Opposing forces of A/T-biased mutations and G/C-biased gene conversions shape the genome of the nematode *Pristionchus pacificus*

Andreas M. Weller¹, Christian Röiedelsperger¹, Gabi Eberhardt¹, Ruxandra I. Molnar¹, and Ralf J. Sommer¹

¹Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

(Dated: January 9, 2014)

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 potential mechanism acting against AT biased mutations. This study provides insight into genome evolution by combining MAL studies with natural variation.

I. INTRODUCTION

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 (Dенвер et al., 2009, 2012) and *C. briggsae* (Baer et al., 2005; Howe et al., 2010), Drosophila melanogaster (Keightley et al., 2009), Daphnia pulex (Seffert et al. 2008), and Arabidopsis thaliana (Ossowski et al., 2010). These studies showed that the base substitution mutation rate (μ), 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 towards 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 will be 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 non-adaptive 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 micro-evolutionary 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 are available (McGaughran et al., 2013). MA lines have been generated in *P. pacificus* (Molnar et al., 2011) and more than 600 wild isolates covering its cosmopolitan distribution are available in the laboratory. Four major clades, called A to 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 re-sequenced, which revealed strong divergence resulting in more than 6 million single nucleotide polymorphisms (SNP).
(Rödelsperger et al., 2014, accompanying paper). 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, accompanying paper). These studies provide a genomic and phylogenetic framework for evolutionary comparisons of mutational patterns.

Here, we re-sequenced 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 \( \mu_{bs} \) for P. pacificus as baseline for reconstruction of the evolutionary history of the species. We used this \( \mu_{bs} \) 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 non-synonymous 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.

II. RESULTS

A. 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 16x. 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.

B. De novo mutations and mutation frequencies

In total, we identified 802 de novo mutations in the nuclear genome of these 22 MA lines (Fig.1A). Mutations appear to be equally distributed over all chromosomes (Fig.2) indicating no evidence for mutation hot spots in P. pacificus. We calculated the rate of base-substitution \( \mu_{bs} \) 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 \( \mu_{bs} \) is \( 2 \times 10^{-9} \) with a non-significant (\( \chi^2 \) test, \( P = 0.27 \)) variation among the 22 MA lines from \( 1.4 - 2.6 \times 10^{-9} \). By analyzing a sequence of 5 MB in 47 representative wild isolates, we used the identified \( \mu_{bs} \) 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 \( 4.3 \times 10^6 \) generations (Fig.1B).

C. Influence of coding state on mutation rate

Next, we evaluated the effect of mutations on gene function (Fig.1A and Table 1). Of the 802 mutations, 127 were found in exons, including 9 premature stop codons, 88 non-synonymous 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 three genes have a C. elegans 1:1 ortholog with unknown function.
The dN/dS rate $\omega$ 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 $\omega$ 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 a $\omega$ 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.

D. 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 natural variants (Fishers exact test, $P < 0.0001$). Next, we analyzed $p_{bs}$ separately for all six possible base substitution types. We found a strong bias towards G:C to T:A transversions and G:C to A:T transitions (Fig.1A) On average, we observed 3.9 mutations towards A/T for every mutation towards G/C. We calculated the AT bias by weighting for GC content in the genome following standard procedures (Lynch, 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).

E. 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 (Fig.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 (Fig.3A). The AT bias does not differ between coding and noncoding regions (Fig.3A). Together, these results suggest that a strong force exists counterbalancing the mutational AT bias.
III. 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 $\mu_{os}$ of $0.8 - 2.1 \times 10^{-9}$ reported for several Caenorhabditis species (DENVER et al., 2012). These findings suggest a stable mutation rate in nematodes over an evolutionary separation time of more than 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 *Drosophila 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 (Rødelperger et al. 2014, accompanying paper) 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). 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, unpublished observation). In addition, the effects of mutations in exons were randomly distributed between stop codons, non-synonymous and synonymous mutations. A formal analysis of dN/dS further showed that there is no evidence for selection (gcBGC) in natural isolates of the species (D. melanogaster) 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 countervailing 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 suggest 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 (Rødelperger et al. 2014, accompanying paper).

While empirical studies in multicellular organisms are sparse, there is ample evidence that GC biased gene conversion (gcBGC) represents one of the major countervailing 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 non-randomly 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 transmis-
sion 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 (Lynch, 2007; Nagylaki, 1983; Walsh, 1983). 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 which 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., 2010, 2007). Nematodes rest on the living beetle in the dauer stage (Weller et al., 2010) and only resume development on the beetle carcass to feed on all type 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 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.

IV. MATERIALS AND METHODS

V. 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 was performed using reagents and the standard protocol of the Illumina TruSeq v2 sample preparation kit as described in (Rödelsperger et al., 2014, accompanying paper). Samples were selected for an insert size of 300-400bp by excision from an agarose gel. Library concentration was finally determined on a Bioanalyzer DNA 1000 chip and normalized to 10nM before pooling into 12-plexes. The libraries were sequenced in-house on an Illumina HiSeq 2000 sequencer.

VI. 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 4x and 34x. Initial SNP calling was performed using samtools and bcftools (version 0.1.17) (Li et al., 2009). In order 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.

VII. MUTATION RATE AND SPECTRUM

The mutation rate was calculated as previously described by (Dennver et al., 2009) as $\mu_{bs} = m/(LnT)$ where $\mu_{bs}$ 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 X2 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:

\[ \Delta T \text{bias} = \frac{m(\rightarrow A/T)}{\Delta GC} \times \frac{m(\rightarrow G/C)}{1 - \Delta GC}^{-1} \]

Where \( m(\rightarrow G/C) \) stands for the fraction of mutations that change an A/T nucleotide into a G/C nucleotide and \( \Delta GC \) 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.

VIII. TMRCA 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 dataset of 5MB. We inferred the topology of the trees under the birth-death speciation model (GERNHAIR, 2008), using the HKY and GTR 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 above 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.

IX. ANALYSIS OF NATURAL VARIANTS

For the comparison of AT bias between MA lines and natural strains, we used a dataset of 104 re-sequenced strains (Rödelsperger et al. 2014, accompanying paper). As described in (Rödelsperger et al. 2014, accompanying paper), 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 dN/dS ratio \( \omega \) in de novo mutations and natural variants separately. \( \omega \) 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 \( \omega \). Codeml was run assuming the F3x4 codon frequency model and a uniform \( \omega \) across branches.

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

References


DRUMMOND, A., and A. RAMBAUT, 2007 BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7: 214–.


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...


J oh n so n , M., I. Z a re t s k a y a , Y. R a y t s e l i s , Y. M e r e z h u k , S. McGinnis, et al., 2008 NCBI BLAST: a better web interface. Nucleic Acids Research 36: W5–W9.


L e t a v a y o v á , L., E. M ar k o v á , K. H e r m a n s k á , V. V l ě c k o vá , D. V l ě s á k o vá , 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.


L u , P., X. H an , J. Q i , J. Y a ng , A. J. W j e r a t n e , 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 Research 22: 508–518.


M c g a u g h e r an , A., K. M o r g a n , and R. J. S o m m e r , 2013 Unraveling the evolutionary history of the nematode Pristionchus pacificus: from lineage diversification to island colonization. Ecol Evol 3: 667–675.

M o l n a r , R. I., G. B a r t e l m e s , I. D i n k e l a c k e r , H. W it t e , and R. J. S o m m e r , 2011 Mutation Rates and Intraspecific Divergence of the Mitochondrial Genome of Pristionchus pacificus. Molecular Biology and Evolution 28: 2317–2326.

M o r g a n , K., A. M c G a u g h e r a n , L. V i l l a t e , M. H e r m a n n , H. W it t e , 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. Molecular Ecology 21: 250–266.


O s s o w s k i , S., K. S c h n e e b e r g e r , J. I. L u c a s -L l e d ó , N. W a r t h m a n n , R. M. C l a r k , et al., 2010 The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis thaliana. Science 327: 92–94.

S e y f e r t , A. L., M. E. A. C r i s t e s c u , L. F r i s s e , S. S c h a a c k , W. K. T h o m á s , et al., 2008 The Rate and Spectrum of Microsatellite Mutation in Caenorhabditis elegans and Daphnia pulex. Genetics 178: 2113–2121.

S o m m e r , R. J., 2009 The future of evo-devo: model systems and evolutionary theory. Nat Rev Genet 10: 416–422.

S u n g , W., M. S. A c k e r m a n , S. F. M i l l e r , T. G. D o a k , and M. L y n c h , 2012 Drift-barrier hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences 109: 18488–18492.

T e e b o r , G. W., R. J. B o o r s t e i n , and J. C a d e t , 1988 The Repairability of Oxidative Free Radical Mediated Damage to DNA: A Review. Int J Radiat Biol 54: 131–150.

U n t e r g a s s e r , A. H., N. C o i v e n , X. R a o , T. B i s s e l l e r , R. G e u r t s , et al., 2007 Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research 35: W71–W74.


W e l l e r , A. M., W. E. M a y e r , R. R a e , and R. J. S o m m e r , 2010 Quantitative Assessment of the Nematode Fauna Present on Geotrupes Dung Beetles Reveals Species-Rich Communities with a Heterogeneous Distribution. Journal of Parasitology 96: 525–531.


Y a n g , Z., and J. P. B i e l a w s k i , 2000 Statistical methods for detecting molecular adaptation. Trends in Ecology & Evolution 15: 496–503.
### Table I Properties of de novo mutations in MA lines

<table>
<thead>
<tr>
<th>Coding state</th>
<th>( p_{bs} )</th>
<th>A:T to G:C</th>
<th>G:C to A:T</th>
<th>A:T to C:G</th>
<th>G:C to T:A</th>
<th>G:C to C:G</th>
<th>A:T to T:A</th>
<th>N</th>
<th>Ts:Tv</th>
<th>AT bias (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon</td>
<td>2 ( \times 10^{-9} )</td>
<td>12</td>
<td>49</td>
<td>3</td>
<td>38</td>
<td>9</td>
<td>16</td>
<td>127</td>
<td>0.92</td>
<td>5.3</td>
</tr>
<tr>
<td>Intron</td>
<td>1.9 ( \times 10^{-9} )</td>
<td>33</td>
<td>108</td>
<td>17</td>
<td>79</td>
<td>22</td>
<td>73</td>
<td>332</td>
<td>0.74</td>
<td>5.9</td>
</tr>
<tr>
<td>Intergenic</td>
<td>2.1 ( \times 10^{-9} )</td>
<td>39</td>
<td>118</td>
<td>17</td>
<td>89</td>
<td>31</td>
<td>49</td>
<td>343</td>
<td>0.84</td>
<td>4.8</td>
</tr>
</tbody>
</table>

\(^a\)Total number in all 22 MA lines.

\(^b\)Weighted by nucleotide composition at the coding state considered

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

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

\(^a\)Fraction of all mutations

\(^b\)Weighted by genomic nucleotide composition