Rapid Evolution of *cis*-Regulatory Sequences via Local Point Mutations

Jonathon R. Stone[1] *and Gregory A. Wray*[2]

Department of Ecology and Evolution, State University of New York at Stony Brook

Although the evolution of protein-coding sequences within genomes is well understood, the same cannot be said of the *cis*-regulatory regions that control transcription. Yet, changes in gene expression are likely to constitute an important component of phenotypic evolution. We simulated the evolution of new transcription factor binding sites via local point mutations. The results indicate that new binding sites appear and become fixed within populations on microevolutionary timescales under an assumption of neutral evolution. Even combinations of two new binding sites evolve very quickly. We predict that local point mutations continually generate considerable genetic variation that is capable of altering gene expression.

Introduction

During the past decade, interest in elucidating the genetic basis of phenotypic evolution has been intense (Raff 1996; Davidson 2001). Attention has been focused on developmental regulatory genes, particularly those encoding transcription factors. Alterations in the expression profiles of these genes or in the binding properties of proteins encoded by them can yield particularly interesting phenotypes involving duplicated structures, homeotic transformations, or novel morphologies (Raff and Kauffman 1983; Gerhart and Kirschner 1997). Many evolutionary changes in the expression of these genes have been identified (e.g., Patel et al. 1989; Grenier et al. 1997), and extensive but indirect evidence for evolutionary changes in their downstream targets has been gathered, manifested in the form of entirely unrelated expression domains within and among species (e.g., Lowe and Wray 1997; Keys et al. 1999).

The genetic basis for many of these evolutionary changes probably resides within the *cis*-regulatory regions that control transcription (Arnone and Davidson 1997; Davidson 2001). (We use the term promoter to denote the entire *cis*-regulatory apparatus). Yet, promoter evolution remains poorly understood for several reasons. First, there is a limited amount of information available concerning promoter sequence variation and its consequences within and among species. The general organization of some promoter sequences has been maintained for 10^7 years (e.g., Damjanovski et al. 1998; Ludwig, Patel, and Kreitman 1998), although functional differences can evolve over comparable or even shorter time intervals (e.g., Franks et al. 1988; Ross, Fong, and Cavener 1994; Wang et al. 1999). Sequence comparisons indicate that single transcription factor binding sites can appear and disappear among relatively closely related species (e.g., Gonzalez et al. 1995; Damjanovski et al. 1998) and even within populations (e.g., Tournamille et al. 1995; Segal, Barnett, and Crawford 1999). However, no obvious relation between degree of sequence divergence and change in gene expression has emerged (e.g., Maduro and Pilgrim 1996). Second, a comprehensive understanding of promoter evolution is confounded by the complex encoding of regulatory information within genomes (Yuh, Bolouri, and Davidson 1998; Davidson 2001). Functional consequences of particular mutations within coding regions, such as the introduction of nonsynonymous substitutions, stop codons, and frameshifts (Gillespie 1991; Li 1997), can often be predicted from sequence data. In contrast, sequence data alone provide little direct information concerning conservation or change of promoter function. Instead, experimental tests in the form of expression assays are required. Finally, there exists no conceptual framework for understanding promoter evolution and guiding empirical studies.

As a first step toward achieving such a framework, we considered the following question: what time period would be required for new transcription factor binding sites to evolve (i.e., appear and become fixed within populations) as a consequence of local point mutations within promoters under an assumption of neutral evolution? Because individual binding sites are the functional elements of promoters, the answer to this question will provide insights into the origin of genetic variation relevant to promoter function and the rate at which transcriptional regulation evolves. Of course, promoters actually consist of a few to more than 50 such sites (Arnone and Davidson 1997; Latchman 1999). Although we acknowledge the plausibility that genomic rearrangements can transfer existing regulatory regions between locations within genomes and thereby comprise an important component of promoter evolution, with our approach we are explicitly testing whether local point mutations can comprise a significant component of promoter evolution by rapidly establishing binding sites. Evidence is accumulating that promoter variants that differ by as little as a single binding site may be subject to selection (e.g., Tournamille et al. 1995; Segal, Barnett, and Crawford 1999; Ludwig et al. 2000). To address the evolutionary origin of individual binding sites, we simulated the evolution of promoters by using a computer program to implement standard mutation models and scanning iteratively for the appearance of particular binding sites, then calculating the likelihood that

[1] Present address: Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada.

[2] Present address: Department of Biology, Duke University.

Key words: computer simulation, enhancer, evolution of development, promoter, transcription factor.

Address for correspondence and reprints: Gregory Wray, Department of Biology, Box 90338, Duke University, Durham, North Carolina 27708-0338. E-mail: gwray@duke.edu.

Mol. Biol. Evol. 18(9):1764–1770. 2001

these binding sites would become fixed using population genetic theory. The results indicate that binding sites capable of altering gene expression can evolve via local point mutations on short timescales without invoking selection.

Materials and Methods

Approach

A two-step approach was adopted. In the first step, computer simulation was used to determine the temporal likelihoods (''waiting times'') for the appearance of particular short nucleotide sequences (representing a variety of transcription factor binding sites) within DNA segments (representing either single modules or enhancers or entire promoters). Important differences exist between calculating the probability of binding-site occurrence (i.e., presence by chance) and calculating the likelihood of binding-site appearance (i.e., introduction via local point mutations) within a DNA segment. Biologically, binding-site occurrence is a phenomenon that pertains to individuals in a single generation and is a consequence of heredity, whereas binding-site appearance transpires within populations over multiple generations and is a consequence of heredity, mutation, recombination, and selection. Analytically, binding-site occurrence is a matter of probability theory, whereas binding-site appearance requires additional implementation of models and data. Consequently, formulas exist for calculating the probability of binding-site occurrence (e.g., Kleffe and Langbecker 1990; Kleffe and Grau 1993), whereas the problem of determining the likelihood of binding-site appearance has eluded analytic solution. We invoked computer simulation to accommodate these differences and provide flexibility concerning aspects of molecular evolution analyses that involve particular assumptions (e.g., mutation models, base compositions) and aspects of binding-site appearance analyses that are manifested above the sequence level of organization (e.g., binding-site interactions). Waiting times (measured in terms of generations and mutations) were tallied for the appearance of new binding sites anywhere within a DNA segment of a haploid genome, without selection preserving partial matches. In the second step, well-established principles of population genetics were used to convert these waiting times into estimated fixation times for new binding sites in populations of reasonable sizes.

Sequences

Eight DNA sequences were obtained from GenBank to represent segments flanking genes that might contain *cis*-regulatory elements (accession numbers L13454; M1022, X00479; M36469; Z4824; U04269; AE001274, AC003011, AC002552, U60409, AF008205, AC002134, AF008206, U70253, AC002305, AF008207, AC003679, AC004018; M99054; X06157). The first 200 and 2,000 bp of these sequences were chosen for analysis, without knowledge of whether they corresponded to exons, introns, or intergenic segments. These 200- and 2,000-bp flanking sequences (''regions'') represent single modules (or enhancers) and entire promoters, respectively. Nine short sequences also were chosen for analysis. The lengths (5–9 bp) and base compositions of these short sequences (''binding sites'') represented actual transcription factor binding sites within well-characterized regions: GAGAG, eukaryote GAGA site; TATAA, eukaryote TATA box; AGGATT, *Endo16* Otx binding site; TCCCCG, *Endo16* GCF1 binding site; ACCAAAA, *Endo16* P binding site; ATCAAAG, *Endo16* CG4 binding site; AAGTGATTA, *Endo16* Z binding site excluding final A; TTTTTAAGA, *even-skipped* stripe 2 enhancer hb binding site 9; TTCCCCGAA, *even-skipped* stripe 2 enhancer DSB2 binding site (Arnone and Davidson 1997; Yuh, Bolouri, and Davidson 1998; Ludwig and Kreitman 1995).

Computer Simulation

A computer program (available at http://www.zoo.utoronto.ca/stone/PPE/ppe.htm) was developed that modified regions according to standard mutation models and scanned for the appearance of specific new binding sites. Each of the 16 regions was paired with each of the 9 binding sites, and each of the 144 region–binding-site pairs was entered into the computer program. The computer program iteratively introduced base changes (''mutations'') into the regions according to any of three different mutation models: a one-parameter (Jukes and Cantor), two-parameter (Kimura 1980), or standing-distribution (Felsenstein 1981) model of nucleotide substitution. In each iteration (''generation''), a pseudorandom number generator was used to determine whether a mutation would occur (mutation rate = 10^{-9} per base per generation; Li 1997); if so, a pseudorandom number generator was used to determine the location at which the mutation would be realized and, according to the mutation model chosen, what the resulting base would be. In each generation, allowance was made for the possibility of a second mutation.

Scanning for the appearance of binding sites was performed in only one direction and prior to every generation, until binding sites were established. Because binding sites within regions are typically position-independent with respect to the basal promoter (Latchman 1999; Davidson 2001), new binding sites were allowed to appear anywhere within the regions. At the end of each run, the computer program returned the location of the binding site along the region (''match site''), the number of generations, the number of mutations, and the minimum possible number of mutations required for establishment of the binding site (''shortest path at match site''). (More formally, the shortest path at a particular site is the minimum number of changes required for establishment of a ''substring'' at that particular position within a ''string,'' given a finite set of symbols—here, the letters A, C, G, and T. For example, consider the segment ACGT at position 4 within the 10-base string GGGACGTCCC. The minimum number of changes required to establish the substring ACAT at that position is 1. Of course, many other, longer, paths that establish ACAT at that position could occur, most involving multiple substitutions at the same site or reversals. The shortest path at a particular site cannot exceed

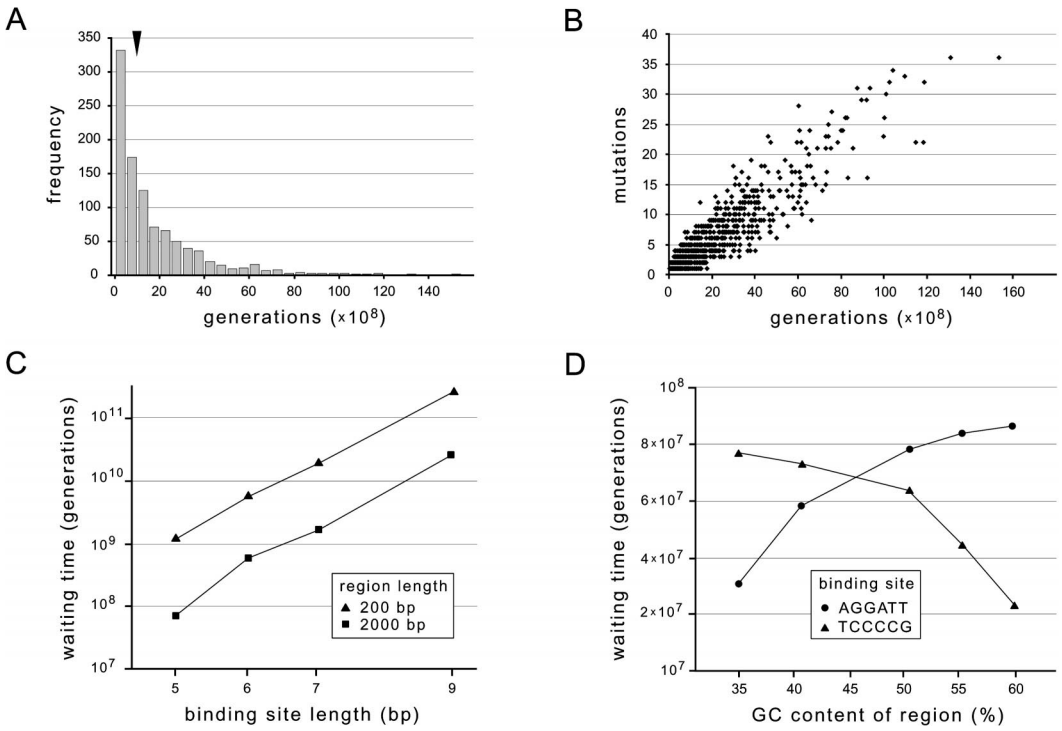

FIG. 1.—Evolutionary dynamics associated with the appearance of a new transcription factor binding site. *A*, Frequency distributions of generations (a proxy for waiting times) were strongly positively skewed (median value 9.5×10^8, indicated by arrowhead; example shown corresponds to 200-bp region L13454 and binding site GAGAG). *B*, Frequency distributions of mutations (another proxy for waiting times) also were strongly positively skewed and consequently were significantly correlated with generations (example shown as in *A*). *C*, The waiting time required for establishment of binding sites increased with increasing binding-site length and decreasing region length (average median values shown; note the logarithmically scaled ordinate). *D*, Waiting time exhibited a dependency on base composition, increasing with GC content of region for GC-poor binding sites and decreasing for GC-rich binding sites (values of GC content in examples shown correspond to 200-bp regions Z4824; U04269; AE001274, AC003011, AC002552, U60409, AF008205, AC002134, AF008206, U70253, AC002305, AF008207, AC003679, AC004018; M99054; X06157); the magnitude of this effect was modest relative to the effects of changes in binding-site length or region length (compare the ordinate scales in *C* and *D*).

the length of the substring.) One thousand replicates were performed for each of the 144 region–binding-site pairs using a one-parameter model of nucleotide substitution. Additional computer simulation was conducted for some region–binding-site pairs using either a two-parameter or standing-distribution model of nucleotide substitution.

Results

Waiting Times: Single Binding Sites

Without exception, frequency distributions of waiting times were strongly positively skewed, as expected with a Poisson process (fig. 1*A*). This indicates that binding-site appearance was usually rapid. Waiting times measured in terms of generations and mutations were positively linearly correlated (fig. 1*B*). Therefore, we consider median waiting times (10^8–10^{11} for generations or 1–1,108 for mutations, omitting exact matches) rather than mean values as summary measures for region–binding-site pairs. As expected, waiting time increased with increasing binding-site length and decreasing region length (fig. 1*C*) again, whether this was measured in terms of generations or mutations. Waiting time was considerably more sensitive to changes in binding-site length than to changes in region length, scaling approximately exponentially with the former and linearly with the latter (fig. 1*C*). These results can be used to predict waiting times involving a variety of binding-site and region lengths.

The manner in which these waiting times were determined was conservative for several reasons. Because the computer program simulated the modification of a region along a single DNA strand, the waiting times corresponded biologically to a lineage of haploid, asexually reproducing individuals. The appearance of new binding sites would be achieved on dramatically shorter timescales in diploid organisms within populations of realistic sizes and that initially were genetically heterogeneous. Supplementary computer simulation in which the effective population size was increased to 10, 50, and 100 decreased generations approximately 10-, 50-, and 100-fold. This approximately linear trend is expected to asymptote for appreciably larger effective population sizes. Because the algorithm used in our computer simulation incorporated reverse mutations, new binding sites appeared without selection preserving partial matches. We therefore assumed that the neutral fixation rate of new promoter alleles was equal to the mutation rate (Gillespie 1991; Li 1997). In addition, binding sites with one or two nucleotides mismatched are often functional, and binding sites typically function in either orientation (Arnone and Davidson 1997; Latch-

man 1999). Invoking selection to preserve partially matching segments and allowing for establishment of binding sites in either direction would significantly reduce waiting times; the first modification would impart a particularly strong effect.

Two additional assumptions involved in our computer simulation might have affected waiting times. First, we assumed that all binding sites within a region could mutate without disrupting promoter function, whereas real promoters contain functional binding sites that are preserved by selection (Segal, Barnett, and Crawford 1999; Ludwig et al. 2000). This would effectively reduce the number of locations within a *cis*-regulatory region that could vary and thereby would increase waiting times. Because binding sites comprise a small proportion of nucleotides within regions (approximately 2%–15% in 5′ flanking sequences; Arnone and Davidson 1997), this increase in waiting times should be modest. Second, we assumed that we could neglect the effects of nonhomologous recombination, because this probably occurs much less frequently than does point mutation. In any case, recombination should introduce no systematic bias in waiting times.

Fixation Times Within Populations

To estimate fixation times for real-world cases, numbers of generations derived from the simulations can be combined with realistic parameter values. For example, with an effective population size of 10^6, a new 6-bp binding site will appear somewhere within the region extending 2 kb 5′ of a given gene in one individual approximately every 2,250 generations. (The average median value corresponding to 6-bp binding sites was 4,506,870,000 generations; given an effective population size of 10^6, the presence of diploidy, and the possibility of the binding site appearing on either DNA strand, the estimated fixation time was (4,506,870,000 generations/(10^6 individuals $\times$ 2 DNA strands)) $\approx$ 2,254 generations.) Typically, this will be the case for each gene in a genome and every 6-bp binding site.

Using realistic but conservative generation times, estimated fixation times can be calculated for a variety of organisms (table 1). Most of these fixation times are less than a millennium, and all are less than 600,000 years. Including realistic parameter values associated with real-world cases, more specific estimated fixation times can be calculated. For example, consider the evolution of a new hunchback protein binding site within the ~600 bp *even-skipped* stripe 2 enhancer in *Drosophila*. Given an effective population size of 10^6, the presence of diploidy, a 6-bp binding site, the possibility of the binding site appearing on either DNA strand, and a generation time of approximately 5 weeks, the estimated fixation time is approximately 75 years. As hunchback can bind to several variants on the consensus binding site, the actual waiting time would be shorter. Thus, we conclude that the evolution of new transcription factor binding sites is a continuous process, occurring on microevolutionary timescales.

Table 1
Estimated Fixation Times[a] (years) for the Appearance of 6-bp Transcription Factor–Binding Sites via Local Point Mutations

Organism	One Site in 200-bp Region	One Site in 2,000-bp Region	Two Sites in 200-bp Region	Two sites in 2,000-bp Region
Saccharomyces	2	0.2	362	48
Arabidopsis	1,127	119	271,215	35,571
Drosophila	226	24	54,244	7,115
Caenorhabditis	31	4	7,233	949
Strongylocentrotus ...	45,069	4,759	10,848,600	1,422,800
Mus	752	80	180,810	23,714
Homo	56,336	5,949	13,560,800	1,778,500

[a] Estimates for single or pairs of specific 6-bp bindings sites within each 200- or 2,000-bp region of a genome, with an effective population size of 10^6 and conservatively long generations times: *Saccharomyces*, 1,500 per year; *Arabidopsis*, 2 per year; *Drosophila*, 10 per year; *Caenorhabditis*, 75 per year; *Strongylocentrotus*, 0.05 per year, *Mus*, 3 per year; *Homo*, 0.04 per year.

Sensitivity Analysis

This general result was unaffected by the modification of several assumptions. Differences in particular base compositions (GC content) among regions or binding sites elicited only minor effects on waiting times. The same was true when various mutation models were compared: one-parameter, two-parameter, and standing-distribution models of nucleotide substitution produced quantitatively similar results. Covariation of GC content between binding sites and regions engendered a small but noticeable effect: as expected, longer waiting times were required to establish AT-rich binding sites within GC-rich regions and vice versa (fig. 1*D*). The most pronounced effects were evoked using a standing-distribution model, which biases mutations to preserve overall base composition, combined with substantial mismatch of GC content between binding sites and regions. However, the magnitudes of these effects were modest relative to the effects of changes in binding-site length or region length (compare the ordinate scales in fig. 1*C* and *D*).

Waiting and Fixation Times: Binding-Site Pairs

Given the importance of interactions among transcription factors while binding to promoters (Latchman 1999; Davidson 2001), it is instructive to calculate estimated fixation times associated with regions containing particular combinations of binding sites. We therefore simulated the simultaneous appearance of two binding sites within 200- and 2,000-bp regions via local point mutations, again without invoking selection, and converted waiting times into estimated fixation times. We conservatively assumed that the two binding sites must reside on the same DNA strand and that selection would occur only after both binding sites had been established. Some of the waiting times, and therefore fixation times (last two columns in table 1), were short on macroevolutionary timescales. For example, given an effective population size of 10^6, the presence of diploidy, the possibility of the binding-site pair appearing on either DNA strand, and a generation time of approximately 5 weeks,

the estimated fixation time associated with two 6-bp binding sites within a 200-bp region in *Drosophila* is approximately 55,000 years (third row, third column in table 1). Thus, even simple combinations of new transcription factor binding sites can evolve on microevolutionary timescales without invoking selection.

Discussion

It is important to recognize that while our computer simulation can be used to predict waiting times for the appearance of a new transcription factor binding site, it cannot be used to predict the functional consequences. It is likely that in many cases, the appearance of a new binding site will have no effect on the expression of a nearby gene. For example, some transcription factors must interact with other DNA-binding proteins to affect gene expression (Gerhart and Kirschner 1997; Latchman 1999), and therefore the appearance of a single new binding site might be functionally neutral. Alternatively, nearby silencer sites (Burcin et al. 1997; Leclerc, Eskild, and Guerin 1997; Kim and Siu 1998; Li et al. 1998) might override any new binding site that appears. Nevertheless, it is plausible that in some cases the appearance of a new binding site will alter gene expression. Because most promoters are inactive by default (Latchman 1999; Davidson 2001), activation of transcription by the appearance of a single new binding site seems less likely than does modulation or restriction of an existing phase of gene expression. Of course, even if a new binding site does affect gene expression, there is no way to predict what, if any, phenotypic consequences might ensue. As with amino acid substitutions within coding regions of genes, we predict that in many cases the consequences of a new binding site appearing within a promoter will be either detrimental or neutral; only in rare cases will it be beneficial. The salient prediction yielded by our computer simulation is that local point mutations will constantly produce new binding sites that in principle are capable of altering gene expression and that such sites may be subject to selection. Furthermore, because a particular change in gene expression could in principle result from a variety of modifications to regulatory sequences, the actual fixation times associated with particular types of transcriptional changes are probably even shorter than the estimated fixation times we predict.

The promoter regions of eukaryotic genes are complex and include approximately a dozen to several dozen transcription factor binding sites (Arnone and Davidson 1997). The likelihood of a dozen binding sites evolving simultaneously without selection is infinitesimally small, as can easily be estimated by extrapolating the trend apparent in table 1. We envision instead that complex regulatory systems are the result of long and complex evolutionary histories involving stepwise assembly and turnover of binding sites. Local point mutations, transposition, and recombination all likely play important roles in this process. For example, changes in transcription could result from the gain (or loss) of a key binding site as a consequence of local point mutations or from the insertion (or deletion) of several new binding sites following transposition (or recombination). There is ample empirical evidence for both processes, the former from sequence comparisons among species that reveal gains and losses of single binding sites (e.g., Gonzalez et al. 1995; Ludwig and Kreitman 1995; Tournamille et al. 1995; Margarit et al. 1998), and the latter from sequences indicative of transpositional origins for binding sites (e.g., Britten 1997; Kidwell and Lisch 1997).

In contrast to transposition and recombination, the establishment of new binding sites via local point mutations would be accomplished incrementally, requiring from one to several independent point mutations. Nevertheless, the results of our computer simulation suggest that the evolution of specific new binding sites in flanking sequences adjacent to each gene in a genome is virtually inevitable in populations of realistic sizes over timescales of months to millions of years, depending on generation time. Although the assembly of simple combinations of new transcription factor binding sites will take much longer (approximately 250-fold longer for two 6-bp binding sites than for a single 6-bp binding site; table 1), even the waiting times associated with such pairwise combinations are usually short intervals in macroevolutionary terms.

It is important to note that these results concern the evolution only of one or two particular 6-bp binding sites within a 200- or 2,000-bp region 5′ of a single gene. Typical metaphyte or metazoan genomes contain many features that could expedite the evolution of single binding sites and combinations of binding sites: $>10^4$ genes, dozens of different sequences that could serve as binding sites, sequence heterogeneity within populations, and other regions where binding sites could function (within introns, 3′ of the gene, and farther 5′). For example, our results suggest that new 6-bp binding sites for real transcription factors will evolve somewhere within 2 kb of the start site of transcription of any gene in humans at a rate of $\sim$0.013 per genome per generation. Within a population of 10^6 individuals, this translates to $\sim$12,500 new 6-bp sites during each generation. (This estimate is highly approximate, as the fraction of the 4,096 possible 6-bp sequences that correspond to the consensus binding sites of real transcription factors is not known for any organism, and as our computer simulation involved several simplifying assumptions; however, the estimate is conservative, as our computer simulations assumed an initially isogenic population, whereas significant genetic heterogeneity exists in noncoding DNA within real populations.) For comparison, an empirically based estimate of the neutral mutation, and therefore fixation rate for point mutations, in humans is $\sim$175 per genome per generation (Nachman and Crowell 2000), or $\sim 175 \times 10^6$ new single-nucleotide polymorphisms in a population of the same size. When corrected for the difference in mutation rates between the two studies (10^{-9} vs. 2.5×10^{-8}), Nachman and Crowell's (2000) estimated rate of SNP origins is $\sim$270-fold higher than our predicted rate of 6-bp binding site origins. Given that the origin of most 6-bp sites will require more than one point mutation (fig. 2*B*), this dif-

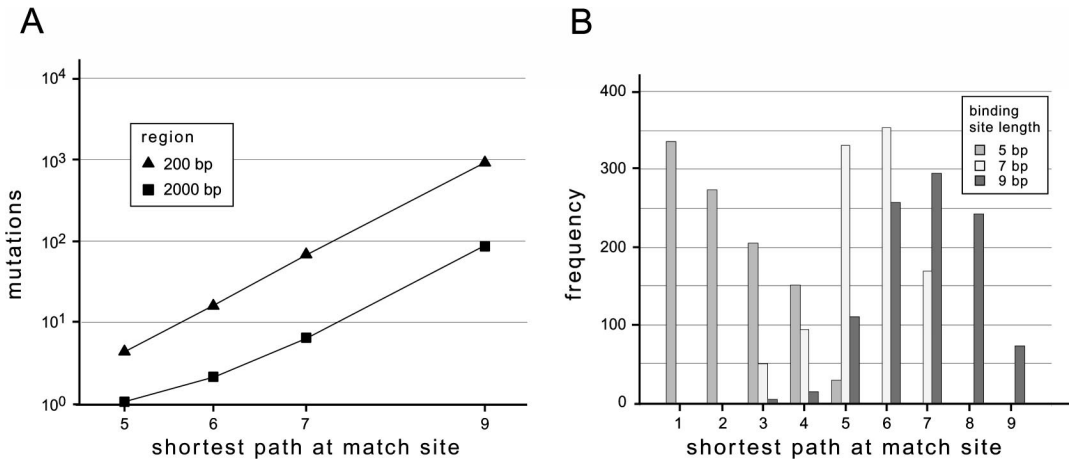

FIG. 2.—Mutations required for the appearance of a new transcription factor binding site. *A,* Numbers of local point mutations that actually occurred in establishing a new binding site exceeded the minimum number possible at that position along the original region (shortest path at match site), especially for longer binding sites and shorter regions (means of values shown for 200- and 2,000-bp regions; to be conservative, the greatest possible values of shortest paths at match sites are plotted). *B,* As binding-site length increased, the skew of the distribution of shortest paths changed from positive, to zero, to negative (the example shown corresponds to 200-bp region L13454 and binding sites GAGAG, ACCAAAA, and TTCCCCGAA).

ference seems reasonable. The fundamental conclusion that may be drawn from these considerations is that local point mutations should continuously generate considerable genetic variation within natural populations that is capable of altering transcription.

The results of our computer simulation provide a basis for interpreting empirical data concerning promoter evolution. One application of the results involves inferring the evolutionary history of a promoter in two closely related species. If the promoter in one species differs from the promoter in the other species by only one transcription factor binding site, then this difference is most plausibly explained as the product of local point mutations, whereas a large, contiguous block of unique binding sites in only one of the two species is more likely the result of transposition or recombination. Our approach provides a basis for identifying cases situated between these two extremes: using specific parameters, it is possible to determine the number and size of binding sites that are most likely to have resulted as a consequence of local point mutations rather than transposition or recombination (in either case, outgroup comparison with related species would be needed to polarize the differences as gains or losses).

Another application of the results involves deducing the evolutionary dynamics of *cis*-regulatory sequences. In our computer simulation, the number of local point mutations that actually occurred in converting the original base sequence at the match site into the binding site was typically greater than the minimum number possible (shortest path at match site) and scaled approximately exponentially with binding-site length (fig. 2*A*). As binding-site length increased, the skew of the distribution of shortest paths at match sites changed from positive (5-bp binding site), to zero (7-bp binding site), to negative (9-bp binding site) (fig. 2*B*). Thus, the efficiency of establishment of binding sites (measured as the number of mutations that occurred) was dependent on binding-site length: shorter binding sites typically involved substitutions at only a few positions, whereas longer binding sites typically involved complete turnover of bases. This accords with intuition, as steps intermediate toward a match can be eliminated by reverse mutations, and the likelihood of this occurring is greater for longer than for shorter binding sites.

The short fixation times that we predict for binding sites in populations of realistic sizes suggest that functional differences in promoter sequences among species require neither extended divergence times nor genomic rearrangements. We encourage researchers interested in understanding the developmental genetic basis for phenotypic evolution to test these predictions with empirical studies of variation in promoter structure and function within populations and among species.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (J.R.S.) and by the National Science Foundation and the National Aeronautics and Space Administration (G.A.W.). We thank D. Dykhuizen for helpful discussions and E. Abouheif, J. Balhoff, A. Bely, E. Knott, J. Kuhn, M. Pizer, and M. Rockman for constructive comments.

LITERATURE CITED

ARNONE, M. I., and E. H. DAVIDSON. 1997. The hardwiring of development: organization and function of genomic regulatory systems. Development **124**:1851–1864.

BRITTEN, R. J. 1997. Mobile elements inserted in the distant past have taken on important functions Gene **205**:177–182.

BURCIN, M., R. ARNOLD, M. LUTZ, B. KAISER, D. RUNGE, F. LOTTSPEICH, G. N. FILIPPOVA, V. V. LOBANENKOV, and R. RENKAWITZ. 1997. Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. Mol. Cell. Biol. **17**:1281–1288.

DAMJANOVSKI, S., M.-H. HUYAH, K. MOTAMED, E. H. SAGE, and M. RINGUETTE. 1998. Regulation of SPARC expression during early *Xenopus* development: evolutionary divergence and conservation of DNA regulatory elements between amphibians and mammals. Dev. Genes Evol. **207**: 453–461.

DAVIDSON, E. H. 2001. Genomic regulatory systems: development and evolution. Academic Press, San Diego, Calif.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

FRANKS, R. R., B. R. HOUGH-EVANS, R. J. BRITTEN, and E. H. DAVIDSON. 1988. Spatially deranged though temporally correct expression of a *Strongylocentrotus purpuratus* actin gene fusion in transgenic embryos of a different sea urchin family. Genes Dev. **2**:1–12.

GERHART, J., and M. KIRSCHNER. 1997. Cells, embryos, and evolution. Blackwell Science, Maldon, England.

GILLESPIE, J. H. 1991. The causes of molecular evolution. Oxford University Press, New York.

GONZALEZ, P., P. V. RAO, S. B. NUNEZ, and J. S. ZIGLER. 1995. Evidence for independent recruitment of Zeta-Crystallin/Quinone Reductase (CRYZ) as a crystallin in camelids and hystriocomorph rodents. Mol. Biol. Evol. **12**:773–781.

GRENIER, J. K., T. L. GARBER, R. WARREN, P. M. WHITTINGTON, and S. CARROLL. 1997. Evolution of the entire arthropod *Hox* gene set predated the origin and radiation of the onychophoran/arthropod clade. Curr. Biol. **7**:547–553.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KEYS, D. N., D. L. LEWIS, J. E. SELEGUE, B. J. PEARSON, L. V. GOODRICH, R. L. JOHNSON, J. GATES, M. P. SCOTT, and S. B. CARROLL. 1999. Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. Science **283**:532–534.

KIDWELL, M. G., and D. LISCH. 1997. Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. USA **94**:7704–7711.

KIM, H. K., and G. SIU. 1998. The notch pathway intermediate HES-1 silences *CD4* gene expression. Mol. Cell. Biol. **18**: 7166–7175.

KIMURA, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KLEFFE, J., and E. GRAU. 1993. The joint distribution of patterns in random sequences with applications to the RC-measures for expressivity. CABIOS **9**:275–283.

KLEFFE, J., and U. LANGBECKER. 1990. Exact computation of pattern probabilities in random sequences generated by Markov chains. CABIOS **6**:347–353.

LATCHMAN, D. 1999. Eukaryotic transcription factors. 3rd edition. Academic Press, San Diego, Calif.

LECLERC, S., W. ESKILD, and S. L. GUERIN. 1997. The rat growth hormone and human cellular retinol binding protein 1 genes share homologous NF1-like binding sites that exert either positive or negative influences on gene expression in vitro. DNA Cell Biol. **17**:951–967.

LI, Q. L., C. A. BLAU, C. H. CLEGG, A. ROHDE, and G. STAMATOYANNOPOULOS. 1998. Multiple epsilon-promoter elements participate in the developmental control of *epsilon-globin* genes in transgenic mice. J. Biol. Chem. **273**:17361–17367.

LI, W.-H. 1997. Molecular evolution. Sinauer, Sunderland, Mass.

LOWE, C. J., and G. A. WRAY. 1997. Radical alterations in the roles of homeobox genes during echinoderm evolution. Nature **389**:718–721.

LUDWIG, M. Z., C. BERGMAN, N. H. PATEL, and M. KREITMAN. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature **403**:564–567.

LUDWIG, M., and M. KREITMAN. 1995. Evolutionary dynamics of the enhancer region of *even-skipped* in *Drosophila.* Mol. Biol. Evol. **12**:1002–1011.

LUDWIG, M. Z., N. H. PATEL, and M. KREITMAN. 1998. Functional analysis of *eve* stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. Development **125**:949–958.

MADURO, M., and D. PILGRIM. 1996. Conservation of function and expression of *unc-119* from two *Caenorhabditis* species despite divergence of non-coding DNA. Gene **183**:77–85.

MARGARIT, E., A. GUILLEN, C. REBORDOSA, J. VIDAL-TABADOADA, M. SANCHEZ, F. BALLESTA, and R. OLIVA. 1998. Identification of conserved potentially regulatory sequences of the *SRY* gene from 10 different species of mammals. Biochem. Biophys. Res. Comm. **245**:370–377.

NACHMAN, M. W., and S. W. CROWELL. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics **156**:297–304.

PATEL, N. H., E. MARTIN-BLANCO, K. G. COLEMAN, S. J. POOLE, M. C. ELLIS, T. B. KORNBERG, and C. S. GOODMAN. 1989. Expression of *engrailed* proteins in arthropods, annelids, and chordates. Cell **58**:955–968.

RAFF, R. A. 1996. The shape of life. University of Chicago Press, Chicago.

RAFF, R. A., and T. C. KAUFFMAN. 1983. Embryos, genes, and evolution. Indiana University Press, Bloomington.

ROSS, J. L., P. P. FONG, and D. R. CAVENER. 1994. Correlated evolution of the *cis*-acting regulatory elements and developmental expression of the *Drosophila GLD* gene in 7 species of from the subgroup Melanogaster. Dev. Genet. **15**: 38–50.

SEGAL, J. A., J. L. BARNETT, and D. L. CRAWFORD. 1999. Functional analysis of natural variation in Sp1 binding sites of a TATA-less promoter. J. Mol. Evol. **49**:736–749.

TOURNAMILLE, C., Y. COLIN, J. P. CARTRON, and C. LE VAN KIM. 1995. Disruption of a GATA motif in the *Duffy* gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat. Genet. **10**:224–228.

WANG, R.-L., A. STEC, J. HEY, L. LUKENS, and J. DOEBLEY. 1999. The limits of selection during maize domestication. Nature **398**:236–238.

YUH, C.-H., H. BOLOURI, and E. H. DAVIDSON. 1998. Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. Science **279**:1896–1902.

STEPHEN PALUMBI, reviewing editor

Accepted May 30, 2001