Promoter evolution is impossible by random mutations

Jean O'Micks

According to a change in evolutionary paradigms, mutations in the regulatory circuitry of genes has been suggested as the major force behind morphological changes during evolution. Even though we cannot observe mutations going on in the regulatory region of genes one by one, these events can still be modelled *in silico* via computer programs, such as PromMute. This program simulates point mutations happening over a set number of generations within a promoter sequence. PromMute models the formation of several target transcription factor binding sites from random sequences. This paper examines how successfully target regulatory modules can form via random mutations with increasing numbers of motifs, which can also be constrained by physical distance as modelled by PromMute. The results indicate that longer and larger numbers of motifs make it more and more difficult, virtually impossible, for regulatory modules to form by random mutation, as required by molecular evolution, especially so with added spatial constraints.

Previous creationist publications have detailed how genetic conservation is not capable of driving molecular evolution. This is because evolution demands the flux of newly generated genetic elements over the course of evolutionary time. Thus, 'evolutionary conservation' is an oxymoron, in that change cannot be affected by conservation and stasis.[1]

Ohno[2] recognized this and realized that instead of newly evolving genes, changes behind the regulation of existing genes are responsible for changes in phenotype, thereby supposedly affecting evolutionary development. This has become a new paradigm within molecular evolution.[3] For example, it is thought that different variants, differing by 1 bp of the palindromic P3 motif TAATYNRATTA reside in the regulatory region of the rhodopsin gene in subsets of *Drosophila* photoreceptors.[4]

The question arises, besides the coding regions of genes, how did the regulatory region of genes arise via random mutations? It is one thing to explain how genes arise via random mutations, which isn't much dealt with, but it is another thing to explain how a functioning regulatory region arises by random mutations. Furthermore, both gene regions have to be present in order for a gene to function properly. An assembly line can have all the necessary mechanical parts in place, but if there is no process which turns on the assembly line and modulates its function, then the assembly line achieves nothing.

The regulatory region of a gene is a complex genetic structure, which includes the promoter, which is that part of the genome just upstream of a given gene, as well as enhancer elements, and the first introns of the gene. Distal enhancers also take part in gene regulation which may even reside on other chromosomes. In general, the promoter can be delineated as the section of the gene –500 to +100 bp around the transcription start site (TSS). Some authors define the promoter length as 1,000 to even 3,000 or 5,000 bp. The promoter is a stretch of DNA which contains so-called transcription factor binding sites (TFBSs), also known as regulatory motifs (or just simply motifs in the rest of this paper). These motifs are the docking sites of transcription factors (TFs), generally proteins, which bind with their surface to the face of the TFBS in a lock and key manner. Motifs are generally between 5–20 bp long, and are characterized by a more or less definitive sequence. Physico-chemical interactions between amino acid side chains on the TF surface and the bases in the DNA are integrated so as to affect gene regulation. Some TFs induce gene expression, some act as repressors. Depending on how the TF binds to its motif, some positions within the motif sequence can be degenerate. For example, the sequence for the ABRE element ACGTGKC is degenerate at position six, where K = G/T. Motifs can also form regulatory modules, where multiple motifs and TFs act in concert to modulate the behaviour of Polymerase II, which initiates gene transcription.[5]

Since motif sequences are relatively short compared to the sequences of whole genes, it would be interesting to examine how motifs could have come into existence via random genetic mutations. Motifs could form via series of single base pair mutations. However, it is impossible to follow single base pair mutations within a population of organisms over hundreds or thousands of generations during evolutionary time, which are needed for the motifs themselves to form. The best we can do to approximately model this process is to simulate promoter evolution *in silico*.

Results

Several candidate promoter mutation simulation programs were examined to use in this analysis, such as ev,[6] PPE,[7] and PromMute.[8] The program PromMute was chosen due to several considerations: as opposed to ev and PPE, PromMute simulated a wider variety of real motifs in a longer promoter segment 1 Kbp long. Position weight matrixes (PWMs) were used to score possible occurrences of target motifs. A PWM is a 4xn matrix, which represents a motif n bp long, and is used to score the occurrence of a motif. Each of the positions of the motif's occurrence is scored based on which of the 4 bases (A, C, G, or T) they match. All positional subscores are added up to give the score for the occurrence of a given motif.

Motifs were simulated to form anywhere within the promoter region. Most importantly, as opposed to ev and PPE, PromMute doesn't allow for the selection of partially formed motifs. This is because natural selection cannot operate on a partially present motif, because a motif either functions, or it doesn't—it either binds its corresponding TF or it doesn't. This follows binary logic. In this respect, regulatory motif modules are irreducibly complex systems. In order for a valid promoter region to evolve in the PromMute program, all target motifs have to be fully functional (with a score of its occurrence at least 90% of the maximum score of its corresponding PWM). In other words, all target motifs have to have undergone a gain of function mutation, resulting in the motif being newly capable of binding its TF. PPE uses real promoters in its simulation, whereas PromMute uses a random sequence, but this is actually ideal to simulate the appearance of islands of meaningful genetic signals in a sea of random, meaningless sequence.

Description of the PromMute program

PromMute simulates point mutations in a promoter of a set size within a set number of organisms over a set number of cycles/generations. In each promoter, the optimal occurrence of each motif is determined in each organism (that is, the closest to the target motif sequence which is the highest-scoring occurrence). Each motif occurrence is scored according to its PWM, and if the ratio of its score divided by the maximum score of that PWM is above the motif cut-off limit, then that occurrence of that motif is deemed functional. The lowest scoring organisms (bottom 50%) are eliminated and their place is taken over by organisms from the top 50% by binary division (this simulates natural selection). All of this happens in one generation, and the goal is to reach a point where all selected target motifs become functional. If all motifs are functional except one, the simulation keeps on running. Either all motifs become functional eventually or the simulation halts after the set number of cycles.

PromMute was run five times with sets of one, two, or three motifs, and the average number of generations were taken to describe the amount of time needed for these motif(s) to form. If the program ran for 10,000 generations, it was assumed that the motif(s) weren't able to form. A selection cut-off of 0.5 was chosen, and motif cut-offs of 0.7, 0.8, 0.9, and 1.0 were analyzed for 100 organisms for single motifs, and a motif cut-off of 0.9 for motif pairs and triplets. PromMute was also developed so as to halt when all target motifs were formed within a specified spacer region of 100 bp. This was done for the analysis of motif pairs and triplets. In this analysis a motif is synonymous with a transcription factor binding site (TFBS). In total five sets of analyses were run: single TFBS, pairs of TFBS with and without spacers, and TFBS triplets with and without spacers.

Single motifs

All 24 motifs from the Promoter Database of *Saccharomyces cerevisiae* (SCPD) were analyzed, and the log10 value of the number of generations needed for each motif to form is depicted in figure 1 for motif cut-off values of 0.7–1.0. Black denotes a log10 value of zero (one generation), whereas light grey denotes a log10 value of four (10,000 generations). We can see that the longer a motif is, and the higher the motif cut-off is, the longer it takes for a

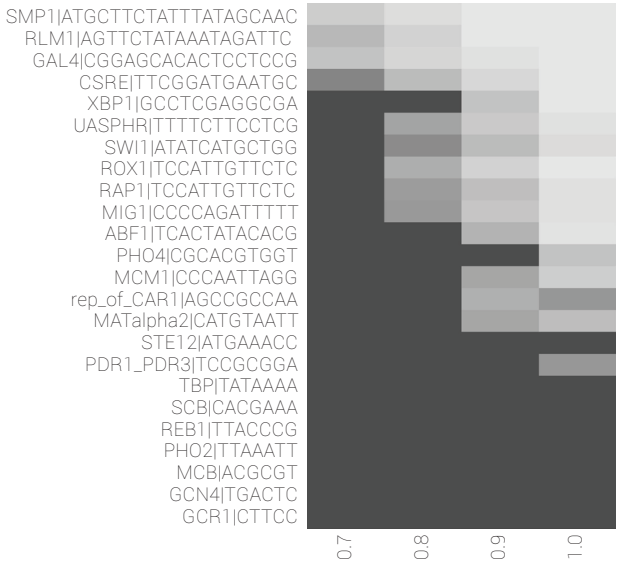

Figure 1. Log10 values of generation times needed for the 24 regulatory motifs to form individually in a proximal promoter as simulated by PromMute for motif cut-off values of 0.7, 0.8, 0.9, and 1.0. Black values correspond to a generation of 1, or a log10 generation time of 0. Light grey values correspond to log10 values of 4, corresponding to 10,000 generations.

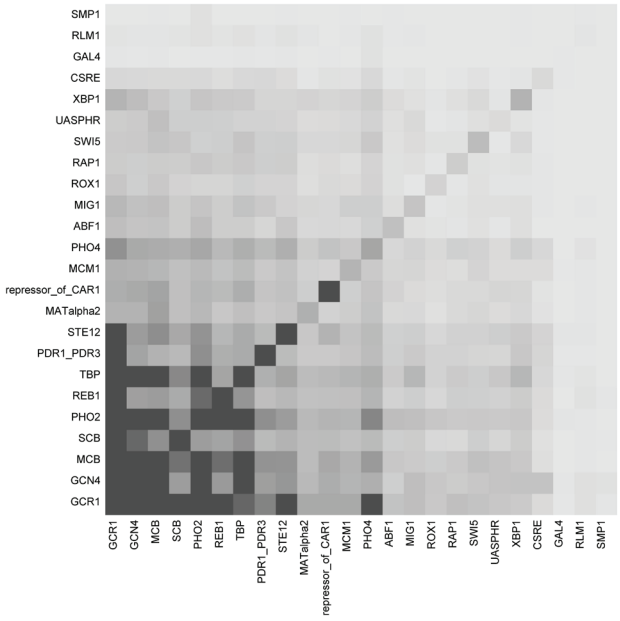

Figure 2. Heat map of log10 generation time for the formation of pairs of motifs. Black values correspond to a generation of 1, or a log10 generation time of 0. Light grey values correspond to log10 values of 4, corresponding to 10,000 generations.

given motif to form. At smaller lengths of 5–8 bp for motifs such as GCR1 (CTTCC) and GCN4 (TGACTC) we can see that these motifs form within short periods of time. This is not surprising, since a motif 5 bp long occurs randomly once every $4^5 = 1{,}024$ bp, which is very close to the size of the proximal promoter that was simulated.

Motifs of length 13 bp and longer do not form by chance at a motif cut-off of 1.0, whereas motifs of length 18–20 bp do not form at a motif cut-off of 0.9–1.0 according to the settings of the PromMute program.

The average log10 generation value (where the generation value was not assumed to be unlimited) is 1.43 for a motif cut-off of 0.9.

Motif pairs

The formation of motif pairs was examined at a motif cut-off value of 0.9. Here 24 x 24 = 576 possible pairs were studied. As we can see in figure 2, darker shades correspond to log10 values of 0 (one single generation), whereas lighter shades correspond to values of 4 (10,000 generations). As we can see, motif pairs each under 10 bp (such as PHO4) in length had no serious difficulty forming. The longer motifs, such as GAL4 (17 bp), RLM1 (18 bp), and SMP1 (20 bp) all showed log10 values of close to four (10,000 generations) when paired up with all other motifs.

Motif triplets

Because of spatial restraints in depicting the results for 13,824 triplet motifs stemming from 24 motifs, five motifs were chosen: GCR1, a very short motif (5 bp); MATalpha2, a short motif (9 bp); CSRE, an average length motif (13 bp); GAL4, a long motif (17 bp); and SMP1, a very long motif (20 bp), each 4 bp longer than the previous motif. This way we can see how motifs of different lengths behave when put together in triplets.

As we can see, the generation time for each triplet-based module has greatly increased (figure 3). The average log10 generation value rose from 2.8 for motif pairs to 3.56 for motif triplets. The difference is 0.75, meaning that it is $10^{(3.56-2.8)} = 10^{0.76} = 5.75$ times more difficult for motif triplets to form than motif pairs. This is the degree of difficulty which arises when one more motif is added for random mutations to form.

Motif combinations with spacer restrictions

The PromMute program was further developed in that it took certain spacer restraints into consideration, since TFs act in concert to influence gene expression. Lu *et al.*[9] demonstrated the distance conservation of transcription regulatory motifs in human promoters. For example, the

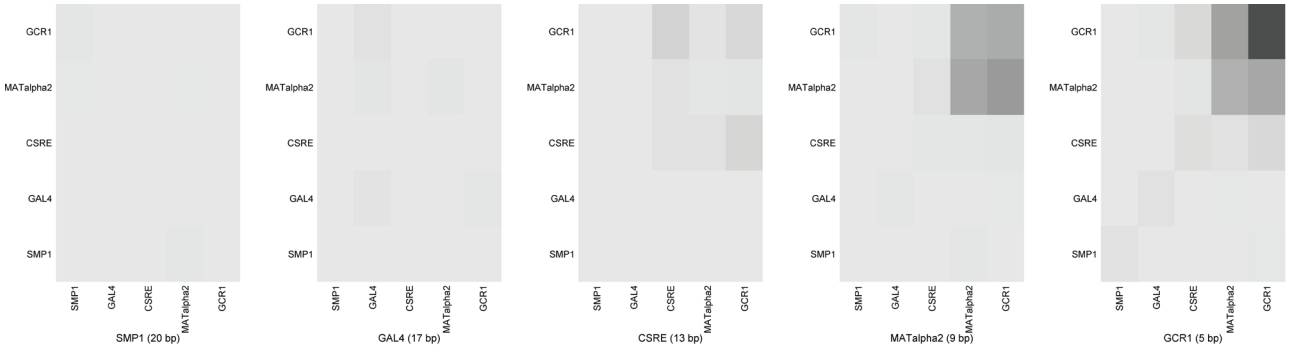

Figure 3. Heat map of log10 generation time for the formation of motif triplets. Black values correspond to a generation of 1, or a log10 generation time of 0. Light grey values correspond to log10 values of 4, corresponding to 10,000 generations.

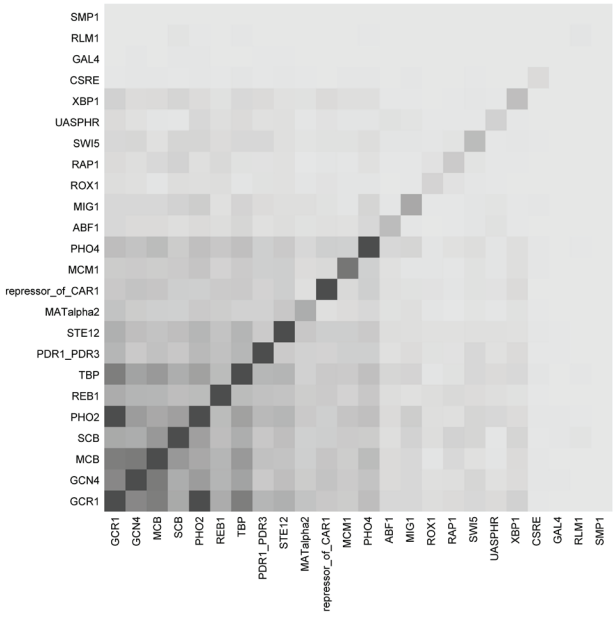

Figure 4. Heat map of log10 generation time for the formation of pairs of motifs constrained by a spacer motif of 100 bp. Black values correspond to a generation of 1, or a log10 generation time of 0. Light grey values correspond to log10 values of 4, corresponding to 10,000 generations.

E1B adenovirus gene diminishes in activity, if the spacing increases between the GC-box and the TATA-box.[10] In order to do this, TFs must also be in each other's spatial vicinity. Thus the motifs themselves must be relatively close to each other in order for their corresponding TFs to dock to each other.

A total 25,976 human proximal promoters 1 Kbp long were downloaded from the Swiss Institute of Bioinformatics, and 127 human PWMs were downloaded from the JASPAR database. All 127 PWMs were scanned in each of the human proximal promoters at a motif cut-off value of 0.9. The average distance between two motifs was found to be 110 bp, but this was approximated in this study with a spacer of 100 bp.

Figure 4 shows the log10 values of the generation number needed for the formation of motif pairs within the spacer limit. As we can see, on a global level, log10 generation values have shifted to higher values as the effect of the spatial constraint of requiring both motifs to be found within 100 bp of each other. Compared to motif pairs without the spacer restraint the average log10 generation values (where the generation value was not assumed to be unlimited) is 3.3 for motif pairs with spacer, and 2.8 without. The difference is 0.5, meaning that it is $10^{(3.3-2.8)} = 10^{0.5} = 3.2$ times more difficult for motif pairs to form with a spacer requirement than pairs unbound by this limitation requirement.

When motif triplets were analyzed with spacers (see figure 5), the average log10 generation value rose from 3.2 to 3.6. Thus it is $10^{(3.6-3.2)} = 10^{0.4} = 2.5$ times more difficult for spaced motif triplets to form than spaced motif pairs. When adding spacers to motif triplets, it is only $10^{(3.6-3.56)} = 10^{0.04} = 1.1$ times more difficult for this kind of regulatory module to form when just considering motif triplets. Figure 6 shows the relationship between the average log10 generation values for single motifs, motif pairs, and motif triplets with and without spacers.

Discussion

We can draw a number of interesting conclusions from this analysis of the *in silico* simulation of motifs in promoter sequences. First of all, it validates the concept of disallowing the selection of partially formed motifs. As illustrated in the original paper, the longer a motif is, the larger surface it exposes to random mutations which can fragment it. If we take a run to be a number of consecutive generations, with its own unique PWM score, then this increased rate of motif fragmentation results in a larger number of runs of shorter lengths. If you have a short motif, then it will hardly be interrupted by any mutations, and its PWM score stays the same for a long time. However, a longer motif fragments more easily, and thus the PWM fluctuates a lot. This means the chances are higher that a random mutation occurs within

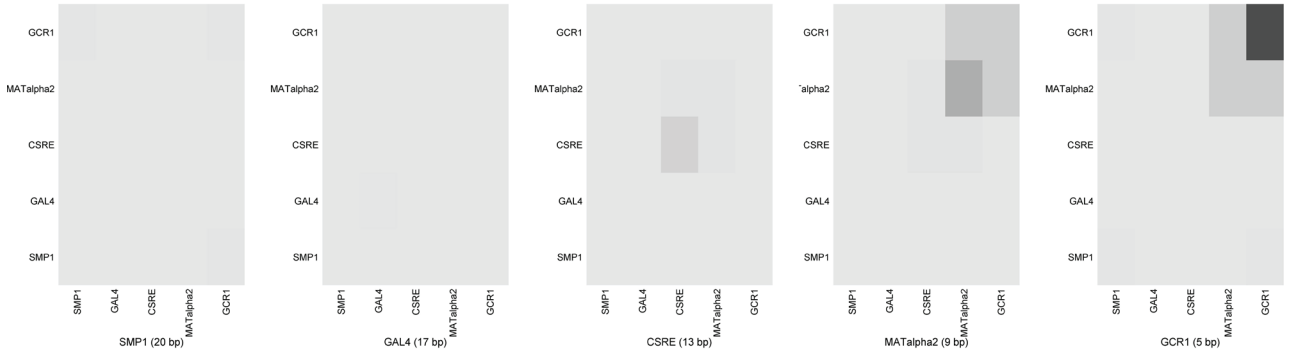

Figure 5. Heat map of log10 generation time for the formation of motif triplets constrained by a spacer motif of 100 bp. Black values correspond to a generation of 1, or a log10 generation time of 0. Light grey values correspond to log10 values of 4, corresponding to 10,000 generations.

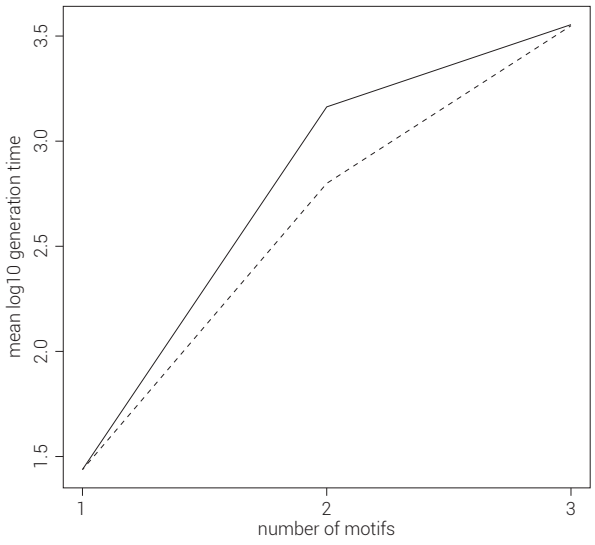

Figure 6. Mean log10 generation time for single motifs, pairs of motifs, and motif triplets with spacer constraint (solid curve), and without such constraint (dashed curve)

the motif, thereby destabilizing it or even causing a loss of function mutation, or lowering its PWM score. Thus, a random mutation can take the motif either closer to its functional, target sequence, but it can also step away from it.

Mathematically speaking, this problem is similar to the random walk of a drunken sailor. Starting from a random motif sequence, the individual positions of the motif must all consecutively mutate to the correct base. This corresponds to a random walk of the drunken sailor stumbling consecutively in one direction (towards build-up of the target sequence). Longer sets of steps going consecutively in one direction are less and less likely.

The whole goal of the PromMute simulation is to demonstrate whether target motifs with a defined sequence can arise via a series of random mutations, called a 'mutational trajectory'. Such a trajectory is depicted in figure 7, namely T,T,T,T,C,T,T,C,C,A,G,C which transforms the random motif GAGGCCTCAACA into the target motif TCCATTGTTCTC. If we have a random motif of length n bp, n/4 bp of this random motif will match the sequence of the target motif without any mutations having happened. If we select a motif cut-off of 0.9, this means that the initial random motif must follow a mutational trajectory, whereby 90% of its bases will match the sequence of the target motif. This corresponds to (0.9–0.25)n = 0.65n mutations happening in the right order for this to happen. In figure 7, this means equals 0.65 x 12 ~ 8 mutations for a motif of length 12. This is similar to the Hamming distance concept of how different two sequences are. At each step along the trajectory, a single nucleotide at a given position has a 1/3 chance of mutating to the proper base. Since the nucleotide is changing, it has to mutate to the correct bp of the remaining three possibilities other than itself, if theoretically, all possible bp mutations have the same probability. This probability is equal to $3^{-0.65n}$, which is approximately $10^{-0.325n}$. For a motif of length 12, this probability is approximately 10^{-4}. If a module of five motifs, each of length 12, is required to form a regulatory module within a promoter, the chance for this to happen at random is 10^{-20}. This probability has been defined by evolutionist Richard Dawkins[11] as low enough for an event to be practically impossible.

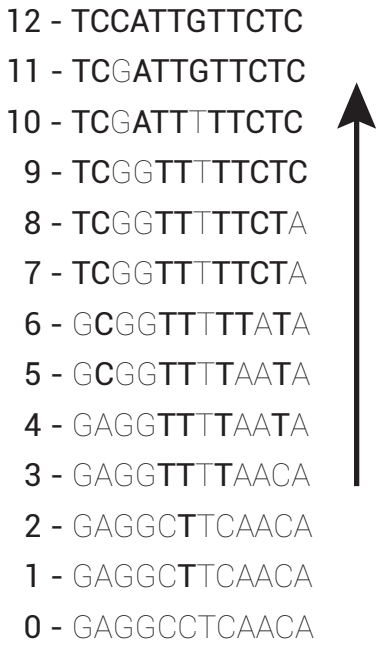

Figure 7. Formation of target motif according to drunken sailor random walk model. Here 3 bp of a random 12-mer may match the target motif sequence at the top by random chance. At a motif cut-off of 0.9, 11 – 3 = 8 bp need to match with at least 90% of the positions of the target sequence. This means that eight consecutive mutations have to occur to match the target, which has a probability of $10^{-0.325x12} \sim 10^{-4}$.

Based on previous analyses, on average, 7.41 motifs were found in each promoter with a range between 1 and 56, a median of 7, and a standard deviation of 3.38. The distribution of the frequency of motifs per proximal promoter can be seen in figure 8. As we can see, it follows a normal distribution with a slight skew to the right. Forming a proximal promoter with seven motifs is hard enough, but one with 56 motifs surely must be insurmountable for molecular evolution to achieve. However, even if this promoter contains redundant regulatory modules, the formation of seven motifs at once is approximately 10^{-28}. As a test, PromMute was run with a spacer value of 100 bp to simulate a real promoter with seven motifs: GCR1 (5bp), MCM1 (10 bp), ABF1 (12 bp), MIG1 (12 bp), CSRE (13 bp), GAL4 (17 bp), and SMP1 (20 bp) as a realistic test. Five runs showed that the TFBS complex could not form.

When studying motif pairs it was observed that the longer motifs, such as GAL4, RLM1, and SMP1, all showed log10 values of about 4

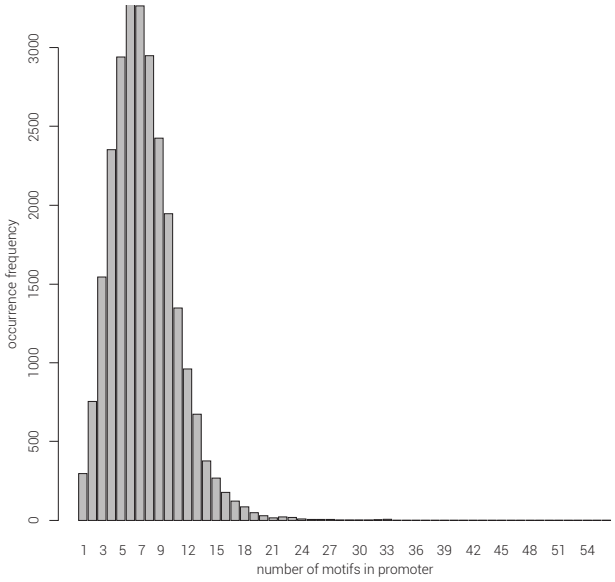

Figure 8. Occurrence frequency of given number of motifs in human proximal promoters (1 Kbp)

(10,000 generations). Since in figure 1 these same motifs also showed log10 values of four, these motifs masked out the effect that any shorter motif might have besides it. This is not surprising, since it is harder for another motif to form if in and of itself a longer motif doesn't form by random chance. Decreases in average log10 generation values after addition motifs is expected to taper off and reach a plateau. This is because it is more and more difficult for additional motifs to form by chance.

We must further note here that besides simulating a select number of target motifs in a hypothetical promoter, we still have not modelled fully realistic motifs. Whereas the formulation of triplet motifs is albeit hard, but still possibly feasible in some cases with shorter motifs, many promoters contain modules of more than three motifs. Based on calculations using 1 Kbp human promoters and PWMs, it was found that there are 7.41 motifs in a given human promoter on average. Since all human motifs are not known, this number could be even higher.

Also, we must take into consideration, that the mutation rate used in PromMute was 10^{-3}, which is three to five times higher than the average mutation rate in nature. Therefore generation times needed for the formation of motifs/modules should be multiplied by 1,000.

Tandem repeats, CpG islands, G-quadruplexes

Gene regulation does not occur exclusively by TFBS. Other regulatory elements include short tandem repeats (STR), G-quadruplexes (G4), and CpG islands (CGI), which all affect gene regulation in their own special way.

STRs are repeats of short stretches of DNA 2–5 bp long. About 17% of human genes and 25% of yeast genes have at least one STR in their regulatory elements.[12,13] STRs generally do not overlap with TFBS. For example, the number of CAG repeats in exon 1 of the SK3 calcium-activated potassium channel influences the severity of schizophrenic symptoms.[14]

G-quadruplexes are guanine-rich tracts where four or more tandem guanine tracts can fold into a four-stranded secondary structure, induced by transcriptional perturbation. G4s can hinder the translocation of helicase and thus cause transcription arrest.[15] CGIs determine the number of PolII binding sites, and promoters with CGIs are characterized by high transcriptional activity.[16]

Besides distance constraints between pairs of motifs, several regulatory motifs are constrained as to how far away they are from the TSS, based on a study of the distribution of 165 TRANSFAC motifs in 600 bp human promoters. Not only this, but motif pairs also show functional association.[17] This is because the TFs which bind to these TFBSs interact with the PolII protein during transcription.

Summary and conclusion

According to a shift in evolutionary paradigms, and also as pointed out in the creationist literature, molecular evolution has a lot to overcome in explaining how the coding regions of genes evolved, but also as to how their regulation also evolved. As two halves of a whole, both the coding region and its promoter are fundamentally necessary to form and start functioning at exactly the same time. All this is needed for hundreds of genes in order for a minimal organism to be viable.[18]

Statistically speaking, a shorter motif has a higher chance of being present in a long enough stretch of sequence. For example, there are $4^5 = 1{,}024$ different kinds of pentamers. A pentamer of any sequence is expected to occur every 1,000 bp, which is the length of the promoter region in the analysis. A 10-mer occurs once every $4^{10} = 1$ Mbp. Therefore random mutation-based evolution would have to work very hard to form such a target 10-mer out of a random stretch of sequence. If random mutations have to form a whole number of motifs, and which also must fall within a certain distance from one another, all these additional complicating circumstances make it even more impossible for molecular evolution to form a regulatory module out of nothing (especially when taking spatial constraints into consideration). If position from the TSS would also be taken into consideration as an added realistic factor, then this hurdle would increase even more.

Multiple TFBSs form a complex module in order to regulate gene expression, with all of them needed to be present at one time and in the same place. Natural selection acts upon fully formed regulatory modules as a whole,

Table 1. Motif name, sequence, and maximum PWM value of the 24 yeast regulatory motifs in the SCPD

Motif	Sequence	Motif length	Maximum PWM score
GCR1	CTTCC	5	7.711
GCN4	TGACTC	6	9.482
MCB	ACGCGT	6	9.783
PHO2	TTAAATT	7	9.542
REB1	TTACCCG	7	11.511
SCB	CACGAAA	7	11.323
TBP	TATAAAA	7	10.663
PDR1/PDR3	TCCGCGGA	8	13.546
STE12	ATGAAACC	8	11.469
MATalpha2	CATGTAATT	9	14.137
repressor_of_CAR1	AGCCGCCAA	9	12.798
MCM1	CCCAATTAGG	10	13.914
PHO4	CGCACGTGGT	10	12.144
ABF1	TCACTATACACG	12	14.804
MIG1	CCCCAGATTTTT	12	15.148
RAP1	ACACCCATACAC	12	15.91
ROX1	TCCATTGTTCTC	12	16.124
SWI5	ATATCATGCTGG	12	13.796
UASPHR	TTTTCTTCCTCG	12	14.823
XBP1	GCCTCGAGGCGA	12	15.052
CSRE	TTCGGATGAATGG	13	16.731
GAL4	CGGAGCACACTCCTCCG	17	19.931
RLM1	AGTTCTATAAATAGATTC	18	22.37
SMP1	ATGCTTCTATTTATAG-CAAC	20	24.945

and not on partial motifs. The formation of functional, information-bearing genetic elements does not happen by random genetic mutations; rather they are design elements which have been specifically created to work together to regulate the gene that they reside in.

Materials and methods

The Windows program PromMute, as described in Cserhati 2012,[8] was rewritten in perl so that a user can run it at a Linux command prompt. A list of one or more target regulatory motifs from the *Saccharomyces cerevisiae* Promoter Database (SCPD)[19] can be supplied as a parameter of the simulation. These motifs are listed in table 1.

Heat maps were created using R version 3.2.4.

1 Kbp human promoters (–1,000 to –1 bp) were downloaded from the Swiss Institute of Bioinformatics (www.isb-sib.ch/). PWMs for human TFBSs were downloaded from the JASPAR database[20] (jaspar.genereg.net/).

Perl scripts and matrixes are available at the following URL at GitHub: github.com/jeanomicks/prommute_scripts.

References

1. Cserhati, M., Creation aspects of conserved non-coding sequences, *J. Creation* **21**(2):101–108, 2007.
2. Ohno, S., *Evolution by Gene Duplication*, Springer-Verlag, Heidelberg, Germany, 1970.
3. Carroll, S.B., Endless forms: the evolution of gene regulation and morphological diversity, *Cell* **101**(6):577–580, 2000.
4. Rister, J., Razzaq, A., Boodram, P. *et al.*, Single-base pair differences in a shared motif determine differential Rhodopsin expression, *Science* **350**(6265): 1258–1261, 2015.
5. Wray, G.A., Hahn, M.W., Abouheif, E. *et al.*, The Evolution of Transcriptional Regulation in Eukaryotes, *Mol. Biol. Evol.* **20**:1377– 1419, 2003.
6. Schneider, T.D., Evolution of biological information, *Nucleic Acids Res.* **28**: 2794–2799, 2000.
7. Stone, J.R. and Wray, G.A., Rapid evolution of cis-regulatory sequences via local point mutations, *Mol. Biol. Evol.* **18**:1764–1770, 2001.
8. Cserhati, M., PromMute—A Promoter Mutation Simulation for Modeling the Evolution of Genetic Regulatory Elements, *J. Comput. Sci. Syst. Biol.* **5**: 074–080, 2012.
9. Lu, J., Luo, L. and Zhang, Y., Distance conservation of transcription regulatory motifs in human promoters, *Comput. Biol. Chem.* **32**(6):433–437, 2008.
10. Grace, M.L., Chandrasekharan, M.B., Hall, T.C. *et al.*, Sequence and spacing of TATA box elements are critical for accurate initiation from the beta-phaseolin promoter, *J. Biol. Chem.* **279**(9):8102–8110, 2004.
11. Dawkins, R., *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*, W.W. Norton, New York, 1996.
12. Bolton, K.A., Ross, J.P., Grice, D.M. *et al.*, STaRRRT: a table of short tandem repeats in regulatory regions of the human genome, *BMC Genomics* **15**;14:795, 2013.
13. Abe, H. and Gemmell, N.J., Abundance, arrangement, and function of sequence motifs in the chicken promoters, *BMC Genomics* **15**;15:900, 2014.
14. Grube, S., Gerchen, M.F., Adamcio, B. *et al.*, CAG repeat polymorphism of KCNN3 predicts SK3 channel function and cognitive performance in schizophrenia, *EMBO Mol. Med.* **3**(6):309–319, 2011.
15. Zhang, C., Liu, H.H., Zheng, K.W. *et al.*, G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix, *Nucleic Acids Res.* **41**(14):7144–7152, 2013.
16. Elango, N. and Yi, S.V., Functional relevance of CpG island length for regulation of gene expression, *Genetics* **187**(4):1077–1083, 2011.
17. Vardhanabhuti, S., Wang, J. and Hannenhalli, S., Position and distance specificity are important determinants of cis-regulatory motifs in addition to evolutionary conservation, *Nucleic Acids Res.* **35**(10):3203–3213, 2007.
18. O'Micks., J., Bacterial genome decay from a baraminological viewpoint, *J. Creation* **29**(2):110–118, 2015.
19. Zhu, J. and Zhang, M.Q., SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*, *Bioinformatics* **15**:607–611, 1999.
20. Mathelier, A., Zhao, X., Zhang, A.W. *et al.*, JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles, *Nucleic Acids Res.* **42**(Database issue):D142–147, 2014.

Jean O'Micks *has a Ph.D. in biology. He has been an active creationist for 15 years and takes a great interest in molecular biology. He has published a number of articles in* Journal of Creation.