Functional Conservation of Nucleosome Formation Selectively Biases Presumably Neutral Molecular Variation in Yeast Genomes

Gregory A. Babbitt*, and C. R. Cotter

School of Biological and Medical Sciences, Rochester Institute of Technology

*Corresponding author: E-mail: gabsbi@rit.edu.

Accepted: 29 November 2010

Abstract

One prominent pattern of mutational frequency, long appreciated in comparative genomics, is the bias of purine/pyrimidine conserving substitutions (transitions) over purine/pyrimidine altering substitutions (transversions). Traditionally, this transitional bias has been thought to be driven by the underlying rates of DNA mutation and/or repair. However, recent sequencing studies of mutation accumulation lines in model organisms demonstrate that substitutions generally do not accumulate at rates that would indicate a transitional bias. These observations have called into question a very basic assumption of molecular evolution; that naturally occurring patterns of molecular variation in noncoding regions accurately reflect the underlying processes of randomly accumulating neutral mutation in nuclear genomes. Here, in *Saccharomyces* yeasts, we report a very strong inverse association ($r = -0.951$, $P < 0.004$) between the genome-wide frequency of substitutions and their average energetic effect on nucleosome formation, as predicted by a structurally based energy model of DNA deformation around the nucleosome core. We find that transitions occurring at sites positioned nearest the nucleosome surface, which are believed to function most importantly in nucleosome formation, alter the deformation energy of DNA to the nucleosome core by only a fraction of the energy changes typical of most transversions. When we examined the same substitutions set against random background sequences as well as an existing study reporting substitutions arising in mutation accumulation lines of *Saccharomyces cerevisiae*, we failed to find a similar relationship. These results support the idea that natural selection acting to functionally conserve chromatin organization may contribute significantly to genome-wide transitional bias, even in noncoding regions. Because nucleosome core structure is highly conserved across eukaryotes, our observations may also help to further explain locally elevated transition bias at CpG islands, which are known to destabilize nucleosomes at vertebrate promoters.

Key words: nucleosome, chromatin, evolution, selection, gene regulation, transition bias.

Mutational Accumulation Does Not Account for Comparative Substitutional Bias

Transition/transversion (Ts:Tv) biases observed in genome comparisons are potentially the result of mutational and/or selective evolutionary processes that influence the alteration of the two fundamental types of nitrogenous bases (i.e., purines and pyrimidines) that comprise DNA. It has long been observed that (purine/pyrimidine) conserving substitutions (i.e., transitions) are generally more common in comparative studies of DNA sequence than (purine/pyrimidine) altering substitutions (i.e., transversions) despite the fact that potential transversions theoretically outnumber potential transitions by two to one.

This unexpected transitional bias observed in comparative genomic studies is traditionally attributed to mutational bias, whereby transitions are either more likely to occur or less likely to be repaired than transversions. Spontaneous base substitution patterns were first explained by tendencies of potential tautomeric shifts involving atypical complementary base pairing (Topal and Fresco 1976), however, this idea now appears largely discredited by empirical data (Vonborstel 1994). Other mutational biases driven by

spontaneous or oxidative deamination of cytosine or 5-methylcytosine and UV light induced mutagenesis at dipyrimidine sites (CC or CT) seem to more realistically explain the predominance of one type of transition (C:G > T:A). However, certain transversions (C:G > A:T) also commonly result from oxidative damage resulting in guanine conversion to 8-oxo-guanine (Friedberg et al. 2006). The reduction of all substitution events by 89% in *Escherichia coli* grown in an anaerobic environment is a powerful testament to the role of oxidative damage in driving these mutational biases (Sakai et al. 2006).

Mutational biases can also be caused by functional constraints on the organization of the genome as well as by patterns of its damage or repair. Molecular evolutionists have long observed that mutational biases are influenced by the structure of the genetic code because the possible transitions at 4-fold degenerate (silent) sites are about twice as common as transversions (Yang 2006). This idea is well supported by observations that transitional bias is significantly larger in vertebrate mitochondrial genes that are highly evolutionarily conserved (Kumar 1996), that transitional bias is somewhat reduced when comparing total coding with noncoding regions in *Drosophila* (Moriyama and Powell 1996), and that ribosomal RNA genes (i.e., srDNA) and pseudogenes both lack a consistent transitional bias (Vawter and Brown 1993; Keller et al. 2007).

Recent advances in high-throughput sequencing have led to the direct analysis of genome-wide mutation accumulation spectra in several model organisms including *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Drosophila melanogaster*, and *Caenorhabditis elegans* (Lynch et al. 2008; Denver et al. 2009; Keightley et al. 2009; Ossowski et al. 2009). These studies are consistent with a possible role of oxidative damage in driving mutational biases, demonstrating a predominance of both C:G > T:A transition (all studies) and C:G > A:T transversion (all but *Arabidopsis*). However, surprisingly, although Ts:Tv ratios are variable in yeast, fly, and worm (about 1.2:1, 1:1, 1:2, respectively), only *Arabidopsis* is highly transitionally biased (2.4:1). Even in this one case, the transitional bias is caused by the predominance of only one type of transition, the deamination of cytosine. These major inconsistencies observed between the frequency of mutation accumulation and existing natural patterns of molecular variation are leading some to challenge one of the most primary assumptions of molecular evolution; that sequence comparisons of natural molecular variation can accurately represent underlying neutrally evolving mutational processes (Lynch et al. 2008; Denver et al. 2009). If true, this is potentially problematic as this central tenet of molecular evolution also forms the theoretical basis underlying current methods of modern phylogenetic reconstruction.

Does Selection Maintaining Chromatin Organization Drive Substitutional Bias?

If the biases observed during mutation accumulation are not responsible for the patterns typically observed in sequence polymorphism, such as the transitional bias usually observed even in noncoding DNA, then what is? Denver et al. (2009) suggest that their findings may indicate some unknown level of natural selection operating to purge transversions from the *C. elegans* genome. Additionally, Rosenberg et al. (2003) have demonstrated that transitional bias is spatially uniform across mammalian genomes after controlling for local differences in CpG hypermutability. This would suggest that if selection was responsible for transitional bias (outside of CpG), it must also act uniformly across most of the genome. Lynch et al. (2008) also note evidence of a possible mechanism in natural populations countering a strong mutational pressure to increase AT content (i.e., GC-biased gene conversion). Because increased AT content is a strong driver of nucleosome positioning, facilitating nucleosome exclusion (Field et al. 2008; Kaplan et al. 2009), it may also be that purifying selection to preserve the packaging of DNA into chromatin could be acting uniformly across the nuclear genome. Chromatin also plays an active role in gene regulation as well as a general role in genome packaging and therefore its molecular evolution may play an equally important role in gene regulatory evolution (Babbitt 2010; Babbitt et al. 2010). This is further supported by the observation that primary chromatin structure and organization (i.e., nucleosome positioning) has recently been demonstrated to be largely sequence dependent (Ioshikhes et al. 2006; Tolstorukov et al. 2007, 2008; Kaplan et al. 2009), subject to natural selection (Babbitt and Kim 2008; Babbitt et al. 2010), and to affect the overall rates of base substitution (Warnecke et al. 2008).

In this study, we use a structurally based model of the deformation energy (DE) required by given DNA sequences to deform to the molecular structure of the nucleosome core (Tolstorukov et al. 2007, 2008) to ascertain whether the pattern of base substitution between the genomes of three very closely related species of *Saccharomyces* yeasts can be explained by natural selection acting to functionally conserve a genome-wide signature of primary chromatin organization (i.e., nucleosome formation). We also compare this with results obtained from both neutral simulation as well as mutation accumulation lines in yeast. Because the energy model is derived from explicitly mechanistic (i.e., energetic) first principles, it is sensitive to many potential features of DNA sequence that allow or inhibit local nucleosome formation. These properties include local AT content associated with translational positioning of the nucleosome (Field et al. 2008; Kaplan et al. 2009) as well as well-known 10–11 bp dinucleotide frequencies favored by the rotational

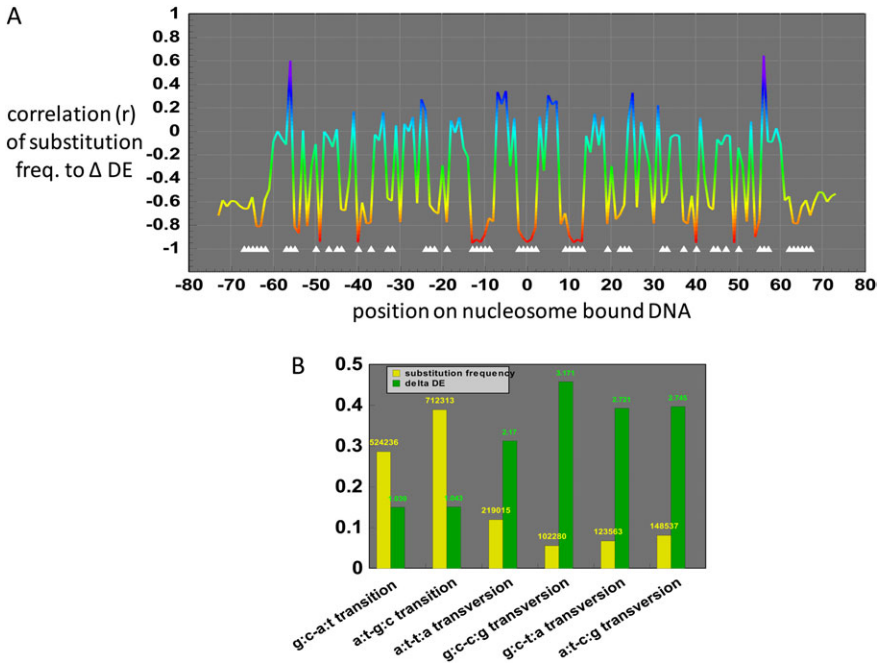

Fig. 1.—The association between the frequency of each substitution type and their average energetic impact on DNA deformation to the molecular structure of the nucleosome core (ΔDE). (*A*) Strongest inverse association (shown in orange and red) is observed when mutations are present at the most deformable sites in the energy model nearest the nucleosome surface (white triangles indicate low energy positions in human α satellite sequence cocrystallized with histone octamer in the best resolved nucleosome core particle structure, i.e., fig. 7*A* in Tolstorukov et al. 2007). (*B*) Also shown are the inverse correlation between substitution frequency and ΔDE at the dyad or center position on the nucleosome (position = 0, $r = -0.951$, $P < 0.004$; strand specificity is ignored in the figure).

phasing of DNA on the nucleosome surface (Ioshikhes et al. 2006; Segal et al. 2006; Tolstorukov et al. 2007), and the A/T asymmetry across the dyad or center of sequence bound to the nucleosome core.

Evidence for Selection Conserving Chromatin Organization in Driving Mutational Biases

Across the entire genome, we find that when substitution events occur at the dyad position on the energy model, the average change in deformation energy (ΔDE) imposed by the 12 possible types of substitution is very strongly inversely associated with their respective frequency throughout the yeast genomes ($r = -0.951$, $P < 0.004$). Upon further investigation of this association at all possible 147 sites on the threading template of the DE model, while utilizing a substantial subset (=10%) of the genome, we find that this strong inverse association between a given substitution type's energetic impact upon nucleosome formation (i.e., ΔDE) and it genomic frequency occurs whenever substitutions fall nearest to crucial low DE sites in the model; sites typically nearest the surface of the nucleosome core (with roughly 10–11 bp periodicity; fig. 1*A*). It was previously demonstrated that the long axis deformability in dinucleotides (i.e., roll and slide) near the surface of the nucleosome core is critical to nucleosome formation/exclusion (Tolstorukov et al. 2007). This negative association between ΔDE and mutation frequency also persists strongly in noncoding regions of the yeast genome ($r = -0.940$, $P < 0.001$ at dyad) indicating that this pattern is not driven by any mutational bias imposed by the genetic code. At model sites demonstrating this strong inverse association, we find that purine–pyrimidine transversions have, on average, roughly twice the energetic impact on nucleosome formation than do purine–purine or pyrimidine–pyrimidine transitions (fig. 1*B*). If these observed patterns were the product of purifying selection acting to maintain nucleosome formation associated with local chromatin organization, we would also predict that it should be absent in mutation accumulation lines. We mined the data presented in table S1 of Lynch et al. (2008) to obtain a list of transitions and transversions accumulating under conditions of neutral genetic drift. We successfully located 147 bp background sequences from the Saccharomyces Genome Database for all but one of these mutations and computed their energetic impacts on nucleosome formation (i.e., ΔDE; table 1). We found no significant correlations between ΔDE and mutational frequency, supporting the idea that mutations in these lines may be accumulating without any regard to their energetic effects on nucleosome formation. We also found no significant difference in ΔDE between transitions and transversions in the

Table 1
Energetic Impacts of Substitution Events under Conditions of Neutral Mutation Accumulation (from Lynch et al. 2008)

Chromosome	Position	Substitution Class	Substitution Type	ΔDE
5	351995	Ts	AG	2.87
11	605009	Ts	CT	0.71
8	262177	Ts	CT	0
10	33149	Ts	GA	0.95
16	834238	Ts	GA	1.07
16	331354	Ts	GA	4.1
15	679548	Ts	GA	0.38
11	239813	Ts	GA	4.56
7	67430	Ts	GA	3.62
3	54214	Ts	TC	0.71
15	541599	Ts	TC	0.95
12	277642	Ts	TC	3.4
7	804473	Tv	AC	1.26
7	625107	Tv	AC	1.26
9	154205	Tv	AC	2.01
8	231499	Tv	AT	2.91
15	986649	Tv	CA	1.71
13	503024	Tv	CA	0.95
13	913509	Tv	CA	0.86
13	824994	Tv	CA	0.41
7	561788	Tv	GC	2.82
2	125366	Tv	CG	0.31
12	617519	Tv	GC	1.03
13	213440	Tv	GC	0.83
2	536163	Tv	GC	1.5
16	804029	Tv	GT	4.05
2	696533	Tv	GT	0.67
4	1148647	Tv	GT	1.26
9	380265	Tv	GT	5.77
12	716670	Tv	GT	2.57
14	688148	Tv	INVERSION	0
4	117354	Tv	DELETE G	10.05

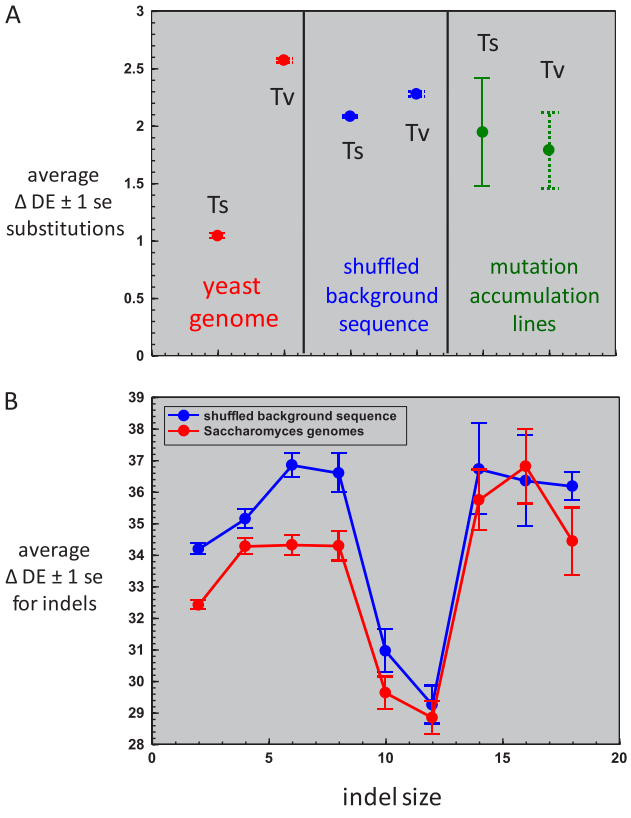

Fig. 2.—The relationship between various classes of mutation and their average energetic impact on DNA deformation to the molecular structure of the nucleosome core (ΔDE). (*A*) Comparison of the average ΔDE for transitions (Ts) and transversions (Tv) observed in the *Saccharomyces* genomes (left), the same substitutions placed in randomly reshuffled sequence backgrounds (center), and substitutions observed in mutation accumulation lines (right; data from Lynch et al. 2008). (*B*) The relationship between size of insertion and deletion events and their average energetic impact on the ability of DNA to deform to the molecular structure of the nucleosome core (ΔDE) in both *Saccharomyces* genome and randomly reshuffled backgrounds. Energetic impact on indels on DE is lowest at 10–12 bp size, which tends to maintain the existing rotational phasing of DNA on the nucleosome core.

mutation accumulation study ($t = 0.271$, $P = 0.782$), thus contrasting the effect observed in the yeast genome (fig. 2*A*). This result appears to also compare favorably with results obtained by neutral simulations where observed (transitions/transversions) in the yeast genomes have significantly (lower/higher) energetic impacts on nucleosome formation than when placed in the randomly reshuffled settings of existing local DNA (Ts: $t = -10.334$, $P < 0.001$ and Tv: $t = -8.192$, $P < 0.001$; fig. 2*A*). This result strongly indicates a role of selection in driving different energetic impacts on nucleosome formation for the two substitutional classes. The lack of any significant difference between transitions and transversions in the mutation accumulation lines may also simply reflect poor statistical power in this relatively small sample size ($n = 12$ Ts and $n = 18$ Tv). To address this potential problem with our interpretation, we bootstrapped our original *t*-test on the comparative genomic data using 10,000 replications of identically sized samples (with replacement). At a sample size matching that of Lynch et al. (2008), we found that only 34% of our bootstrap *t*-tests indicate a significant difference in ΔDE when comparing transition with transversion events (table 2). Given the parameters of our genomic data, we estimate that a sample size of about 80 neutrally accumulated substitutions would provide adequate statistical power (i.e., over 95% rejection of null). We performed a similar bootstrap analysis regarding the relatively small mean difference in ΔDE observed between transition and transversion events in the Lynch et al. (2008) study ($\Delta DE_{Tv} - \Delta DE_{Ts} = 0.16$). Here, we found that less than 1% of our bootstrap *t*-tests would be predicted to demonstrate a mean difference this small (table 2). Taken together, these results of our bootstrap *t*-test indicate that although the apparent lack of significant difference we observed in the data set of Lynch et al. (2008) may

Table 2
Bootstrap *t*-test of Whole Genome Comparing Mean ΔDE of Transition with Transversion Events Using Small Sample Sizes Similar to Those Obtained Mutation Accumulation Line Studies

Bootstrap Sample Size	Frequency Significant	Frequency Mean Difference < 0.16
20	0.239	0.0212
30	**0.341**	**0.0081**
40	0.524	0.0021
50	0.708	0.0011
60	0.854	0.0007
70	0.935	0.0001
80	0.976	0
90	0.994	0
100	0.999	0

NOTE.—Sample size of Lynch et al. (2008) is in bold underlined text.

be due to the relatively small sample size, the mean difference we observe is probably not. Therefore, the mutation accumulation line investigation of the energetic impacts of these two classes of substitutions on nucleosome formation seems generally supportive of the idea that during yeast evolution, purifying selection may have purged many substitutions that significantly altered chromatin organization, and these purged substitutions would be predicted to be largely dominated by transversion events.

We also analyzed the energetic impacts of insertion/deletion events (i.e., indels) on nucleosome formation. As one would expect due to the greater impact of indels on both DNA sequence as well their greater ability to shift the rotational phasing of DNA on the surface of nucleosome core, we found that indels tend to have much larger average ΔDE than substitutions (*y* axis; fig. 2*B*). The 10–11 bp phasing is also evidenced by the dramatic drop in ΔDE in indels with similar size (i.e., 10–11 bp in length; fig. 2*B*). We also find a signature of potential purifying selection on indels acting to maintain existing local levels of DE. As with substitutions, the average ΔDE for indels observed in the yeast genomes was significantly lower than those observed in neutrally evolving random simulation ($t = -5.038$, $P < 0.001$). This observation supports a previous suggestion that indels may be quite important in the evolution of dynamic chromatin structure of regulatory regions (Babbitt et al. 2010; He et al. 2010).

Concluding Remarks

Overall, our findings indicate a significant role for the functional conservation of chromatin organization in the molecular evolution of eukaryotic genomes. Our results are supportive of genome-wide purifying selection acting to preserve chromatin organization in yeasts by purging many naturally occurring purine–pyrimidine transversions. We believe this selective bias for substitutions that minimally affect ΔDE may help explain current inconsistencies observed between comparative estimates of substitution frequency and rates of mutation accumulation observed in recent high-throughput sequencing studies of model organisms. More specifically, the reduced energetic impact of transitions compared with transversions on sites that are crucial in nucleosome formation may have significantly contributed to the transitional bias observed in comparative studies of yeast genomes. Because the positioning of nucleosome on DNA is usually translationally dynamic (i.e., nucleosomes are able to slide to some degree on most sequences), any purely neutral accumulation of mutation is probably unlikely outside of highly bottlenecked inbred populations because the movement of nucleosomes on DNA sequence will potentially place all sites on the surface of the sliding histone cores. Additionally, the apparent accumulation of transitions that are less energetically disruptive to chromatin than those observed in random simulation as well as the apparent accumulation of substitutions without regard to energetic impact on nucleosome in a well-designed mutation accumulation line, lends further support to this role of purifying selection in maintaining chromatin organization.

It has been recently suggested that chromatin evolution is a primarily neutral process (Tirosh et al. 2010). This conclusion was based upon a lack of empirical correspondence between gene expression divergences in hybridized yeast and divergences in experimentally mapped nucleosome positions. However, recent studies demonstrate a significant role of chromatin structure in explaining expression divergences between duplicated genes (Li et al. 2010) and the evolution of gene regulation (Tsankov et al. 2010) in yeast. Here, we have presented another molecular evolutionary analysis that would also seem to contradict the recent hybridization findings. We suggest that the conclusion of Tirosh et al. (2010) may be premature, especially if chromatin's role in gene expression is not simply a matter of occupancy but is more biophysically complex and dynamic. The propensity of nucleosomes to move translationally (i.e., slide) along DNA dictates that the majority of nucleosome positioning is statistical rather than static. Therefore, occupancy may be expected to be a poor predictor of the functioning of dynamic chromatin around many promoters (Babbitt 2010; He et al. 2010; Sha et al. 2010). Indeed, a fundamental challenge in our work was presented by the fact that any given mutations affect on DE could not be modeled exactly because its position on the nucleosome core could not be predicted exactly, which is why we resort to analyzing the effect of each mutation on all possible sites of the nucleosome model template (as shown in fig. 1*A*). At any rate, further investigation of the role of natural selection in governing chromatin organization will be needed before its importance in gene regulatory evolution can be truly specified. The further development of molecular evolutionary inference tests to infer natural selection acting upon chromatin, in addition to comparative studies conducted both within and between populations will eventually help in attaining this goal.

Lastly, the relationship between transition bias and nucleosome formation that we describe here may have particular relevance to the genomic occurrence of CpG islands, which appear in conjunction with increased transition bias at the evolutionary time of mammalian radiation (Arndt and Hwa 2004). It has been known that CpG islands, regions associated with promoters and transcription start sites of noncoding RNAs and dominated by a high density of unmethylated CG dinucleotides (Illingworth et al. 2008; Gibney and Nolan 2010), also demonstrate unusually high local rates of transition (Rosenberg et al. 2003). More recently, it has been demonstrated that these methylation states of CpG are also associated with the compaction and stabilization of nucleosomes (Singh 2009; Choy et al. 2010), the rotational setting of DNA on the surface of the nucleosome core (Hebert and Crollius 2010), and the interaction of nucleosomes with the sequences of Alu insertions (Salih et al. 2008). Additionally, hypermethylation of CpG islands is also associated with abnormal gene regulation in most cancers (Claes et al. 2010). Taken as a whole, this line of evidence strongly suggests that the evolutionary conservation of molecular dynamics of nucleosomes functioning in normal regulation of mammalian genomes, which depend heavily on CpG-mediated gene regulation, may also result in even higher localized regions of transition bias. In future population genomics studies in humans, it will be interesting to see if genome-wide substitution patterns indicative of potentially altered nucleosome dynamics (e.g., elevated transversion rates) can be directly linked to the incidence of age or environmentally associated disease.

Materials and Methods

We utilized the multiple alignments of *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. bayanus* previously published by Kellis et al. (2003). To ensure accurate homology, we restricted our analysis to only the first three species, then we reconstructed ancestral sequences to each node on the phylogeny using the basml program implemented in PAML (Yang 2006). Strand specificity was inferred from pairwise comparison of each extant species to its ancestral sequence. Flanking sequences to each mutation event (substitution, insertion, and deletion) were recovered for each of the extant species and used to construct a 147 bp sequence used in the DE analysis with the mutation placed at the dyad or centered within a hypothetical nucleosome position. Thus, for a mutation event at site m on a given sequence, flanking sequences upstream and downstream of substitutions were defined ($m-73 \ldots m-1$) and ($m+1 \ldots m+73$), respectively. Flanking sequences upstream of indels (insertions and deletions) were defined ($m-73 \ldots m-1$), whereas downstream flanking sequences were defined either as ($m \ldots m+73+d$) where d = deletion length or ($m \ldots m+73-i$) where i = insertion length. The energy required for the given sequences to deform to the molecular structure of the nucleosome core (i.e., DE) was computed using the computational model implemented by nuScore software (Tolstorukov et al. 2008). DE of base pair steps in a given 147 bp DNA sequence is theoretically "forced" to deform according to a spatial path of nucleosome-bound DNA. The deformation of adjacent base pairings of nucleotides (i.e., dinucleotides) is defined spatially by six possible base pair step orientations (three translational variables: slide, shift, and rise and three rotational variables: roll, tilt, and twist; see Schlick 2010, p. 154). The DNA structure from nucleosome core particle resolved by X-ray crystallography was used as a structural template (Davey et al. 2002). Tolstorukov et al. (2007) demonstrated that sequence deformation to the nucleosome core structure depends largely upon the long axis deformations, roll, and slide. The equation for DE is given below:

$$\mathrm{DE} = \sum_{n=1}^{L}\left(\frac{1}{2}\sum_{i=1}^{6}\sum_{j=1}^{6} f_{ij}(\mathrm{MN})\Delta\theta_i^n\Delta\theta_j^n\right),$$

where $\Delta\theta_i^n = \theta_i^n - \theta_i^0(\mathrm{MN})$ is the imposed deviation of the ith dinucleotide step parameter θ_i^n at the nth step of the nucleosome template from the rest state value $\theta_i^0(\mathrm{MN})$ of the step MN. The $f_{ij}(\mathrm{MN})$ are DNA stiffness constants that depend upon the sequence and L is the number of base pair steps in the nucleosome template (see Olson et al. 1998).

The "energetic impact" of a given mutation event is defined here as the absolute difference in DE or (ΔDE) caused by the mutational change on the sequence of the extant species or

$$\Delta\mathrm{DE} = |\mathrm{DE}_{\mathrm{prior}} - \mathrm{DE}_{\mathrm{posterior}}|$$

ΔDE was computed and averaged for each of the 12 substitution types observed across the three yeast genomes and then correlated to the overall substitution counts to assess the association between the mutational frequency and the energetic impact of various substitution types (result in fig. 1*B*). To assess to what degree the results were model driven, the average ΔDE for each substitution type was also determined through a neutral simulation. The 146 bp sequences flanking every existing mutation (i.e., $-73 \ldots -1, +1 \ldots +73$) were randomized with a simple card shuffling algorithm, thus preserving local AT content while removing any sequence-based contextual information associated with each given mutation. A shuffled background for each mutation was used to compute a new set of ΔDE values for comparison to real genome. The average ΔDE of transitions and transversions in the simulations was compared with those in yeast genomes with *t*-test (result in fig. 2*A*). The average ΔDE of indels in random simulation and yeast genomes was also compared similarly (result in fig. 2*B*).

Most importantly, to assess whether the relationship between substitution frequency and subsequent energetic impacts of substitutions on nucleosome formation was driven by the sites most critical for deformation of sequences to the structure of the nucleosome core (i.e., sites nearest the core surface), the ΔDE of each substitution event observed on chromosome 4 (Chr D), the largest in the yeast genome, was analyzed at all possible 147 sites on the energy model. Here, ΔDE was calculated for a given mutation positioned at each site on the nucleosome structural template (i.e., flanking sequences for substitutions are $(m - 73 - i \ldots m - 1 - i)$ and $(m + 1 - i \ldots m + 73 - i)$ where $i = \{-73, -72, -71 \ldots 71, 72, 73\}$. A correlational curve of the positional energetic impacts of substitutions was produced by correlating the average ΔDE of the 12 substitution types to their genomic frequencies at all 147 sites on the nucleosome structural template of the DE model (results in fig. 1*A*). This analysis was quite computationally expensive, and because the resulting correlational curve was stabilized after analyzing only about 100 genes, we stopped after completion of chr 4, the largest chromosome in the yeast genome.

Acknowledgments

We would like to thank Dr Larry Buckley and two anonymous reviewers for helpful commentary and Dr Michael Tolstorukov for providing nuScore, a software program for the calculation of DNA DE required to conform to the molecular structure of the nucleosome core. G.A.B. thanks Dr Yuseob Kim and Sudhir Kumar for earlier direction in this emerging field of study.

Literature Cited

Arndt PF, Hwa T. 2004. Regional and time-resolved mutation patterns of the human genome. Bioinformatics 20(10):1482–1485.

Babbitt GA. 2010. Relaxed selection against accidental binding of transcription factors with conserved chromatin contexts. Gene doi: 10.1016/j.gene.2010.07.002.

Babbitt GA, Kim Y. 2008. Inferring natural selection on fine-scale chromatin organization in yeast. Mol Biol Evol. 25(8): 1714–1727.

Babbitt GA, Tolstorukov MY, Kim Y. 2010. The molecular evolution of nucleosome positioning through sequence-dependent deformation of the DNA polymer. J Biomol Struct Dyn. 27(6): 765–780.

Choy JS, et al. 2010. DNA methylation increases nucleosome compaction and rigidity. J Am Chem Soc. 132(6):1782–1788.

Claes B, Buysshaert I, Lambrechts D. 2010. Pharmaco-epigenomics: discovering therapeutic approaches and biomarkers for cancer therapy. Heredity 105(1):152–160.

Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 angstrom resolution. J Mol Biol. 319:1097–1113.

Denver DR, et al. 2009. A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. Proc Natl Acad Sci U S A. 106(38):16310–16314.

Field Y, et al. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLOS Comp Biol. 4(11): e1000216.

Friedberg EC, et al. 2006. DNA repair and mutagenesis. Washington (DC): ASM Press.

Gibney ER, Nolan CM. 2010. Epigenetics and gene expression. Heredity 105(1):4–13.

He HH, et al. 2010. Nucleosome dynamics define transcriptional enhancers. Nat Genet. 42(4):343–347.

Hebert C, Crollius HR. 2010. Nucleosome rotational setting is associated with transcriptional regulation in promoters of tissue-specific genes. Genome Biol. 11:R51.

Illingworth R, et al. 2008. A novel CpG island set identifies tissue specific methylation at developmental gene loci. PLoS Biol. 6:e22.

Ioshikhes IP, Albert I, Zanton SJ, Pugh BF. 2006. Nucleosome positions predicted through comparative genomics. Nat Genet. 38: 1210–1215.

Kaplan N, et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458:362–366.

Keightley PD, et al. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. 19:1195–1201.

Keller I, Bensasson D, Nichols RA. 2007. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. PLoS Genet. 3(2):e22.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241–254.

Kumar S. 1996. Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. Genetics 143(1):537–548.

Li J, Yuan Z, Zhang Z. 2010. Revisiting the contribution of cis-elements to expression divergence between duplicated genes: the role of chromatin structure. Mol Biol Evol. 27(7):1461–1466.

Lynch M, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci U S A. 105(27):9272–9277.

Moriyama EN, Powell JR. 1996. Intraspecific nuclear DNA variation in Drosophila. Mol Biol Evol. 13(1):261–277.

Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci U S A. 95:11163–11168.

Ossowski S, et al. 2009. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327: 92–94.

Rosenberg MS, Subramanian S, Kumar S. 2003. Patterns of transitional biases within and among mammalian genomes. Mol Biol Evol. 20(6):988–993.

Sakai A, Nakanishi M, Yoshiyama K, Maki H. 2006. Impact of reactive oxygen species on spontaneous mutagenesis in Escherichia coli. Genes Cells 11(7):767–778.

Salih F, Salih B, Kogan S, Trifonov EN. 2008. Epigenetic nucleosomes: Alu sequences and CG as nucleosome positioning element. J Biomol Struct Dyn. 26(1):9–15.

Schlick T. 2010. Molecular modeling and simulation: an interdisciplinary guide, 2nd ed. New York: Springer. pp. 154.

Segal E, et al. 2006. A genomic code for nucleosome positioning. Nature 442:772–778.

Sha K, et al. 2010. Distributed probing of chromatin structure *in vivo* reveals pervasive chromatin accessibility for expressed and non-expressed genes during tissue differentiation in *C. elegans*. BMC Genomics 11:465.

Singh H. 2009. Teeing up transcription on CpG islands. Cell 138(1):14–16.
Tirosh I, Nadejda S, Barkai N. 2010. Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences. Mol Syst Biol. 6:365.
Tolstorukov MY, Choudhary V, Olson WK, Zhurkin VB, Park PJ. 2008. nuScore: a web-interface for nucleosome positioning predictions. Bioinformatics 24(12):1456–1458.
Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB. 2007. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. J Mol Biol. 371:725–738.
Topal MD, Fresco JR. 1976. Complementary base pairing and the origin of substitution mutations. Nature 263:285–289.
Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. 2010. The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol. 8(7):e1000414.
Vawter L, Brown WM. 1993. Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics 134(2):597–608.
Vonborstel RC. 1994. Origins of spontaneous base substitutions. Mutat Res. 307(1):131–140.
Warnecke T, Batada NN, Hurst LD. 2008. The impact of the nucleosome code on protein coding sequence evolution in yeast. PLoS Genet. 4(11):e1000250.
Yang Z. 2006. Computational molecular evolution. New York: Oxford University Press.

Associate editor: Bill Martin