these genomes (GROGAN et al. 2001). The similar genomic rates and mutational spectra observed across such organismal diversity suggest that all these genomes are subject to similar selective forces. These forces can be interpreted formally as a balance between the costs of achieving replication accuracy (such as limits on the efficiency of DNA repair and the rate of DNA synthesis) and its benefits (such as maintaining the fitness of descendants over many generations). However, the precise nature of these costs and benefits remains unclear, and detailed analysis of spontaneous mutation across a range of microbial diversity offers a useful approach to identifying them in molecular terms.
The molecular analysis of replication errors in vivo requires target genes in which rare spontaneous mutations can be recovered efficiently for quantification and identification. The *S. cerevisiae* orotidylate decarboxylase (*URA3*) gene provides an example of such a mutational target, and corresponding UMP biosynthetic genes have provided favorable mutational targets in archaea. Use of these related targets has yielded dramatically different results in different organisms, however. In yeast, the mutational spectrum was dominated by base-pair substitutions (86% of the total) (Lee et al. 1998), and yielded a genomic mutation rate similar to those of other mesophilic genomes (Drake, et al., 1998). In the thermoacidophilic crenarchaeote *Sulfolobus acidocaldarius* (optimal growth near 80° and pH 3), loss-of-function mutations in the orotate:phosphoribosyl transferase gene, *pyrE* (Grogan et al. 2001), indicated a genomic mutation rate slightly below the consensus rate for diverse mesophilic genomes, whereas the orotidylate decarboxylase gene, *pyrF*, yielded an even lower estimate. In these targets, only about a third of the detected mutations were base-pair substitutions (BPSs) compared to about two-thirds in most other microbial DNA genomes (Grogan et al. 2001). The mesophilic, extremely halophilic euryarchaeote *Haloferax volcanii* yielded *pyrE* mutants at a still lower rate (Mackwan et al. 2007), although this probably resulted from long phenotypic lag or other effects of the unusually high copy number of its genome (Breuert et al. 2006). Furthermore, the mutations recovered from *H. volcanii pyrE2*, the orotate:phosphoribosyl transferase gene, were almost exclusively in-frame deletions templated by short direct repeats (Mackwan et al. 2007); this kind of mutation has not been observed in *S. acidocaldarius* (Grogan et al. 2001; Grogan and Hansen 2003). The fact that mutational patterns of the two archaea (one a thermoacidophile, the other an extreme halophile) differ from each other as well as from the norm of evolutionarily diverse mesophiles raises the possibility that patterns of spontaneous mutation in micro-organisms may correlate more strongly with environmental parameters than with phylogeny. A similar hypothesis has drawn support from the observed base-substitution patterns of orthologous gene pairs within mesophilic versus thermophilic genera of both archaea and bacteria (Friedman et al. 2004).
Quantitative molecular analysis of spontaneous mutation in the extremely thermophilic bacterium *T. thermophilus* would provide a test of the hypothesis that adaptation to high growth temperatures is accompanied by low BPS rates. We therefore established conditions for selecting loss-of-function mutations in two UMP-biosynthetic genes, *pyrE* and *pyrF*, and used this selection to analyze spontaneous mutation both quantitatively and qualitatively. The results indicate that these *T. thermophilus* genes have BPS rates well below those typically observed in target genes of mesophiles but similar to those of the orthologous *S. acidocaldarius* genes.

**MATERIALS AND METHODS**

**Growth conditions:** *T. thermophilus* strain HB27 (Oshima and Imahori 1974) was obtained from the American Type Culture Collection as BAA-163. Liquid growth medium contained (g/l) NaCl, 3.0; sodium citrate, 0.3; CaCl$_2$·2H$_2$O, 0.03; MgSO$_4$·7H$_2$O, 0.2; K$_2$HPO$_4$, 0.17; and 0.1 ml/l of trace-mineral solution (5% FeCl$_3$ plus 0.5% each CuCl$_2$, CoCl$_2$, MnCl$_2$ and ZnCl$_2$ in 1 M HCl). This mineral base was supplemented with 5 g/l enzymatic digest of casein (Sigma) and the pH of the resulting complete medium was adjusted to 7.5 with NaOH. For plating, the tryptone medium was supplemented with an additional 33 mM MgSO$_4$ and 4 mM CaCl$_2$ and solidified with 7 g/l Gellan gum. Uracil (20 mg/l) was routinely added to support the growth of auxotrophs. In some experiments, the growth medium was further supplemented with 0.1% yeast extract (Sigma), with no measurable effect on mutation rates. All plates and liquid cultures were incubated aerobically at 76˚.

Growth of liquid cultures was routinely monitored by turbidity measurements at 600 nm. Viable counts were determined by serial dilution in sterile SM buffer (SAMBROOK and RUSSELL 2001) and plating on solid medium; colonies typically reached 1 mm diameter in 2–3 days. The minimal inhibitory concentration of 5-fluoro-orotic acid (FOA) was determined by incubating serial 1:2 dilutions of FOA in liquid growth medium inoculated with about 10$^7$ cells/ml. Under these conditions, the minimal inhibitory concentration was ≤1.25 g/l. However, FOA
concentrations of 0.5–0.6 g/l blocked all growth and colony formation of wild-type HB27 on solid medium and were used in selective plates.

In order to determine whether mutants resistant to FOA grow more slowly than the wild type, we performed two tests on a set of mutants of independent origin. In a reconstruction test, eight binary mixtures, consisting of liquid cultures of one mutant mixed with a 10-fold excess of wild-type cells, were grown for several hours in non-selective medium, during which time the populations expanded by factors of about 100. For each mixture, the numbers \( N \) of mutant and total cells were measured by plating before and after incubation. The ratio of growth rate constants \( \frac{k^m}{k^w} \) was calculated under the model of exponential growth, where the superscripts \( m \) and \( w \) represent mutant and wild-type strains, respectively. The ratios of growth-rate constants (averaged over several trials and arranged monotonically) were 0.83, 1.02, 1.15, 1.18, 1.21, 1.24, 1.36, and 1.62 (mean 1.20 ± 0.23). In a turbidometric test, non-selective medium (3.0 ml in 16-mm glass tubes) was inoculated to yield about \( 10^7 \) cells/ml and incubated. Absorbance at 600 nm was read at hourly intervals and the resulting values (corrected for the absorbance of the growth medium) were used to calculate \( \frac{k^m}{k^w} \) for each interval between two absorbance measurements. The wild-type strain was represented by three cultures whose absorbance values were averaged at each time point. Six independent comparisons (representing six 1-h intervals of growth) were made of each of nine different mutant cultures to the triplicate-average of the wild type, yielding an average relative growth rate of 0.98 ± 0.29. Thus, the mutants grew at rates indistinguishable from that of the wild type.

In efforts to measure a delay in the appearance of resistant clones (phenotypic lag), exponentially growing cultures were treated with a minimally toxic dose of mutagen and sampled frequently to detect any abrupt increase in mutant frequency. Among N-methyl N’-nitro N-nitroso guanidine (MNNG), butadiene diepoxide, mechlorethamine, N-nitroso-N-methyl urea, and cyclophosphamide, only MNNG appeared to stimulate forward mutation in preliminary tests. However, in experiments extending over several generations, MNNG concentrations that permitted uninterrupted exponential growth generated no detectable increases in mutant
frequency (data not shown).

**Mutation rates and spectrum:** Fluctuation tests used series of liquid cultures started from small inocula. A single wild-type colony was inoculated into non-selective medium and allowed to grow to about $10^7$ cells/ml. It was then diluted with fresh medium and dispensed into micro-dilution trays to yield $10^3$ cells in 200 µl in 10 to 15 wells. The trays were incubated until the cultures reached a density of about $10^8$ cells/ml, and any (small) volume losses due to evaporation were replaced with distilled water. The average number of viable cells in five wells per series was then determined by plating on non-selective medium. In order to determine the number of FOA-resistant mutants, the remaining cultures were plated in their entirety as two aliquots (50 and 150 µl) on selective medium (uracil + FOA). The number of FOA-resistant colonies varied by over 100-fold, and this variance was not affected by varying the size of the inocula by 10-fold, indicating that the inocula did not contain pre-existing mutants (LURIA and DELBRÜCK 1943). The number of mutant colonies arising from the two aliquots of each culture showed no effect of cell density on the efficiency of recovering mutants, and the individual 50- and 150-µl counts were therefore combined. The cultures were ranked according to the number of mutants; the first, second and third quartiles of the resulting distribution were then used to calculate three corresponding estimates of the mutation rates using the numerical solutions of KOCH (1982).

The accumulation method (DRAKE et al. 1998; MACKWAN et al. 2007) used 25-ml cultures, each started from separate small inocula and grown to total populations of about $10^{10}$ cells. Several samples of each culture were plated on selective and non-selective media. The resulting colony counts were used to determine the mutant frequency ($f$) and the population size ($N$) from which the mutation rate was calculated as $\mu = f/\ln(N\mu)$. The median value from this series was taken as the best estimate of the mutation rate (ROSCHE and FOSTER 2000).

Mutation rates are given ± 1 standard deviation (SD). The SD for the median value from the accumulation test was calculated according to Maritz and Jarrett (1978).

FOA-resistant colonies were confirmed to be uracil auxotrophs by inoculation of liquid
medium with and without uracil supplementation. Clonally pure derivatives were stored frozen at –70° in growth medium supplemented with dimethyl sulfoxide (9% w/v). Genomic DNAs, purified from liquid cultures essentially as described (Boom et al. 1990), served as templates in separate standard PCR reactions to amplify pyrF and pyrE, which are separated by a GGGGGG spacer [GenBank Accession nos. AE017221 and AE017222 (Henne et al. 2004)]. The amplification primer upstream of pyrF was 5′-AAAGCGGGAGGAAGGGGGA-3′ and the primer downstream of pyrE was 5′-GAAGGGTGGGCAGGTGAAGAGGA-3′. The PCR reaction was adjusted for a GC-rich template and included a 2-min denaturing step at 95° with Platinum® Taq DNA polymerase, PCRx Enhancer Solution, MgSO4, and PCRx Amplification Buffer (Invitrogen, Carlsbad, CA). The reaction was then run for 35 cycles of 30 sec at 95°, 30 sec at 60°, and 168 sec at 68°, followed by a hold temperature of 4°. The PCR products were purified using the QIAquick purification kit (Qiagen, Valencia, CA). The BigDye-Terminator Cycle Sequencing Kit was used to sequence the purified PCR products using the primers (F = forward, R = reverse) F5′-GGTGCTTTCGCGCTTTT-3′, F5′-AGGGGAGGGCCTTCTT-3′, R5′-AAAGAAGCCCTCCCCCT-3′ and R5′-AAGAAGGGCAGCGAGGAA-3′. When these were inadequate, we used the alternative primers F5′-CCCTGGCTCCGCTTCTCC-3′, R5′-AACGCAGGAGGAAGTGGC-3′, F5′-GCCACTTCCTCCTGCCTT-3′ F and F5′-CTTCTTCCAGGGCGCTT-3′. The sequencing cycle consisted of 2 min at 96° followed by 25 cycles of 10 sec at 96°, 5 sec at 60°, and 4 min at 60°, followed by a hold temperature of 4°.

RESULTS

**Genic mutation rates:** Mutation-rate assays were based on the ability of growth medium containing FOA plus uracil to select pyrimidine auxotrophs lacking either of two UMP-biosynthetic enzyme activities, orotate:phosphoribosyl pyrophosphate transferase or orotidylate decarboxylase, encoded by the pyrE and pyrF genes, respectively (Yamagishi et al. 1996). Efforts to develop other selections for spontaneous *T. thermophilus* mutants were unsuccessful.
For example, in preliminary tests, rifampicin, L-canavanine, L-ethionine, D-cycloserine, naladixic acid, novobiocin, and 5-fluorouracil either failed to inhibit growth at reasonable concentrations, or failed to select spontaneous mutants with sufficiently elevated resistance to the inhibitor (R. R. Mackwan, S. Teredesai and D. W. Grogan, unpublished results). To our knowledge, therefore, the FOA selection provides the only quantitative assay of forward mutation currently feasible for *Thermus* spp.

To determine whether physiological properties of the FOA-resistant mutants would affect the mutation assays, control experiments evaluated the fitness of FOA-selected pyrimidine auxotrophs, whereas possible effects of phenotypic lag were evaluated by comparing different mutation-assay methods. By these criteria, the mutants showed no significant reproductive disadvantage (see MATERIALS AND METHODS) and no detectable impact of phenotypic lag (see below). On average, 20% of the FOA-resistant colonies from selection plates scored as prototrophs (*i.e.*, did not require uracil for growth). In *Sulfolobus solfataricus* strain P2 and *Haloferax volcanii*, all such FOA-resistant uracil prototrophs were confirmed to have intact *pyrE* and *pyrF* genes (Redder and Garrett 2006; Mackwan et al. 2007). Therefore, we based our mutation-rate calculations only on pyrimidine-auxotrophic mutations that could be localized to *pyrE* or *pyrF* by DNA sequencing (see below).

Fluctuation-test mutation rates are shown in Table 1. In Koch’s quartile analysis (Koch 1982), a fluctuation test yields three estimates of the mutation rate, each derived from the corresponding quartile of the frequency distribution. This approach can reveal systematic distortions caused by processes such as phenotypic lag, to which the first quartile is most sensitive and the third quartile is least sensitive. The substantial variation in the rates among the 10 trials was consistent with the small numbers of cultures in each trial (5–10) assayed for mutants. The three quartiles yielded similar mean values, thus providing no hint of phenotypic lag. The genic rate for *pyrE* + *pyrF*, obtained by averaging across the three quartiles and 10 trials, was $(4.34 \pm 2.92) \times 10^{-7}$ mutations per cell division.

From the accumulation method, mean mutation rates per $10^7$ cell divisions for each of the six
cultures (arranged monotonically) were 1.61, 1.80, 2.11, 2.41, 3.46, and 3.59. The appropriate measure of this distribution is the median (Drake et al. 1998; Rosche and Foster 2000), which was \((2.26 \pm 0.66) \times 10^{-7}\). The accumulation method is intrinsically less sensitive to phenotypic lag than is the fluctuation test, so that the similarity of the rates determined by the two methods again suggests the absence of significant phenotypic lag. Because the fluctuation-test mean and the accumulation-method median produced statistically indistinguishable values, we adopted the mean of the two, \((3.30 \pm 2.12) \times 10^{-7}\), as the most reliable first measure of the genic rate. This rate must be adjusted slightly to reflect the fact that two of the 75 sequenced FOA-resistant uracil auxotrophs contained no mutation in the target genes. Thus, the corrected genic rate for the combined target \((pyrE + pyrF)\) is \((73/75)(3.30 \pm 2.12) \times 10^{-7} = (3.21 \pm 2.06) \times 10^{-7}\).

**Mutation spectrum:** The mutational alterations in the sequenced mutants are listed in Table 2, their sequence contexts are shown in Figure 1, and the molecular nature of the mutations is summarized in Table 3. Half of the mutations arose at a single hotspot, a run of eight Gs in \(pyrE\), which accounted for more mutations in the 552-nt \(pyrE\) gene (57) than in the 774-nt \(pyrF\) gene (26). Transitions outnumbered transversions by 16 to 3, a not particularly unusual bias. However, +1 indels outnumbered −1 indels by 37 to 10 including the hotspot, or 7 to 3 ignoring it, which represents a bias not commonly observed *in vitro* (GARCIA-DIAZ and KUNKEL 2006). We observed no complex mutations, transposable-element insertions, or multiple mutations.

**Genomic mutation rate:** Most mutations are deleterious, some are neutral, and few are advantageous. Because the load of deleterious mutations depends primarily upon the rate and kinds of spontaneous mutations, a primary goal of this study was to estimate the genomic mutation rate for *T. thermophilus*. The calculation involves correcting the genic rate for mutation-detection efficiency, followed by scaling the resulting value to the size of the entire genome. Most mutation-reporter genes detect indels with high efficiency, reflecting the gross disruption of the encoded protein, but they typically underestimate the frequencies of BPSs by several-fold, reflecting the predominance of phenotypically subtle or silent BPSs. This phenotypic under-representation requires some form of correction in order to estimate the total
BPS frequency. The first approach uses an average correction factor of 4.726 (± 2.778, 14 samples) determined historically for numerous genes (GROGAN et al. 2001). One potential weakness of this approach is the large variance around the average, while another is the evidence that extreme thermophiles experience stronger purifying selection on BPSs than do the mesophiles from which this average is derived (FRIEDMAN et al. 2004). The latter result suggests that the average missense mutation is more deleterious in an organism growing at higher temperatures, and failure of the historical ratio to account for this stronger selection would cause the BPS rate in *T. thermophilus* to be over-estimated. Another approach estimates the BPS rate from the number of chain-termination mutations (CTs) recovered. The total number of CTs is divided by the total number of pathways to CTs in the mutation-reporter sequence, and is then multiplied by 3 because there are three possible base substitutions per base pair. However, this approach has two limitations. First, A·T → G·C mutations cannot generate CTs. Second, the number of CTs in a spectrum can be small; in the present case, for example, it was only 2 (Table 2).

Starting from the observed average mutation rate for *pyrE* + *pyrF* of (3.21 ± 2.06) × 10⁻⁷ based on 73 mutations (54 indels + 19 BPSs), the genic rate corrected using the historical factor is (3.21 ± 2.06)(10⁻⁷)(54 + 4.726 × 19)/73 = (6.32 ± 4.06) × 10⁻⁷ with a 95% confidence interval (CI) from (3.21 ± 2.06) × 10⁻⁷ to (11.0 ± 7.06) × 10⁻⁷. (Note that the interval boundaries themselves have SDs.) The mutation-reporter consists of 1326 nt, and the genome size, G, equals 2,127,482 nt, yielding a genomic rate of (6.32 ± 4.06)(10⁻⁷)(1/1326)(2.13 × 10⁶) = 0.00101 ± 0.00065 with a 95% CI from 0.00052 ± 0.00033 to 0.00177 ± 0.00113. This overall rate is composed of an indel component of 0.00038 and a BPS component of 0.00063.

There are 103 pathways to CTs in the mutational target, including multiple pathways within a codon and multiple pathways at a site, and 2 CTs were observed. Thus, the BPS rate is (2/73)(3.21 ± 2.06)(10⁻⁷)(3/103) = (2.56 ± 1.64) × 10⁻¹⁰ per nt or (2.56 ± 1.64)(10⁻¹⁰)(2.13 × 10⁶) = 0.00054 ± 0.00035 per genome, close to the rate calculated by the alternative method. Adding the indel genomic rate of 0.00038 gives a total genomic rate of 0.00093.
Because of their close correspondence, we averaged the values derived by the two methods to obtain a consensus genomic rate of 0.00097 ± 0.00052 [with a 95% CI from 0.00073 ± 0.00034 to 0.00135 ± 0.00084] mutations per replication, representing an indel rate of 0.00038 and a BPS rate of 0.00059. This total genomic rate is distinctly lower than the value of 0.0035 ± 0.0007 observed for DNA genomes of diverse mesophiles (see Discussion), primarily because of the low value for the BPS rate. Using a parametric bootstrap method, the thermophile-mesophile difference has \( P < 0.0001 \) for the lower CI boundary, \( P = 0.0002 \) for the central value, and \( P < 0.005 \) for the upper CI boundary.

**DISCUSSION**

To our knowledge, the present study provides the first quantitative molecular assessment of DNA replication accuracy *in vivo* at an extremely high temperature in a bacterium (as opposed to an archaeon). Furthermore, use of the same selection and orthologous mutational targets in *T. thermophilus* and *S. acidocaldarius* (GROGAN et al. 2001) facilitates comparison of the mutational propensities of two highly divergent prokaryotes, each adapted to geothermal environments. The mutation assays gave no hint of phenotypic lag in *T. thermophilus*, either in the rates estimated separately from the three quartiles in the fluctuation tests or in the accumulation-test estimate, which was only slightly lower than the value determined from the fluctuation tests, rather than substantially higher as would have resulted from phenotypic lag. The forward (loss-of-function) mutations recovered using FOA selection included the types of events typical of spontaneous mutation in other systems: a minority of BPSs, many small indels, and a few larger tandem duplications and deletions. Most of the latter two types of mutations arose between short repeated sequences, which is seen in most target genes of mesophiles, although not those of the hyperthermophilic archaeon *S. acidocaldarius* (GROGAN et al. 2001). This result can be compared with 86% BPSs among loss-of-function mutations of the *S. cerevisiae* URA3 gene recovered by the same selection (LEE et al. 1998).
The question sometimes arises of the reliability of a genomic mutation rate obtained by extrapolation from mutation reporters which may comprise 0.1% or less of the genome. Even in the most sophisticated genetic systems, few target genes can assay spontaneous forward mutation via selection, and the choices remain particularly limited in *T. thermophilus*. However, empirical data relate directly to the question of whether spontaneous mutations are expected to be significantly less frequent in the *pyrEF* genes than in the average genomic interval of similar size in *T. thermophilus*, which is the aspect of variation most relevant for interpretations of the present study. First, the mutant reporter genes must be sequenced for the data to be considered robust, due to the influence of mutational hotspots on the genic rate and spectrum. These hotspots almost always consist of simple repeats, which are prone to slippage during DNA replication. As a consequence, contributions from these sites are readily distinguished from BPSs by sequence analysis, which allows BPS rates to be calculated with minimal distortion from the irregular distribution of hotspots among genes.

Second, the *T. thermophilus* rate is based on two genes with only one indel hotspot between them. This hotspot is moderate in strength and, as a mononucleotide run, conforms to the classical slipped-strand model of frameshift mutation (STREISINGER and OWEN 1985). The rates of spontaneous frameshifts at these runs have been analyzed in diverse mismatch-deficient and proficient systems; the rates increase significantly with run length for a given mononucleotide, and are significantly higher for G(C) vs. A(T) runs of equal length (SAGHER *et al.* 1999; GRAGG *et al.* 2002). These intrinsic properties allow the quantitative impact of the *pyrE* hotspot to be compared to that of mononucleotide runs in the rest of the *T. thermophilus* genome. We analyzed the complete HB27 genome sequence and identified one occurrence each of 8A(T) and 9A(T), 174 occurrences of 8G(C), five of 9G(C), and two of 10G(C). Thus, A(T) runs represent a relatively minor source of spontaneous mutation for the *T. thermophilus* genome overall. Furthermore, only about 9% of *T. thermophilus* genes have 8G(C) runs expected to match the mutational strength of the *pyrE* hotspot, and only about 0.4% have longer runs expected to be more mutable than this. Corresponding repeats of dinucleotides, which do not occur in the
pyrEF sequences, are only about one-tenth as abundant as mononucleotide runs throughout the
*T. thermophilus* genome. Thus, the frameshift hotspot found in *pyrE* is one of the few in *T.
thermophilus* having both the strength and abundance to affect the genomic rate of spontaneous
mutation, and its distribution predicts that the target genes would tend to over-estimate, rather
than under-estimate, this genomic rate.

Third, the set of mesophiles that form a basis for comparison have genomic mutation rates of
0.0025, 0.0027, 0.0027, 0.0030, 0.0038, 0.0040, 0.0045 and 0.0046 (mean = 0.0035 ± 0.00097)
for *Escherichia coli*, *Saccharomyces cerevisiae*, Herpes simplex virus 1, *Neurospora crassa*,
bacteriophage λ, bacteriophage T2/T4, *Schizosaccharomyces pombe*, and bacteriophage M13,
respectively, estimated using the historical correction factor (DRAKE et al. 1998; FRASER et al.
2003; DRAKE and HWANG 2005). All but one of these values are based on only one or two
reporter genes but the range of values resulting from the same methods of calculation is
nevertheless less than twofold. The similarity of these individual genomic mutation rates drawn
from diverse mesophiles, and the deviation of the corresponding *T. thermophilus* rate from this
norm, thus provide empirical evidence for the robustness of the conclusion that the *T.
thermophilus* genomic rate is lower than the characteristic mesophile rate.

Because indel hotspots were more the exception than the rule in the reporter genes used to
obtain the mesophilic genomic rates listed above, it is informative to calculate the residual *T.
thermophilus* mutation rates in the absence of the 8G hotspot. The genomic indel rate would
decrease from 0.00038 to 0.000056. Using the historical factor to calculate the BPS rates, the
total genomic rate would decrease from 0.00101 to 0.00069. Using the CT method, it would
decrease from 0.00093 to 0.00060. The corresponding change in average genomic rate for the
reference set of mesophiles is quite small, so that the contrast between the thermophilic and the
average mesophilic genomic rates increases substantially when the thermophile hotspot is
discounted.

The similarity of the mutational properties of *T. thermophilus* and *S. acidocaldarius* contrasts
with the deep evolutionary and biochemical divergence between them. We note, for example,
that *T. thermophilus* and *S. acidocaldarius* share aerobic and heterotrophic modes of growth, extremely high optimal growth temperatures, and genomes of about 2.2 Mbp. However, they have dramatically different DNA base compositions (69 vs. 38 mol% G+C) and represent different domains of cellular life (bacteria vs. archaea), resulting in different molecular contexts of genome replication. Thus, *T. thermophilus* employs a family-A DNA polymerase for genome replication (BULLARD et al. 2002), whereas *Sulfolobus* spp. and other archaea possess only family-B polymerases capable of genome replication (SHE et al. 2001; KAWARABASHI et al. 2001). Archaeal family-B polymerases have the peculiar characteristic of stalling just before template deoxyuridine residues (LASKEN et al. 1996; SHUTTLEWORTH et al. 2004) although, to our knowledge, the functional significance of this property has not been confirmed *in vivo*. *T. thermophilus* and *S. acidocaldarius* have different types of DNA-binding proteins (ZIERER and CHOLI 1990; DU and PÈNE 1999) and different numbers of replication origins (LUNDGREN et al. 2003; HENNE et al. 2004). In addition, strain HB27 lacks an intact gene for reverse DNA gyrase, which is otherwise highly conserved among bacteria and archaea from geothermal habitats (FORTERRE 2002), including *S. acidocaldarius* and *T. thermophilus* strain HB8 (HENNE et al. 2004; CHEN et al. 2005). With respect to DNA repair, *T. thermophilus* encodes complete, bacterial-type systems for nucleotide excision repair (UvrABC) and DNA mismatch repair (MutHSL), whereas *S. acidocaldarius* encodes an apparently incomplete nucleotide-excision-repair system of the eukaryotic type (XPF, XPG) and no identifiable mismatch-repair genes (GROGAN 2004). Conversely, *T. thermophilus* apparently lacks any DNA polymerase of the Y family, whereas *S. acidocaldarius* encodes at least one such enzyme, designated Dbh (BOUDSOCQ et al. 2004). One predicted consequence of this latter difference is weaker induced mutation in *T. thermophilus* compared to *S. acidocaldarius*, which is consistent with our observation that chemical mutagenesis in *T. thermophilus* appeared to be less effective than similar treatment of *S. acidocaldarius* (REILLY and GROGAN 2002).

In summary, quantitative molecular analysis of spontaneous mutations occurring *in vivo* indicates that DNA replication in *T. thermophilus* growing at 76°, far from being error-prone
(CASTÁN et al. 2003), generates mutations at an unusually low rate per genome replication. Comparison of the gene inventories of *T. thermophilus* and *S. acidocaldarius* further suggests that these two thermophiles achieve their low mutation rates by distinct sets of biochemical strategies. Overall, the available data suggest that cellular organisms in geothermal environments may be subject to a selective pressure against substitution mutations which is more stringent than the corresponding selection exerted on mesophiles, with corresponding implications for the rates of molecular evolution in these lineages.

We thank Marilyn Diaz and Tom Kunkel for critical readings of the manuscript. This work was supported by National Science Foundation grants MCB 9733303 and MCB 0543910 to D.W.G., and by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.
LITERATURE CITED

BOOM, R., C. J. SOL, M. M. SALIMANS, C. L. JANSEN, P. M. WERTHEIM-VAN DILLEN et al., 1990

BOUDSOCQ, F., R. J. KOKOSKA, B. S. PLOSKY, A. Vaisman, H. LING et al., 2004
Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis
and translesion replication. J. Biol. Chem. 279: 32932-32940.

BREUERT, S., T. ALLERS, G. SPOHN and J. SOPPA, 2006

BRÜGGEMANN, H., and C. CHEN, 2006
Comparative genomics of Thermus thermophilus:

BULLARD, J. M., J. C. WILLIAMS, W. K. ACKER, C. JACOBI, N. JANJIC et al., 2002
DNA polymerase III holoenzyme from Thermus thermophilus. Identification, expression,
purification of components, and use to reconstitute a processive replicase. J. Biol. Chem. 277: 13401-13408.


CHEN, L., K. BRUGGER, M. SKOVGAARD, P. REDDER, Q. SHE et al., 2005

COX, M. M., and J. R. BATTISTA, 2005

DRAKE, J. W., and C. B. C. HWANG, 2005
On the mutation rate of herpes simplex virus type 1. Genetics 170: 969-970.

DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH and J. F. CROW, 1998

DU, X., and J. J. PÈNE, 1999
Identification, cloning and expression of p25, an AT-rich DNA-binding protein from the extreme thermophile, Thermus aquaticus YT-1. Nucl. Acids Res.


KAWARABAYASI, Y., Y. HINO, H. HORIKAWA, K. JIN-NO, M. TAKAHASHI et al., 2001 Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, Sulfolobus tokodaii strain 7. DNA Res. 8: 123-140.


LASKEN, R. S., D. M. SCHUSTER and A. RASHTCHIAN, 1996 Archaebacterial DNA polymerases
tightly bind uracil-containing DNA. J. Biol. Chem. 271: 17692-17696.


Shuttleworth, G., M. J. Fogg, M. R. Kurpiewski, L. Jen-Jacobson and B. A. Connolly,
2004 Recognition of the pro-mutagenic base uracil by family B DNA polymerases from archaea. J. Mol. Biol. 337: 621-634


**TABLE 1**

Fluctuation-test mutation rates per $10^7$ cell divisions

<table>
<thead>
<tr>
<th>Trial</th>
<th>Quartile 1</th>
<th>Quartile 2</th>
<th>Quartile 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.11</td>
<td>2.36</td>
<td>2.22</td>
</tr>
<tr>
<td>2</td>
<td>1.99</td>
<td>2.17</td>
<td>1.88</td>
</tr>
<tr>
<td>3</td>
<td>4.39</td>
<td>7.92</td>
<td>9.47</td>
</tr>
<tr>
<td>4</td>
<td>6.67</td>
<td>6.93</td>
<td>6.93</td>
</tr>
<tr>
<td>5</td>
<td>0.022</td>
<td>0.022</td>
<td>0.11</td>
</tr>
<tr>
<td>6</td>
<td>1.03</td>
<td>2.71</td>
<td>2.54</td>
</tr>
<tr>
<td>7</td>
<td>7.79</td>
<td>6.72</td>
<td>6.45</td>
</tr>
<tr>
<td>8</td>
<td>2.58</td>
<td>6.32</td>
<td>6.32</td>
</tr>
<tr>
<td>9</td>
<td>3.14</td>
<td>3.06</td>
<td>2.27</td>
</tr>
<tr>
<td>10</td>
<td>7.97</td>
<td>9.11</td>
<td>7.97</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>3.67 ± 2.91</td>
<td>4.73 ± 3.02</td>
<td>4.62 ± 3.16</td>
</tr>
</tbody>
</table>

The global mean ± SD for all 10 fluctuation tests was $(4.34 ± 2.92) \times 10^{-7}$. 
<table>
<thead>
<tr>
<th>pyr</th>
<th>Position&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Mutation&lt;sup&gt;b&lt;/sup&gt;</th>
<th>No.</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>12/13</td>
<td>+A</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>79-186</td>
<td>−108</td>
<td>1</td>
<td>GGCCCCGAG&lt;sup&gt;(99)&lt;/sup&gt;GGCCCCGAG</td>
</tr>
<tr>
<td></td>
<td>86-94</td>
<td>TD7</td>
<td>1</td>
<td>CCCATATCC</td>
</tr>
<tr>
<td></td>
<td>88-108</td>
<td>TD20</td>
<td>1</td>
<td>CATATCCGCGCTACACCTC</td>
</tr>
<tr>
<td></td>
<td>163-168</td>
<td>−4</td>
<td>1</td>
<td>TTCTTT → TT (three ways)</td>
</tr>
<tr>
<td></td>
<td>177-179</td>
<td>+G</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>204</td>
<td>G→A</td>
<td>1</td>
<td>trp → UGA</td>
</tr>
<tr>
<td></td>
<td>249/250</td>
<td>+A</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>A→G</td>
<td>1</td>
<td>lys → glu</td>
</tr>
<tr>
<td></td>
<td>376</td>
<td>C→T</td>
<td>1</td>
<td>gln → UAG</td>
</tr>
<tr>
<td></td>
<td>425</td>
<td>C→A</td>
<td>1</td>
<td>ser → tyr</td>
</tr>
<tr>
<td></td>
<td>547</td>
<td>G→A</td>
<td>2</td>
<td>gly → glu</td>
</tr>
<tr>
<td></td>
<td>604-607</td>
<td>−C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>611-616</td>
<td>−C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>736/737</td>
<td>+C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>49</td>
<td>C→T</td>
<td>1</td>
<td>his → tyr</td>
</tr>
<tr>
<td></td>
<td>57/58</td>
<td>+A</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>73</td>
<td>C→T</td>
<td>1</td>
<td>his → tyr</td>
</tr>
<tr>
<td></td>
<td>76</td>
<td>T→C</td>
<td>2</td>
<td>ser → pro</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>C→T</td>
<td>1</td>
<td>ser → phe</td>
</tr>
<tr>
<td></td>
<td>116-119</td>
<td>+CCCT</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>Change</td>
<td>Count</td>
<td>New Amino Acid</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>-------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>137</td>
<td>G→A</td>
<td>3</td>
<td>gly → glu</td>
<td></td>
</tr>
<tr>
<td>199-203</td>
<td>-GG</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>227</td>
<td>C→T</td>
<td>1</td>
<td>ala → val</td>
<td></td>
</tr>
<tr>
<td>241</td>
<td>G→C</td>
<td>2</td>
<td>ala → pro</td>
<td></td>
</tr>
<tr>
<td>272-279</td>
<td>8G→9G</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>272-279</td>
<td>8G→7G</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>316-335</td>
<td>TD18</td>
<td>1</td>
<td><strong>GACCGCTTTCTGCGGTGGA</strong></td>
<td></td>
</tr>
<tr>
<td>335</td>
<td>A→G</td>
<td>1</td>
<td>glu → gly</td>
<td></td>
</tr>
<tr>
<td>391-400</td>
<td>TD5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>513-517</td>
<td>-C</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>519</td>
<td>C→C2C</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>521</td>
<td>G→GG</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>A→G</td>
<td>1</td>
<td>UAG → trp</td>
<td></td>
</tr>
</tbody>
</table>

*a* Relative to the first base of the initiation codon. “/” denotes an insertion between the indicated positions.

*b* “+” or “−” followed by a base indicates an insertion or deletion, respectively, within a short homopolymeric run or, followed by a number, indicates the number of bases inserted or deleted. “TD” followed by a number indicates a tandem duplication of that many bases. Flanking repeats are underlined in the last column, one copy of the repeat being included in the duplication or deletion. The deletion at *pyrF* 163-168 can be generated by removing any consecutive four bases from the indicated sequence.
### TABLE 3

Types of mutations arising spontaneously in *Thermus thermophilus*

<table>
<thead>
<tr>
<th>Type</th>
<th><em>pyrE</em></th>
<th><em>pyrF</em></th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>57</td>
<td>16</td>
<td>73</td>
</tr>
<tr>
<td>BPSs</td>
<td>13</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>Transitions</td>
<td>11</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Transversions</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Indels</td>
<td>44</td>
<td>10</td>
<td>54</td>
</tr>
<tr>
<td>+1</td>
<td>34</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>(HS)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>–1</td>
<td>6</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>(HS)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>±(2-108)</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

<sup>a</sup> The contribution from indels at the 8G hotspot.

<sup>b</sup> Tandem duplications.
Figure 1.—Spectrum of *pyrF* and *pyrE* mutations. Initiating and terminating codons are bold-face. Insertions between adjacent bases are written directly above the two bases. Insertions or deletions within mononucleotide repeats are shown as + or – above the lightly underlined repeat. Tandem duplications are heavily underlined, the flanking repeats by dotted underlines. The single large *pyrF* deletion is not shown; its flanking repeats are at positions 79-87 and 178-186.