Evolution of simple sequence repeats as mutable sites

in *Tandem Repeat Polymorphisms: Genetic Plasticity, Neural Diversity and Disease*,
edited by Anthony Hannan ©2011 Landes Bioscience and Springer Science+Business Media.

by David G. King
Department of Anatomy and Department of Zoology
Southern Illinois University Carbondale, U.S.A.
dgking@siu.edu

Summary
Because natural selection is commonly presumed to minimize mutation rates, the discovery of mutationally unstable simple sequence repeats (SSRs) in many functional genomic locations came as a surprise to many biologists. Whether such SSRs persist in spite of or because of their intrinsic mutability -- whether they constitute a genetic burden or an evolutionary boon -- remains uncertain. Two contrasting evolutionary explanations can be offered for SSR abundance. First, suppressing the inherent mutability of repetitive sequences might simply lie beyond the reach of natural selection. Alternatively, natural selection might indirectly favor SSRs at sites where particular repeat-number variants have provided positive contributions to fitness. Indirect selection could thereby shape SSRs into "tuning knobs" that facilitate evolutionary adaptation by implementing an implicit protocol of incremental adjustability. The latter possibility is consistent with deep evolutionary conservation of some SSRs, including several in genes with neurological and neurodevelopmental function.

Keywords :
adaptive function, function of SSRs, coding SSRs (SSRs in exons), conservation (of SSR function), evolutionary function, evolvability, indirect natural selection, microsatellite, minisatellite, mutation (definition. mutation rate, evolution of mutation rate, mutation of small effect, mutation pressure), noncoding SSRs, protocol (implicit protocol of incremental adjustability), SSRs (simple sequence repeats), transposable elements (TEs), tuning knob

Introduction

"No one expected that DNA sequences could be so unstable or behave as these do" (Jean-Louis Mandel, quoted in *Science*[1]). The initial discovery that triplet-repeat expansion was responsible for several neurological diseases surprised many geneticists. Perhaps even more surprising has been the subsequent discovery that repeat-number variation can also have nonpathological yet biologically significant effects (e.g., ref. 2-5). Functional consequences attend repeat-number variation in a wide diversity of simple sequence repeats (SSRs; the term encompasses both microsatellite and minisatellite DNA, i.e. tandem repetitive sequences with motifs ranging from mono-, di- and tri-nucleotides up to several tens of basepairs in length). These surprisingly unstable repetitive stretches are so profusely distributed throughout eukaryotic genomes that many genes, perhaps most, include one or more variable SSRs within regulatory and/or coding domains. The complete total remains unknown. Although most attempts to count SSRs have been restricted by motif length, number of repeats or functional domain, two recent surveys of human[6] and *Daphnia pulex*[61] have catalogued hundreds of thousands of SSRs in each genome. The distribution of particular motif classes within a genome can vary substantially among different species.

The sheer abundance of SSRs raises an intriguing question. Why has evolution permitted such prolific sources of genetic instability to become so prevalent? Or, in slightly less teleological language, how do these highly mutable genetic patterns escape elimination or suppression by natural selection? Two contrasting answers can be suggested.

One intriguing possibility is that genetic patterns which confer special modes of mutability are serving an "evolutionary function." In this view, the unexpected prevalence, diversity and high mutation rates of SSRs support a hypothesis that appropriately constrained mutability can be evolutionarily beneficial. If so, then pathological expansion of SSRs is more than just a clinical curiosity. Just as other diseases throughout history have stimulated investigation of basic biological processes, repeat-expansion pathologies may be revealing a previously unsuspected role for a ubiquitous feature of normal genetic organization.

But a more conventional and apparently more parsimonious explanation is that natural selection has but limited ability to eliminate mutation. According to a widely accepted principle of evolutionary biology, mutations of any sort occur not because variation is necessary for adaptation but simply because total suppression of mutation is not feasible. Hence SSRs' surprising instability represents nothing more than an accidental consequence of the replication slippage which inevitably accompanies sequence repetition. Each particular example of repeat-number variation, as documented throughout this volume, may be interesting in itself for its effect on a particular gene. But there should be no reason to expect such mutable sites to provide any novel insight into evolutionary processes.

This chapter discusses both possibilities, beginning with a brief historical review of the conventional argument, sharpened repeatedly over the past century, that all mutations are essentially accidents. Some inadequacies of this argument will then be considered in light of the less familiar "evolutionary function" hypothesis. While evidence for each explanation remains inconclusive, this essay will advocate the proposal that SSRs are common precisely because their particular style of mutation facilitates evolutionary adaptation and has therefore been favored, albeit indirectly, by natural selection. The distinctive properties of SSRs, which initially

appeared so surprising, accord neatly with this proposed evolutionary function. SSRs might even have a special role in behavioral evolution through their effects on genes involved in nervous system development and function.

A brief history of the "mutation" concept

Several interrelated questions have concerned geneticists for much of the past century. Which aspects of genetic variation should be defined as "mutations"? What is the fundamental nature of mutational mechanisms? Are mutation rates optimized to ensure evolutionary adaptability? Or are all mutations essentially accidental errors in DNA replication? Satisfactory answers to these questions remain elusive, although as Darwin[7] himself noted, "Some authors believe it to be as much the function of the reproductive system to produce individual differences, or very slight deviations of structure, as to make the child like its parents."

The precise meaning of "mutation" has evolved as the word itself was assimilated into the language of genetics. For Hugo de Vries,[8] one of the pioneering rediscoverers of Mendel's laws at the start of the twentieth century, a mutation was a saltational jump leading to a new species. But by 1919 Calvin Bridges[9] was applying the term more broadly, with "no restrictions of degree, covering the most extreme as well as the slightest detectable inherited variation." Bridges,[9] who worked with Thomas Hunt Morgan at Columbia University's famous *Drosophila* laboratory, also shared with many modern geneticists an intuitive understanding that deleterious mutations must vastly outnumber beneficial ones:

> "Any organism as it now exists must be regarded as a very complex physicochemical machine with delicate adjustments of part to part. Any haphazard change made in this mechanism would almost certainly result in a decrease of efficiency.... Only an extremely small proportion of mutations may be expected to improve a part or the interrelation of parts in such a way that the fitness of the whole organism for its available environments is increased."

Bridges simply presumed that mutations are "haphazard," with the extreme unlikelihood of beneficial mutations being a self-evident corollary. But by 1937, Alfred H. Sturtevant (another member of Morgan's *Drosophila* group at Columbia) had confirmed "accidental" as a defining attribute of mutation. Sturtevant[10] reasoned that selection should favor the lowering of mutation rates to reduce the loss of reproductive potential due to deleterious mutation. He then considered a possible tendency in the opposite direction based on the necessity of mutations for evolutionary adaptation:

> "It seems at first glance that there should be a counter-selection, due to the occurrence of favorable mutations. It is true that favorable mutations furnish the only basis for improvement of the race, and must be credited with being the only raw material for evolution. It would evidently be fatal for a species, in the long run, if its mutation rate fell to zero, for adjustment to changing conditions would then not long remain possible."

But Sturtevant[10] rejected this possibility:

> "While this effect may occur, it is difficult to imagine its operation. It is clear that the vast majority of mutations are unfavorable . . . [F]or every favorable mutation, the preservation of which will tend to increase the number of genes in the population that raises the mutation rate, there are hundreds of unfavorable mutations that will tend to lower it. Further, the unfavorable mutations are mostly highly unfavorable, and will be more effective in influencing the rate than will the relatively slight improvements that can be attributed to the rare favorable mutations."

Sturtevant[10] then asked, rhetorically, "why does the mutation rate not become reduced to zero?" To this critical question, he gave a famous reply: "No answer seems possible at present, other than the surmise that the nature of genes does not permit such a reduction. In short, mutations are accidents, and accidents will happen" (emphasis added).

Three decades later, in his classic 1966 text *Adaptation and Natural Selection*, evolutionary theorist George C. Williams[11] responded to what was still a frequent assertion, "that natural selection will not produce too low a mutation rate because that would reduce the evolutionary plasticity of the species," with a conclusion even stronger than Sturtevant's:

> "[N]atural selection of mutation rates has only one possible direction, that of reducing the frequency of mutation to zero. That mutations should continue to occur ... requires no special explanation. It is merely a reflection of the unquestionable principle that natural selection can often produce mechanisms of extreme precision, but never of perfection.... Evolution has probably reduced mutation rates to far below species optima, as the result of unrelenting selection for zero mutation rate in every population. Mutation is, of course, a necessary precondition to continued evolutionary change. So evolution takes place, not so much because of natural selection, but to a large degree in spite of it."

This same basic argument continues to be reiterated into our present century. For example, Sniegowski et al.[12] write:

> "[I]t can be appealing to suppose that the genomic mutation rate is adjusted to a level that best promotes adaptation. Most mutations with phenotypic effects are harmful, however, and thus there is relentless selection within populations for lower genomic mutation rates. Selection on beneficial mutations can counter this effect by favoring alleles that raise the mutation rate, but the effect of beneficial mutations on the genomic mutation rate is extremely sensitive to recombination and is unlikely to be important in sexual populations."

As Sniegowski et al.[12] explain, it is the cost of accurate DNA replication, not a need for evolutionary plasticity, that determines mutation rates:

> "The physiological cost of reducing mutation below the low level observed in most populations may be the most important factor in setting the genomic mutation rate in sexual and asexual systems, regardless of the benefits of mutation in producing new adaptive variation. Maintenance of mutation rates higher than the minimum set by this 'cost of fidelity' is likely only under special circumstances."

A recent authoritative review of mutation rate evolution (ref. 13) again echoed Sturtevant's[10] argument and reaffirmed that "the cost of fidelity is the generally accepted explanation for non-zero mutation rates in multicellular eukaryotes."

This prevailing view of mutation, as exemplified by the quotations above, has changed little over the past century in spite of a tremendous increase in our understanding of DNA metabolism with its associated diversity of mutational mechanisms. Mutations continue to be regarded as accidental errors such that the vast majority must be deleterious, albeit with some acknowledged exceptions (below).

A brief critique of mutations as accidents

One reason why subsequent authors still rehearse the essentials of Sturtevant's argument lies in the tenacity of a contrary narrative in which higher-than-minimal mutation rates really are maintained because of their past contribution to adaptive evolution. Although Sturtevant's and Williams' arguments have dominated genetics for several decades, this contrary view is resurging: "Increasing numbers of biologists are invoking 'evolvability' to explain the general significance of genomic and developmental phenomena affecting genetic variation" (ref. 14).

Early interpretations, naively attributing evolvability (also variously called "evolutionary plasticity," "evolutionary potential," etc.) to selection for the future good of the species, can be dismissed on grounds that "natural selection has no foresight" (e.g., ref. 15). But just as implications of "design" are hard to avoid when discussing the function of complex adaptive structures, so also are implications of "foresight" hard to avoid when speaking of genomic patterns that generate novel variation (e.g., ref. 16, 17). The critical issue here lies not with foresight but with the production of hereditary variation as a proper biological function, i.e., as an advantageous trait that selection has favored over preceding generations. The philosophical foundations for evolutionary theory do not require that mutations be "accidental" or "haphazard," only that they be "random" with respect to current adaptive needs. Otherwise mutation itself rather than natural selection would direct the process of adaptation (e.g., ref. 18).

The conviction that mutations are haphazard persists largely through repeated assertion in textbooks and prominent publications -- e.g., "It is common sense that most mutations that alter fitness at all will lower it" (ref. 19); "the vast majority of mutations with observable effects are deleterious" (ref. 13). Yet although many studies have measured the accumulation of deleterious mutations, there remains even now remarkably little experimental evidence regarding the proportion of mutations that increase or decrease fitness to some degree (e.g., ref. 20) and none that effectively distinguishes among different classes of mutation. In the absence of such evidence, the classic "mutations are accidents" argument becomes essentially circular: Because mutations are accidental, if they affect fitness at all they must mostly be deleterious. Because fitness-affecting mutations are mostly deleterious, selection cannot favor mutability. Because selection cannot favor mutability, mutations must occur only as accidents.

But if mutation is defined simply and broadly -- i.e., any change in inherited genetic information, with "no restrictions to degree"[9] -- then it clearly embraces the consequences of several highly organized processes that are hardly accidental. The most familiar example is meiotic recombination, whereby novel gene sequences can be created by precise reciprocal exchange between alleles that differ at more than one site. Sexual reproduction normally assures that every gamete has a unique haploid genotype, randomly generated from a vast number of viable possibilities. Yet even though any particular genotype is an unpredictable chance event, the label "accidental error" is inappropriate (except for inviable aneuploids). Although the selection pressures responsible for maintaining sex and recombination in most plant and animal populations remain controversial, most theories nevertheless recognize variation in one form or another as the principal overriding advantage (e.g., ref. 21). Reconciling these well known facts with the "mutations are accidents" argument has necessitated, as routine practice, that the products of recombination be explicitly excluded from the definition of "mutation" (e.g., ref. 22: "mutation An error in replication of a nucleotide sequence or any other alteration of the genome that is not manifested as reciprocal recombination").

Also often set apart from the "mutation" category is "programmed gene rearrangement," a source for highly structured variation used by parasitic trypanosomes to alter expression of surface antigens as they reproduce within a host, thereby facilitating evasion of the host's immune response (ref. 23). Additional strategies for the active generation of internally organized variation are known in prokaryotic organisms. For example, mutation-prone "contingency genes" (ref. 24) are recognized among microbial geneticists as having a legitimate evolutionary role, predictably generating mutations of particular types that help assure survival of some descendents even if current conditions change. Mutations produced by contingency genes are still "random" (i.e., they occur whether needed or not, and only in chance individuals), but they are no more accidental errors than are particular results from shuffling cards or rolling dice in an orderly game of chance. Such mutational mechanisms are presumably shaped by recurring shifts in selection pressure over many preceding generations. Explaining genetic patterns that have such evolutionary functions "requires a change in our attitude towards the sources of genetic variation, which until recently have largely been thought to rely on errors and accidents happening to DNA" (ref. 25).

Routinely excluding such manifestly nonaccidental sources of genetic variation from consideration as mutation has hindered recognition that the concept of heritable genetic change embraces several highly constrained mechanisms in addition to those "errors" that are patently accidental. The resulting semantic confusion is exacerbated by common reference to "the genomic mutation rate" as if this were a single parameter characterizing a well-defined unitary process. Even apart from the special exceptions above, "mutation" remains a composite concept that encompasses a number of disparate mechanisms (e.g., nucleotide substitution, replication slippage, transposable element activity, etc.). This diversity needs to be disaggregated. "Mutation rate" should never be described by a single statistic (ref. 26). Instead, each separate source of DNA sequence modification should be analyzed on its own terms.

Once the significance of several distinct sources of hereditary variation is acknowledged, the simplistic conclusion of relentless selection for lower mutation rates becomes far less compelling. Furthermore, classical analyses of mutation rate evolution (e.g., ref. 10-12) have generally assumed that particular genes determine an average genome-wide mutation rate by influencing the overall fidelity of nucleotide base-pairing. Any "mutator allele" that increases this mutation rate must reduce fitness by causing haphazard errors throughout the genome. Meanwhile, "the effect of beneficial mutations on the genomic mutation rate is extremely sensitive to recombination,"[12] and in sexually reproducing populations any fortuitous beneficial mutant would have only a small probability of close linkage to the mutator. Standing in sharp contrast to such analyses are the parameters that could allow

indirect selection to favor increased mutability -- i.e., a relatively low likelihood for deleterious fitness effects together with reliable linkage between beneficial mutant alleles and a cause for increased mutability. Remarkably, these are exactly the parameters that characterize the mutability of SSRs.

SSRs as sources of "tuning knob" variation

Although low background rates for nucleotide substitution appear consistent with Williams'[11] "unrelenting selection for zero mutation rate," rates for repeat-number mutations at SSR sites can be several orders of magnitude higher. The resulting variation in repeat number is so pervasive it can be used for DNA fingerprinting. Because phenotypic effects are seldom evident, such variation has been "generally assumed to evolve neutrally" (ref. 27). But an unqualified assumption that repeat number variants have no significant effect on evolutionary fitness can no longer be justified. Evidence has been accumulating for almost three decades that variation in repeat number can and does exert small-scale, quantitative effects on many aspects of gene function (e.g., ref. 2-5). Even if the percentage of SSRs that do influence phenotype is quite small, SSRs are so numerous that repeat-number variants must still make a substantial contribution to overall phenotypic variation.

Functional effects of repeat-number variation are not limited to rare cases of pathological expansion, nor even to SSRs that directly encode amino acid repeats. So-called "noncoding" SSRs with a variety of different motifs are also found in introns, in UTRs and in upstream and downstream regulatory regions of many genes. (The adjective "noncoding" is potentially misleading, as it typically refers to any sequence that does not directly encode peptide sequences with canonical triplet codons. As ironically noted in a recent article in a prominent journal,[28] "many functions are encoded ... in the noncoding portion of the genome.") Early on, evidence that such mutation-prone SSR sites could provide an abundant supply of small-scale quantitative genetic variation led to speculation that these sites function as "evolutionary tuning knobs" (ref. 29, 30). SSRs would thus embody an "implicit protocol" (cf. ref. 31) for incremental adjustability.

In fact, classical evolutionary theory has long held that "mutations of small effect" can improve fitness with a probability approaching fifty percent (ref. 32), especially in natural conditions where selection pressures vary over space and time. And as early as the 1960s Levins[33, 34] had demonstrated how changing or heterogenous environments can lead to increased mutation rates. But prior to discovery of SSRs, hardly anyone had imagined how unstable DNA sequences such as SSRs could evade Bridges'[9] intuitive expectation for a very low proportion of beneficial mutations. SSRs demonstrate how readily a simple mechanism can preferentially yield mutations whose characteristically small effect on phenotype could carry a non-negligible probability of selective advantage as well as a low probably for harm. Although any newly-arising allele of small effect, even if beneficial, can be readily lost through genetic drift before weak selection can increase its prevalence in a population (e.g., ref. 35), nevertheless high rates of repeat-number mutation guarantee a continued resupply of such alleles.

Thus the presumption that deleterious mutations must vastly outnumber beneficial ones has become quite doubtful for repeat-number variants at SSR sites. Even at those sites associated with repeat expansion diseases, pathological expansion arises only from rare "premutation" alleles at one extreme of a normal, nonpathogenic range. Most variation falls within this relatively safe range, but if selection should favor a shift in repeat number then the high mutation rate of SSRs assures that appropriate new variation will be quickly forthcoming.

Empirical evidence for nonpathological phenotypic effects of naturally occurring SSR variation, especially for noncoding SSRs, remains quite limited relative to the multitude of SSRs found throughout most eukaryotic genomes. Nevertheless, several cases already include circumstantial evidence that SSR variants have supported adaptive differentiation among natural populations (ref. 36; more recently ref. 37-39). Vinces et al.[40] have provided the strongest experimental evidence to date that SSRs can indeed serve an evolutionary function in eukaryotes, reporting not only functional effects of repeat-number variation within promotor regions but also establishing that this variation could be the basis for evolutionary adaptation in laboratory populations of yeast (*Saccharomyces cerevisiae*).

Not only should the characteristic mutability of SSRs carry a fair probability for adaptive advantage, this mutability is also inextricably associated with sequence repetition at each individual SSR site. At any SSR, each repeat-number allele retains the inherent site-specific mutability by which it arose. Thus selection favoring any advantageous repeat-number variant also favors the site's potential for incremental adjustability. Since recombination cannot separate cause (i.e., the site-specific adjustability protocol) from consequence (particular alleles), this intrinsic linkage should strongly dispose SSRs toward indirect site-by-site selection for adaptively appropriate mutation rates (ref. 36, 41-44).

Selective shaping of local mutation rates requires some hereditary variation in those rates. In the case of SSRs, the rate for repeat-number mutation can be lower or higher depending on the presence or absence of interruptions or imperfect repeats (e.g., ref. 45). Once mutation rate variation exists at a particular site, indirect selection acts through direct selection upon individual repeat-number mutant alleles. Indirect selection against a higher rate is relatively inefficient, since any copy of an allele with a higher mutation rate can be eliminated only after that copy gives rise to a deleterious repeat-number mutant. In contrast, indirect selection favoring a high-mutation-rate allele can be much more effective. Once a beneficial repeat-number mutation appears at a single copy of such an allele, direct selection that increases the frequency of the beneficial mutant necessarily increases the frequency of the high mutation rate as well, since this mutation rate is retained by each copy of the beneficial mutant. Thus, as long as any beneficial variants appear within a population before every copy of a high-mutation-rate site is eliminated by direct selection against a long series of deleterious mutants, the higher rate (i.e., the "tuning knob" protocol of incremental adjustability) will prevail at that site.

Origin and maintenance of SSRs

As long as all SSRs were believed to lie in nonfunctional intergenic domains, neither their mutability nor their abundance posed any special theoretical difficulty. Since their prevalence in functional domains has become more widely appreciated, however, simply dismissing them as meaningless genetic junk is no longer adequate. A satisfactory explanation should answer two separate questions. First, by what mechanisms do SSRs originate de novo? Second, once any particular SSR has arisen, how is its presence maintained over time?

Explanations for the origins of SSRs remain, at best, incomplete (cf. ref. 45). Minisatellite SSRs require some mechanism for initial duplication of a lengthy motif. In contrast, microsatellite SSRs with their shorter motifs can arise easily by chance nucleotide substitution in previously nonrepetitive DNA. In contrast, microsatellite SSRs with their shorter motifs can arise either by chance nucleotide substitution in previously nonrepetitive DNA or by short insertions that duplicate adjacent sequence.[62] Microsatellite SSRs of several different motifs can also be created in abundance through the action of transposable elements (TEs). Moreover, microsatellite SSRs can in turn promote the activity of TEs (for examples, see ref. 42, 45). This association suggests the intriguing possibility of a synergistic coevolution between SSRs and TEs, especially since TEs have also been proposed as major contributors to evolvability (e.g., ref. 46, 47).

Regardless of how SSRs originate, a complete explanation for their abundance in functional domains should consider not only the speculative evolutionary "tuning knob" function (above) but also two alternative hypotheses of "adaptive function" and "mutation pressure."

Adaptive function. If sequence repetition is necessary for some essential adaptive function, then natural selection might retain SSRs in spite of the intrinsic mutability that attends sequence repetition. For SSRs in each of several motif classes and functional domains, this hypothesis requires that sequence repetition offers sufficient immediate functional advantage to offset the presumed liability of frequent mutations. Yet it is far from obvious why sequences with greater stability could not function equally well at most SSR loci. For example, a repeating amino acid stretch can be encoded by a DNA sequence in which codon repetition is interrupted by alternative codon usage, thereby reducing the propensity toward replication slippage. Just such stabilizing interruptions are indeed found in some sequences that encode amino acid repeats. Presumably an immediate adaptive role for any other SSR class could also be served by functionally equivalent but nonrepetitive sequences. Thus, although this "adaptive function" hypothesis may apply in special cases, it seems doubtful that all roles occupied by SSRs require essential sequence repetition. The only function that is plainly shared by all SSRs is that of mutability itself.

Mutation pressure. SSRs might also be self-perpetuating through their own intrinsic mutability, if this mutation pressure were sufficient to resist spontaneous degradation of sequence repetition by nucleotide substitition (ref. 45). By this hypothesis, once a repetitive sequence exceeds a threshhold number of repeats at which replication slippage becomes frequent, then repeat expansion can reverse any reduction in repeat number. (The threshhold length for replication slippage remains inadequately characterized for most motif classes [ref. 45, but see ref. 48, 61].) At the same time, sequential bouts of expansion and contraction can purge mutations that would otherwise interrupt motif repetition. Thus replication slippage alone might explain the persistence of "junk" SSRs in nonfunctional intergenic regions. Nevertheless, its adequacy for explaining SSRs in functional domains remains open to question. For this "mutation pressure" hypothesis by itself to explain the persistence of functional SSRs, one must presume for a wide range of distinct SSRs with differing motifs and functional roles that selection pressure against mutability at each SSR site is too weak to overcome the mutation pressure. This hypothesis also requires a tacit assumption that cost-effective molecular mechanisms for suppressing replication slippage have proven altogether inaccessible to the evolutionary process.

Evolutionary function. Unlike either the "adaptive function" or the "mutation pressure" hypothesis on its own, the "evolutionary function" hypothesis imposes no requirement for overcoming the putative cost of deleterious accidental mutations. Indeed, the "evolutionary function" hypothesis proposes that SSRs persist in functional domains because of their advantageous mutations rather than in spite of deleterious ones. This hypothesis is supported primarily by a close correspondence between the peculiar properties of SSRs and the special conditions needed to sustain indirect selection of mutability (see "SSRs as sources of 'tuning knob' variation", above). By this hypothesis, SSRs are selected for their "tuning knob" role as efficient suppliers of potentially adaptive variation, providing an abundant and practically inexhaustible supply of reversible, quantitative variation that can facilitate evolutionary adaptation (ref. 36, 41-44).

Nevertheless, even though the "evolutionary function" hypothesis directly contradicts the conventional view that "mutations are accidents," it remains compatible with both "adaptive function" and "mutation pressure" hypotheses. If sequence repetition should be directly advantageous for any particular SSR, then any indirect benefit from incremental adjustability would simply reinforce direct selection. And if mutation pressure can promote the abundance of nonfunctional SSRs, it should also assist the evolutionary function of SSRs by maintaining these sequences through periods when they are not yielding beneficial variants. But without an evolutionary function for SSRs, both of these more conventional hypotheses lack persuasive power for SSRs in functional domains. Thus acknowledging an "evolutionary function" for SSRs may create a more robust explanation for the prevalence of SSRs across their full range of motif classes and genomic locations.

Even though many questions regarding SSR origins and maintenance remain to be addressed by future research, a scenario such as the following may be readily imagined based on the above considerations. Once an SSR appears at a particular site (by whatever mechanism), repeat-number variation will begin to accumulate. If the SSR resides in a truly nonfunctional region of the genome, the ensuing variation should have no impact on fitness. The SSR may then shrink or grow at the whim of replication slippage and genetic drift, perhaps thereby maintaining itself over an indefinite number of generations (cf. ref. 45) while incidentally preserving its low-risk potential for some future contribution to adaptation. If an SSR initially emerges at a site where its sequence fits into a pre-existing functional role,

or if a novel role becomes established in its region of influence, then repeat-number variation will inevitably have some effect on that role. If that variation happens to be consistently deleterious, selection will favor mutations that suppress replication slippage by shortening or interrupting the repeat, eventually eliminating sequence repetition at the site. But if incremental adjustability is at least occasionally advantageous, then indirect selection will preserve the beneficial variants and with them the site-specific mutational mechanism by which they arose. The mere presence of a variable SSR at any particular functional location would then imply that current or recent adaptation had exploited repeat-number variation at that site.

Evolutionary conservation of SSRs

Buschiazzo and Gemmel[6] have recently analyzed microsatellite SSRs in alignments of the human genome with genomes of 16 other vertebrate species. They report that the extent of SSR conservation between human and other species declines exponentially with increasing phylogenetic distance, paralleling the declining proportion of alignable genome sequence. This is consistent with prior observation that SSRs at particular locations are often not shared among related species (e.g., ref. 49). But Buschiazzo and Gemmel[6] also report a surprisingly high level of conservation over deep evolutionary time. Almost 200,000 microsatellite SSRs are shared between human and at least one non-primate species, with over 10,000 conserved between human and opossum (*Monodelphus domestica*). Chicken (*Gallus gallus*), frog (*Xenopus tropicalis*), zebrafish (*Danio rerio*) and pufferfish (*Tetraodon nigroviridis*) each share with human over 1000 microsatellites. These latter numbers represent SSRs enduring for several hundred million years.

Unfortunately, little can be safely inferred from sequence conservation alone, without additional information or assumptions. Long term sequence conservation is commonly taken as evidence of an important sequence function, at least for protein-coding sequences. However the reliability of such inference for SSRs, especially for noncoding SSRs, should not be presumed without further analysis. Exponential decline in the number of SSRs conserved over time since evolutionary divergence could be consistent with either of two quite different interpretations.

One possibility is that SSRs persist even without any enduring function, simply because replication slippage can plausibly shield an SSR (to an unknown extent) from routine degradation by nucleotide substitution and genetic drift (see "Mutation pressure", under "Origin and maintenance of SSRs", above). They would arise spontaneously and then eventually disappear, with a "life cycle" whose duration or "half life" is determined by the distribution over time of competing types of mutations (ref. 45). In this case, exponential decline in SSR conservation, including the appearance of exceptional conservation for a few SSRs, could be nothing more than a simple statistical expectation of random decay across a very large array of SSRs. In other words, the phylogenetic lability of SSRs among related species might simply reinforce the conventional assumption (e.g., ref. 27) that these sequences "evolve neutrally" under the influence of their intrinsic mutability.

On the other hand, this same pattern of exponential decline in SSR conservation might obtain because specific selection pressures vary extensively across a phylogeny. If SSRs are preserved by indirect selection (see "Evolutionary function", under "Origin and maintenance of SSRs", above), then the observed decline in proportion of conserved SSRs could result from a decreasing fraction of adjustable sites that are shared over time by diverging species. After all, increasing phylogenetic distance between species is commonly accompanied by increasing divergence of adaptive traits, which must be accomplished through patterns of sequence divergence that remain largely unexplored. In this case, exceptional sequence conservation would indeed indicate an important function. But lack of conservation would not necessarily indicate any absence of function. An SSR could serve a temporarily important function during a particular episode of adaptation, only to be superseded by other SSRs as adaptive divergence continued.

Buschiazzo and Gemmel[6] also reported that the extent of conservation declines more rapidly for noncoding SSRs than for those located in exons. If SSRs do serve an evolutionary function, then this observation suggests that noncoding SSRs experience weaker functional constraint and hence may be serving more labile "tuning knob" roles. This in turn appears consistent with current understanding that much adaptive evolution, at least at the level of morphology and behavior, occurs through changes in regulatory sequences where many noncoding SSRs are found. Deep conservation of any particular nonexonic SSR would then suggest an especially persistent locus for regulatory adjustment. Unfortunately, in such domains we currently have little basis for predicting either the type or the degree of sequence divergence that accompanies adaptive divergence.

A peculiar style of SSR conservation, shared by a small set of 22 human genes, was recently discovered by Riley and Krieger.[50, 51] The transcript for each of these genes includes in its untranslated region a long uninterrupted dinucleotide SSR whose upstream flanking sequence is highly conserved between human and opossum (*Monodelphis domestica*). Alignments of these genes with homologues in 17 nonhuman vertebrate species revealed that the human dinucleotide SSRs were frequently replaced by other SSRs with alternative motifs. Thus these sites reveal an evolutionary history during which each site's character as an SSR has been retained even while its specific sequence has been extensively remodelled. Something more than simple mutation pressure has evidently constrained evolution at these sites, to retain a basic SSR pattern in spite of mutational churning sufficient to transform the sites. A constraint based on immediate adaptive function should be expected to minimize the extent of sequence remodelling, while mutation pressure sufficient to remodel the site should not, by itself, readily recreate a different SSR. Thus an evolutionary function that constrains the site as an SSR while exploiting the mutational flexibility of simple sequence repetition appears especially plausible for these sites.

Apart from these few intriguing examples, patterns of SSR conservation in both coding and noncoding domains remain poorly characterized. Future research that associates conserved SSRs with particular functional domains and gene ontologies may help discriminate among alternative explanations for their abundance.

SSRs with neurological significance

For the purpose of this volume, the relationship between SSR variation and nervous system function has special relevance. An intuitive expectation that the evolution of adaptive behavior must require exquisite adjustment of innumerable parameters of neuronal anatomy and physiology suggests, to this writer at least, the possibility that incremental adjustability supplied by SSRs may play a special role in nervous system evolution. Several observations appear consistent with such a possibility.

First of all is the remarkable predominance of neurological disorders in the list of human repeat expansion pathologies (e.g., ref. 52, other chapters in this volume). Whether such a functional bias is meaningful or just a statistical fluke is not yet clear. Nevertheless, at least until evidence comes to light that other systems are equally prone to pathological triplet repeat expansions, one might entertain an ad hoc speculation that relatively recent and rapid human evolution not only has utilized triplet repeats in many genes with neurological function but also has pushed several of these to the limit of their functional capacity, dangerously near the edge of the "premutation" repeat-number range where further expansion can become progressively pathological.

Prompted by early reports of an association between triplet repeats and neurological disorders, many laboratories began seeking additional examples of trinucleotide repeats in protein-coding domains. As early as 1994, Gerber et al.[3] reported homopeptide stretches (which can be encoded by either perfect or imperfect triplet DNA repeats) in several transcription factors. A few years later, a search by Karlin and Burge[53] for proteins containing multiple homopeptide stretches found a preponderance of developmental proteins, including many involved in nervous system development. Huntley et al.[54] confirmed that SSRs are overrepresented in developmental proteins but also found that apart from some polyhistidine sequences SSRs are not especially enriched in genes expressed in brain and nervous system. Łabaj et al.[55] recently reported that polyleucine is overrepresented in signal peptides, transient regions soon cleaved and degraded from growing protein chains. Phenotypic effects of repeat variation at most such sites have yet to be demonstrated. Nevertheless, a few recent studies have emphasized the possibility that such SSRs play evolutionary roles involving a variety of regulatory mechanisms. For example, Huntley and Clark,[56] analyzing amino acid repeats in 12 species of the fly genus *Drosophila*, found such sequences to be especially common in genes encoding developmental, signaling and regulatory factors. They also report that "the presence of repeats is associated with an increase in evolutionary rate upon the entire sequence in which they are embedded." In a different fly species (*Teleopsis dalmanni,* "stalk-eyed" flies with bizarre head shape), Birge et al.[57] found that genes encoding glutamine repeats were overrepresented among genes expressed in developing head, including nervous tissues, with several of these genes showing repeat-number variation that was correlated with variation in head shape.

Coding SSRs with minisatellite motifs (i.e., motif sequences longer than six basepairs) have also been implicated in nervous system evolution. Tompa[58] reports that evolution by SSR expansion has shaped a number of intrinsically unstructured proteins, including at least four with known neurological function: neural zinc finger factor-1 (with a repeating motif of 44 basepairs), neuromodulin bt (an 11 basepair repeat), neurofilament-H (a hexanucleotide repeat, at the upper end of the microsatellite range) and prion protein (an octanucleotide repeat). Tompa[58] concludes, "these repeat regions carry important functions and, thus, their inherent genetic instability and the structurally/functionally permissive nature of unstructured proteins provide a unique combination for rapid and advantageous evolutionary changes." Tyedmers et al.[59] have independently hypothesized the prion PSI^{+} "as a capacitor to promote evolvability," because of its ability to reveal cryptic genetic variation (at least in yeast) and thus promote survival in fluctuating environments.

As the sample above indicates, most studies associating SSRs with gene functions have concentrated on those, especially triplet repeats, that occur in exons. Although many genes also contain SSRs in noncoding domains, ontologies for such genes remain poorly characterized. One exception is the set of human genes found by Riley and Krieger[50, 51] to contain transcribed but untranslated dinucleotide SSRs flanked by deeply conserved sequences (above). Of these 22 genes, 19 have known functions. Remarkably, all but one of these 19 genes have critical roles in the embryonic nervous system. Thus this newly described genomic pattern, possessing both a highly conserved feature (SSRs in transcribed but untranslated regions) and a highly variable feature (motif patterns in the SSR sites), appears essential for several neurodevelopmental functions that evidently entail repeated evolutionary remodelling of the included SSR.

This miscellany of observations (also see ref. 52 as well as other chapters in this volume) implicates a wide variety of mutationally variable SSRs in neurological as well as other functions. But apart from the evident role of repeat expansion in several human neurological disorders, definitive conclusions regarding a special or widespread role for SSRs in behavioral evolution remain tantalizingly out of reach. Nevertheless such observations are surely sufficient to warrant some attention to the possibility of an evolutionary "tuning knob" role for any SSR that is found anywhere near a gene with any neurological function.

Conclusion

George C. Williams, who argued so strongly in 1966, that "natural selection of mutation rates has only one possible direction, that of reducing the frequency of mutation to zero,"[11] also admitted in the same volume that "our current picture of evolutionary adaptation is, at best, oversimplified and naive."[60] SSRs, by exemplifying how high mutation rates may prevail when the probability of deleterious variation is sufficiently low and beneficial mutants are directly linked to the mutational mechanism, may thus guide our understanding of mutation beyond Sturtevant's[10] dismissive dictum that "accidents will happen."

At the very least, it has become evident that mutationally unstable SSRs can have important biological functions. The characteristic properties, abundant distribution and phylogenetic conservation of SSRs are consistent with multiple roles for these surprisingly mutable sequences, including the production of potentially advantageous variation. A complete explanation for SSRs will surely include an evolutionary role, with their mutability shaped by indirect selection to provide an implicit "tuning knob"

protocol of incremental adjustability. If so, then we should be attentive to the possibility that repeat-number variation is influencing the function of practically any gene, including most especially those that guide the development and function of nervous tissue.

References

1 Morell V. The puzzle of the triplet repeats. Science 1993; 260:1422-1423.

2 Hamada H, Seidman M, Howard BH et al. Enhanced gene expression by the poly(dT-dG) · poly(dC-dA) sequence. Mol Cellular Biol 1984; 4:2622-2630.

3 Gerber H-P, Seipel K, Georgiev O et al. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 1994; 263:808-811.

4 Kashi Y, King DG, Soller M. Simple sequence repeats as a source of quantitative genetic variation. Trends Genet 1997; 13:74-78.

5 Li Y-C, Korol AB, Fahima T et al. Microsatellites within genes: Structure, function, and evolution. Mol Biol Evol 2004; 21:991-1007.

6 Buschiazzo E, Gemmel NJ. Conservation of Human Microsatellites across 450 Million Years of Evolution. Genome Biol Evol 2010; 2:153–165.

7 Darwin CR. On the Origin of Species by Means of Natural Selection. London: John Murray, 1859:131 (Facsimile edition by Harvard University Press, Cambridge, Massachusetts, 1964).

8 de Vries H. The origin of species by mutation. Science 1902; 15:721-729.

9 Bridges CB. Specific modifiers of eosin eye color in *Drosophila melanogaster*. J Exp Zool 1919; 28:337-384.

10 Sturtevant AH. Essays on evolution. I. On the effects of selection on mutation rate. Q Rev Biol 1937; 12:464-467.

11 Williams GC. Adaptation and Natural Selection. Princeton: Princeton University Press, 1966:139-141,270.

12 Sniegowski PD, Gerrish PJ, Johnson T et al. The evolution of mutation rates: separating causes from consequences. BioEssays 2000; 22:1057-1066.

13 Baer CF, Miyamoto MM, Denver DR. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nature Rev Genet 2007; 8:619-631.

14 Sniegowski PD, Murphy HA. Evolvability. Current Biology 2006; 16:R831-R834.

15 Ayala FJ. One hundred fifty years without Darwin are enough! Genome Res 2009; 19:693-699.

16 Darlington CD. A diagram of evolution. Nature 1978; 276:447-452.

17 Lennox JG. Teleology. In: Keller EF, Lloyd EA, eds. Keywords in Evolutionary Biology. Cambridge, Massachusetts: Harvard University Press, 1992:324-333.

18 Gould SJ. Darwinism and the expansion of evolutionary theory. Science 1982; 216:380-387.

19 Maynard Smith J. Evolutionary Genetics. Oxford: Oxford University Press, 1989:55.

20 Shaw RG, Shaw FH, Geyer C. What fraction of mutations reduces fitness? Evolution 2003; 57:686-689.

21 Hadany L, Comeron JM. Why are sex and recombination so common? Ann NY Acad Sci 2008; 1133:26-43.

22 Futuyma DJ. Evolutionary Biology, 3rd ed. Sunderland, Massachusetts: Sinauer Associates, 1998: 769.

23 Barry JD. Implicit information in eukaryotic pathogens as the basis of antigenic variation. In: Caporale LH, ed. The Implicit Genome. Oxford: Oxford University Press, 2006:91-106.

24 Bayliss CD, Moxon ER. Repeats and variation in pathogen selection. In: Caporale LH, ed. The Implicit Genome. Oxford: Oxford University Press, 2006:54-76.

25 Arber W. Gene products with evolutionary functions. Proteomics 2005; 5:2280-2284.

26 King DG, Kashi Y. Mutation rate variation in eukaryotes: evolutionary implications of site-specific mechanisms. Nature Rev Genet 2007; 8. doi:10.1038/nrg2158-c1.

27 Ellegren H. Microsatellites: Simple sequences with complex evolution. Nat Rev Genet 2004; 5:435-445.

28 Parker SCJ, Hansen L, Abaan HO et al. Local DNA topography correlates with functional noncoding regions of the human genome. Science 2009; 324:389-392.

29 Trifonov EN. The multiple codes of nucleotide sequences. Bull Math Biol 1989; 51:417–432.

30 King DG, Soller M, Kashi Y. Evolutionary tuning knobs. Endeavour 1997; 21:36-40.

31 Doyle J, Csete M, Caporale L. An engineering perspective: The implicit protocols. In: Caporale LH, ed. The Implicit Genome. Oxford: Oxford University Press, 2006:294-298.

32 Fisher RA. The Genetical Theory of Natural Selection. Oxford: Oxford University Press, 1930.

33 Levins R. Theory of fitness in a heterogeneous environment. VI. The adaptive significance of mutation. Genetics 1967; 56:163-178.

34 Levins R. Evolution in Changing Environments: Some Theoretical Explorations. Princeton: Princeton University Press, 1968.

35 Orr HA. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. Evolution 1998; 52:935-949.

36 Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. Trends Genet 2006; 22:253-259.

37 Sawyer LA, Sandrelli F, Pasetto C et al. The *period* gene Thr-Gly polymorphism in Australian and African *Drosophila melanogaster* populations: Implications for selection. Genetics 2006; 174:465-480.

38 Lindqvist C, Laakkonen L, Albert VA. Polyglutamine variation in a flowering time protein correlates with island age in a Hawaiian plant radiation. BMC Evol Biol 2007; 7:105. doi:10.1186/1471-2148-7-105.

39 Johnsen A, Fidler AE, Kuhn S et al. Avian Clock gene polymorphism: evidence for a latitudinal cline in allele frequencies. Mol Ecol 2007; 16:4867–4880.

40 Vinces MD, Legendre M, Caldara M et al. Unstable tandem repeats in promoters confer transcriptional evolvability. Science 2009; 324:1213-1216.

41 King DG, Soller M. Variation and fidelity: The evolution of simple sequence repeats as functional elements in adjustable genes. In: Wasser SP, ed. Evolutionary Theory and Processes: Modern Perspectives. Dordrecht: Kluwer Academic Publishers, 1999:65-82.

42 King DG, Trifonov EN, Kashi Y. Tuning knobs in the genome: Evolution of simple sequence repeats by indirect selection. In: Caporale LH, ed. The Implicit Genome. Oxford: Oxford University Press, 2006:77-90.

43 Kashi Y, King DG. Has simple sequence repeat mutability been selected to facilitate evolution? Isr J Ecol Evol 2006; 52:331-342.

44 King DG, Kashi Y. Indirect selection for mutability. Heredity 2007; 99:123-124.

45 Buschiazzo E, Gemmel NJ. The rise, fall and renaissance of microsatellites in eukaryotic genomes. BioEssays 2006; 28:1040–1050.

46 Jurka J, Kapitonov VV, Kohany O et al. Repetitive sequences in complex genomes: Structure and evolution. Annu Rev Genomics Hum Genet 2007; 8:241-259.

47 Oliver KR, Green WK. Transposable elements: powerful facilitators of evolution. BioEssays 2009; 31:703–714.

48 Kelkar YD, Strubczewski N, Hile SE et al. What is a microsatellite: A computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats. Genome Biol Evol 2010; 2:620–635.

49 Barbará T, Palma-Silva C, Paggi GM et al. Cross-species transfer of nuclear microsatellite markers: potential and limitations. Molec Ecol 2007; 16:3759–3767.

50 Riley DE, Krieger JN. Embryonic nervous system genes predominate in searches for dinucleotide simple sequence repeats flanked by conserved sequences. Gene 2009; 429:74-79.

51 Riley DE, Krieger JN. UTR dinucleotide simple sequence repeat evolution exhibits recurring patterns including regulatory sequence motif replacements. Gene 2009; 429:80-86.

52 Fondon JW, Hammock EAD, Hannan A, King DG. Simple sequence repeats: Genetic modulators of brain function and behavior. Trends Neurosci 2008; 31:328-334.

53 Karlin S, Burge C. Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. Proc Natl Acad Sci USA 1996; 93:1560–1565.

54 Huntley MA, Mahmood S, Golding BG. Simple sequence in brain and nervous system specific proteins. Genome 2005; 48:291–301.

55 Łabaj PP, Leparc GG, Bardet AF et al. Single amino acid repeats in signal peptides. FEBS Journal 2010; 277:3147–3157.

56 Huntley MA, Clark AG. Evolutionary analysis of amino acid repeats across the genomes of 12 *Drosophila* species. Mol Biol Evol 2007; 24:2598–2609.

57 Birge LM, Pitts ML, Baker RH et al. Length polymorphism and head shape association among genes with polyglutamine repeats in the stalk-eyed fly, *Teleopsis dalmanni*. BMC Evol Biol 2010; 10:227. doi:10.1186/1471-2148-10-227.

58 Tompa P. Intrinsically unstructured proteins evolve by repeat expansion. BioEssays 2003; 25:847–855.

59 Tyedmers J, Madariaga ML, Lindquist S. Prion switching in response to environmental stress. PLoS Biol 2008; 6(11): e294. doi:10.1371/journal.pbio.0060294.

60 Williams GC. Adaptation and Natural Selection. Princeton: Princeton University Press, 1966:270.

References added at proof stage:

61 Sung W, Tucker A, Bergeron DR et al. Simple sequence repeat variation in the *Daphnia pulex* genome. BMC Genomics 2010; 11:691. doi:10.1186/1471-2164-11-691.

62 Zhu Y, Strassmann JE, Queller DC. Insertions, substitutions, and the origin of microsatellites. Genet Res Camb 2000; 76:227-236.