# Mutational Processes

**Michael T. Clegg**
**Peter L. Morrell**
*University of California, Riverside, California, U.S.A.*

## INTRODUCTION

Mutation is a permanent heritable change in the DNA of an individual organism. It is the ultimate source of all the biological diversity on earth. Life could not adapt and diversify without the substrate of heritable variation. In agriculture the exploitation of novel heritable variants has been the engine of plant and animal improvement. This article outlines the mechanisms that are responsible for heritable changes in DNA.

## NUCLEOTIDE SUBSTITUTION EVENTS

Many different categories of mutation affect DNA molecules. The most regular form of mutational change is when one nucleotide is incorrectly replaced by a different nucleotide during replication (Fig. 1). These nucleotide substitutions are assumed to be stochastically regular in time. (Like radioactive decay, each mutation is assumed to arrive randomly, but with a constant probability in time.) Nucleotide substitution events are also assumed to occur with the same probability at every nucleotide site, independent of the nucleotide state at the mutated site or at neighboring sites. Much of our direct knowledge of the nucleotide substitution processes comes from the study of bacteria or viruses where many generations and very large populations of organisms can be studied in the laboratory. These studies indicate that the DNA replication process has a remarkably high fidelity, with a nucleotide substitution error rate on the order of one error per one billion replications per site (or $10^{-9}$ errors per site per generation).[1] These studies have also tended to validate the assumptions of constancy and independence previously mentioned. Estimates for eukaryotic organisms like crop plants and animals cannot be obtained from direct observation and must be calculated indirectly. Indirect calculations are based on comparisons between homologous gene sequences separated for some period of evolutionary time. Statistical models of the substitution process must be invoked, together with various assumptions about the absence of natural selection, to arrive at estimates of mutation rates. Nevertheless, indirect estimates from, for example, large numbers of mammalian genes are consistent with those from bacteria, indicating an average nucleotide substitution rate of roughly $4 \times 10^{-9}$ per site per generation.[1]

## INSERTION/DELETION EVENTS

Another major category of mutation is the insertion or deletion of stretches of DNA (indels). Indels are most frequently observed in areas of untranslated DNA (DNA that does not code for a protein). Much of the genome of eukaryotic organisms is made up of untranslated regions, so on a genomic scale, a substantial fraction of mutation events are likely to be indels. Indels arise from a number of causes, including slipped-strand-mispairing in replication, unequal crossover events, or transposable element insertions. In some cases indels have been shown to depend on local nucleotide context, so the assumption of independence over DNA region cannot be invoked.[2] Moreover, the assumption of stochastic regularity in time is also of questionable validity, at least for some forms of transposable element insertions. Finally, the alignment of indels is difficult when sequences have diverged for a substantial length of time, because overlapping events cannot be distinguished. All of these factors severely limit our ability to use mathematical models and to make statistical calculations based on indel data.

### Slipped-Strand Mispairing

Despite the fact that slipped-strand mispairing events are context-dependent and are usually inferred indirectly rather than observed, it is desirable to get some sense of the rate at which these kinds of events accumulate. The effects of replication slippage or slipped-strand mispairing generally cannot be observed directly and must be inferred based on the distribution of repetitive nucleotide sequence or base composition surrounding indels (Fig. 2). Estimates from *Drosophila* nucleotide sequence data suggested approximately 0.16 indels per nucleotide substitution.[3,4] Estimates based on chloroplast data from grasses suggested that indels and nucleotide substitutions contributed nearly equal proportions of total mutations; however, the authors point out that the relative contribution of indels



*Encyclopedia of Plant and Crop Science*
DOI: 10.1081/E-EPCS 120017011

GACACGCCG
GACACGCCG
GGCACCCCC

**Fig. 1** An alignment of DNA sequence data. Three nucleotide substitutions are underlined.

appears to diminish among more highly diverged species because of apparent superimposition of indels on previous indels.[5] Slipped-strand mispairing appears to play a major role in indel generation, and may be particularly important in the proliferation of repetitive sequences, such as the short two- and three-base motifs that make up microsatellites.[6]

### Unequal Crossover Events

At tandemly repeated genes there is potential for out-of-register recombination, which can result in duplication, deletion, or truncation of a chromosomal region. Such events may be especially common in tandemly repeated arrays of genes such as nuclear ribosomal DNA.

### The Interaction Between Mutation and Recombination

Genetic recombination is the process that disassociates mutations along a chromosome, typically through cross over—the breaking and reannealing of homologous portions of parental chromosomes during meiosis. A second form of recombinational exchange is commonly observed between different tandemly repeated gene family members. When such an exchange occurs, a mutant site on one repeated gene copy is transmitted to an adjacent gene copy, causing the repeated copies to become more similar in sequence than would otherwise be the case. This process is called concerted evolution. A related form of recombinational exchange that generates new alleles is observed in microsatellite loci where unequal exchange between tandemly arranged di- or trinucleotide repeats (or higher-order repeats) causes an increase or decrease in repeat number. Finally, unequal exchanges between repeated elements at different chromosomal locations can cause duplications and deficiencies of larger DNA regions.

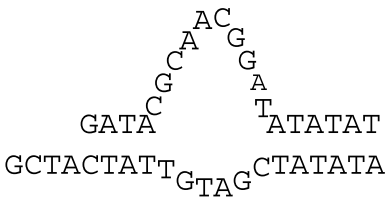
## TRANSPOSON-INDUCED MUTATIONS

Mobile elements (transposons) are a ubiquitous feature of plant and animal genomes. There are two broad categories of mobile elements: class I elements that replicate from an RNA intermediate via reverse transcriptase, and class II or DNA elements that replicate via a cut-and-paste mechanism. Both classes of elements are relatively common in plant and animal genomes. Class II elements tend to be associated with elevated transposition rates, but both categories are clearly implicated as causal agents in many mutations, owing to their ability to insert into or adjacent to genes, and thereby disrupt or alter gene function or gene expression.

One important aspect of transposon-mediated mutation is the ability to alter gene expression patterns. This may occur because the element has inserted into a 5′ UTR (untranslated region in the 5′ region of a gene) and has thereby disrupted transcription factor-binding sequence motifs, or it may occur because sequence motifs within the insertion act as novel sites for transcription factor binding. This latter capacity has been documented in a number of cases in plants and animals, and illustrates the creative acquisition of new gene expression patterns through the transposition of appropriate sequence features around the genome. The fact that transposable elements can move entire sequence motifs into new genic environments and induce novel patterns of expression tells us transposon mutation is modular and acts at a level beyond the individual nucleotide site.

## GENE DUPLICATION

One of the important discoveries of the genomics era is the fact that many genes are redundant in plant and animal




CGATGATACGCAACGGATATATATA
GCTACTATGCGTTGCCTATATATAT

GATAACATCGATATAT
CTATTGTAGCTATATA

**Fig. 2** Slipped-strand mispairing between noncontiguous repeats. Excision of the shorter single-stranded loop results in the sequence shown in the second alignment. Excision of the longer loop results in the third alignment. In the second and third alignments, the changed segments are underlined.

genomes. The processes that lead to the duplication of genes range from polyploidy (the doubling of entire genomes owing to the union of unreduced gametes, discussed next), to unequal crossover, to transposon-mediated transposition of genes. There are very few estimates of rates of gene duplication, but one recent estimate suggests a duplication rate of 0.01 per gene per million years.[7]

Transposons are also implicated in the movement of genes or fragments of genes around the genome. This may occur as a consequence of reverse transcription of mRNA molecules or because a gene or gene fragment has been acquired within a mobile element. In either case it represents a kind of higher-level mutation that may provide adaptive flexibility by placing genes into new regions, possibly in association with entirely different expression signals.

## POLYPLOIDY

In addition to gene duplication, polyploid formation can potentially induce or accelerate many of the mutational processes outlined in the foregoing discussion. Polyploidy is very common in flowering plants. Crop plants with polyploid origins include bananas, cotton, peanut, and wheat. In addition to extant polyploids, evidence of one or more rounds of ancient polyploid formation have been identified in species such as *Arabidopsis*[8,9] and maize.[10] Some of the ancient polyploid events appear to have occurred in lineages ancestral to the majority of flowering plants.[8,9]

Polyploids are frequently the result of hybridization between well differentiated parental species. Thus, there is a potential for recombination, gene-conversion, and concerted evolution to occur between partially homologous chromosomes from the parental species. Transposons from one parental genome may also invade the genome of the other; with gene duplication, there is the potential for loss of duplicate gene function or even the elimination of coding DNA sequence.[11]

## CONCLUSION

Agriculture developed and flourished because humans learned how to exploit and maintain useful hereditary variants in a wide diversity of plant and animal species. Since the rediscovery of Mendelian genetics, a major research goal has been the acquisition of a detailed understanding of the mechanisms that generate hereditary variation. Our knowledge of the mutational process has grown enormously in recent years, revealing that mutation is a complex phenomenon with many mechanisms. One important lesson has been that mutation operates at various levels, ranging from nucleotide changes to transposition to gene duplication to the doubling of entire genomes. This rich variety of biological process has led to the wealth of mutational diversity that we seek to exploit in adapting plants and animals to human needs.

## ARTICLES OF FURTHER INTEREST

*Bioinformatics*, p. 125
*Molecular Evolution*, p. 748
*Polyploidy*, p. 1038

## REFERENCES

1. Li, W.-H. *Molecular Evolution*; Sinauer Associates, Inc.: Sunderland, MA, 1997; 181.
2. Morton, B.R.; Clegg, M.T. A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). Curr. Genet. **1993**, *24* (4), 357–365.
3. Pritchard, J.K.; Schaeffer, S.W. Polymorphism and divergence at a *Drosophila* pseudogene locus. Genetics **1997**, *147* (1), 199–208.
4. Petrov, D.A.; Hartl, D.L. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol. Biol. Evol. **1998**, *15* (3), 293–302.
5. Golenberg, E.M.; Clegg, M.T.; Durbin, M.L.; Doebley, J.; Ma, D.P. Evolution of a noncoding region of the chloroplast genome. Mol. Phylogenet. Evol. **1993**, *2* (1), 52–64.
6. Levinson, G.; Gutman, G.A. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol. Biol. Evol. **1987**, *4* (3), 203–221.
7. Lynch, M.; Conery, J.S. The evolutionary fate and consequences of duplicate genes. Science **2000**, *290* (5494), 1151–1155.
8. Vision, T.J.; Brown, D.G.; Tanksley, S.D. The origins of genomic duplications in Arabidopsis. Science **2000**, *290* (5499), 2114–2117.
9. Bowers, J.E.; Chapman, B.A.; Rong, J.K.; Paterson, A.H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **2003**, *422* (6930), 433–438.
10. Gaut, B.S.; d' Ennequin; Peek, A.S.; Sawkins, M.C. Maize as a model for the evolution of plant nuclear genomes. Proc. Natl. Acad. Sci. U. S. A. **2000**, *97* (13), 7008–7015.
11. Wendel, J.F. Genome evolution in polyploids. Plant Mol. Biol. **2000**, *42*, 225–249



Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016