

Opposing Forces of A/T-Biased Mutations and G/C-Biased Gene Conversions Shape the Genome of the Nematode *Pristionchus pacificus*

Andreas M. Weller, Christian Rödelsperger, Gabi Eberhardt, Ruxandra I. Molnar, and Ralf J. Sommer¹
Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

ABSTRACT Base substitution mutations are a major source of genetic novelty and mutation accumulation line (MAL) studies revealed a nearly universal AT bias in *de novo* mutation spectra. While a comparison of *de novo* mutation spectra with the actual nucleotide composition in the genome suggests the existence of general counterbalancing mechanisms, little is known about the evolutionary and historical details of these opposing forces. Here, we correlate MAL-derived mutation spectra with patterns observed from population resequencing. Variation observed in natural populations has already been subject to evolutionary forces. Distinction between rare and common alleles, the latter of which are close to fixation and of presumably older age, can provide insight into mutational processes and their influence on genome evolution. We provide a genome-wide analysis of *de novo* mutations in 22 MALs of the nematode *Pristionchus pacificus* and compare the spectra with natural variants observed in resequencing of 104 natural isolates. MALs show an AT bias of 5.3, one of the highest values observed to date. In contrast, the AT bias in natural variants is much lower. Specifically, rare derived alleles show an AT bias of 2.4, whereas common derived alleles close to fixation show no AT bias at all. These results indicate the existence of a strong opposing force and they suggest that the GC content of the *P. pacificus* genome is in equilibrium. We discuss GC-biased gene conversion as a potential mechanism acting against AT-biased mutations. This study provides insight into genome evolution by combining MAL studies with natural variation.

BASE substitution mutations are a major source of genetic novelty and provide the raw material on which evolutionary forces will act. An accurate picture of the processes generating *de novo* mutations and their frequencies is fundamental for the understanding of evolution (Lynch 2007), as well as for mutational processes causing diseases (Lu *et al.* 2011). Mutation accumulation (MA) line analysis became the method of choice to study the frequencies and spectra of *de novo* mutations (Mukai 1964). Most recently, MA lines have been used in model species in combination with whole-genome sequencing. Among multicellular organisms, this included *Caenorhabditis elegans* (Denver *et al.* 2009, 2012) and *C. briggsae* (Baer *et al.* 2005; Howe *et al.* 2010), *Drosophila melanogaster* (Keightley *et al.* 2009), *Daphnia pulex* (Seyfert *et al.* 2008), and *Arabidopsis thaliana* (Ossowski

et al. 2010). These studies showed that the base substitution mutation rate (μ_{bs} , measured in mutations per site per generation), while stable to one order of magnitude, is positively correlated with genome size, but negatively correlated with the effective population size across organisms (Lynch 2007). In contrast, the mutation spectra vary considerably between species. In all studies done so far, *de novo* mutations from G/C to A/T outweigh mutations from A/T to G/C, a phenomenon that has been described as AT bias (Lynch 2007). Such a mutational bias creates a pressure toward higher AT content of genomes. However, the actual AT content of most genomes is far from what would be predicted by these mutational forces alone, indicating the existence of other forces that operate in the opposite direction (Lynch 2007).

The comparison of MA line-derived mutational spectra with genome composition in organisms covering all domains of life has been most powerful in providing insight into genome evolution. One additional fruitful line of investigation is the correlation of MA line-derived patterns of mutation frequencies and spectra with those patterns observed from natural variation studies. Variation observed in natural populations has already been affected by adaptive and nonadaptive

Copyright © 2014 by the Genetics Society of America
doi: 10.1534/genetics.113.159863

Manuscript received November 17, 2013; accepted for publication January 2, 2014;
published Early Online January 10, 2014.

Sequence data from this article have been deposited under GenBank accession no.
SRA067565.

¹Corresponding author: Max Planck Institute for Developmental Biology,
Spemannstrasse 37, BW 72076 Tübingen, Germany. E-mail: ralf.sommer@tuebingen.
mpg.de

forces and can therefore capture the result of mutational processes and evolutionary forces more accurately than a single reference genome sequence of the organism in question. Such studies will be most informative for species with a comprehensive collection of wild isolates that cover the complete geographic range and microevolutionary time frame of diversification of the organisms. Specifically, the comparison of MA line-derived patterns with those of rare and common alleles, the latter of which are close to fixation and thus of presumptive older age, might provide valuable new insight into mutational processes and their influence on genomes and evolution.

One candidate for such studies is the nematode *Pristionchus pacificus*. This species has recently been developed as a model system for integrative studies in evolutionary biology (Sommer 2009). *P. pacificus* has well-developed functional tools for genetic, genomic, and transgenic studies. Detailed knowledge about its ecology based on its tight association with scarab beetles, its population genetics, and natural history is available (McGaughran *et al.* 2013). MA lines have been generated in *P. pacificus* (Molnar *et al.* 2011) and >600 wild isolates covering its cosmopolitan distribution are available in the laboratory. Four major clades, called A–D, can be distinguished within *P. pacificus*, with two clades (B and C) being endemic to the Mascarene Islands La Reunion and Mauritius in the Indian Ocean (Morgan *et al.* 2012). Of particular interest for genomic studies is *P. pacificus* clade C that underwent a massive radiation on La Reunion (Morgan *et al.* 2012). A subset of 104 *P. pacificus* wild isolates had their genome of 169 Mb resequenced, which revealed strong divergence, resulting in >6 million single-nucleotide polymorphisms (SNPs) (Rödelsperger *et al.* 2014). Furthermore, the patterns of divergence and global dispersal between the four *P. pacificus* clades and the repeated invasion of the geologically young volcanic island of La Reunion (Herrmann *et al.* 2010; McGaughran *et al.* 2013) indicated that the *P. pacificus* genome is shaped by weakly deleterious mutations and linked selection (Rödelsperger *et al.* 2014). These studies provide a genomic and phylogenetic framework for evolutionary comparisons of mutational patterns.

Here, we resequenced the full genome of 22 independent MA lines and followed the mutational patterns along three different avenues of analysis. First, we studied *de novo* mutations in MA lines and provide estimates of μ_{bs} for *P. pacificus* as a baseline for reconstruction of the evolutionary history of the species. We used this μ_{bs} estimate to calculate the time to the most recent common ancestor (TMRCA) of the *P. pacificus* clades. Second, we investigated the role of chromosome location and coding state on mutation rates. By comparing signatures of selection against nonsynonymous substitutions between MA line mutations and natural variants, we can test the assumption of neutral mutation accumulation in MA lines. Finally, we studied the mutation spectrum in terms of transition–transversion ratios and AT bias. We observed strong differences in the mutation spectra between MA line-derived *de novo* mutations and natural variants. In addition, the mutational bias in

natural variants is negatively correlated with their age, suggesting the result of selection processes in nature that balance mutation spectra.

Materials and Methods

MA line creation and library preparation

The creation of mutation accumulation lines has been described in detail by Molnar *et al.* (2011). Briefly, 100 lines were initiated from the offspring of a single individual of the strain *P. pacificus* PS312. The lines were propagated for 142 generations under benign conditions by transferring a single random hermaphrodite per generation. The 82 surviving lines were frozen in liquid nitrogen (Molnar *et al.* 2011). We later thawed 22 lines and extracted DNA for Illumina sequencing, using the MasterPure DNA purification kit. DNA was first sheared using the Covaris S2 system. End repair, adenylation, and ligation of one of the 12 index adapters were performed using reagents and the standard protocol of the Illumina TruSeq v2 sample preparation kit as described in Rödelsperger *et al.* (2014) (accompanying article in this issue). Samples were selected for an insert size of 300–400 bp by excision from an agarose gel. Library concentration was finally determined on a Bioanalyzer DNA 1000 chip and normalized to 10 nM before pooling into 12-plexes. The libraries were sequenced in house on an Illumina HiSeq2000 sequencer.

Sequencing and variant calling

Illumina reads were aligned to the reference genome (*P. pacificus* Hybrid Assembly) by stampy (version 1.0.13) (Lunter and Goodson 2011). Duplicate reads were removed using the samtools (version 0.1.17) (Li *et al.* 2009) “rmdup” command. After quality filtering, coverage per line was between 4× and 34×. Initial SNP calling was performed using samtools and bcftools (version 0.1.17) (Li *et al.* 2009). To remove low-quality mutations as well as mutations called in several of the lines (which indicated a mutation already present in the founder animal), we employed a custom Python script for further conservative filtering. Putative mutations were accepted as candidates if they (a) were covered by between 5 and 30 reads, (b) had a minimum phred quality score of 10, and (c) had at least an FQ value of 40 (indicating a homozygous position). Candidates were then compared to the founder animal and all other MA lines. Mutations were accepted as unique only if (a) not a single read in any other line indicated the same alternative base and (b) at least 10 other lines had a high-quality reference consensus at the given position. We used classic Sanger sequencing to verify the accuracy of our mutation calling. Primers for 21 randomly selected mutations were designed using a custom pipeline implementing Primer3 (Untergasser *et al.* 2007) for initial primer finding and NCBI Blast+ (Johnson *et al.* 2008) to confirm the uniqueness of primer candidates in the reference sequence. After sequencing, reads were aligned by NCBI Blast+ (Johnson *et al.* 2008) and mutations were called using a custom Python script.

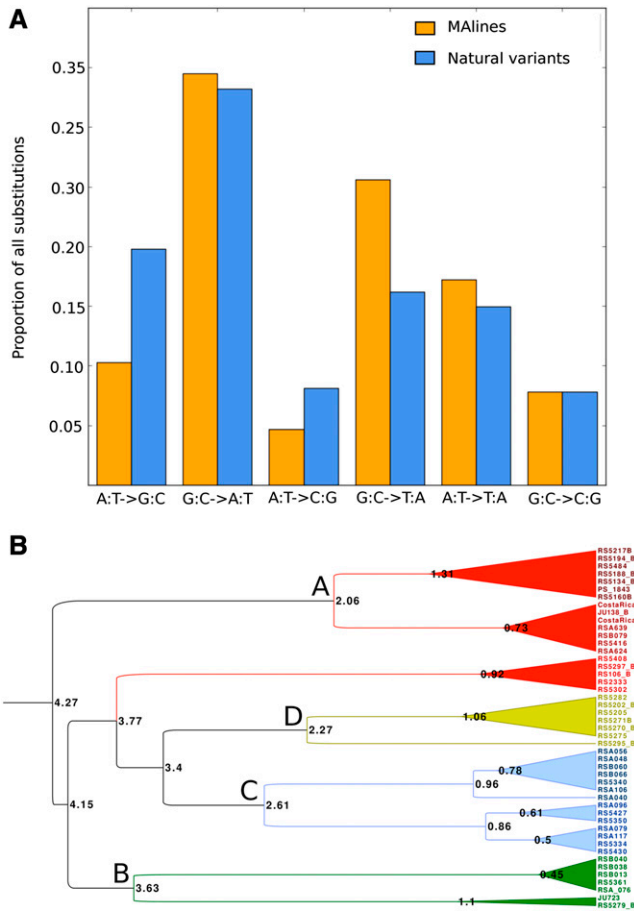


Figure 1 (A) Comparison between the base substitution types in MA line mutations and natural variants as a percentage of all substitutions. (B) Phylogeny of *P. pacificus* derived from the mutation rate estimated in this study and 5 Mb of sequence. Number at nodes represents time to the most recent common ancestor of the node in millions of generations.

Mutation rate and spectrum

The mutation rate was calculated as previously described by Denver *et al.* (2009) as $\mu_{bs} = m/(LnT)$ where μ_{bs} is the base substitution rate per generation, m is the number of mutations, L is the number of lines, n is the number of sites that passed our criteria for mutation calling, and T is the number of generations. We used chi-squared tests to evaluate the randomness of mutations between lines, chromosomes, and functional categories. The expected random distribution was calculated based on the null expectation that, *e.g.*, the total number of mutations would be uniformly distributed among the lines in accordance with the number of sites actually considered in each line. A custom Python script identified the coding context of each mutation, using the set of genes described previously (Borchert *et al.* 2010). All mutations that did not fall between the boundaries of a known gene were classified as intergenic. Three mutations that hit two different genes, on the sense- and antisense strand, were counted for both hits. We followed the procedure from Lynch (2007, p. 125) to calculate the rate bias,

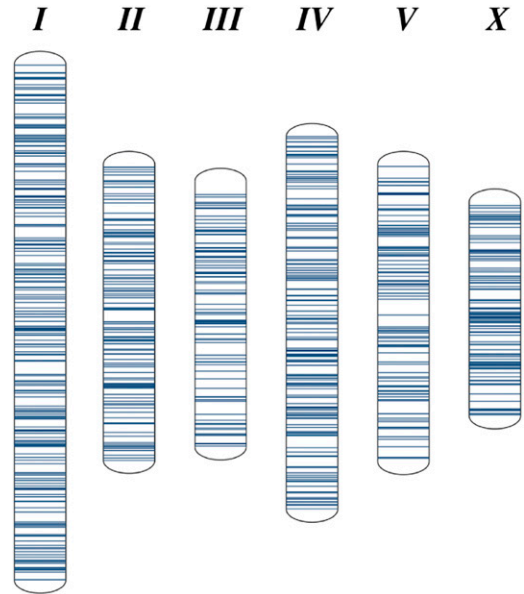


Figure 2 Chromosome location of *de novo* mutations in the MA lines. Each vertical bar represents a single mutation. Chromosome sizes are drawn to scale according to their genetic size.

$$ATbias = \frac{m(\rightarrow A/T)}{\Delta GC} * \left(\frac{m(\rightarrow G/C)}{1 - \Delta GC} \right)^{-1},$$

Where $m(\rightarrow G/C)$ stands for the fraction of mutations that change an A/T nucleotide into a G/C nucleotide and ΔGC is the fraction of G/C nucleotides in the genome.

The expected equilibrium AT composition is calculated as $m/(1 + m)$, where m is the AT bias. The Ts/Tv rate is simply the ratio of the total number of transitions to transversion mutations.

TMRC estimation

We inferred relative divergence times for the phylogeny of the 47 *P. pacificus* strains, using BEAST (v.1.7.4) (Drummond *et al.* 2012) on a randomly chosen data set of 5 Mb. We inferred the topology of the trees under the birth-death speciation model (Gernhard 2008), using the Hasegawa-Kishino-Yano and generalized time-reversible substitution models and assuming a strict molecular clock on all branches with the mutation rate found in the MA lines. We ran five independent runs for 10 million generations, sampling every 1000th generation. We verified the convergence of runs by examining the effective sample size of the likelihood and posterior probability parameters for each analysis and by visual inspection in Tracer (v.1.5) (Drummond and Rambaut 2007). The trees and parameter estimates from the five runs were combined using LogCombiner (v.1.7.4) (Drummond *et al.* 2012). The results were considered reliable only when the effective sampling size of all parameters was >100 . Using TreeAnnotator (v.1.7.4) (Drummond *et al.* 2012), the samples from the posterior were summarized on the maximum credibility tree, with the posterior probability limit set to 0.8 and summarizing mean node heights.

Table 1 Properties of *de novo* mutations in MA lines

Coding state	μ_{bs}	A:T to G:C ^a	G:C to A:T	A:T to C:G	G:C to T:A	G:C to C:G	A:T to T:A	<i>N</i>	Ts:Tv	AT bias ^b
Exon	2×10^{-9}	12	49	3	38	9	16	127	0.92	5.3
Intron	1.9×10^{-9}	33	108	17	79	22	73	332	0.74	5.9
Intergenic	2.1×10^{-9}	39	118	17	89	31	49	343	0.84	4.8

^a Total number in all 22 MA lines.

^b Weighted by nucleotide composition at the coding state considered.

Analysis of natural variants

For the comparison of AT bias between MA lines and natural strains, we used a data set of 104 resequenced strains (Rödelsperger *et al.* 2014). As described in Rödelsperger *et al.* (2014), ancestral alleles were determined based on maximum-likelihood comparison to all 104 sequenced strains and the outgroup species *P. expectatus*. In total, a derived allele in at least one strain was present at 655,705 natural variant sites in the genome. The codeml program from the PAML package (Yang 1997) was used to infer a maximum-likelihood estimate of the d_N/d_S ratio ω in *de novo* mutations and natural variants separately. ω was averaged over all exons of the 118 genes carrying an exonic *de novo* SNP in the MA lines. The same exons were used in the natural variants to obtain a comparable estimate of ω . Codeml was run assuming the F3x4 codon frequency model and a uniform ω across branches.

Results

Whole-genome sequencing of MA lines

P. pacificus MA lines were propagated for 142 generations and first analyzed for mutation accumulation and estimates of the TMRCA in the mitochondrial genome (Molnar *et al.* 2011). Here, we sequenced the nuclear genome of 22 MA lines, using next generation sequencing technology (see *Materials and Methods*). In total, 746 million reads were generated, resulting in mean genome coverage of 16 \times . After alignment to the reference, a total genome sequence of 49–141 Mb per MA line passed the quality cutoff for mutation calling. To avoid false-positive mutation calls due to assembly errors, we used a conservative pipeline that selected for high-quality putative mutations and compared them between lines to filter for duplicate sites. We evaluated 21 mutations by PCR and Sanger sequencing and confirmed all of them as valid calls, indicating that the number of false-positive mutation calls is limited.

De novo mutations and mutation frequencies

In total, we identified 802 *de novo* mutations in the nuclear genome of these 22 MA lines (Figure 1A). Mutations appear to be equally distributed over all chromosomes (Figure 2), indicating no evidence for mutation hot spots in *P. pacificus*. We calculated the rate of base substitution μ_{bs} as the number of mutations divided by the number of generations and sites. Thus, the line-specific base substitution rate was

derived by taking the number of sites actually considered in each line into account. The mean μ_{bs} is 2×10^{-9} with a nonsignificant (χ^2 test, $P = 0.27$) variation among the 22 MA lines from 1.4 to 2.6×10^{-9} . By analyzing a sequence of 5 Mb in 47 representative wild isolates, we used the identified μ_{bs} to calculate the TMRCA of the species. Following an approach described elsewhere (Molnar *et al.* 2011), we estimated the TMRCA for all *P. pacificus* strains at 4.3×10^6 generations (Figure 1B).

Influence of coding state on mutation rate

Next, we evaluated the effect of mutations on gene function (Figure 1A and Table 1). Of the 802 mutations, 127 were found in exons, including nine premature stop codons and 88 nonsynonymous and 30 synonymous changes. This distribution does not differ from an assumed random distribution (Fisher's exact test, $P = 0.76$). Five of the genes with stop codons were orphan genes without a *C. elegans* 1:1 ortholog, while 3 genes have a *C. elegans* 1:1 ortholog with unknown function. The d_N/d_S rate ω serves as a measurement of the strength of selection acting on a sequence, with ratios >1, 1, and <1 indicating positive, neutral, and purifying selection, respectively (Yang and Bielawski 2000). We used a maximum-likelihood approach to calculate ω as 1.06 over all 118 genes with exonic mutations in the MA lines. Interestingly, variants found in the same 118 genes in the genome of the available 104 natural isolates have an ω of only 0.17. These results suggest that mutations observed in MA lines follow neutral expectations, whereas natural variation within the same genes indicates the action of purifying selection.

Mutation spectra and AT bias

To obtain insight into the processes generating mutations, we analyzed the mutation spectra in the 22 sequenced MA lines. We observed a ratio of transitions to transversions (Ts/Tv) in MA lines and natural populations of 0.8 and 1.1, respectively, which is a significant difference (Fisher's exact test, $P < 0.0001$). In addition, it is higher in exons than in introns and intergenic regions, both in MA lines and in natural variants (Fisher's exact test, $P < 0.0001$). Next, we analyzed μ_{bs} separately for all six possible base substitution types. We found a strong bias toward G:C to T:A transversions and G:C to A:T transitions (Figure 1A). On average, we observed 3.9 mutations toward A/T for every mutation toward G/C. We calculated the AT bias by weighting for GC content in the genome, following standard procedures (Lynch

Table 2 Base substitution types, Ts:Tv ratio, and AT bias across species

Species	A:T to G:C ^a	G:C to A:T	A:T to C:G	G:C to T:A	G:C to C:G	A:T to T:A	<i>N</i>	Ts:Tv	AT bias ^b
<i>P. pacificus</i>	0.105	0.344	0.046	0.258	0.078	0.173	802	0.81	5.3
<i>C. elegans</i> (Denver <i>et al.</i> 2009)	0.105	0.205	0.113	0.281	0.077	0.22	391	0.448	2.235
<i>D. melanogaster</i> (Keightley <i>et al.</i> 2009)	0.19	0.305	0.115	0.155	0.115	0.121	174	0.977	1.509
<i>A. thaliana</i> (Ossowski <i>et al.</i> 2010)	0.118	0.588	0.071	0.094	0.071	0.059	85	2.4	3.625
<i>Saccharomyces cerevisiae</i> (Lynch <i>et al.</i> 2008)	0.103	0.286	0.057	0.312	0.182	0.06	1250	0.638	3.74
<i>Homo sapiens</i> (Lynch 2010b)	0.221	0.408	0.078	0.110	0.125	0.067	3336	1.697	1.732

^a Fraction of all mutations.

^b Weighted by genomic nucleotide composition.

2007). We found an AT bias of 5.3, indicating a strong decrease in GC content of the genome as a consequence of *de novo* mutations. This value represents one of the highest AT bias values based on whole-genome sequencing known to date (Table 2).

AT bias and population structure

Finally, we compared the mutation spectra between MA lines and natural variants to infer the mode by which selective forces act on newly arising natural variants. In total, 655,705 natural variants were considered. An overwhelming majority of these variants were present in only a few strains (Figure 3). In contrast to the AT bias of *P. pacificus* MA line mutations, the mean AT bias of the natural variants is only 2.4. In addition, the AT bias of derived alleles is negatively correlated with their frequency among the sampled strains; *i.e.*, alleles close to fixation show no bias compared to rare alleles (Figure 3A). The AT bias does not differ between coding and non-coding regions (Figure 3A). Together, these results suggest that a strong force exists, counterbalancing the mutational AT bias.

Discussion

In this study we analyzed base substitution mutations in 22 MA lines and compared the mutation patterns and spectra with natural variants obtained in 104 wild isolates of *P. pacificus*. The mutation rate of *P. pacificus* is surprisingly similar to the μ_{bs} of $0.8\text{--}2.1 \times 10^{-9}$ reported for several *Caenorhabditis* species (Denver *et al.* 2012). These findings suggest a stable mutation rate in nematodes over an evolutionary separation time of >200 million years, the expected time to the TMRCA of *Pristionchus* and *Caenorhabditis* (Dieterich *et al.* 2008). As the mutation rate found in MA lines in *D. melanogaster* (Keightley *et al.* 2009) is also very similar, our result is another indicator that mutation rates can be generalized even among higher animal taxa. This result lends credence to phylogenetic analysis based on molecular clocks, such as our estimation of the TMRCA of the *P. pacificus* clades, which is at the same order of magnitude as previous mtDNA-based analyses (Molnar *et al.* 2011).

The finding of equal mutation rates for nematodes with very different effective population sizes (Rödelsperger *et al.* 2014) and gene numbers (Dieterich *et al.* 2008) is not predicted

by the mutation-barrier framework. The genomic content of coding DNA and the effective population size are argued to be the lower limit for an ever-decreasing mutation rate driven by natural selection, as the overwhelming majority of mutations are deleterious (Lynch 2010a; Sung *et al.* 2012). A caveat to be considered here is the recent bottleneck in most *C. elegans* populations (Andersen *et al.* 2012), which might lead to an underestimation of the effective population size in this species.

We did not find a deviance from random expectations between individual MA lines, chromosomes, or coding states. Also, there was no significant correlation between mutation in MA lines and recombination patterns based on a previous linkage map of *P. pacificus* (www.pristionchus.org; our unpublished observation). In addition, the effects of mutations in exons were randomly distributed between stop codons and nonsynonymous and synonymous mutations. A formal analysis of d_N/d_S further showed that there is no evidence for selection in the MA lines, while there is a clear trace of strong purifying selection in the natural variants. These results follow the *a priori* expectation that the benign laboratory environment leads to an unbiased accumulation of mutations in MA lines. Together, the ω of 1 and the random distribution of mutations are in accordance with the central premise of MA experiments, observing *de novo* mutations without the strong influence of selection.

The ratio of transitions to transversions shows some interesting patterns. First, the Ts/Tv ratio is higher in exons compared to noncoding regions, which is the opposite pattern from the one observed in *C. elegans* (Denver *et al.* 2012). A role of transcription in the Ts/Tv ratio runs counter to the argument made in Babbitt and Cotter (2011) that the Ts/Tv ratio is shaped by the energetics of nucleosome architecture leading to stronger purifying selection against transversions, which should be independent of coding state. Second, we would *a priori* expect the two types of AT driving mutations (G:C to A:T transitions and G:C to T:A transversions) to be equally changed by GC-biased gene conversion in the natural variants. We observe a much stronger effect of GC-biased gene conversions in G:C to T:A transversions, but it is unclear why this is the case.

The spectrum of *de novo* mutations in *P. pacificus* shows an AT bias of 5.3, one of the highest values observed by high-throughput sequencing to date (Table 2) (Lynch 2010b). This result allows two important conclusions. First, the nearly

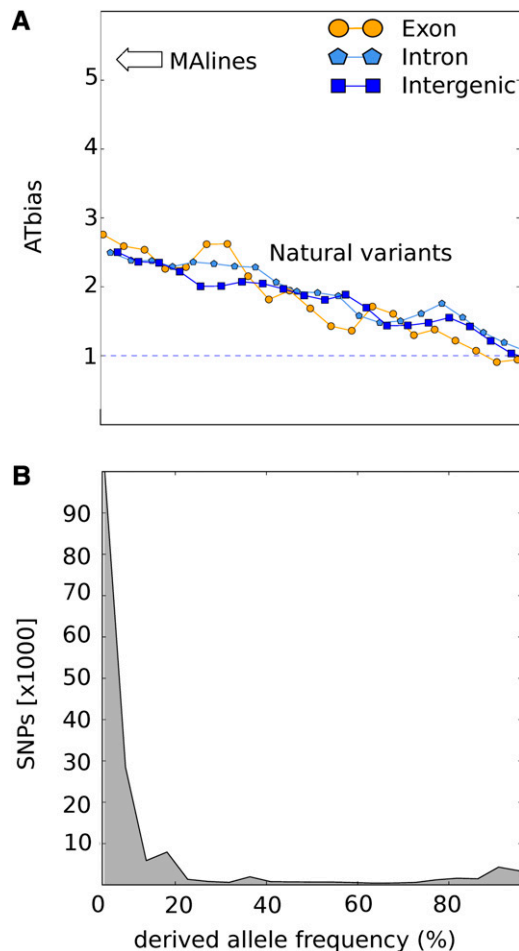


Figure 3 (A) AT bias in the natural variants of *P. pacificus* divided by derived allele frequencies in the population. MA line-derived AT bias is included for comparison. (B) Number of natural variants studied per derived allele frequency.

identical mutation rates of *P. pacificus* and *C. elegans* result from different mutation spectra as *C. elegans* has an AT bias of only 2.2. One possible explanation for these findings could be that *P. pacificus* and *C. elegans* MA lines have been exposed to different mutagenic forces, although the laboratory conditions of the MA lines should have been reasonably similar. Given the observed differences in the mutational forces acting on *P. pacificus* and *C. elegans*, the identical mutation frequencies might therefore result from a selection-like process. It should be noted that the proofreading activities of the DNA replication machinery have not been subject to direct investigations and therefore it remains unknown whether and to what extent proofreading activities are altered and locally optimized during the course of evolution.

The second major conclusion deriving from the high AT bias observed in MA lines is the missing correlation with the GC content of the *P. pacificus* genome. If *de novo* mutations were the only factor influencing nucleotide composition, one would expect a GC content of only 16% for the *P. pacificus* genome (Lynch 2007). However, the observed GC content is

42%, is stable over all studied natural isolates of the species (Rödelsperger *et al.* 2014), and is thus far from a pure mutation-based equilibrium. This strong discrepancy suggests the existence of powerful forces acting in the opposite direction. Indeed, the analysis of the AT bias in natural isolates provides strong support for the existence of such counterbalancing forces. First, the mean AT bias in natural isolates is 2.4 and already much lower than what is seen in MA lines. Derived alleles that are close to fixation and therefore have been exposed to selection for a longer time period show no AT bias at all. This observation strongly suggests that AT-driving variants are eliminated from the genome and that this tendency increases with the age of the variant. Furthermore, observing no AT bias in those variants that are closest to fixation suggests that the GC content of the *P. pacificus* genome is in an equilibrium state. This conclusion is also supported by the observations that the sister species *P. exspectatus* has a similar GC content (Rödelsperger *et al.* 2014).

While empirical studies in multicellular organisms are sparse, there is ample evidence that GC-biased gene conversion (gcBGC) represents one of the major counterbalancing forces acting opposite to the AT bias (Lynch 2007). gcBGC is a mechanism to resolve double-strand breaks by using a homologous region (*e.g.*, from the sister chromatid) as a repair template. Mismatches between the sequences at heterozygous sites are nonrandomly resolved in favor of G:C nucleotides, which leads to an increase of genomic GC content (Lynch 2007). Even though natural selection influences survival rates of organisms as a whole, while gcBGC acts through the transmission ratio of individual alleles, both processes would lead to the same outcome of increased frequencies of G:C alleles over time. Given the currently available technology, it is virtually impossible to distinguish between the effects of natural selection and gcBGC (Nagylaki 1983; Walsh 1983; Lynch 2007). Several observations suggest the frequency of gcBGC to be magnitudes higher than mutations, with values of 10^{-5} in multicellular species and up to 10^{-2} in fungi (Marais 2003). Given that the AT bias in *P. pacificus* is not different between coding and noncoding regions, we assume gcBGC to be a likely candidate for the observed phenomenon, although there are no direct observations on gcBGC in *P. pacificus*.

A possible explanation for the increased rate of both AT bias by mutation and GC bias by gcBGC might be in an increased role of oxidative stress in *P. pacificus*. Oxidative stress boosts AT-biased mutations by deamination of cytosine and guanine (Teebor *et al.* 1988). At the same time, it causes double-strand breaks that might possibly be repaired via gcBGC (Letavayová *et al.* 2006; Lynch 2010a). Thus, oxidative stress might induce opposing forces on the genomic GC content that cancel each other out, making them effectively invisible if only fixed natural variants are studied. Potential support for a role of oxidative stress comes from the known ecosystem of *P. pacificus* and related nematodes. Unlike *C. elegans*, *Pristionchus* lives in tight association with

scarab beetles in a necromenic interaction (Herrmann *et al.* 2007, 2010). Nematodes rest on the living beetle in the dauer stage (Weller *et al.* 2010) and resume development on the beetle carcass only to feed on all types of microbes (McGaughran *et al.* 2013). Consistent with this stressful environment the *P. pacificus* genome shows a strong increase in the number of enzymes potentially involved in the detoxification of xenobiotics. For example, *P. pacificus* contains 198 gene predictions encoding for cytochrome P450 enzymes, whereas *C. elegans* has only 59 such gene predictions (Dieterich *et al.* 2008). While the ecosystem and the genome of *P. pacificus* strongly suggest a life under oxidative stress, direct evidence for these environmental circumstances on mutational processes is not yet available and awaits future analysis.

Data availability

All reads were submitted to the NCBI Sequence Read Archive. The reference genome and linkage maps for *P. pacificus* are available at <http://www.pristionchus.org>.

Acknowledgments

We thank R. Neher, F. Zanini, and A. Künstner for many fruitful discussions and for critically reading the manuscript. This study was funded by the Max Planck Society.

Literature Cited

Andersen, E. C., J. P. Gerke, J. A. Shapiro, J. R. Crissman, R. Ghosh *et al.*, 2012 Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat. Genet.* 44: 285–290.

Babbitt, G. A., and C. R. Cotter, 2011 Functional conservation of nucleosome formation selectively biases presumably neutral molecular variation in yeast genomes. *Genome Biol. Evol.* 3: 15–22.

Baer, C. F., F. Shaw, C. Steding, M. Baumgartner, A. Hawkins *et al.*, 2005 Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. *Proc. Natl. Acad. Sci. USA* 102: 5785–5790.

Borchert, N., C. Dieterich, K. Krug, W. Schütz, S. Jung *et al.*, 2010 Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res.* 20: 837–846.

Denver, D. R., P. C. Dolan, L. J. Wilhelm, W. Sung, J. I. Lucas-Lledó *et al.*, 2009 A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc. Natl. Acad. Sci. USA* 106: 16310–16314.

Denver, D. R., L. J. Wilhelm, D. K. Howe, K. Gafner, P. C. Dolan *et al.*, 2012 Variation in base-substitution mutation in experimental and natural lineages of *Caenorhabditis* nematodes. *Genome Biol. Evol.* 4: 513–522.

Dieterich, C., S. W. Clifton, L. N. Schuster, A. Chinwalla, K. Delehaunty *et al.*, 2008 The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat. Genet.* 40: 1193–1198.

Drummond, A., and A. Rambaut, 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut, 2012 Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29: 1969–1973.

Gernhard, T., 2008 The conditioned reconstructed process. *J. Theor. Biol.* 253: 769–778.

Herrmann, M., W. E. Mayer, R. L. Hong, S. Kienle, R. Minasaki *et al.*, 2007 The nematode *Pristionchus pacificus* (Nematoda: Diplogastridae) is associated with the oriental beetle *Exomala orientalis* (Coleoptera: Scarabaeidae) in Japan. *Zoolog. Sci.* 24: 883–889.

Herrmann, M., S. Kienle, J. Rochat, W. E. Mayer, and R. J. Sommer, 2010 Haplotype diversity of the nematode *Pristionchus pacificus* on Réunion in the Indian Ocean suggests multiple independent invasions. *Biol. J. Linn. Soc. Lond.* 100: 170–179.

Howe, D. K., C. F. Baer, and D. R. Denver, 2010 High rate of large deletions in *Caenorhabditis briggsae* mitochondrial genome mutation processes. *Genome Biol. Evol.* 2: 29–38.

Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis *et al.*, 2008 NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36: W5–W9.

Keightley, P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar *et al.*, 2009 Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* 19: 1195–1201.

Letavayová, L., E. Marková, K. Hermanská, V. Vlčková, D. Vlasáková *et al.*, 2006 Relative contribution of homologous recombination and non-homologous end-joining to DNA double-strand break repair after oxidative stress in *Saccharomyces cerevisiae*. *DNA Repair* 5: 602–610.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Lu, P., X. Han, J. Qi, J. Yang, A. J. Wijeratne *et al.*, 2011 Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by re-sequencing *Landsberg erecta* and all four products of a single meiosis. *Genome Res.* 22: 508–518.

Lunter, G., and M. Goodson, 2011 Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 21: 936–939.

Lynch, M., 2007 *The Origins of Genome Architecture*, Ed. 1. Sinauer Associates, Sunderland, MA.

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

Lynch, M., 2010b Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. USA* 107: 961–968.

Lynch, M., W. Sung, K. Morris, N. Coffey, C. R. Landry *et al.*, 2008 A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. USA* 105: 9272–9277.

Marais, G., 2003 Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* 19: 330–338.

McGaughran, A., K. Morgan, and R. J. Sommer, 2013 Unraveling the evolutionary history of the nematode *Pristionchus pacificus*: from lineage diversification to island colonization. *Ecol. Evol.* 3: 667–675.

Molnar, R. I., G. Bartelmes, I. Dinkelacker, H. Witte, and R. J. Sommer, 2011 Mutation rates and intraspecific divergence of the mitochondrial genome of *Pristionchus pacificus*. *Mol. Biol. Evol.* 28: 2317–2326.

Morgan, K., A. McGaughran, L. Villate, M. Herrmann, H. Witte *et al.*, 2012 Multi locus analysis of *Pristionchus pacificus* on La Réunion Island reveals an evolutionary history shaped by multiple introductions, constrained dispersal events and rare out-crossing. *Mol. Ecol.* 21: 250–266.

Mukai, T., 1964 The genetic structure of natural populations of *Drosophila melanogaster*. *Genetics* 50: 1–19.

Nagylyaki, T., 1983 Evolution of a large population under gene conversion. *Proc. Natl. Acad. Sci. USA* 80: 5941–5945.

Ossowski, S., K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark *et al.*, 2010 The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327: 92–94.

- Rödelsperger, C., R. A. Neher, A. M. Weller, G. Eberhardt, H. Witte *et al.*, 2014 Characterization of genetic diversity in the nematode *Pristionchus pacificus* from population-scale resequencing data. *Genetics* 196: 1153–1166.
- Seyfert, A. L., M. E. A. Cristescu, L. Frisse, S. Schaack, W. K. Thomas *et al.*, 2008 The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. *Genetics* 178: 2113–2121.
- Sommer, R. J., 2009 The future of evo-devo: model systems and evolutionary theory. *Nat. Rev. Genet.* 10: 416–422.
- Sung, W., M. S. Ackerman, S. F. Miller, T. G. Doak, and M. Lynch, 2012 Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci. USA* 109: 18488–18492.
- Teebor, G. W., R. J. Boorstein, and J. Cadet, 1988 The repairability of oxidative free radical mediated damage to DNA: a review. *Int. J. Radiat. Biol.* 54: 131–150.
- Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts *et al.*, 2007 Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35: W71–W74.
- Walsh, J. B., 1983 Role of biased gene conversion in one-locus neutral theory and genome evolution. *Genetics* 105: 461–468.
- Weller, A. M., W. E. Mayer, R. Rae, and R. J. Sommer, 2010 Quantitative assessment of the nematode fauna present on *Geotrupes* dung beetles reveals species-rich communities with a heterogeneous distribution. *J. Parasitol.* 96: 525–531.
- Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13: 555–556.
- Yang, Z., and J. P. Bielawski, 2000 Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15: 496–503.

Communicating editor: D. Begun