

Mutational Biases

Lev Y Yampolsky, *East Tennessee State University, Johnson City, Tennessee, USA*

Arlin Stoltzfus, *National Institute of Standards and Technology, Gaithersburg, Maryland, USA*

Based in part on the previous version of this Encyclopedia of Life Sciences (ELS) article, Mutational Biases by Lev Y Yampolsky.

Mutational biases refer to systematic asymmetries or nonuniformities in the occurrence of mutations, heritable changes that take place in an individual organism. Mutational biases arise by asymmetries in damage, repair and replication of the genetic material. Many such biases are known: familiar examples include transition:transversion bias, the CpG effect and deletion:insertion bias. Mutation biases are biologically significant to the extent that they modulate the probabilities of heritable diseases and of evolutionary changes.

Nature of Mutation and Mutational Biases

Mutation (genetic mutation) is a process of heritable change; the outcome or product of this process, which may happen in germline or somatic tissue, is a mutation or mutant. The process of genetic mutation occurs in all living systems. It can be detected and studied under controlled conditions in populations of laboratory organisms, in cultured cells and in cell-free replication and repair systems.

In practice, biologists frequently apply the term 'mutation' to higher-level outcomes or processes that go beyond mutation, but that implicate specific mutational events, e.g. evolved differences between species and population variants, often are referred to as 'mutations' to the extent that they implicate specific mutational events. In the present context, such broad usage would be confusing: here, 'mutation' refers specifically to the process that happens in individual organisms; a mutation bias is a bias inherent to this process.

Typically mutation is described as 'random' without any clear explanation of what this means. Given that, the chance of success for predicting an individual mutational outcome is extremely low, mutation may be said to be 'random' in the sense of 'unpredictable'. However, the process of mutation is not 'random' in the sense of having uniformly distributed outcomes, i.e. the sense of 'random' that applies to flipping a coin or drawing a card from a well-shuffled deck. Individual mutations may have vastly different rates, as shown in early studies of bacteriophage T4 that

revealed rates of specific point mutations from 10^{-4} to 10^{-11} , a range of 7 orders of magnitude (Drake, 1970). More typically, because specific mutations are so rare and thus difficult to detect, most available knowledge of mutation bias is in regard to differences in the *mean rates* for different classes of mutation, averaged over many specific mutations of the same type. Differences such as this may be described as *mutational nonuniformities, asymmetries or biases*. Sometimes the term *mutation spectrum* is used to refer to the entire distribution of mutations or mutation rates relevant to some segment of deoxyribonucleic acid (DNA), or to some disease phenotype.

Although most current knowledge of mutation comes from nonhuman systems, and the details of mutation may differ, the basic concept of mutation bias – as a reflection of the biochemistry and enzymology of DNA damage, repair and replication – is universal. Presumably for any category of mutation, or any dimension of measurement of mutational effects, a detailed study would reveal asymmetries or biases in mutation. However, few such biases have been characterized. This article deals with a few examples of mutational biases, the molecular mechanisms responsible for these biases and the significance of mutation bias. **See also:** [DNA Repair](#); [Gene Conversion](#); [Molecular Evolution](#); [Neutral Theory](#); [Mutagenesis](#); [Mutagenesis Mechanisms](#); [Mutations and New Variation](#); [Overview](#)

Sources of information about mutation

Owing to the rareness of mutational events, knowledge of mutation rates and patterns is difficult to acquire by direct means. Most knowledge about mutation rates is based on (i) using laboratory detection systems by which large numbers of possible variants can be screened (or selected) based on phenotypic effects of mutations (e.g. Bebenek and Kunkel, 1995); (ii) comparing sequences that have diverged evolutionarily for very large numbers of generations (e.g. Andersson and Andersson, 1999) and (iii) analysing naturally occurring genetic variation sampled from a population of organisms (e.g. Kondrashov, 2003).

Advanced article

Article Contents

- Nature of Mutation and Mutational Biases
- Biases in Point Mutations
- Other Types of Mutational Biases
- Evolutionary Significance

Online posting date: 15th July 2008

ELS subject area: Evolution and Diversity of Life

How to cite:

Yampolsky, Lev Y; and, Stoltzfus, Arlin (July 2008) Mutational Biases. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0001729.pub2

An experimental screening system widely used by mutation researchers is based on the phenomenon of ‘ α complementation’, whereby the α fragment of the LacZ enzyme can restore activity to a fragment corresponding to the rest of the enzyme. In the mutation assay, a DNA sequence encoding LacZ α is mass-produced to serve as a mutational target, then copies are subjected to mutation, expressed and assayed for complementation (loss of activity signifies a mutation). Mutant copies are sequenced to identify the specific mutation involved. In other experimental systems, detection relies not on screening, but on a *selection* scheme in which mutations in the target sequence provide a growth advantage over nonmutants. Experimental systems of both types often are subjected to unnatural conditions that enhance the rate of mutation, e.g. chemical mutagens, or the use of a strain with aberrant DNA repair enzymes (Schaaper and Dunn, 1987), conditions that limit the usefulness of results.

To convert observations of detected mutants into rates of mutational events per generation is difficult, for various reasons. The outcome of a single mutation event might be reflected in several separately detected mutants, and this effect of propagation must be taken into account with appropriate statistics (Rosche and Foster, 2000). The set of detectable mutations in the target sequence may not be completely known. The detectable mutants may not be uniform in detectability, resulting in ascertainment biases.

The analysis of evolutionary divergence, or of population variation, allows a systematic study of mutation under more natural conditions, but must take into account effects of natural selection. For studies of divergence, typically this is accomplished by choosing only those sequences for which the fates of mutant alleles would be unaffected, or largely unaffected, by natural selection, i.e. sequences thought to evolve wholly or predominantly by neutral evolution such as pseudogenes or intergenic regions. In the case of neutral evolution, the rate of evolution is numerically equal to the

neutral mutation rate (Kimura, 1983). Thus, Andersson and Andersson (1999) compared metK pseudogenes in closely related *Rickettsia* species, finding evidence of a strong bias of deletions over insertions and of transitions over transversions, and a weaker bias from GC (guanine, cytosine) to AT (adenine, thymine) base pairs. Likewise, Nachman and Crowell (2000) studied primate pseudogenes and discovered differences in transition versus transversion rates and a strong CpG bias, followed by a number of similar studies (see **Table 1** below). In some cases, information about mutational patterns is inferred from the analysis of human genetic diseases (Krawczak *et al.*, 1998), using generalized databases such as the Human Gene Mutation Database (Stenson *et al.*, 2003).

Biases in Point Mutations

Mutations that do not alter the overall organization of a chromosome, but affect only a single localized point on the genetic map are called ‘point mutations’. Point mutations, though perhaps not always the most important mutations, are the most well studied.

Transition–transversion bias

Replacements of one nucleotide for another typically are the most common point mutations. Among nucleotide replacements, ‘transitions’ are replacements of a purine by another purine, or of a pyrimidine by another pyrimidine; other replacements are ‘transversions’. If all rates of mutation were equal, then, given four possible transitions and eight transversions (**Figure 1**), one would expect transversion mutations twice as often, but this is not observed. In fact, transitions typically are more numerous due to transition:transversion bias in mutation rates.

Table 1 Substitution biases in hominid genome according to several recent estimates

	Transition/ transversion bias ^a	CpGv bias (transversions)	CpGi bias (transitions)	Data type	Reference
μ	$\mu_i:\mu_v$	$\mu_{\text{CpGv}}:\mu_v$	$\mu_{\text{CpGi}}:\mu_i$		
0.55 ^b	4.4	8.0	13.3	Divergence in pseudogenes	Nachman and Crowell (2000)
–	4.2	2.2	5.4	Divergence in pseudogenes	Zhang and Gerstein (2003)
0.76 ^b	2.6	11.4	28.2	Divergence in introns	Yampolsky <i>et al.</i> (2005)
–	3.6	8.2	21.8	Divergence in random sequences ^c	Ebersberger <i>et al.</i> (2002)
–	2.4	2.6	10.8	SNPs in introns	Yampolsky <i>et al.</i> (2005)
0.39	3.8	2.8	15.3	Mendelian diseases	Kondrashov (2003)

Notes: μ_v , basal mutation rate (non-CpG transversions), 10^{-8} ; μ_i , non-CpG transition rate; μ_{CpGv} , rate of transversions at CpG sites; μ_{CpGi} , rate of transitions at CpG sites and SNPs, single nucleotide polymorphisms. Divergence data since common ancestor with chimpanzee.

^aCalculated per individual pair of source and destination nucleotides, not per nucleotide site.

^bAssuming 5 million years since common ancestor and 20 years generation time.

^cExons excluded.

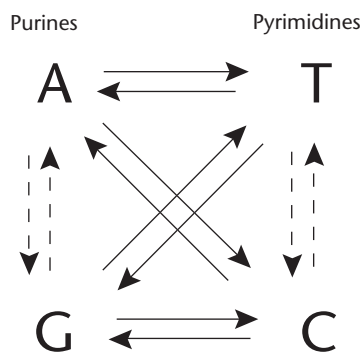


Figure 1 Nucleotide transversions (solid lines) and transitions (dashed lines) typically occur with different rates (A, adenine; G, guanine; C, cytosine; T, thymine).

Spontaneous transitions during DNA replication are believed to occur through noncanonical pairing of nucleotides, due to a shift of one of the nucleotides to a rare tautomeric form. For instance, the imino tautomer of A can pair with a C instead of a T. During replication, either the template or the incoming nucleotide can be tautomeric for such a mispairing to occur. In the next cycle of replication, an A:C mispair that has escaped DNA repair will give rise to one A:T pair of the original type, and a mutant G:C pair representing an A→G transition mutation.

Spontaneous transversions during replication are believed to arise through a combination of two events – tautomerization and base rotation. If an A shifts to its rare tautomeric (imino) form, it can pair with another A if this second A undergoes rotation around its glycosidic bond. Such rotation causes the so-called *syn*-conformation of DNA, also known as Z-DNA. If the aberrant A:A pair survives to the next replication cycle, it will give rise to an A:T pair of the original type, along with a mutant T:A pair representing a transversion mutation. Only approximately 10% of all DNA is in *syn*-conformation at any given time, which may explain why the rate of transversions is lower than that of transitions.

The combination of transition:transversion and CpG bias (see below) results in particularly high rates of C→T and G→A transitions (**Figure 2**) apparent in human-to-chimpanzee pseudogene comparisons (Nachman and Crowell, 2000) and in human hereditary diseases (Krawczak *et al.*, 1998). There is a good agreement on the value of transition/transversion bias in human genome: most studies based on a variety of data estimate it as close to 4-fold (when calculated per pair of source and destination nucleotides, which reflects the fact that there are eight possible transversions and four possible transitions) or to 2-fold, when calculated per nucleotide site (**Table 1**). **See also:** [Base Pairing in DNA: Unusual Patterns](#); [DNA Structure: A-, B- and Z-DNA Helix Families](#)

CpG bias

CpG dinucleotides are well-known ‘hot spots’ for C→T and G→A transition mutations (**Figures 2 and 3**), a

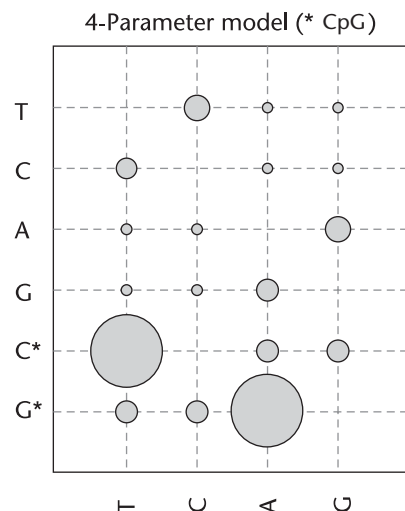


Figure 2 A bubble plot showing the relative rates of different types of DNA substitution mutations typical for mammalian genomes (based on results from Hwang and Green, 2004). The size of the bubble represents the relative rate of mutation of a nucleotide shown on the vertical axis to the nucleotide shown on the horizontal axis (asterisks, CpG context).

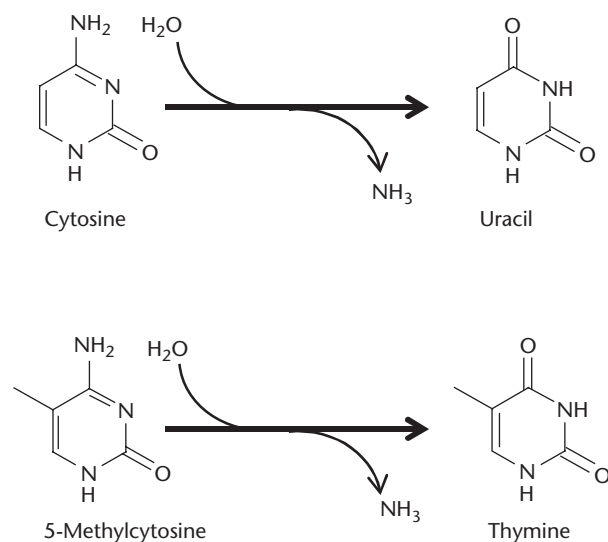


Figure 3 While deamination of C results in uracil, recognized by repair enzymes as an error, deamination of 5-methylcytosine results in T, which is not recognized as an error because it is normally present in DNA. In vertebrates, methylation of Cs occurs preferentially at sites followed by a G residue, leading to an unusual mutation pattern at CpG dinucleotides.

phenomenon known as CpG bias (the ‘p’ in ‘CpG’ refers to the phosphodiester bond connecting adjacent C and G nucleotides on the same DNA strand). Approximately one-quarter of evolved single-nucleotide differences in human and chimpanzee lineages have occurred at these sites (Nachman and Crowell, 2000). **See also:** [DNA Methylation and Mutation](#)

The cause of this bias is reasonably well known to involve a cascade of events in DNA modification, damage and repair. In vertebrates (and to a lesser extent, other animals), C residues that are followed by a G residue are much more likely to be modified by addition of a methyl group, resulting in 5-methyl-cytosine. Although oxidative deamination of C results in uracil, which is excised efficiently from DNA by repair enzymes, allowing for subsequent repair, oxidative deamination of 5-methyl-cytosine results in T, which is a normal DNA base and is not excised as efficiently (Figure 3). As a result, C residues in the CpG context are much likely to undergo a transition to T, with the complementary effect that G residues in a CpG context are much more likely to undergo a transition to A. The effect is on the order of 10-fold in humans (Table 1), although different studies give a wide range of estimates. Thus, DNA methylation, which typically protects DNA and also may regulate gene expression, radically changes mutation rates, an effect mediated by cellular processes of DNA damage and repair. Evolutionary modelling suggests that the rate of CpG transitions increased 4–8-fold in conjunction with the mammalian radiation 90 million years ago (Arndt *et al.*, 2003). **See also:** [DNA Methylation](#)

GC to AT bias

Forty years ago, Cox and Yanofsky (1967) isolated a mutant strain of *Escherichia coli* with a biased pattern of mutation favouring GC over AT base pairs. This kind of bias, whether towards G:C and C:G pairs, or in the reverse direction, towards A:T and T:A pairs, is called a GC to AT bias (sometimes just a 'GC' bias). GC to AT bias often is invoked to account for the fact that genomes of prokaryotes show large systematic differences in their genomic GC content (portion of G:C and C:G pairs), from 0.25 to 0.75 (Sueoka, 1988). This range is even broader if only third positions of codons are considered (0.05–0.97).

Eukaryotic genomes differ less widely in their GC content, but some eukaryotes, notably vertebrates, exhibit a remarkable subgenomic organization into GC-rich 'isochores' distinct from background regions (Sueoka and Kawanishi, 2000). A fascinating evolutionary hypothesis explored by Fryxell and Zuckerkandl (2000) is that deamination of methylated C (see earlier) may have played a key role in the evolution of isochores. C deamination occurs at much higher rates in partly melted (temporarily single-stranded) DNA. Partial DNA melting is much less likely in GC-rich regions of DNA than in GC-poor regions because G:C base pairs are stronger than A:T pairs. Thus, C deamination, which causes a majority of GC-reducing C to T transitions in mammals (see above), occurs at much lower rates in GC-rich areas than in GC-poor areas. These dynamics represent a positive feedback loop: a genomic region that is somewhat GC-poor will tend to further decrease its GC content, perhaps leading to the 'isochore' pattern. **See also:** [Isochores](#)

However, the role of GC mutation bias in isochore evolution is unresolved; some studies implicate biased

gene conversion (Meunier and Duret, 2004). DNA repair mechanisms acting on mismatches in heteroduplex regions during gene conversion tend to favour CG over AT base pairs (Marais, 2003). In humans, gene conversion bias towards the G allele over the T allele in GT mismatches may be as high as 24 times (Marais, 2003, Box 2). **See also:** [Base Composition Patterns](#); [Biased Gene Conversion and its Impact on Human Genome Evolution](#); [Gene Conversion](#)

Strand-specific bias

Chargaff's rule, based on canonical pairing rules for nucleotides, holds that in double-stranded DNA, the frequencies of complementary nucleotides are equal ($A = T$; $C = G$). If the two strands are evolving neutrally without mutation bias, the same equality will hold *within a single strand*. In addition, from Chargaff's rule it follows that $(A + C) = (G + T) = 50\%$ within each strand. Deviations from these relationships within a neutrally evolving region would indicate that the two strands are not equal with respect to mutation rates. For instance, if TG and GT mismatches are more likely than AC and CA ones, the strand that is more mutation-prone will accumulate Gs and Ts. **See also:** [Chargaff](#), [Erwin](#)

Replication is one source of strand-specific bias, with the asymmetry arising from the fact that, during replication, the leading strand is synthesized continuously in the direction of movement of the replication fork, whereas the lagging strand is synthesized in pieces in the opposite direction. The pattern generated by replication-associated mutation biases would be asymmetric with respect to the origin of replication that determines whether a given strand in a flanking region is replicated in leading-strand or lagging-strand mode. Precisely this association with replication origins is found in bacterial genomes (Lobry, 1996). **See also:** [DNA Replication: Mammalian](#); [DNA Replication: Prokaryotes and Yeast](#)

Transcription is another factor in strand-asymmetric mutation, possibly more important than replication. The rate of AG transitions exceeds the rate of TC transitions in transcribed regions of human genes, while these rates are equal in nontranscribed regions; this asymmetry is particularly strong for A to G transitions (Green *et al.*, 2003). This leads to an excess of G + T over A + C on the coding strand, a pattern observed in a number of human genes. This pattern is consistent with the action of transcription-mediated repair, which allows mismatches another chance to be repaired in transcribed portions of genes before the next round of DNA replication. Transcription-mediated repair is strand-biased, as mismatched area is excised from the transcribed strand and repaired using the coding strand as the template. This mechanism does not generate any new strand asymmetry; however, it merely makes any asymmetries in base misinsertion apparent when viewed as changes observed on the coding strand (Green *et al.*, 2003).

Other Types of Mutational Biases

Although most of the systematic information available on mutation biases refers to point mutations (see earlier), in many other cases there is clear, though less systematic, evidence of mutation biases. For instance, transposon insertions are a kind of mutation, and such insertions often occur with ‘target site’ preferences that depend on the nucleotide sequence immediately upstream and downstream of the site of insertion. Presumably large deletions and duplications also exhibit biases in mutation. **See also:** [Retrotransposition and Human Disorders](#)

More information is available in the case of short insertions and deletions, often called ‘indels’ when it is not possible or practical to distinguish the insertions from the deletions. **Figure 4** shows data for ‘microindels’, insertions and deletions of 20 base pairs or less, based on human gene mutations. There is a marked deletion:insertion bias, and also a strong tendency for the rate of mutation to decrease with length, for both deletions and insertions.

Short indels may emerge during replication (or replicative repair) by a form of ectopic strand pairing called ‘slippage’, illustrated in **Figure 5**. Repetitive DNA areas such as microsatellite loci are particularly vulnerable to template slippage, which presumably explains the high rate of insertions and deletions in microsatellites.

The unusual mutational propensities of triplet repeats, in particular, are recognized as key causal factors in a number of neurological degenerative diseases in humans, including fragile X syndrome, spinobulbar muscular atrophy, myotonic dystrophy and Huntington disease (Sinden *et al.*, 2002), as well as many cancers. Typically the severity of

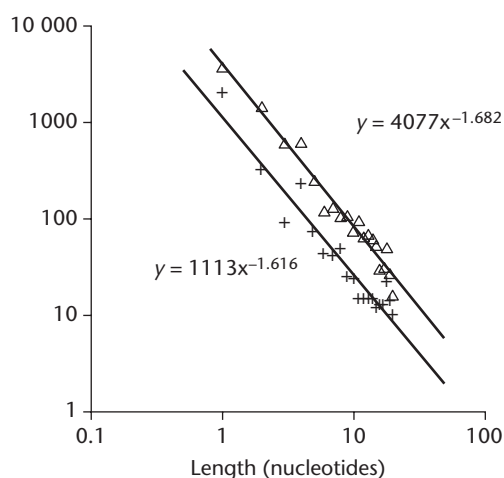


Figure 4 Numbers of short insertion (plus symbols) or deletion (triangles) mutants in the Human Gene Mutation Database (Chuzhanova *et al.*, 2003), shown on a double-log scale. Given reasonable assumptions (most such indels are null alleles; the chance of ascertainment is not significantly affected by length), variation in frequencies of different types (insertion or deletion) and lengths of indels directly reflects differences in rates of mutation. Thus, deletion mutations occur at a rate about 4 times that of insertions (from the slopes of the regression lines). The rates fall off sharply with length, according to a power law (grey regression lines).

effects depends on the number of repeats: individuals with large numbers of repeats have a hereditary susceptibility, with pathological effects that are aggravated as the number of repeats expands further. Nucleotide triplets such as CTG (complement, CAG) are particularly unstable.

Although the molecular causes of these diseases are diverse (e.g. chromosomal breaks in fragile X syndrome; accumulation of an abnormal protein in Huntington disease), they all typically exhibit anticipation (aggravation of symptoms in consecutive generations) and are most likely caused by identical mutagenesis mechanism. This mechanism, known as trinucleotide repeat expansion, is similar to polymerase slippage illustrated in **Figure 5**, but is thought to be enhanced by the formation of alternative DNA structures within triplet repeats, such as triplexes, quadruplexes and slipped-strand structures on double-stranded DNA and single-strand hairpins, which form upon strand separation. The longer is the run of trinucleotide repeats, the more pronounced are the unusual helical properties of the DNA. **See also:** [Expanding Mutations/Genetic Anticipation](#); [Repetitive Elements and Human Disorders](#); [Trinucleotide Repeat Expansions: Disorders](#); [Trinucleotide Repeat Expansions: Mechanisms and Disease Associations](#)

Evolutionary Significance

The importance of mutation bias in molecular evolution has been recognized for several decades (Stoltzfus, 2006). Transition–transversion bias, for instance, is a key feature of sequence evolution that reflects transition–transversion bias in mutation (Wakeley, 1996). Systematic differences in GC content within and among genomes often are attributed to systematic biases in mutation (Frank and Lobry, 1999). In *Drosophila*, deletions are more likely than insertions, and this bias has been invoked to explain the compact genome of *Drosophila*, which has relatively few and short noncoding regions (Petrov, 2002).

The role of mutation bias has a larger significance for evolutionary theory, but not in regard to the Lamarckian theory, which supposes that mutations that improve adaptation occur preferentially. The mutation biases discussed earlier are *not* of this type and, indeed, Lamarckian mutations are not reliably known to occur in any system. The main issue for contemporary evolutionary biologists is the sufficiency of the Modern Synthesis (neo-Darwinian) view in which selection is assumed to be the ultimate source of explanation in evolution, with mutational sources being unimportant on the grounds that each species has a ‘gene pool’ of available variation sufficient to meet any challenge, without new mutations. The ‘mutationist’ view, by contrast, allows the kind of dependence of evolution on new mutations that seems necessary to account for the relationship between mutational tendencies and evolutionary tendencies (Stoltzfus, 2006). **See also:** [Molecular Evolution: Patterns and Rates](#); [Mutations and New Variation: Overview](#)

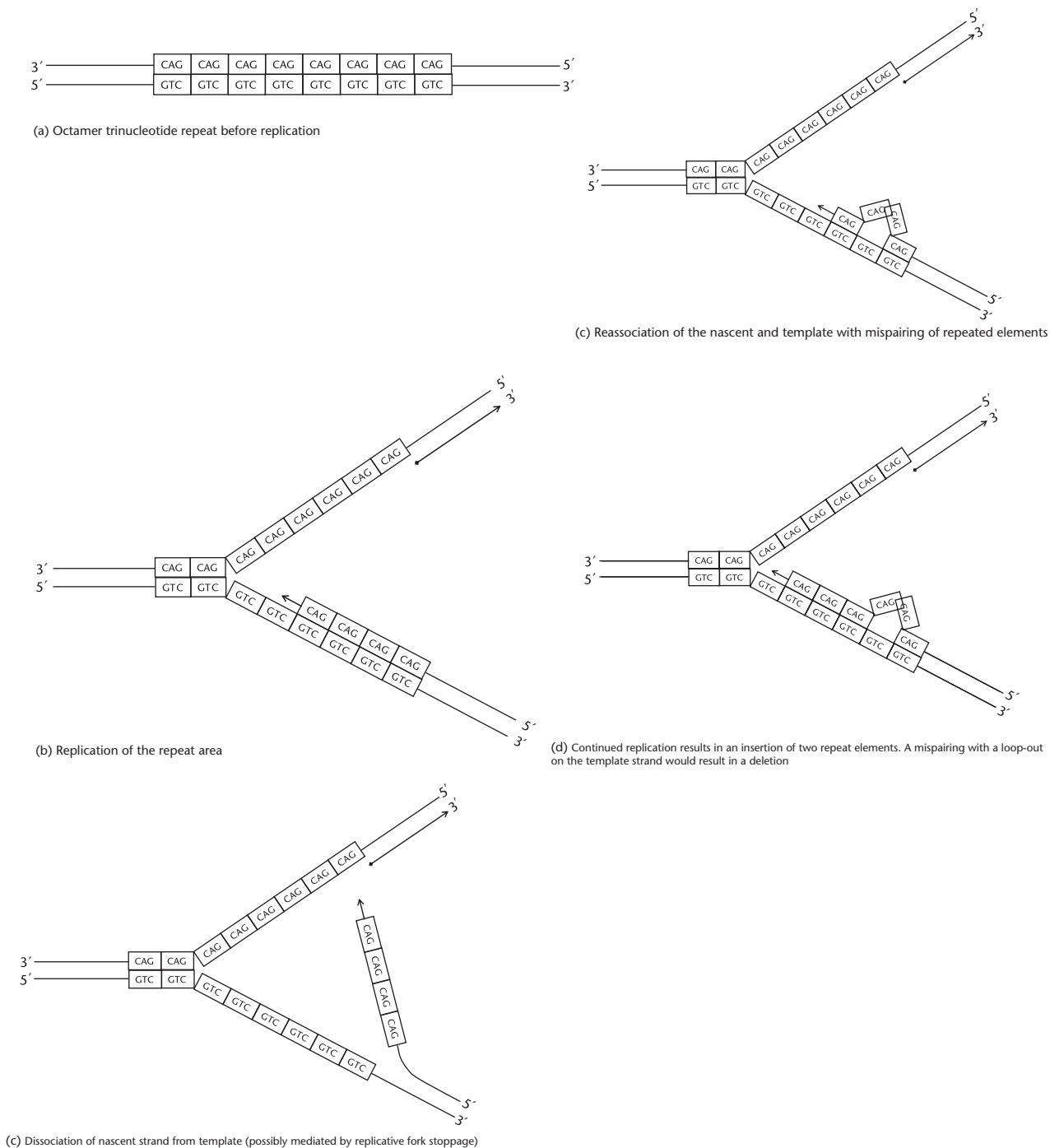


Figure 5 Slipped-strand mispairing (DNA polymerase slippage) causes insertions and deletions during replication of repetitive DNA.

References

Andersson JO and Andersson SG (1999) Genome degradation is an ongoing process in *Rickettsia*. *Molecular Biology and Evolution* **16**: 1178–1191.

Arndt PF, Petrov DA and Hwa T (2003) Distinct changes of genomic biases in nucleotide substitution at the time of

mammalian radiation. *Molecular Biology and Evolution* **20**(11): 1887–1896.

Bebenek K and Kunkel TA (1995) Analyzing fidelity of DNA polymerases. *Methods in Enzymology* **262**: 217–232.

Chuzhanova NA, Anassis EJ, Ball EV, Krawczak M and Cooper DN (2003) Meta-analysis of indels causing human genetic disease: mechanisms of mutagenesis and the role of

- local DNA sequence complexity. *Human Mutation* **21**(1): 28–44.
- Cox EC and Yanofsky C (1967) Altered base ratios in the DNA of an *Escherichia coli* mutator strain. *Proceedings of the National Academy of Sciences of the USA* **58**: 1895–1902.
- Drake JW (1970) *The Molecular Basis of Mutation*. San Francisco: Holden-Day.
- Ebersberger I, Metzler D, Schwarz C and Pääbo S (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. *American Journal of Human Genetics* **70**(6): 1490–1497.
- Frank AC and Lobry JR (1999) Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* **238**: 65–77.
- Fryxell KJ and Zuckerkandl E (2000) Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Molecular Biology and Evolution* **17**: 1371–1383.
- Green P, Ewing B, Killer W *et al.* (2003) Transcription-associated mutational asymmetry in mammalian evolution. *Nature Genetics* **33**(4): 514–517.
- Hwang DG and Green P (2004) Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proceedings of the National Academy of Sciences of the USA* **101**: 13994–14001.
- Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Human Mutation* **21**: 12–27.
- Krawczak M, Ball EV and Cooper DN (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *American Journal of Human Genetics* **63**: 474–488.
- Lobry JR (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biology and Evolution* **13**: 660–665.
- Marais G (2003) Biased gene conversion: implications for genome and sex evolution. *Trends in Genetics* **19**: 330–338.
- Meunier J and Duret L (2004) Recombination drives the evolution of the GC-content in the human genome. *Molecular Biology and Evolution* **21**(6): 984–990.
- Nachman MW and Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**: 297–304.
- Petrov DA (2002) DNA loss and evolution of genome size in *Drosophila*. *Genetica* **115**: 81–91.
- Rosche WA and Foster PL (2000) Determining mutation rates in bacterial populations. *Methods* **20**: 4–17.
- Schaaper RM and Dunn RL (1987) Spectra of spontaneous mutations in *E. coli* strains defective in mismatch correction: the nature of in vivo DNA replication errors. *Proceedings of the National Academy of Sciences of the USA* **84**: 6220–6224.
- Sinden RR, Potaman VN, Oussatcheva EA *et al.* (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *Journal of Bioscience* **27**: 53–65.
- Stenson PD, Ball EV, Mort M *et al.* (2003) Human Gene Mutation Database (HGMD): 2003 update. *Human Mutation* **21**(6): 577–581.
- Stoltzfus A (2006) Mutationism and the dual causation of evolutionary change. *Evolution & Development* **8**: 304–317.
- Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the USA* **85**: 2653–2657.
- Sueoka N and Kawanishi Y (2000) DNA G+C content of the third codon position and codon usage biases of human genes. *Gene* **261**: 53–62.
- Wakeley J (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in Ecology and Evolution* **11**: 158–162.
- Yampolsky LY, Kondrashov FA and Kondrashov AS (2005) Distribution of the strength of selection against amino acid replacements in human proteins. *Human Molecular Genetics* **14**(21): 3191–3201.
- Zhang Z and Gerstein M (2003) Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Research* **31**(18): 5338–5348.

Further Reading

- Duret L, Eyre-Walker A and Galtier N (2006) A new perspective on isochore evolution. *Gene* **385**: 71–74.
- Maki H (2002) Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annual Review of Genetics* **36**: 279–303.
- Montoya-Burgos JI, Boursot P and Galtier N (2003) Recombination explains isochores in mammalian genomes. *Trends in Genetics* **19**(3): 128–130.
- Petrov DA and Hartl DL (1999) Patterns of nucleotide substitution in *Drosophila* and mammalian genomes. *Proceedings of the National Academy of Sciences of the USA* **96**: 1475–1479.
- Qu HQ, Lawrence SG, Guo F, Majewski J and Polychronakos C (2006) Strand bias in complementary single-nucleotide polymorphisms of transcribed human sequences: evidence for functional effects of synonymous polymorphisms. *BMC Genomics* **7**: 213.
- Rubinsztein DC, Amos B and Cooper G (1999) Microsatellite and trinucleotide-repeat evolution: evidence for mutational bias and different rates of evolution in different lineages. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **354**: 1095–1099.
- Schmeigner C, Hoegel J, Vogel W and Assum G (2007) The rate, not the spectrum, of base pair substitutions changes at a GC-content transition in the human NF1 gene region: implications for the evolution of the mammalian genome structure. *Genetics* **175**(1): 421–428.