On the Possibility of Constructive Neutral Evolution

Arlin Stoltzfus*

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, B3H 4H7 Canada

Abstract. The neutral theory often is presented as a theory of "noise" or silent changes at an isolated "molecular level," relevant to marking the steady pace of divergence, but not to the origin of biological structure, function, or complexity. Nevertheless, precisely these issues can be addressed in neutral models, such as those elaborated here with regard to scrambled ciliate genes, gRNA-mediated RNA editing, the transition from self-splicing to spliceosomal splicing, and the retention of duplicate genes. All of these are instances of a more general scheme of "constructive neutral evolution" that invokes biased variation, epistatic interactions, and excess capacities to account for a complex series of steps giving rise to novel structures or operations. The directional and constructive outcomes of these models are due not to neutral allele fixations per se, but to these other factors. Neutral models of this type may help to clarify the poorly understood role of nonselective factors in evolutionary innovation and directionality.

Key words: Neutral evolution — Scrambling — RNA editing — Spliceosomal introns — Gene duplication — Complexity

Introduction

Neutral evolution is often seen as a conservative or "silent" process. Although this may be the most common

---
*Present address: CARB, 9600 Gudelsky Drive, Rockville, MD 20850, USA
E-mail: arlin@carb.nist.gov

mode of neutral evolution, it is not the only conceivable mode. Neutral models may be applied to what is here called "constructive" evolution and what is elsewhere referred to as "innovation" or the evolution of "novelty" or sometimes "new functions." Neutral models of constructive evolution already exist, the most relevant example being the "codon capture" phase of the Osawa–Jukes model for changes in a genetic code (Osawa et al. 1992). In the preliminary "codon loss" phase of this model, the influence of a mutation bias results in the gradual disappearance of one type of codon, allowing subsequent loss of the corresponding tRNA by random fixation of a null mutant [see Osawa et al. (1992) for possible biological precedents]. If the mutation bias is later relaxed, the stage is set for "codon capture" and a new genetic code: the lost codon may arise (by mutation in some gene), but such events will produce only deleterious alleles until such time as an appropriate tRNA (one that reads the lost codon) appears. Because spontaneous duplications of tRNA genes and mutation that change tRNA specificities occur naturally (Osawa et al. 1992; Robeson et al. 1980), such a tRNA is expected to appear given sufficient time. When this happens, instances of the lost codon may accumulate (by mutation and random genetic drift) and the new genetic code may become entrenched.

Covello and Gray (1993) have sketched a neutral model for the evolution of RNA pan-editing that is in some ways similar to "codon capture" and that is elaborated further below. In both the Covello–Gray model and the codon capture phase of the Osawa–Jukes model, the outcome of a series of neutral changes is an increase in the number of parts, operations, or interactions (among

parts) contributing operationally to fitness. Both models can be understood as instances of a more general model of constructive neutral evolution invoking (i) the presence of excess capacity in biological systems (e.g., a gratuitous duplicate tRNA gene), (ii) biases in the production of variants (e.g., mutation biases leading to global changes in codon frequencies), and (iii) a compounding of selective constraints due to epistatic interactions with neutrally evolving sites (e.g., constraints against loss of a previously gratuitous tRNA due to instances of a previously lost codon).

The terms "constructive" and "neutral" require some clarification, since the former term is unfamiliar in an evolutionary context, and the latter, though familiar, is a well-known source of confusion. First, the term "constructive" is not meant in the vernacular sense of "sympathetic" or "positive" but is intended merely as a descriptive term, useful in conjunction with "reductive" and "conservative" to refer to increase, decrease, and lack of significant change (respectively) in the complexity of features that contribute operationally to organismal fitness. For instance, in the Osawa–Jukes model for genetic code changes, the codon loss phase (loss of a codon and iso-accepting tRNA) is reductive, the codon capture phase (gain of a tRNA and reappearance of the codon) is constructive, and the overall codon reassignment is conservative (to the extent that the two codes are simply arbitrarily different ways to map 21 meanings onto 64 words). In precise terms, an attribute can be said to "contribute operationally to fitness" to the extent that its removal or diminution would reduce fitness. Attributes with this property are usually said to have a "function" or to be "functional," and when "function" is used in this sense, "contructive evolution" is synonymous with "evolution of new functions" or "increase in functional complexity."

Second, misunderstandings of "neutrality" are the basis for a great variety of erroneous reports of the death or collapse of the neutral theory of molecular evolution (e.g., Berry 1982, p 30). Although the word "neutral" may in other contexts indicate a null model or a "random" model, in the context of the neutral theory of molecular evolution (Kimura 1983), neutrality is based on the susceptibility of alleles to fixation by random genetic drift: two alleles are neutral ("effectively neutral") by comparison if their selection coefficients differ by much less than the reciprocal of the effective population size (Kimura 1983, p 44; Li 1997, p 55). Somewhat larger fitness effects are allowed for the fixation of very slightly deleterious ("nearly neutral") alleles (Ohta 1996). The necessity of comparing selection coefficients and population size in this manner has been understood by population geneticists for 70 years (e.g., Haldane 1932).

Nevertheless, various maverick notions of "neutrality" that depart from this definition persist: the notion that there is an absolute definition of neutrality (i.e., fitness difference of zero) and that any departure from this unachievable ideal vitiates the neutral theory (e.g., as implied by Wallace 1991), or that a convenient rule-of-thumb exists by which "neutral" really means something akin to "biochemically similar," or that "neutral" applies to isolated characters (rather than to a comparison of character states) according to a criterion of "lacking function," or that neutral evolution must involve "silent" changes that depend on "degeneracy" of the relationship between genotype and phenotype (e.g., Cronin 1991). Instead, the only necessary "degeneracy" in the concept of neutrality is with respect to fitness: neutral evolution is a transition between states with approximate parity of fitness, there being no restriction on how a given degree of fitness is achieved, so that changes in phenotypic and "functional" characters are fair game, including everything from morphological changes documented in the fossil record (Lande 1976) to molecular changes that alter enzyme activities (Hartl et al. 1985).

A related use of "neutral evolution" refers by exclusion to any change that is not itself an adaptation by natural selection (as by Nitecki and Hoffman 1987). In this broader sense, "neutral" changes would include not only random fixations and genetic hitchhiking (fixation of an allele tightly linked to a selected allele), but also pleiotropic effects of selective allele fixations. If a selective allele fixation takes place on the basis of the advantage conferred by attribute X, other attributes, Y and Z, are fixed in the population simultaneously if they are developmental or structural consequences of the same genetic change. Attributes Y and Z, considered in isolation, may be neutral *sensu stricto* or deleterious (though any deleterious effect of Y and Z will not have outweighed the advantage of X). The concepts of allometry and "spandrels" (Gould and Lewontin 1979), as well as many invocations of "developmental constraints" (Maynard Smith et al. 1985), are based on this implication of pleiotropy, which might be thought of as "developmental hitchhiking." Occasionally, pleiotropic effects are seen explicitly as an alternative to random fixation as a mechanism of "neutral" change (e.g., as discussed by Nei 1987, p 387). The proportion of phenotypic changes that are "neutral" in this broad sense is not known and could be quite large if genetic changes with significant fitness effects on one character typically affect many other characters.

Thus, neutral evolution is not restricted to "silent" or "nonfunctional" changes—indeed, it is not restricted to any category of outcome excluding changes biased toward increased fitness—and thus it may be said that constructive neutral evolution is possible, in theory. One may allow this possibility in theory without giving it serious attention in practice, therefore the aim of this work is to advance this possibility to the level of detailed
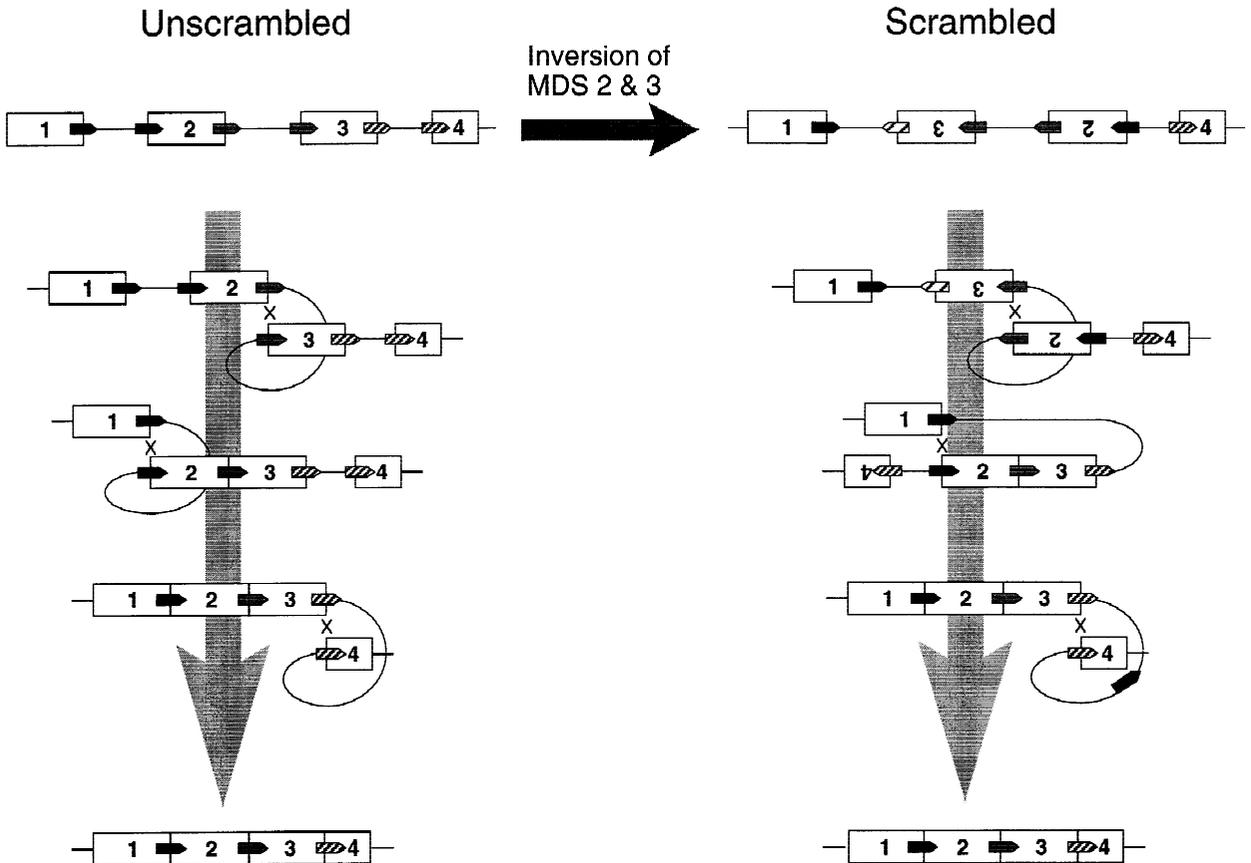
Fig. 1. Evolutionary scrambling and developmental unscrambling. Hypothetical genes with four MDS segments (*gray boxes*) and three IES segments (*interconnecting lines*) are shown. Corresponding MDS/IES and IES/MDS boundaries have short matching sequences indicated by the smaller *arrow-shaped boxes*. Evolutionary change (from unscrambled to scrambled) is indicated by the *rightward arrow* leading from a nonscrambled gene to the scrambled gene that would result from inverting a block including MDS 2 and 3. Steps in developmental processing are indicated by the *downward arrows* leading from micronuclear genes to macronuclear ones. For both the unscrambled and the scrambled versions, the same three precise crossovers (at points marked with an "X") complete the rearrangement process: the only difference

is that, for the scrambled gene, the topological outcome of the second crossover is an inversion rather than a deletion. The order of recombination events does not matter in the simple hypothetical case shown here. Observed configurations of scrambled genes are considerably more complex and seem to be patterned in ways that reflect the avoidance of configurations in which nonproductive ordering of recombination events is possible; observed configurations also could reflect aspects of chromosome structure (Prescott and Greslin 1992), presumably by influencing the likelihood of events of developmental or mutational rearrangement. Nonetheless, the bias noted in the text will be strong even if the set of neutral, readily realizable configurations is only a tiny fraction of the conceivable configurations.

hypotheses regarding specific cases. Four case studies are presented: (i) a more complete version of the Covello–Gray model (which did not initially address the issue of biased variation as a driving force) applied to gRNA-mediated RNA pan-editing, as well as schemes to address (ii) gene scrambling in hypotrichous ciliates, (iii) the evolution of the eukaryotic spliceosomal splicing system, and (iv) the retention of duplicate gene loci. Subsequently, the conceptual basis of these models is clarified and considered in a broader context.

Case Studies in a General Model

The Evolution of Scrambled Ciliate Genes

Ciliates are unicellular eukaryotes with two types of nuclei: the somatic "macronucleus" contains genetic mate-

rial copied and processed from the germ-line "micronucleus" following sexual reproduction. In hypotrichous ciliates, the micronuclear genome has interrupted genes, mosaics of MDS (macronuclear-destined sequence) and IES (internal eliminated sequence) segments (Prescott 1997). Such interruptions are possible because micronuclear genes are not expressed. During macronuclear development, IESs are spliced out, and flanking MDSs are ligated together to make an operational gene. The IES elements appear to have arisen from transposons that duplicate the local sequence into which they are inserted; the duplicate sequences flank the inserted element and, apparently, provide the specificity for developmental excision by crossing-over (Fig. 1, left).

What is curious about this system is not the presence of IES elements—which can be seen as one of many forms of selfish DNA that inhabit eukaryotic genomes—

but the fact that micronuclear MDS segments are often scrambled in order and orientation relative to the macronuclear (expressed) copy of the gene (Prescott 1997). This scrambling has been considered a mystery, for lack of an apparent adaptive benefit.

However, scrambling may have evolved under these conditions even without an adaptive benefit. The reliance of the excision mechanism on specific duplicate sequences would appear to result in an unsolicited capacity for unscrambling of a rearranged gene during macronuclear development, as illustrated in Fig. 1. To the extent that rearranged genes (arising by micronuclear mutations) are unscrambled effectively, they will be neutral alleles capable of drifting to fixation.

Given such a buffer against the otherwise adverse effects of micronuclear gene rearrangements, a long-term net increase in scrambling would be expected, simply because there are many more scrambled than unscrambled configurations. More specifically, a set of n MDS segments will have $2^{n-1}\, n!$ distinct configurations, only one of which would be in the "unscrambled" category: for example, for $n = 8$, there is 1 unscrambled configuration and 5,160,959 scrambled variants, 35 of which could be reached from the unscrambled state by a single inversion. Given sufficient time, if a scrambled gene can evolve (by mutation and drift), it will evolve, and by subsequent changes it will be less likely to revert than to wander more deeply into the morass of scrambled configurations.

The Evolution of gRNA-Mediated RNA Pan-Editing

In trypanosome mitochondria, gene expression proceeds through an eleborate stage in which insertion of U (uracil) nucleotides (nt) occurs at thousands of specific sites in protein-coding regions of gene transcripts (Alfonzo et al. 1997). This process does not destroy the meaning of the genetic message but, instead, creates it: the unedited transcripts are encoded in bizarre "cryptogenes" (e.g., see Fig. 2A), but the edited transcripts encode operational proteins. The editing process is guided by a large number of short (50- to 80-nt) RNA molecules called "guide RNAs" or gRNAs, which have "anchor" regions of complementarity to the transcript to be edited and "guide" regions that guide the editing process by base-pairing interactions (see Fig. 2A). Once a gRNA is anchored to its template by base-pairing, editing of a short block of transcript (usually 30–40 nt) proceeds by cycles of cleavage, U-insertion, and ligation (Alfonzo et al. 1997).

The evolution of this "apparently superfluous detour" (Malek et al. 1996) can be understood as an accumulation of (i) edited sites, where U-insertion sites in RNA represent T (thymine) nucleotide deletions at the corresponding gene sites; and (ii) gRNAs, which (because of their complementarity to gene transcripts) must have arisen initially by duplication and antisense transcription of a gene segment. The gRNAs are often said to allow "correction" of "mistakes" in transcripts, but the implied evolutionary sequence of events (given explicitly by Cavalier-Smith 1997) is itself a mistake: if gRNAs arise by duplication and antisense transcription of a gene segment, a gRNA gene that arises *after* a T deletion will lack this same nucleotide position and, thus, has no capacity for correction. Instead, the duplication that gives rise to a gRNA gene must precede the T deletion. This raises the possibility that the chance expression of a prospective gRNA gene renders a subsequent T deletion tolerable, such that it is a neutral allele subject to random fixation. The T deletion would represent a harmful frameshift without the gRNA, thus the occurrence of a deletion imposes a constraint that stabilizes the gRNA (as well as the enzymatic capacity for editing) against subsequent evolutionary loss (see Fig. 2B) (Covello and Gray 1993).

Although such an initial T deletion could be reverted by a T insertion (allowing subsequent loss of a gRNA), an increase in editing is far more likely, for two reasons. First, DNA polymerases typically delete single nucleotides several times more frequently than they insert them (e.g., Thomas et al. 1991) so that individual neutral changes may be biased toward deletions. Second, and more importantly, a gRNA is typically long enough to provide the (initially gratuitous) capacity to guide editing at not just one site, but one or two dozen sites (Fig. 2). Thus, while reversion from one to zero editing sites requires a T insertion at one specific position, an increase from one to two sites may occur by a T deletion at any one of $n - 1$ positions, where n is the number of possible editing sites (i.e., U residues in the mature transcript region corresponding to the gRNA). Taken together, these two biases render an increase from one to two sites $r(n - 1)$ times more likely than a reversion from one to zero, where r is the deletion:insertion mutation bias. The further increase from two to three sites is again favored, and so on, until equilibrium is reached when the ratio of edited to nonedited sites is r. Figure 2A suggests $r \approx 10$. At equilibrium, the completely nonedited configuration would be a fraction $(r + 1)^{-n}$ of the total, or 10^{-9} to 10^{-24} for n in the range of 12–24 and r in the range of 4–10 (i.e., the expected equilibrium distribution is just a case of the binomial probability distribution, over n events, where $r = p/q$).

All of this presupposes an enzymatic machinery of editing, which must have predated the first edited site. The editing machinery exhibits four biochemical activities found commonly in other cells: RNA helicase (to wind or unwind duplexes), endoribonuclease (to cleave an RNA strand), terminal uridyl transferases (to add uridine residues to a strand), and RNA ligase (to ligate strands) (Alfonzo et al. 1997). Prior to participating in

173

A F C D L F W L R L L L W L Y Y C V W S R L C F V V Y F N C L L L I F D F L L

uuuuGuGAuuuuAUUUuGGuuGCGuuuAuuuAUUAuGAuuuAuAuuuAuuGuGuGuGAuCCAGGuuGuGuuuuGuuGuGuAuuuuuAAUUGuuuAuuAuuuGAuAuuuuGAuuuuuuuAuuuA

AGAGUAACGUAUGAAGUUGGUGGAUAGUGACUGUAAGUUAAGAA

UAGUAGUAUGUGUUGGGUUUGAUAUAAGGUAGUACAUAGAAUU

UAGAGUUGAUGUAGGUGAUGAUAUUGAUAUGGUAGCACGUGUUG

GAGUGAAAUUAGCUUAGAUAGUGAUGUUGAUAUAGUAGCACGU

B

No gRNA;
No editing

Excess gRNA;
No editing

Constrained gRNA;
1 edited site

Constrained gRNA;
2 edited sites

DNA

add gRNA / lose gRNA

delete T / insert T

delete T / insert T

RNA

Protein

Fig. 2. Evolution of gRNA-mediated RNA pan-editing. **A** A pan-edited RNA transcript. Shown here are 114 protein-coding nucleotides of the mature (edited) *Trypanosoma cruzi* MURF4 mRNA (5′–3′ orientation), the amino acid translation (*above*), and four overlapping gRNAs (*below*; 3′–5′ orientation) (Avila and Simpson 1995). *Capital letters* indicate the nucleotides (52 of the 114 total nucleotides) that are transcribed from a DNA template; *lowercase* "u" nucleotides (62 of the 69 uridines) result from editing. In general, gRNAs consist of a 5′ "anchor" region that can pair with an RNA transcript, a "guide" region that provides the sequence information utilized in inserting or deleting U residues, and a poly(U) "tail" (not shown), whose significance is unclear. **B** The spread of editing, *Horizontal arrows* show transitions between possible states; *vertical arrows* indicate steps in gene expression. Novel prospective gRNAs may arise by chance at a very low rate and face immediate loss. However, once a T nucleotide is deleted in the corresponding region of the protein-coding gene, loss of the gRNA is no longer possible without first reverting the change by insertion, a step that is extremely unlikely due to biases discussed in the text.

RNA editing, some or all of these activities may have operated together in other contexts. A possible precursor would be a system analogous to the small-nucleolar-RNA-mediated cleavage of rRNA transcripts, a process that exhibits (in common with gRNA-mediated editing) ribonuclease-catalyzed cleavage of a transcript within a region paired to a small RNA (Maxwell and Fournier 1995). In general, mitochondrial genomes are transcribed in long pieces that require multiple post-transcriptional cleavage events (Gillham 1994): possibly trypanosomes had, or still have, a system that utilizes gRNA-like molecules to guide cleavage events (without editing) in non-translated spaces. Indirect evidence also suggests that gRNA gene elements (which may occur interspersed with crytogenes, in repetitive arrays, or on separate chromosomes called "minicircles") have a structural role relating to the amount or concentration of organellar DNA. Such roles do not suggest a raison d'être for RNA fragments with sequences complementary to coding regions (i.e., fragments such as gRNAs) but, instead, provide a reason to suspect the prior existence of a replication–expression system facilitating the spontaneous production of gratuitous RNAs subject to entrainment in editing, a step that otherwise would be strongly limiting on the rate of evolution of gRNA-mediated editing.

To summarize, given insertions and deletions of single nucleotides, as well as a source of prospective gRNAs, pan-editing will arise inevitably, albeit slowly, if there is an enzymatic machinery that corrects mismatches in paired RNA duplexes. This process would be favored strongly by a systemic bias due to the initial state of exact complementarity of gRNA and transcript, a state from which the system will tend to depart by accumulating T deletions. The machinery can be proposed to have existed fortuitously (in the form of a combination of enzymes performing other operations) and to have been entrained in editing, though this issue must remain relatively intractable because it involves a singular event rather than a recurrent one.

The Evolution of Spliceosomal Splicing

Introns are spliced from eukaryotic protein-coding genes by a "spliceosome," a complex assembly of RNA and protein. Biochemically purified spliceosomes include at least several dozens of different proteins and five small (100- to 200-nt) RNA molecules known as "spliceosomal snRNAs" [small nuclear RNAs (Lamond 1993)]. However, spliceosomal splicing appears to have arisen from a simpler form of self-splicing homologous to that seen in modern group II self-splicing introns, as argued from the relevant phylogenetic distributions (Cavalier-Smith 1991), as well as from extensive similarities between the RNA components of the spliceosomal and group II systems (Copertino and Hallick 1993). Some confirmation of this idea has been provided recently, by experiments showing that U5 snRNA can partially complement a deletion of the corresponding region of a group II intron (Hetzer et al. 1997).

The transition from a system with many self-splicing group II-like introns to a system with a spliceosomal splicing machinery and many passive introns can be envisioned in terms of three component processes:

Fragmentation of a Catalytic Intron into Trans-Acting snRNAs. For a coherent self-splicing intron to evolve into a set of snRNAs, it must be split multiply, and the pieces must act in *trans.* Such a change may not have required any prior specialization of the properties of the intron: studies of group II splicing in vitro show that artificially split introns can reassociate and that deletions of specific structural regions can be complemented by supplying the deleted portion in *trans* (Hetzer et al. 1997; Jarrell et al. 1988). This in vitro observation is paralleled in vivo by the observation that some plant organelle genes are split, within their group II introns, into separately transcribed upstream and downstream fragments that associate to complete the splicing reaction (Bonen 1993). To the extent that splitting of the intron can occur with little disruption of splicing, split rearrangements may be neutral alleles capable of drifting to fixation. Given the possibility that intron splitting is tolerated, a net increase in splitting can be anticipated simply because the introns start out unsplit and there are many more ways to arrange genes into sets of discontinuous pieces than into continuous ones.

A further step in fragmentation is represented by the *psaA* gene (of *Chlamydomonas* chloroplasts), which has a group II intron split into at least *three* parts, with a separately transcribed central fragment encoded by the *tscA* gene (Goldschmidt-Clermont et al. 1991). The *tscA* fragment is thus analogous to an snRNA, being a small RNA that operates in *trans* in a splicing reaction. Whether or not the intron fragment is recycled in the splicing of multiple introns is not known at present, but the occurrence of such an interaction can be anticipated on the grounds that RNA helicases are generally available in cells.

Loss of Self-Splicing in Most Introns. A further step in the direction of similarity to the spliceosomal system would be for introns present in the same compartment as a fragmented intron to lose inherent self-splicing ability, becoming dependent on the intron fragments (i.e., the incipient snRNAs). In the presence of such a *trans*-acting fragment, other introns are free to suffer splicing defects compensated by it, a process that might be expected to occur (by mutation and drift) to the extent that spontaneous mutations are more likely to degrade local structural elements than to preserve or augment them. The "group III" introns found in *Euglena* chloroplast genes seem to represent a biological precedent: these group-II-related introns lack most of the catalytic structure expected for group II splicing, and do not self-splice, but are suspected to rely on separately encoded trans-acting RNA fragments (Copertino and Hallick 1993).

The Addition of Dozens of Proteins to the Spliceosome. The many proteins that operate in the spliceosome might be explained adaptively if they enhance the speed or stability of splicing. However, available evidence suggests that spliceosomal splicing is not faster, but slower than group II splicing (Baurén and Wieslander 1994; Beyer and Osheim 1988; Lang and Spritz 1987; Schmidt et al. 1996); and rapid rates of group II splicing have been measured at temperatures that disrupt spliceosomal splicing (Schmidt et al. 1996; Yost et al. 1990). That some spliceosomal proteins may have arisen neutrally can be suggested by analogy with a scheme for the origin of splicing factors proposed by Lambowitz and Perlman (1990) with regard to the CYT18 protein of the fungus *Neurospora crassa.* The CYT18 protein, which is the mitochondrial tyrosyl-tRNA synthetase, facilitates the splicing of some (but not all) mitochondrial group I introns and does so in *N. crassa* but not in most other fungi (Lambowitz and Perlman 1990). Intron-specific splicing factors with restricted phylogenetic distributions have appeared several times in fungal evolution, and this, along with the variable patterns of natural occurrence of the introns (which are mobile elements), suggests that ad hoc dependencies on protein factors have arisen repeatedly (Lambowitz and Perlman 1990).

To explain such a process, Lambowitz and Perlman (1990, p 444) suggest that "after the introns were acquired, they would interact with a variety of cellular RNA-binding proteins, such as aminoacyl-tRNA synthetases. If these interactions fortuitously stabilize structures required for splicing, the intron may then lose the ability to self-splice, thus fixing the interaction." That is, the multitude of RNA-binding proteins in the cellular milieu creates the conditions under which an accidental dependency may arise. Any protein with an affinity for the native state of an intron stabilizes it, by Le Chatelier's principle. The intron then may lose some of its inherent structural stability, resulting in a protein dependency. Rather than replacing catalytically active RNA moieties,

or enhancing speed or stability, the protein would replace RNA determinants of structural stability [a relationship perhaps exemplified by the CBP2 splicing factor (Weeks and Cech 1996)]. The existence of gratuitous affinities between RNA-binding proteins and structural RNAs is readily argued from observations such as the following: several *E. coli* ribosomal proteins facilitate splicing of bacteriophage T4 group I introns in vitro (Coetzee et al. 1994), though *E. coli* strains generally do not harbor group I introns; the *E. coli* ribosomal proteins S4 and S12 have specific binding affinities for rRNA regions other than the ones that they bind in the active ribosome (Stern et al. 1989); and the same *E. coli* S12 protein facilitates the activity of the hammerhead ribozyme, a small RNA engineered from a plant virus (Herschlag et al. 1994).

In summary, the elaborate spliceosomal system may have evolved from a simpler self-splicing system by fragmentation and *trans* action of one or a few active introns, loss of self-splicing ability in the remaining introns, and accretion of many proteins. Each process can be argued on the basis of biological precedents that involve splicing systems and that are not associated with obvious adaptive benefits. Steps such as these also occur in other contexts: apparently gratuitous fragmentation of genes whose products, nevertheless, associate after expression is seen for ribosomal RNA genes (Nedelcu 1997; Wilson and Williamson 1997); reduction or loss of an activity encoded in an element when this activity can be supplied by another copy of the element is seen in nonautonomous transposons (Iida et al. 1983); and the accretion of multiple proteins to catalytic RNAs that are presumed to have operated ancestrally without proteins is suspected for ribosomes and ribonuclease P (Inoue 1994).

The Dilemma of Duplicate Gene Retention

Ohno (1970) argued that the profusion of duplicate isozyme loci in eukaryotes can be traced largely to genome doublings initiated by events of auto- or allotetraploidization. Evidence of such events occurring in the last few hundred million years is readily detectable in patterns of total DNA content, karyotype, and isozyme expression in various groups of organisms, including ferns, salmonids, catostomids, loaches, sturgeons, salamanders, and frogs (Buth 1983; Ohno 1970). The long-term evolutionary stability of duplicate loci in such cases presents something of a paradox (for review, see Li 1980). In the case of salmon and its relatives, tetraploidization is hypothesized to have occurred 50–100 MYA (million years ago), and the proportion of loci for which both duplicate copies are still detectable (by electrophoresis and enzyme-activity-based staining), is 50%; for carp and relatives, 47% of duplicate loci persist after perhaps 16 MY; for loaches, 25% of loci persist after 15–40 MY (Li 1980). Recent work based on sequence data (e.g.,

Nadeau and Sankoff 1997) tends to confirm the conclusion that a substantial proportion of duplicates is retained in evolution.

Since many duplicates are ultimately lost, many or most duplicates must have been redundant initially. Neutral fixation of null mutations readily accounts for the lost loci, but what about the duplicates that are retained? Neutral loss due to the random fixation of null mutations occurs so rapidly that practically none of the duplicates would survive more than a few million years (Li 1980). Yet adaptive divergence implies that "a distinct function is waiting for each daughter gene" produced by the duplication (Hughes 1994), which seems reasonable (and well evidenced) in isolated instances but rather unreasonable as an explanatioin for perhaps 15,000–30,000 pairs of duplicate loci retained over the long term in a fish genome, particularly since the opportunity for adaptive diversification would be restricted to the brief few million years before all of the redundant duplicates are lost by fixation of nulls.

However, the general scheme of constructive neutral evolution that is now familiar suggests a third possibility: (i) excess capacity exists in the form of redundant duplicate genes; (ii) reductions in the activity of each gene would impose a selective constraint preventing the loss of the other duplicate copy; and (iii) such reductions are favored by a bias in the production of variants, since activity-reducing changes will be more common than activity-increasing ones. Previous neutral models do not include activity-reducing mutations, only null mutations. Yet the majority of spontaneous mutations are not gene-inactivating deletions or insertions, but nucleotide substitutions (e.g., Sommer and Ketterling 1994), and only a minor fraction (perhaps 10%) of substitutions that alter protein sequences results in a reduction of activity by more than one or two orders of magnitude (e.g., Rennell et al. 1991).

When activity-reducing mutations are included in a neutral model, stabilization of a duplicate pair is a common outcome: computer simulations suggest that roughly 10–30% of initially redundant locus pairs will be stably established within a few million generations by random fixation of activity-decreasing changes at both loci, as shown in Fig. 3 (A. Stoltzfus and O.C. Feeley, unpublished). This model can account for long-term retention of a substantial proportion of duplicate isozyme loci and is consistent with (i) the observation that most retained isozyme pairs exhibit patterns of expression consistent with simple quantitative changes in gene or enzyme activity (Ferris and Whitt 1979) and (ii) the evidence that, in general, duplicate genes experience nonconservative changes at a rate that is heightened initially (Hughes 1994; Ohta 1994) and that is roughly equal for the two copies (Hughes and Hughes 1993).

Such a model obviously will not account for all known aspects of duplicate gene divergence and persis-
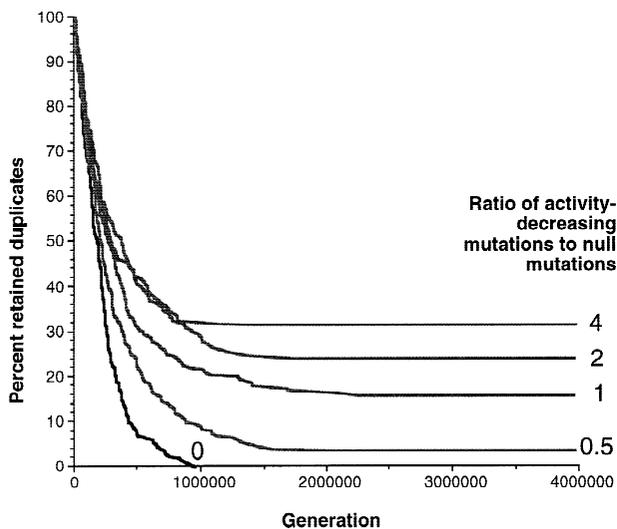
Fig. 3. Simulated retention of duplicate loci under a neutral model. Duplicate loci are lost rapidly due to null mutations (*black curve*), unless activity-decreasing mutations are also considered (*gray curves*). As the rate of activity-reducing mutations increases, more duplicates are retained, due to fixation of activity-reducing changes at both loci. Each curve is based on 200 replicates with an initial population of 100 haploid individuals with two loci, each contributing an activity of 1*X*, for a total of 2*X,* and facing threshold selection on survival such that only individuals with a combined activity of at least 1*X* survive and reproduce. Genes are subject to null mutations at a rate of 2×10^{-6} per gene per generation, and activity-reducing mutations at a rate of 0, 0.5, 1, 2, or 4 times the null rate. An activity-reducing mutation multiplies activity by a factor *r,* where *r* is a uniform random deviate in [0,1]. Biologically reasonable ratios of activity-decreasing: null mutations are probably in the range of 1–3. Activity-increasing mutations (not included here) do not seem to alter the outcomes substantially as long as they are less common and less extreme than activity-decreasing mutations. Preliminary results from a diploid model (not shown) are qualitatively similar.

tence, yet it has several interesting properties, two of which may be noted. First, although the model presented here allows only simple quantitative divergence in activity, this limitation is owing to its simplistic genotype–phenotype–fitness relationship: more complex modes of divergence would be possible given a multifaceted model in which the role of the duplicate gene products may be subdivided in many qualitative ways, rather than a single quantitative way, resulting in specialization or "functional" divergence. Second, to the extent that this model explains the persistence of some substantial proportion of duplicate loci following a genome doubling (or a subchromosomal duplication that includes some redundant duplicates), it expands enormously the opportunity for subsequent specialization (adaptive or otherwise), presumably proportional to the areas under the curves in Fig. 3.

Discussion

> *It is worth noting that presumably no serious biologists think that other evolutionary mecha-*

> *nisms [i.e., other than natural selection], such as drift or pleiotropy, can produce complex and intricate traits that appear to be adaptations.*

> Brandon (1990, p 175)

For most biologists, features that are complex or coordinated, that figure prominently in the biology of an organism, and that can only have arisen by a long series of changes will "appear to be adaptations." The common assumption (e.g., as in the passage quoted above) is that such traits arise by natural selection, usually by the classic model of a series of successive, small modifications, each of which is beneficial for some reason relating to the "function" or current utility of the trait. Clearly, the traits addressed in the case studies above would qualify as "complex and intricate." They also "appear to be adaptations" in the sense of eliciting proposals of hidden adaptive benefits [most prominently for the case of RNA editing (e.g., Hajduk et al. 1993; Landweber 1992; Stuart 1991; Stuart et al. 1997; Weissman et al. 1990)].

In the models outlined here, complex and intricate traits arise, not by the classical model of beneficial refinements but, instead, by a repetition of neutral steps. The fundamental sequence of events is that a novel attribute appears initially as an excess capacity and later becomes a contributor to fitness, due to a neutral change at some other locus that creates a dependency on it. This fundamental sequence of events may occur just once (as in the Lambowitz–Perlman model to explain the CYT18 splicing factor) or a similar sequence of events may be repeated many times at different loci (as in the other models discussed here).

A feature that results from such steps may appear complex if it has arisen from some other attribute with a long evolutionary history (e.g., as for a novel gene created by gene duplication) and if it involves the long-term accumulation of many similar changes. Given inevitable purifying selection, any novel attribute that arises in this manner is likely to be coordinated, rather than in conflict, with its biological milieu: it will be an "aptation" in the sense of Gould and Vrba (1982) and "polite" in the sense of Zuckerkandl (1992). Whether such features should be termed "adaptations" is a matter of semantics. The products of constructive neutral evolution would be adaptations if an adaptation is defined as an attribute that contributes operationally to fitness (e.g., Hecht and Hoffman 1986; Zuckerkandl 1992) but not if an adaptation must be "built by selection," that is, by the model of successive selective refinements (e.g., Gould and Vrba 1982; Williams 1966).

Constructive neutral evolution, then, would differ from the classical model of adaptive refinement in three respects. Directionality or the recurrence of similar steps is due not to a common adaptive benefit to which the steps accrue, but to a common bias in the production of variants. A novel attribute is not immediately beneficial

but, instead, appears initially as excess capacity. A "function" or contribution to fitness ultimately ascribed to the attribute thus does not arise by an immediate beneficial effect whose subsequent loss would be deleterious but, instead, by a subsequent change at an interacting site that renders loss of the attribute deleterious. In the sections that follow, these three aspects of the model are explained more fully and related to similar concepts encountered in discussions of novelty, complexity, and directionality in evolution. Brief comments are made on the subject of testability.

Interactions That Alter Selective Constraints

Though it is generally agreed that traits may interact in complex ways, the implications of the common simplifying assumption that loci evolve independently have not been fully explored, partly because empirically based concepts of gene interaction are poorly developed (Phillips 1998), and abstract models with rich interactions (e.g., Kauffman 1993) are difficult to interpret in biological terms. In a neutral model rich in interactions, such as the covarion model (Fitch 1972), complex interactions between sites reduce to a more tractable issue of whether a particular variant in its genetic context is neutral (and thus subject to fixation by neutral kinetics), or deleterious (not subject to fixation). The set of deleterious variants for a particular site (in its genetic context) is referred to in the language of "selective constraints" (see Kimura 1983, Chap. 7). For example, an amino acid site in a protein is selectively constrained to be glycine if all other amino acids may arise as variants and face extinction due to adverse fitness effects. This concept is not the same as the rather broadly defined "developmental constraints" of Maynard Smith and colleagues (1985), which include selective constraints as well as effects of pleiotropy, biases in the production of variants, effects arising from modes of transmission genetics, direct implications of physical and chemical laws, and so on.

To the extent that complex interactions prevail, selective constraints on any given site will be subject to change as a function of genetic context. A fundamental aspect of the models presented here is that changes at one genetic site have developmental implications that alter the selective consequences of other possible changes. Interactions of this type explain how a capacity that is initially gratuitous may acquire a role in the maintenance of fitness that ensures its persistence, owing to neutral changes at other sites. For instance, in the pan-editing model, the advent of a prospective gRNA renders tolerable a set of possible T deletions that would otherwise represent deleterious frameshifts: the subsequent occurrence of a T deletion prevents the loss of the (previously gratuitous) gRNA gene. In the splicing model, fragmentation of one intron allows other introns to lose inherent self-splicing ability, but the fragmentation of the first intron is no longer reversible when other introns become dependent on it. Such effects may be compounded by additional changes of the same type (e.g., further fragmentation of an intron or further accumulation of edited sites).

The specific notion that subsequent changes might "lock in" a previously variable feature, as well as the more general notion that the longer a feature has persisted (for whatever reason), the greater the chance that constraints preventing future changes will accumulate, both have been suggested many times (e.g., Bull and Charnov 1985; Maynard Smith and Szathmary 1995; Riedl 1978). The concept invoked here is not much different from the "contingent irreversibility" of Maynard Smith and Szathmary (1995), except that the constraint is imposed on a previously gratuitous attribute.

Excess Capacity

An attribute may be said to represent an excess capacity to the extent that (i) its presence represents the capacity to carry out some operation and (ii) its absence (or diminution) would not be opposed by purifying selection (thus, the concept does not apply to any kind of plasticity or canalization that actually contributes to fitness). In a neutral model, the most interesting excess capacities are those that represent some qualitatively new capacity (rather than a quantitative excess) and whose advent represents the relaxation of selective constraints, opening up the possibility of some previously forbidden change.

The specific excess capacities invoked above include the unsolicited capacity for developmental unscrambling that precedes the appearance of scrambled genes, the ability of a continuous intron to reassociate and splice when fragmented, the ability of one gene of a redundant pair to compensate for a reduction in activity of the other, the ability of the ancestral CYT18 protein to bind a group I intron, and so on. In most instances these excess capacities are proposed as initial conditions: sometimes this proposal is merely a post hoc assumption, though in most cases it is an inference (e.g., some duplicate loci resulting from genome doubling must have been redundant, because so many were lost subsequently; gRNAs must precede the edited sites that are dependent on them) or is suggested more directly by empirical evidence (e.g., proteins often bind structured RNAs fortuitously, and split introns can reassociate in vitro).

Because it is not possible to atomize organisms completely to yield a one-to-one relationship between capacities and discrete material parts (or genetic determinants), gratuitous capacities do not inevitably correspond to expendable parts (or genetic determinants), though this sometimes may be the case. For instance, unutilized gRNAs or duplicate genes are spare parts that can be lost, but the capacity of an intron to reassociate when split is distributed throughout the intron.

In some cases it may seem more felicitous to refer to "buffering," "tolerance," or "unrealized potential" than to "excess capacity," but the choice of words is not as important as the concept (as defined generically above), which is widely invoked in discussions of evolutionary novelty, as by Galis (1996), who refers to "excess structural capacity" in describing the evolution of musculoskeletal systems (see also Frazetta 1975), and by Gould and Lewontin (1979) in the concept of "spandrel" or Gould and Vrba (1982) in the concept of "preaptation," and in the "tolerance" invoked by Dover (1992). Multiple examples of innovations based on the coopting of "spandrels" or "preaptations" have been given recently (Armbruster 1996; Gould 1997). In molecular evolution, there are many cases in which mobile elements or repetitive DNA (surely an unarguable source of excess capacity with respect to the host) are invoked in the origin of novelty, as in the involvement of introns in exon shuffling (Patthy 1987), of satellite DNA in centromere formation (Csink and Henikoff 1998), of retroposons in *Drosophila* telomere maintenance (Biessmann and Mason 1997), and of Alu elements in various schemes of gene regulation (Britten 1996).

Biases in the Production of Variants

Mutation is often said to be "random," but such statements refer not to a proposed uniformity in the spontaneous production of variation, but to a logical restriction on causal models of microevolution, to the effect that selection acts subsequent to the origin of variation and cannot influence it directly. For any means of measuring or categorizing the outcomes of variation, biases (meaning simply "nonuniformities" or "asymmetries") are to be expected. In a neutral model, unless other factors intervene, such a bias will bias the direction of evolutionary change, resulting in parallel changes or directional trends.

A distinction between two entirely different sources of bias is useful. The more immediately obvious type is a "mutational" bias, an inequality in the rates of mutational change between specific genetic states that arise from specific aspects of the machinery for replication, repair, and transmission of genetic material. Detailed molecular studies invariably reveal such nonuniformities: some nucleotide sites are more mutable than others, DNA polymerases cause deletions more frequently than insertions, mobile elements show insertion site preferences, some chromosomal rearrangements occur more readily than others, and so on. However, even if all rates of mutation between specific genetic states were equal, a second source of bias would exist, because some categories of possible variants will be populated by more genetic states than others, that is, some phenotypic categories are widely distributed in a locally accessible region of conceptual "genotype space." Such "systemic" biases

do not arise from the properties of mutational mechanisms themselves, but from aspects of the organization and interaction of parts in a developmental–genetic system.

With this distinction in mind, the sources of bias invoked above may be listed. In the gene-scrambling and RNA pan-editing cases, and in the fragmentation of introns, the initial state of the system (unscrambled, unedited, unfragmented) is unique or rare with regard to some extensive set of combinatorial possibilities (scrambled, edited, fragmented) that may be reached by mutation and (possibly neutral) fixation. The resulting systemic bias drives a departure from the improbable initial state to one of many alternative states. In the editing model, a deletion:insertion mutational bias plays a subsidiary role. In the gene duplication model, as well as in the explanation for loss of self-splicing and for the origin of protein dependencies in splicing, it is assumed that mutations that reduce activity or affinity or stability are much more common than those with the opposite effect [a bias that plays a prominent role in discussions of the complexification of regulatory networks by Zuckerkandl (1997)]. The resulting directionality consists in duplicate genes undergoing reductions in activity, and introns losing self-splicing ability, becoming dependent on available proteins as well as *trans*-acting intron fragments. In both cases, the biases are systemic and result from a history of selection such that the initial state of the system (genes with highly specific activities, introns with independent splicing ability) is unusual with respect to possible alternative states. In all cases, biases are invoked as causes of directionality, with systemic biases playing a much more prominent role.

Outside of studies of neutral evolution, biases in the production of variants are only rarely viewed explicitly (Vrba and Eldredge 1984) or implicitly (Thomson 1985) as biases on the expected course of evolution. More commonly, biases in the production of variation are denied any such influence, or when they are identified as evolutionary factors, they are invoked as "developmental constraints" (Maynard Smith et al. 1985), with considerable confusion about what this terminology actually implies about evolutionary processes (Amundson 1994; Antonovics and van Tienderen 1991). The empirical pattern to be explained is clear enough, though, at least in studies of molecular evolution, where it is commonly observed that homoplasies, directional change, and patterned rate differences reflect known or strongly suspected mutational biases, as in the case of transition:transversion bias (Gojobori et al. 1982; Golding 1987), GC bias (Foster et al. 1997; Gu et al. 1997), deletion:insertion bias (de Jong and Ryden 1981), point mutations templated by replication slippage or other ectopic pairings (Cunningham et al. 1997; Golding 1987; Macey et al. 1997), the effect of repeat-unit length on the mutability of di-, tri-, or tetranucleotide repeats (Schug et al.

1998), and regional composition effects on nucleotide substitution biases (Wolfe 1989). Morphological examples may also be found, such as the analysis by Alberch and Gale (1985) of evolutionary trends reflecting developmental variation in digital skeletal elements.

A Note on Testability

To question the testability of neutral models may seem strange to some molecular evolutionists; nevertheless, the notion that all nonselective factors fall into an intractable category of "chance" is common:

> When one attempts to determine for a given trait whether it is the result of natural selection or of chance (the incidental byproduct of stochastic processes), one is faced by an epistemological dilemma. Almost any change in the course of evolution might have resulted by chance. Can one ever prove this? Probably never. (Mayr 1983)

Thus, in Mayr's defense of the "adaptationist program" (from which the above quotation is taken), nonselective factors (e.g., mutation, development, environment) are recognized, yet assigned to "chance," not because this is the way the world works—these "chance" processes have physical causes and potentially predictable outcomes—but because nonselective factors are (in this view) so poorly understood or so rarely important that it is impossible to erect testable hypotheses of their influence on the course of evolution. This pragmatic position, to the extent that it is not a self-fulfilling prophecy, must ultimately succumb to the advance of knowledge. Indeed, the black box of "chance" is already being dissected in studies of molecular evolution, as suggested by conceptual advances such as the codon capture model or the covarion model and by empirical results (cited above) suggesting the importance of biases in mutation in explaining patterns of evolutionary divergence.

For the sake of example, a few of the testable implications of the models outlined above can be mentioned. The duplicate gene model gives a specific-time course for gene loss following redundant duplication (Fig. 3) and a mean reduction (converging on a twofold reduction) in per-locus activity among retained duplicates. The pan-editing model implies that r (the deletion bias) can be estimated from the distribution of edited sites: this value can be compared to the (presently unknown) mutation bias of the relevant polymerase; furthermore, the rarity of U-deletional editing (not mentioned earlier), which occurs by the same mechanism as U-insertional editing (Cruz-Reyes and Sollner-Webb 1996), appears from the present model to be explained entirely as an effect of r and, thus, suggests that there will not be a large inherent difference in the efficacy of U-insertional and U-deletional editing.

Summary and Prospectus

To summarize, interactions among evolving sites, excess capacities, and biases in the production of variants may bring about the evolution of complex and aptive features, without the necessary involvement of selective allele replacements. Neutral models based on these concepts have been devised to account for the evolution of RNA pan-editing, duplicate gene families, and so on. For the case of duplicate gene retention, it has been possible to implement a conceptual model in the form of a rigorous computer simulation that suggests that a substantial fraction of duplicate loci might be retained by a neutral process of mutual compensation.

A final issue of interest is whether nonselective factors such as complex interactions among evolving sites, excess capacities, and biases in the production of variants are limited in their importance to cases of neutral evolution, or even more limited to a few molecular curiosities such as gene scrambling. From the foregoing discussion it is apparent, at least, that similar evolutionary factors are invoked commonly in treatments of novelty and directionality that do not specifically address molecular features or neutral models. Possibly, factors such as biases in variation might operate in qualitatively similar ways in either a neutral or an adaptive model. If so, models of constructive neutal evolution may be interpreted broadly as attempts to understand the influence of diverse evolutionary factors separately from the proximate cause of allele replacement (selection or drift), or more narrowly as hypotheses of neutral evolution in the strict sense, as they are ostensibly given above.

Note Added at Proof

Force et al. (1999, Genetics 151:1531–1545) have recently presented a neutral model for the retention of duplicate genes that is conceptually related to the model presented here.

References

Alberch P, Gale EA (1985) A developmental analysis of an evolutionary trend: Digital reduction in amphibians. Evolution 39:8–23

Alfonzo JD, Thiemann O, Simpson L (1997) The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. Nucleic Acids Res 25:3751–3759

Amundson R (1994) Two concepts of constraint: Adaptationism and the challenge from developmental biology. Philos Sci 61:556–578

Antonovics J, van Tienderen PH (1991) Ontoecoenophyloconstraints? The chaos of constraint terminoloy. Trends Ecol Evol 6:166–168

Armbruster WS (1996) Exaptation, adaptation, and homoplasy: Evolution of ecological traits in Dalechampia vines. In: Sanderson MJ, Hufford L (eds) Homoplasy: The recurrence of similarity in evolution. Academic Press, San Diego, pp 227–243

Avila HA, Simpson L (1995) Organization and complexity of minicircle-encoded guide RNAs in Trypanosoma cruzi. RNA 1: 939–947

Baurén G, Wieslander L (1994) Splicing of balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. Cell 76:183–192

Berry RJ (1982) Neo-Darwinism. Edward Arnold, London

Beyer AL, Osheim YN (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev 2:754–765

Biessmann H, Mason JM (1997) Telomere maintenance without telomerase. Chromosoma 106:63–69

Bonen L (1993) Trans-splicing of pre-mRNA in plants, animals and protists. FASEB J 7:40–46

Brandon RN (1990) Adaptation and environment. Princeton University Press, Princeton, NJ

Britten RJ (1996) DNA sequence insertion and evolutionary variation in gene regulation. Proc Natl Acad Sci USA 93:9374–9377

Bull JJ, Charnov EL (1985) On irreversible evolution. Evolution 39: 1149–1155

Buth DG (1983) Duplicate isozyme loci in fishes: Origins, distribution, phyletic consequences, and locus nomenclature. In: Rattazzi MC, Scandalios JG, Whitt GS (eds) Isozymes: Current topics in biological and medical research. Alan R. Liss, New York, pp 381–400

Cavalier-Smith T (1991) Intron phylogeny: A new hypothesis. Trends Genet 7:145–148

Cavalier-Smith T (1997) Cell and genome coevolution: Facultative anaerobiosis, glycosomes and kinetoplastan RNA editing. Trends Genet 13:6–9

Coetzee T, Herschlag D, Belfort M (1994) Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. Genes Dev 8:1575–1588

Copertino DW, Hallick RB (1993) Group II and group III introns of twintrons: Potential relationships with nuclear pre-mRNA introns. Trends Biochem 18:467–471

Covello PS, Gray MW (1993) On the evolution of RNA editing. Trends Genet 9:265—268

Cronin H (1991) The ant and the peacock. Cambridge University Press, Cambridge

Cruz-Reyes J, Sollner-Webb B (1996) Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. Proc Natl Acad Sci USA 93:8901–8906

Csink AK, Henikoff S (1998) Something from nothing: The evolution and utility of satellite repeats. Trends Genet 14:200–204

Cunningham CW, Jeng K, Husti J, et al. (1997) Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. Mol Biol Evol 14:113–116

de Jong WW, Ryden L (1981) Causes of more frequent deletions than insertions in mutations and protein evolution. Nature 290:157–159

Dover GA (1992) Observing development through evolutionary eyes: A practical approach. Bioessays 14:281–287

Ferris SD, Whitt GS (1979) Evolution of the differential regulation of duplicate genes after polyploidization. J Mol Evol 12:267–317

Fitch WM (1972) Rate of change of concomitantly variable codons. J Mol Evol 1:84–96

Foster PG, Jermiin LS, Hickey DA (1997) Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J Mol Evol 44:282–288

Frazetta TH (1975) Complex adaptations in evolving populations. Sinauer Associates, Sunderland, MA

Galis F (1996) The application of functional morphology to evolutionary studies. Trends Ecol Evol 11:124–129

Gillham NW (1994) Organelle genes and genomes. Oxford University Press, New York

Gojobori T, Li W-H, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 17:245–250

Golding GB (1987) Nonrandom patterns of mutation are reflected in evolutionary divergence and may cause some of the unusual patterns observed in sequences. In: Loeschcke V (ed) Genetic constraints on adaptive evolution. Springer-Verlag, Berlin, pp 151–172

Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD (1991) A small chloroplast RNA may be required for trans-splicing in Chlamydomonas reinhardtii. Cell 65:135–143

Gould SJ (1997) The exaptive excellence of spandrels as a term and prototype. Proc Natl Acad Sci USA 94:10750–10755

Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist program. Proc Roy Soc London B 205:581–598

Gould SJ, Vrba ES (1982) Exaptation—A missing term in the science of form. Paleobiology 8:4–15

Gu X, Hewett-Emmett D, Li W-H (1998) Directional mutational pressure affects the amino acid composition and hydrophobicity of proteins in bacteria. Genetica 102/103:383–391

Hajduk SL, Harris JE, Pollard VW (1993) RNA editing in kinetoplastid mitochondria. FASEB J 7:54–63

Haldane JBS (1932) The causes of evolution. Longmans, Green and Co., New York

Hartl DL, Dykhuizen DE, Dean AM (1985) Limits of adaptation: The evolution of selective neutrality. Genetics 111:655–674

Hecht MK, Hoffman A (1986) Why not neo-Darwinism? A critique of paleobiological challenges. Oxford Surv Evol Biol 3:1–47

Herschlag D, Khosla M, Tsuchihashi Z, Karpel RL (1994) An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. [Erratum appears in EMBO J 13(16): 3926, 1994]. EMBO J 13:2913–2924

Hetzer M, Wurzer G, Schweyen RJ, Mueller MW (1997) Tran-activation of group II intron splicing by nuclear U5 snRNA. Nature 386:417–420

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B 256:119–124

Hughes AL, Hughes MK (1993) Adaptive evolution in the rat olfactory receptor gene family. J Mol Evol 36:249–254

Iida S, Meyer J, Arber W (1983) Prokaryotic IS elements. In: Shapiro JA (ed) Mobile genetic elements. Academic Press, New York

Inoue R (1994) Time to change partners. Nature 370:99–100

Jarrell KA, Dietrich RC, Perlman PS (1988) Group II intron domain 5 facilitates a trans-splicing reaction. Mol Cell Biol 8:2361–2366

Kauffman SA (1993) The origins of order: Self-organization and evolution. Oxford University Press, New York

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Lambowitz AM, Perlman PS (1990) Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. Trends Biochem Sci 15:440–444

Lamond AI (1993) The spliceosome. Bioessays 15:595–603

Lande R (1976) Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334

Landweber LF (1992) The evolution of RNA editing in kinetoplastic protozoa. BioSystems 28:41–45

Lang KM, Spritz RA (1987) In vitro splicing pathways of pre-mRNAs containing multiple intervening sequences. Mol Cell Biol 7:3428–3437

Li W-H (1980) Rate of gene silencing at duplicate loci: A theoretical study and interpretation of data from tetraploid fishes. Genetics 95:237–258

Li W-H (1997) Molecular evolution. Sinauer, Sunderland, MA

Macey JR, Larson A, Ananjeva NB, Papenfuss TJ (1997) Replication slippage may cause parallel evolution in the secondary structures of mitochondrial transfer RNAs. Mol Biol Evol 14:30–39

Malek O, Lättig K, Hiesel R, Brennicke A, Knoop A (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J 15:1403–1411

Maxwell ES, Fournier MJ (1995) The small nucleolar RNAs. Annu Rev Biochem 64:897–934

Maynard Smith J, Szathmary E (1995) The major transitions in evolution. W.H. Freeman, Oxford

Maynard Smith J, Burian R, Kauffman S, et al. (1985) Developmental constraints and evolution. Q Rev Biol 60:265–287

Mayr E (1983) How to carry out the adaptationist program? Am Nat 121:324–334

Nadeau JH, Sankoff D (1997) Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 147:1259–1266

Nedelcu AM (1997) Fragmented and scrambled mitochondrial ribosomal RNA coding regions among green algae: a model for their origin and evolution. Mol Biol Evol 14:506–517

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nitecki MH, Hoffman A (1987) Neutral models in biology. Oxford University Press, New York

Ohno S (1970) Evolution by gene duplication. Springer-Verlag, New York

Ohta T (1994) Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics 138:1331–1337

Ohta T (1996) The current significance and standing of neutral and nearly neutral theories. Bioessays 18:673–677

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264

Patthy L (1987) Intron-dependent evolution: Preferred types of exons and introns. FEBS Lett 214:1–7

Phillips PC (1998) The language of gene interaction. Genetics 149:1167–1171

Prescott DM (1997) Origin, evolution, and excision of internal elimination segments in germline genes of ciliates. Curr Opin Genet Dev 7:807–813

Prescott DM, Greslin AF (1992) Scrambled actin I gene in the micronucleus of Oxytricha nova. Dev Genet 13:66–74

Rennell D, Bouvier SE, Hardy LW, Poteete AR (1991) Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol 222:67–87

Riedl R (1978) Order in living organisms. John Wiley & Sons, New York

Robeson JP, Goldschmidt RM, Curtiss R, III (1980) Potential of *Escherichia coli* isolated from nature to propagate cloning vectors. Nature 283:104–106

Schmidt U, Podar M, Stahl U, Perlman PS (1996) Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing in vitro and in vivo: Phenotypes and suppressor mutations. RNA 2:1161–1172

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster.* Mol Biol Evol 15:1751–1760

Sommer SS, Kettterling RP (1994) How precisely can data from transgenic mouse mutation-detection systems be extrapolated to humans? Lesions from the human factor IX gene. Mutat Res 307:517–531

Stern S, Powers T, Changchien L-M, Noller HF (1989) RNA-protein interactions in 30S ribosomal subunits: Folding and function of 16S rRNA. Science 244:783–790

Stuart K (1991) RNA editing in mitochondrial mRNA of trypanosomatids. TIBS 16:68–72

Stuart K, Allen TE, Heidmann S, Seiwert SD (1997) RNA editing in kinetoplastid protozoa. Microbiol Mol Biol Rev 61:105–120

Thomas DC, Roberts JD, Sabatino RD, et al. (1991) Fidelity of mammalian DNA replication and replicative DNA polymerases. Biochemistry 30:11751–11759

Thomson KS (1985) Essay review: The relationship between development and evolution. Oxford Surv Evol Biol 2:220–233

Vrba ES, Eldredge N (1984) Individuals, hierarchies and processes: Towards a more complete evolutionary theory. Paleobiology 10:146–171

Wallace B (1991) The manly art of self-defense: On the neutrality of fitness components. Q Rev Biol 66:455–465

Weeks KM, Cech TR (1996) Assembly of a ribonucleoprotein catalyst by tertiary structure capture. Science 271:345–348

Weissman C, Cattaneo R, Billeter MA (1990) Sometimes an editor makes sense. Nature 343:697–699

Williams GC (1966) Adaptation and natural selection: A critique of some current evolutionary thought. Princeton University Press, Princeton, NJ

Wilson RJ, Williamson DH (1997) Extrachromosomal DNA in the Apicomplexa. Microbiol Mol Biol Rev 61:1–16

Wolfe KH (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285

Yost HJ, Petersen RB, Lindquist S (1990) RNA metabolism: Strategies for regulation in the heat shock response. Trends Genetics 6:223–227

Zuckerkandl E (1992) Revisiting junk DNA. J Mol Evol 34:259–271

Zuckerkandl E (1997) Neutral and nonneutral mutations: The creative mix—Evolution of complexity in gene interaction systems. J Mol Evol 44:S2–S8