Genetic Variation and Molecular Evolution

Werner Arber
Biozentrum, University of Basel, Klingelbergstrasse 70, Basel, Switzerland

Keywords

Biological Evolution
A nondirected, dynamic process of diversification resulting from the steady interplay between spontaneous mutagenesis and natural selection.

DNA Rearrangement
Results from mostly enzyme-mediated recombination processes, which can be intra- or intermolecular.

Evolution Gene
Its protein product acts as a generator of genetic variations and/or as a modulator of the frequency of genetic variations.

Gene Acquisition
Results from horizontal transfer of genetic information from a donor cell to a receptor cell. With bacteria, this can occur in transformation, conjugation, or phage-mediated transduction.

Natural Selection
Results from the capacity of living organisms to cope with the encountered physicochemical and biological environments. Largely depending on its genetic setup and physiological phenotype, each organism may have either a selective advantage or a selective disadvantage as compared to the other organisms present in the same ecosystem.

Spontaneous Mutation
Defined here as any alteration of nucleotide sequences occurring to DNA without the intended intervention of an investigator. The term mutation is used here as a synonym of genetic variation.

Transposition
DNA rearrangement mediated by a mobile genetic element such as a bacterial insertion sequence (IS) element or a transposon.

Variation Generator
An enzyme or enzyme system whose mutagenic activity in the generation of genetic variation has been documented.

The comparison of DNA sequences of genes and entire genomes offers interesting insights into the possible evolutionary relatedness of genetic information of living organisms. Together with a relatively rich database from experimental microbial genetics, conclusions can be drawn on the molecular mechanisms by which genetic variations are spontaneously generated. A number of different specific mechanisms contribute to the overall mutagenesis. These mechanisms are here grouped into three natural strategies of the spontaneous generation of genetic variations: local changes of DNA sequences, intragenomic rearrangement of DNA segments, and acquisition of foreign DNA by horizontal gene transfer. These three strategies have different qualities with regard to their contributions to the evolutionary process. As a general rule, none of the known mechanisms producing genetic variants is clearly directed. Rather, the resulting alterations in the inherited genomes are more random. In addition, usually only a minority of resulting variants provide a selective advantage. Interestingly, in most of the molecular mechanisms involved, the products of so-called evolution genes are involved as generators of genetic variation and/or as modulators of the frequencies of genetic variation. Products of evolution genes work in tight collaboration with nongenetic factors such as structural flexibilities and chemical instabilities of molecules, chemical and physical mutagens, and random encounter. All of these aspects contributing to the spontaneous generation of genetic variations together form the core of the theory of molecular evolution. This theory brings neo-Darwinism to the molecular level. In view of the increasing evidence coming particularly from microbial genetics, knowledge of molecular evolution can be seen as a confirmation of Darwinism at the level of biologically active molecules, in particular, nucleic acids and proteins. Philosophical and practical implications of this knowledge will be briefly discussed.

1 Introduction

Evolutionary biology has traditionally devoted its major attention to the comparison of phenotypical traits of higher organisms, both of those actually living and of those that have been extinct (paleontological fossils). The resulting theory of descent is, together with other criteria, at the basis of the systematic classification of living organisms. Darwin's theory of natural selection brought a new element into the understanding of the long-term development of forms of life. Natural selection is the result of organisms coping with the encountered living conditions that are dependent on both the environmental physicochemical conditions and the activities of all living forms in a particular ecological niche. The Darwinian theory of evolution also postulated that intrinsic

properties of life are not entirely stable and principally identical for all organisms of a given species. In the so-called modern synthesis, in which evolutionary biology and genetics became integrated, transmissible phenotypic variations representing the substrate of natural selection were explained as due to genetic variations (or mutations). Shortly thereafter, deoxyribonucleic acid (DNA) was identified as the carrier of genetic information. DNA is thus also the target for mutagenesis. Within the last few decades a rapid development of molecular genetics with novel research strategies leading to genomics, sequencing of entire genomes, and functional studies of genes and their products paved the way for hitherto inaccessible knowledge on the basis of life and its multiple manifestations. This also relates to the process of molecular evolution. A synoptical insight into the various molecular mechanisms contributing to the generation of genetic variations represents a molecular synthesis between the neo-Darwinian theory and molecular genetics. This synthesis can confirm the Darwinian evolution at the molecular level.

2
Principles of Molecular Evolution

The principles of molecular evolution to be outlined here are founded on

1. the neo-Darwinian theory of evolution with its three pillars of genetic variation, natural selection, and isolation;
2. the solidly established microbial genetics database;
3. DNA sequence comparisons with bioinformatic tools;
4. physicochemical knowledge on the reactivity, conformational flexibility, and chemical stability of biologically active molecules.

2.1
Evolutionary Roles of Genetic Variation, Natural Selection, and Isolation

The long-term maintenance of any form of life requires a relatively high stability of its genetic information. However, rare occasional genetic variations occur in all organisms. This gives rise to mixed populations of organisms with the parental genome and organisms having one or more alterations in their genome. These populations are steadily submitted to natural selection. The experience shows that, in general, favorable genetic variations are considerably less frequent than unfavorable ones. The latter provide a selective disadvantage. Indeed, genetic variants with unfavorable genetic alterations will sooner or later get eliminated from propagating populations, which will become enriched for organisms carrying favorable genetic variations. It should be noted that by far not all alterations in the nucleotide sequences of a genome will lead to a change in the phenotype of the organism. However, such silent and neutral mutations may later become physiologically relevant in conjunction with still other, upcoming DNA sequence changes.

Natural selection is by no means a constant element. It varies both in time and in space. This is due to variations in the physicochemical environmental conditions as well as to variations of the life activities of all the different organisms present in a particular ecological niche and forming an ecosystem. Since a genetic variation may also affect the influence that the organism exerts on the other organisms present in the same ecosystem (e.g. think of weeds and pathogenicity

effects, but also of beneficial, synergistic effects), any novel mutation may not only influence the life of the concerned organism itself but also the lives of other cohabitants of the same ecological niche.

The third pillar of biological evolution – besides genetic variation and natural selection – is isolation. Evolutionary biologists define two different aspects of isolation. One of these is *geographic isolation*, which may seriously reduce the number of potential habitats for an organism. The other type of isolation is called *reproductive isolation*. For example, two distantly related diploid organisms may not be fertile in sexual reproduction. But reproductive isolation can also be seen in a wider definition to seriously limit the possibility of horizontal transfer of segments of genetic information between two different kinds of organisms.

As summarized in Fig. 1, genetic variation drives biological evolution. Complete genetic stability would render any

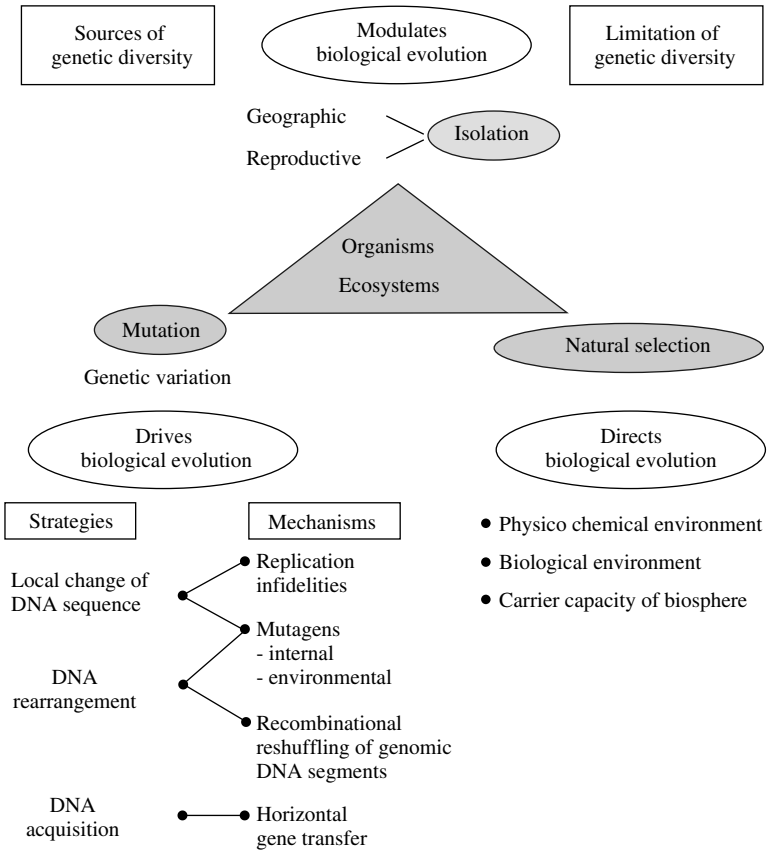

Fig. 1 Synoptical presentation of major elements of the theory of molecular evolution. A number of specific mechanisms, each with its own characteristics, contribute to the four groups of mechanisms of genetic variation listed. Each of the specific mechanisms follows one and sometimes more than one of the three principal, qualitatively different strategies of genetic variation.

evolutionary process impossible. A very high frequency of genetic variation would rapidly lead to the extinction of the concerned organisms because of the stated prevalence of unfavorable mutations in the spontaneous generation of genetic variations. It is natural selection together with the available sets of genetic variants that determines the direction of biological evolution, or in other words, the directions in which the branches of the tree of evolution grow. Finally, the geographic and reproductive isolations modulate the evolutionary process.

2.2 Molecular Mechanisms of the Generation of Genetic Variation

The concept to be presented here requires the reader to question some long-established textbook knowledge, such as the claim that spontaneous mutations would largely result from errors, accidents, and illegitimate processes. We defend here the alternative view that living nature actively cares for biological evolution

1. by making use of intrinsic properties of matter such as a certain degree of chemical instability and of structural flexibility of molecules, and
2. by having developed genetically encoded systems, the products of which are involved in the generation of genetic variations and in modulating the frequencies of genetic variation.

It might be relevant to mention here that the term mutation is differently defined in classical genetics and in molecular genetics. In classical genetics, a mutant is any variant of a parental form, showing in its phenotypic properties some alteration that becomes transmitted to the progeny. In contrast, it has become a habit in molecular genetics to call any alteration in the parental nucleotide sequence of the genome a mutation, whether it has phenotypic consequences or not. There is good reason to assume that in most spontaneously occurring mutagenesis events, the specific mechanism involved will not pay attention to whether the sequence alteration at the involved target site will cause a phenotypic alteration or not. Therefore, for studies on mechanisms and on the statistics of their occurrence, it is indicated to follow the molecular genetics definition of the term mutation, which we use here as a synonym of genetic variation. We will use the term *spontaneous mutation* to label any type of DNA sequence alteration unintended by the investigator. This definition says nothing about whether a mutation relates to a phenotypic change.

At present, the research on genetic variation mainly follows two strategies. First, the increasing availability of entire and partial genome sequences offers excellent means to compare regulatory sequence elements, specific domains of genes, entire genes, groups of genes, and entire genomes with regard to DNA sequence homologies and genome organization. Within a given species, this can reveal a genetic polymorphism. Between more or less related species, it can give a reliable measure for evolutionary relatedness. For example, the molecular clock – an indicator of evolutionary relatedness – is based on single nucleotide alterations. Sequence comparisons can often suggest how sequence alterations could have occurred in the course of past evolution. Secondly, a more reliable insight into the generation of genetic variations can be gained by the observation of individual processes of mutagenesis. In view of the large size of genomes and of the rare and random

occurrence of spontaneous DNA sequence alterations, this approach is relatively difficult. However, a rich database is already available from microbial genetics, particularly from bacteria and their viruses and plasmids. Their relatively small genomes are haploid so that phenotypic effects caused by genetic variation normally become rapidly manifested. With appropriate selection and screening techniques, this allows one to identify occasional, functionally relevant mutations in populations. On the other hand, investigations on structural alterations in the genomes of individual bacterial colonies, for example, by the study of restriction fragment length polymorphism, can reveal when and where on the genome a DNA rearrangement must have occurred.

A quite solid, general result of this type of experimental investigations reveals that in the spontaneous generation of genetic variations, it is not just a single mechanism at work. Rather, a number of mechanistically different processes contribute to the overall mutagenesis. We will discuss selected examples below. Interestingly, a critical evaluation of the situation shows that the specific mechanisms of mutagenesis often depend both on nongenetic elements and on activities of specific enzymes, the products of the so-called evolution genes. The multitude of thus identified, distinct mechanisms contributing to the formation of genetic variations can be grouped into three qualitatively different natural strategies (Fig. 1). These are

1. local changes in the DNA sequences,
2. rearrangement of DNA segments within the genome,
3. the acquisition by horizontal transfer of a DNA segment originating in another kind of organism.

Selected examples for each of these strategies will be discussed in Sects. 4, 5, and 6.

3 Genetic Variation in Bacteria

Several seminal discoveries, largely based on work carried out with microorganisms between 1943 and 1953, were essential for the later development of molecular genetics.

1. It was realized that bacteria and bacteriophages have genes that can mutate, and that spontaneous mutations in microorganisms normally arise independently of the presence of selective agents. It was also learned that the genetic information of bacteria and of some bacteriophages is carried in DNA molecules rather than in other biological macromolecules such as proteins.
2. The newly discovered phenomena of DNA transformation, bacterial conjugation, and bacteriophage-mediated transduction revealed natural means of horizontal gene transfer between different bacterial cells.
3. It was seen that horizontal gene transfer has natural limits, including systems of host-controlled modification, which are today known as *DNA restriction–modification systems*.
4. Mobile genetic elements were identified as sources of genetic instability and were seen to represent mediators of genetic rearrangements. While such rearrangements are often caused by transposition, they can also result from the integration of a bacteriophage genome into the genome of its bacterial host strain, which is thereby rendered lysogenic.

It is at the end of this fruitful period of discoveries, in 1953, that structural analysis of DNA molecules led to the double-helix model. The filamentous structure of DNA molecules made it clear as to how genetic information could be contained in the linear sequences of nucleotides. The double helical nature of the model also offered an understanding of semiconservative DNA replication at the molecular level and thus of information transfer into progeny.

Many classical microbial genetic investigations were carried out with *Escherichia coli* K-12. Its genome is a single circular DNA molecule (chromosome) of about 4.6×10^6 bp. In periods of growth, the rate of spontaneous mutagenesis is about 10^{-9} per bp and per generation. This represents one new mutation in every few hundred cells in each generation. *E. coli* has several well-studied bacteriophages and plasmids. This material facilitates investigations on life processes in these bacteria.

Under good growth conditions, the generation time of *E. coli*, measured between one cell division and the next, is very short, on the order of 30 min. Upon exponential growth, this leads to a multiplication factor of 1000 every 5 h. Thus, a population of 10^9 cells representing 30 generations is reached from an inoculum of a single cell in only 15 h. This rapid growth rate greatly facilitates population genetic studies and thus facilitates investigations on the evolutionary process.

On the filiform DNA molecules of *E. coli* and its bacteriophages and plasmids, the genetic information is relatively densely stored as linear sequences of nucleotides or base pairs. Genes depend on the presence of continuous sequences of base pairs (reading frames) that encode specific gene products, usually proteins, and of expression control signals that ensure the occurrence of gene expression at the relevant time with the needed efficiency. Mutations can affect reading frames as well as control signals, both of which represent specific DNA sequences. In addition, some specific DNA sequences relate to the control of the metabolism of the DNA molecules, in particular, their replication.

In the following Sects. 4, 5, and 6, we will describe selected examples of mechanisms contributing each in its specific manner to the formation of genetic variants. We will group them into the already mentioned three major natural strategies (Fig. 1), although, as we will later see, some of the specific mechanisms involve more than one of these strategies.

4
Local Changes in the DNA Sequences

The process of DNA replication is one of the important sources of genetic variation, which may depend both on structural features of the substrate DNA and on functional characteristics of the replication fork.

Some of the ''infidelities'' of DNA replication are likely to depend on tautomeric forms of nucleotides, that is, a structural flexibility inherent in these organic compounds. Base pairing depends on specific structural forms. Conformational changes of nucleotides can result in a mispairing if short-living, unstable tautomeric forms are ''correctly'' used in the synthesis of the new complementary strand upon DNA replication. For this reason, we do not consider mutations resulting in this process as errors and call them *infidelities*. DNA replication is indeed one of the sources of nucleotide substitution, and this plays an

important role in the evolutionary development of biological functions.

An inherent low degree of chemical instability of nucleotides represents another source of mutagenesis. For example, cytosine can undergo oxidative deamination to become uracil. Upon replication, this gives rise to an altered base pairing and results also in nucleotide substitution.

Other replication infidelities that may relate to slippage in the replication fork can result in either the deletion or the insertion of one or a few nucleotides in the newly synthesized DNA strand. If such mutations occur within reading frames for protein synthesis, the phenotypic effect may be drastic. This is the case when, in the protein synthesized from a gene affected by a frameshift mutation, the amino acid sequence downstream of the site of mutation strongly differs from that of the nonmutated product. In addition, the size of such a mutated protein is usually altered depending on the chance occurrence of an appropriate stop codon in the new reading frame.

Proofreading devices and other enzymatic repair systems prevent replication infidelities from producing mutations at intolerably high rates. Generally, they act rapidly after replication by screening for imperfect base pairing in the double helix. Successful repair thereby requires specific means to distinguish the newly replicated DNA strand from its template, the complementary parental strand. Because of these correction activities, many primary mispairings are removed before they have the opportunity to become fixed as mutations. DNA repair systems modulate the frequencies of mutagenesis.

Genetic information of some viruses, and sometimes also segments of genetic information of chromosomal origin, may pass through RNA molecules, which may later become retrotranscribed into DNA. No efficient repair systems are known to act at the level of RNA. Indeed, RNA replication shows a higher degree of infidelity than DNA replication. In consequence, genetic information that becomes replicated as RNA molecules generally suffers increased mutation rates.

A relatively large number of internal and environmental chemicals exert mutagenic effects by means of molecular mechanisms that in many cases are well understood and often cause local DNA sequence changes. Some intermediate products of the normal metabolic activities of a cell may be mutagenic and may thus contribute to spontaneous local mutagenesis. The mutagens of the environment include not only a multitude of chemical compounds but also ultraviolet radiation and some physicochemical constraints such as elevated temperature, which influence the chemical stability of nucleotides. Each of these mutagens and mutagenic conditions contributes in a specific way to the generation of genetic variations.

Again, many of the potential sequence alterations brought about by internal and environmental mutagens are efficiently repaired by enzymatic systems. However, since the efficiency of such repair is rarely 100%, evolutionarily relevant mutations persist.

5
Intragenomic DNA Rearrangements

Various recombination processes are well known to mediate DNA rearrangements, which often result in new nucleotide sequences.

While in haploid organisms general recombination is not essential for propagation, it influences genetic stability at

the population level in various ways as a generator of new sequence varieties. For example, it can bring about sequence duplications and deletions by acting at segments of homology that are carried at different locations in a genome. General recombination is also known to act in the reparation of damage caused to DNA by ionizing radiation. In this case, an intact genome can become assembled from undamaged segments of sister copies of the chromosome by homologous recombination. The best-known contribution of general recombination to genetic diversity is meiotic recombination bringing about random recombinants between the paternal and the maternal chromosomes in diploid organisms.

Two other widely spread types of recombination systems are dealt with in more detail in Sects. 5.1 and 5.2: site-specific recombination and transposition. Both are known to contribute to genetic variation. Still other recombination processes, such as the one mediated by DNA gyrase can, for the time being, perhaps best be grouped as illegitimate recombinations. This group may contain several different molecular mechanisms that act with very low efficiency and have remained at least in part unexplained.

5.1 Site-specific DNA Inversion at Secondary Crossover Sites

Genetic fusions represent the results of joining together segments of two genes (gene fusions) or of two operons (operon fusions) that are not normally together. An *operon* is a set of often functionally related genes that are copied into messenger RNA (i.e. transcribed) as a single unit. As a result of this organization, those genes are coordinately regulated; that is, they are turned on or off at the same time. Therefore, in an operon fusion, one or more genes are put under a different transcription control, but the genes *per se* remain unchanged. In contrast, gene fusion results in a hybrid gene composed of sequence motifs and often of functional domains originating in different genes.

In site-specific DNA inversion, a DNA segment bordered by specific DNA sequences acting as sites of crossing-over becomes periodically inverted by the action of the enzyme DNA invertase. Depending on the location of the crossover sites, DNA inversion can give rise to gene fusion or to operon fusion. The underlying flip-flop system can result in microbial populations composed of organisms with different phenotypic appearances: if, for example, the DNA inversion affects the specificity of phage tail fibers, as is the case with phages P1 and μ of *E. coli*, phage populations with two different host ranges will result.

Occasionally, a DNA sequence that deviates considerably from the efficiently used crossover site, a so-called *secondary crossover site*, can serve in DNA inversion, which thus involves a normal crossover site and a secondary crossover site. This process results in novel DNA arrangements, many of which may not be maintained because of lethal consequences or reduced fitness; but if a few new sequences are beneficial to the life of the organism, these may be selectively favored. This DNA rearrangement activity can thus be looked at as evolutionarily important. Since many different DNA sequences can serve in this process as secondary crossover sites, although at quite low frequencies, site-specific DNA inversion systems act as variation generators in large populations of microorganisms. I have thus postulated that this evolutionary role of DNA inversion systems may be more important than

their much more efficient flip-flop mechanism, which can at most help a microbial population to more readily adapt to two different, frequently encountered environmental conditions. As a matter of fact, other strategies could be used as well for the latter purpose.

Computer-aided comparison of DNA sequences quite often reveals that independent genes may consist of a particular domain with high homology and of other DNA sequences showing no significant signs of relatedness. DNA inversion using secondary sites of crossing-over is a potential source of such mosaic genes. DNA inversion can span over relatively large distances in DNA molecules and has the advantage of not loosing any DNA sequences located between the two sites of crossing-over. Deletion formation represents another source for gene fusion, but it has the disadvantage that the DNA sequences between the sites of crossing-over are usually eliminated.

5.2 Transposition of Mobile Genetic Elements

Nine different mobile genetic elements have been found to reside, often in several copies, in the chromosome of *E. coli* K-12 derivatives. This adds up to occupation of about 1% of the chromosomal length by such insertion sequences, also called *IS elements*. At rates on the order of 10^{-6} per individual IS element and per cell generation, these mobile genetic elements undergo transpositional DNA rearrangements. These include simple transposition of an element and more complex DNA rearrangements such as DNA inversion, deletion formation, and the cointegration of two DNA molecules. Because of different degrees of specificity in the target selection upon transposition, the IS-mediated DNA rearrangements are neither strictly reproducible nor fully random. Transposition activities thus also act as variation generators. In addition to DNA rearrangements mediated by the enzyme transposase, which is usually encoded by the mobile DNA element itself, other DNA rearrangements just take advantage of extended segments of DNA homologies at the sites of residence of identical IS elements at which general recombination can act. Altogether, IS elements represent a major source of genetic plasticity of microorganisms.

Transposition occurs not only in growing populations of bacteria but also in prolonged phases of rest. This is readily seen with bacterial cultures stored at room temperature in stabs (little vials containing a small volume of growth medium in agar). Stabs are inoculated with a drop of a bacterial culture taken up with a platinum loop, which is inserted (''stabbed'') from the top to the bottom of the agar. After overnight incubation, the stab is tightly sealed and stored at room temperature. Most strains of *E. coli* are viable in stabs during several decades of storage. That IS elements exert transpositional activities under these storage conditions is easily seen as follows.

A stab can be opened at any time, a small portion of the bacterial culture removed, and the bacteria well suspended in liquid medium. After appropriate dilution, bacteria are spread on solid medium. Individual colonies grown upon overnight incubation are then isolated. DNA from such subclones is extracted and fragmented with a restriction enzyme. The DNA fragments are separated by gel electrophoresis. Southern hybridization with appropriate hybridization probes can then show whether different subclones reveal restriction fragment length polymorphisms

(RFLPs), which are indicative of the occurrence of mutations during storage.

If this method is applied to subclones isolated from old stab cultures, and if DNA sequences from residential IS elements serve as hybridization probes, an extensive polymorphism is revealed. None or only little polymorphism is seen with hybridization probes from unique chromosomal genes. Good evidence is available that transposition represents a major source of this genomic plasticity observed in stabs, which at most allow for a very residual growth at the expense of dead cells. One can conclude that the enzymes promoting transposition are steadily present in the stored stabs. Indeed, the IS-related polymorphism increases linearly with time of storage for periods as long as 30 years. In a culture of *E. coli* strain W3110 analyzed after 30 years of storage, each surviving subclone had suffered on the average about a dozen RFLP changes as identified with hybridization probes from eight different residential IS elements, of which IS5 was the most active. Lethal mutations could of course not be identified in this study.

Lethal mutations that affect essential genes for bacteriophage reproduction can be accumulated in the prophage state of the phage genome in its lysogenic host. Such mutants can be screened for their inability to produce infective phage particles upon induction of phage reproduction. Experiments were carried out with *E. coli* lysogenic for a phage P1 derivative grown in batch cultures at 30 °C for about 100 generations allowing for alternative periods of growth and rest. Most of the independent lethal mutants could thereby be identified to be caused by the transposition of an IS element from the host chromosome into the P1 prophage that is maintained in its host as a plasmid. In these experiments, IS2 was the most active element and it mainly inserted into a few preferred regions of the P1 genome, but each time at another site. The used insertion targets did not show any detectable homology or similarity with each other. This is another good example for an enzymatically mediated variation generator. The experiment as such identifies IS transposition as a major source of lethal mutagenesis.

There is no evidence available that bacterial mobile genetic elements would play an essential role in the bacterial life span extending from one cell division to the next. However, these elements are major players in the evolution of bacterial populations. As we have seen here, they contribute to intragenomic DNA rearrangements. Depending on the target sequences involved, the resulting mutations may often be lethal by interrupting essential reading frames or expression control regions. Favorable mutations may be relatively rare, but these can contribute to evolutionarily advantageous developments of the genome. In Sect. 6, we will see that mobile genetic elements also play important roles in the natural strategy of DNA acquisition.

6 DNA Acquisition

While the mutagenesis mechanisms belonging to the strategies of local changes in the DNA sequences (Sect. 4) and of intragenomic DNA rearrangements (Sect. 5) are exerted within the genome and can affect any part of the genome, an additional strategy of spontaneous sequence alterations depends on an external source of genetic information. In DNA or gene acquisition, genetic information indeed originates from an organism other than

the one undergoing mutagenesis. In bacteria, DNA acquisition can occur by means of transformation, conjugation, or virus-mediated transduction. In the latter two strategies of horizontal gene transfer, a plasmid or a virus, respectively, acts as natural gene vector.

The association and dissociation of chromosomal genes with natural gene vectors often arises from transpositional activities and from general recombination acting at IS elements that are at different chromosomal locations. These mechanisms have been well studied with conjugative plasmids and with bacteriophage genomes serving in specialized transduction. For example, composite transposons, which are defined as two identical IS elements flanking a segment of genomic DNA (often with more than one gene unrelated to the transposition process), are known to occasionally transpose into a natural gene vector and, after their transfer into a receptor cell, to transpose again into the receptor chromosome. Hence, together with other mechanisms, such as site-specific and general recombination, transposition represents an important promoter of horizontal gene transfer.

Several natural factors seriously limit gene acquisition. Transformation, conjugation, and transduction depend on surface compatibilities of the bacteria involved. Furthermore, upon penetration of donor DNA into receptor cells, the DNA is very often confronted with restriction endonucleases. These enzymes cause a fragmentation of the invading foreign DNA, which is subsequently completely degraded. Before fragments become degraded, however, they are recombinogenic and may find a chance to incorporate all or part of their genetic information into the host genome. Therefore, we interpret the role of restriction systems as follows: they keep the rate of DNA acquisition low, and at the same time they stimulate the fixation of relatively small segments of acquired DNA to the receptor genome. This strategy of acquisition in small steps can best offer microbial populations the chance to occasionally extend their biological capacities without extensive risk of disturbing the functional harmony of the receptor cell by acquiring too many different functions at once. These considerations have their relevance at the level of selection exerted on the hybrids resulting from horizontal DNA transfer. This selection represents one of the last steps in the acquisition process.

DNA acquisition by horizontal gene transfer is a particularly interesting source of new genetic information for the receptor bacterium because the chance that the acquired DNA exerts useful biological functions is quite high – most likely, it has already assumed the same functions in the donor bacterium.

An interesting hypothesis links the universality of the genetic code with the important role played by horizontal gene transfer in the evolutionary development of the living world. According to this view, those organisms using the most common genetic language would, in the long term, be able to profit best from the increasing worldwide pool of genetic functions under the pressure of adapting to changing living conditions.

7
The Three Natural Strategies Generating Genetic Variations Contribute Differently to the Evolutionary Process

Biological evolution is a systemic process. As outlined above, many different specific mechanisms contribute to generate genetic diversity that represents at any time

the substrate for natural selection. The building up of functional complexity is a stepwise process, in which many random attempts of genetic alterations become rapidly rejected, while relatively few novel sequences are approved as favorable by natural selection and are maintained and amplified. The genome can thus be seen as a cabinet in which key information from favorable historical developments is stored. Stepwise, additional favorable information is added. In the context of changing selective conditions, stored information, having lost its functionally beneficial relevance, may be deleted or favorably altered. As we have seen, a multitude of different mechanisms are behind this dynamic process. For a better understanding of the events, we have grouped the identified mechanisms into three major natural strategies of genetic variation: local changes in the DNA sequences, intragenomic DNA rearrangements, and DNA acquisition. These three strategies have different qualities with regard to their contributions to biological evolution.

The local DNA sequence change is probably the most frequently involved strategy of genetic variation. Indeed, its frequency, which depends primarily on intrinsic properties of matter, chemical instability and conformational flexibility, would be intolerably high if it would not be modulated by efficient enzymatic systems of DNA repair. Local sequence changes bring about nucleotide substitution, the deletion and the insertion of one or a few base pairs, or a local scrambling of a few base pairs. These sequence changes can contribute to a stepwise improvement of a biological function. It must be kept in mind that the functional test for such improvement is carried out by natural selection. In principle, a long series of stepwise local sequence changes could also be expected to bring about a novel biological function. However, this kind of long-term process can gain efficiency only once natural selection starts to be exerted on such upcoming function.

In contrast, the reshuffling of DNA segments within the genome can be considered as a tinkering with existing elements, whereby favorable gene fusions and operon fusions may occasionally result. DNA rearrangement can also be the source of gene duplication and higher amplification, which are widely recognized contributions to the evolutionary progress.

In Sect. 6, we have already pointed to the evolutionarily high efficiency of horizontal gene transfer. As a matter of fact, DNA acquisition allows the recipient organism to share in the success of evolutionary developments made by others. In drawing the evolutionary tree of bacteria, DNA acquisition should be accounted for by more or less randomly adding temporal horizontal shunts between individual branches. It must be kept in mind that usually only small DNA segments flow through such shunts in horizontal gene transfer.

Several of the specific mechanisms of genetic variation employ, strictly speaking, more than one of the three strategies shown in Fig. 1. In transposition of IS elements, for example, a chromosomal DNA segment consisting of the mobile genetic element can undergo a translocation and thereby become inserted at a new target site. As a rule, the target sequence thereby gets duplicated, which usually involves a few nucleotides. Thus, this transposition event will consist of both a DNA rearrangement and a local sequence change.

As far as we know, most of the well-studied microbial strains use in parallel each of the three natural strategies for the generation of genetic variations. In

addition, bacteria very often use not only one, but several different specific mechanisms for mutagenesis by each of the strategies. Dissimilar specific mechanisms often work with different efficiencies as reflected by their contribution to the overall mutation rate. For any given strategy, it might be less relevant which specific mechanism is at work than the fact that the particular strategy finds its application with an evolutionarily useful efficiency. In other words, specific mechanisms may substitute for each other within a strategy, at least to some degree. This rule does not apply between the strategies because of the difference in the qualities of their evolutionary contributions.

The efficiency displayed by a given specific mechanism of spontaneous mutagenesis may depend on both internal (e.g. availability of enzymes that mediate mutagenesis) and external factors (e.g. environmental stress). It is also to be noted that some mechanisms may act more or less randomly along a DNA molecule, while other mechanisms may show regional or site preferences for their activities. In view of these considerations, we tend to assume that an evolutionarily fit (or well prepared) organism should best be able to use a few specific mutagenesis mechanisms for each of the three strategies to generate genetic variations. In Sect. 8, we will explain what we mean by evolutionary fitness.

8 Evolution Genes and Their Own Second-order Selection

The attentive reader will have seen in the description of some specific mechanisms contributing to the spontaneous mutagenesis that besides a number of nongenetic factors, specific products of genes are very often at work. These gene products can belong to systems for repair of DNA damage and will, in this case, modulate the frequency of mutagenesis. Similarly, restriction enzymes seriously reduce both the chance of DNA acquisition and the size of a DNA segment that may eventually be acquired by the recipient cell. Other gene products such as transposases and other mediators of DNA recombination act as generators of genetic variations. Since variation generators and modulators of the frequencies of genetic variation are evolutionary functions, we call the underlying genetic information *evolution genes*. In the microbial world, these evolution genes generally play no essential role in the physiology of individual lives going by cell division from one generation to the next. Under laboratory conditions, neither restriction–modification systems nor enzymes for DNA rearrangements are needed for the propagation of bacteria. The role of such enzyme systems is primarily evolutionary and becomes manifest at the level of populations.

We assume that evolutionary genes are themselves submitted to selective pressure. However, such selection cannot follow the rules of direct selection for improvements of essential functions such as those of housekeeping genes. Rather, the selection for the presence and improvement of a variation generator will be exerted at the level of populations. Clearly, it will also be an individual that may one day undergo a mutation, which improves an evolutionary function. This function will also be exerted in its progeny in which appropriate genetic variants of genes for directly selected products will be either more or less abundant. Any genetic alteration that affects an evolution gene and proves in the long term to be of higher evolutionary

value will be maintained and will provide an evolutionary benefit to the carrier of the involved gene. In the long run, this will lead to fine-tuning of the evolutionary functions of both variation generators and of modulators of the frequency of genetic variation. The underlying indirect selection based on the cells ability to provide genetic variants at a well-balanced level is called *second-order selection*.

We must be aware that some gene products may exert their essential functions for the benefit of both the life of the individual and the evolutionary progress of the population. In these cases, we assume that evolutionary selection is exerted for both kinds of functions and will eventually bring the gene to a fine-tuned state to optionally carry out its functions for each of the different purposes. However, as we have already mentioned, a number of gene products involved in genetic variation are inessential for the lives of individual bacterial cells. Similarly, the products of many housekeeping genes are inessential for biological evolution.

9 Arguments for a General Relevance of the Theory of Molecular Evolution for All Living Organisms

Largely on the basis of evidence from microbial genetics, we have so far postulated that the products of a number of evolution genes contribute, each in its specific way, to the generation of genetic variants at evolutionarily useful frequencies. Thereby, the sources of mutagenesis may relate either to the activity of the evolution gene product itself (e.g. a transposase) or to a nongenetic factor (e.g. a chemical mutagen or an intrinsic structural flexibility of a nucleotide). In many cases, nongenetic factors and products of evolution genes cooperate in the formation of genetic variants at physiologically tolerable and evolutionarily beneficial levels. This is, for example, the case in spontaneous mutagenesis by an environmental mutagen when some of the primary damage on the DNA gets successfully repaired while some other damage leads to a fixed mutation.

The theory of evolution postulates that life on Earth started almost four billion years ago with primitive, unicellular microorganisms. It is in the first two billion years that microbes must have developed the basis for the actual setup of evolutionary strategies and the underlying evolution genes. One can postulate that the acquired evolutionary capacities could have allowed some microbial populations to undergo a division of labor in more and more stable associations of cells. This development might later have led to multicellular organisms. In this kind of development, the evolutionary fitness of the involved organisms might have been an important precondition. At still later stages of further evolutionary development, the three natural strategies for generating genetic variations (Sects. 4 to 7) must have continued exerting their evolutionary influence together with some other factors such as the formation of symbiotic associations. As a matter of fact, we attach considerable evolutionary relevance to endosymbiosis of higher organisms with bacteria. Such situations of cohabitation may form an ideal condition for occasional horizontal gene transfer between the close associates.

A scientifically justified quest for further experimental proof of the postulates of the theory of molecular evolution remains quite difficult to be answered. Clearly, there is a need for research on

the spontaneous generation of genetic variation in higher organisms, ideally at the level of the genomes. While this is already quite difficult in microorganisms, it is of increased perplexity with the much larger genomes of higher organisms. However, sequence comparisons now offer fruitful ways to search for sequence homologies, sequence similarities, single nucleotide polymorphisms, as well as for traces of intragenomic DNA rearrangements and of horizontal transfer of genetic information. Data so far available are in support of the principles of molecular evolution outlined in Sect. 2, which are likely to be valid for any kind of living organism.

Some of the general evolutionary strategies developed in microorganisms must have also turned out to be useful for the developmental and physiological processes at somatic levels of higher organisms. The generation of antibody diversity in the immune systems of vertebrates by genetic rearrangements and so-called *somatic mutagenesis* is a good example. Another example is the enzymatic repair of DNA damage caused in somatic cells by external mutagens such as UV irradiation.

These considerations illustrate that whatever gene function may prove to be useful for whatever particular purpose, it may be evolutionarily retained and in the course of time further fine-tuned. We have already encountered this principle in the microbial world, where we have postulated multifunctional enzymes (such as working both for the physiology of the cells and for an evolutionary task) to become evolutionarily improved both by direct and by second-order selection for the various functions. This may also be the case in higher organisms.

10
Conceptual Aspects of the Theory of Molecular Evolution

With reference to Sect. 2.2, it is fair to again explicitly state that for the time being, evolution genes and evolution functions are a concept rather than a fully proven fact. This concept is based on a particular way to interpret numerous available experimental data. We will briefly analyze the difficulties to clarify the situation in a scientific debate. This will be followed by pointing to philosophical and more practical values of a deeper understanding of the molecular processes that drive biological evolution.

10.1
Pertinent Scientific Questions

In the history of scientific investigations, biologists have often searched for evidence that living organisms would be able to specifically modify, or adapt, their genetic information in order to better cope with upcoming changes in the living conditions. Most of these attempts have failed to give the expected response. In other cases, where a certain degree of specific adaptation could be observed, specific causal explanations have sometimes been found upon deeper investigation. However, there is at present no good scientific evidence for a general rule that genetic alterations would always be directed toward a specific goal. This situation favors the view that spontaneous mutations affect DNA more or less randomly, which is in line with the general observation that only a minority of spontaneous mutants prove to be favorable under the encountered living conditions.

The postulate of evolution genes that act as generators of genetic variations may be a surprise in this context. This has to do with a widely followed concept of genetic

information as a strict program for the fulfillment of a specific task. This definition does apply to many housekeeping genes, the products of which efficiently catalyze a reaction that reliably always yields the same product. This is not how a variation generator that does not work efficiently and that yields a different product from case to case functions. A good example is the transposition of mobile genetic elements. However, not all scientists see the primary function of a mobile genetic element in genetic variation. Rather, some colleagues interpret IS translocation, which often goes along with the replication of the element, as a selfish activity. This interpretation considers mobile genetic elements as parasites with the argument that their activity would often harm their host cell.

This discussion shows that the concept of evolution genes cannot easily be defended by referring to scientific evidence. Rather, the concept reflects an attitude of the observer of natural events. According to the view defended in this article, nature actively cares for biological evolution. The products of evolution genes are actively involved in generating different kinds of genetic variants at frequencies insuring both a certain genetic stability required for maintaining the concerned form of life and a low frequency of genetic variations as the driving force of evolution. This interpretation recognizes biological evolution as an essential principle of self-organization of life on Earth.

Another pertinent question that cannot find an easy scientific answer relates to the evolutionary function of viruses. At least some viruses are clearly identified to sometimes act as gene vectors in horizontal gene transfer. Some of them also temporarily integrate their genome into the host genome. This relates to the lysogenic state of bacteria as well as to endogenous viruses such as retroviruses that reside in many higher organisms. Again, one may wonder if these viruses primarily fulfill evolutionary functions for the evolutionary development of their hosts or whether they should rather be looked at as parasites that may carry out some evolutionary function by accident.

While prokaryotic organisms have genomes that are relatively densely packed with functional genes, many higher organisms have extended segments of intergenic DNA sequences, some of which are highly repetitive. Some of these noncoding sequences are highly homogeneous with regard to their nucleotide composition. While the biological roles played by noncoding regions are still not well understood, it has been postulated that compositional constraints may influence natural selection. These aspects have not been covered in this article and they may more specifically relate to the molecular evolution of higher organisms, in addition to the principles outlined here largely on the basis of evidence from microbial genetics.

10.2 Philosophical Values of the Knowledge on Molecular Evolution

One of the central questions of human curiosity is to know where life – and more specifically human life – comes from. The Darwinian theory opposed to the idea of a specific act of creation for each particular form of life, the alternative explanation of a steady evolutionary development implying the descent of the actual species from common ancestors. The directions of evolution are thereby given by natural selection acting steadily on all available forms of life

including all present variants. Until recently, the sciences could not specifically explain how genetic variants are generated. With recently developed research strategies, molecular genetics can now fill this gap. The branch of science called *molecular evolution* explains that there is not just a single source for genetic variants. Rather, many different specific mechanisms contribute to the generation of genetic variants at low frequencies. These mechanisms follow one and sometimes more than one basic strategies of evolutionary development. These are local changes in the DNA sequences, rearrangements of DNA segments within the genome, and acquisition of segments of foreign DNA by horizontal gene transfer. As a general rule, spontaneous mutagenesis is not specifically directed; it is at least to some extent random, so that only a minor fraction of spontaneous genetic variants turn out to be favorable for the concerned organism and thus provide it with a selective advantage. Nevertheless, new knowledge on the precise molecular mechanisms of the generation of genetic variations provides strong evidence that, in many cases, specific enzymes are involved – the products of evolution genes. These products work in tight collaboration with nongenetic factors that can be intrinsic properties of matter or environmental conditions. This view of the evolutionary process represents the core of a theory of molecular evolution and it can be seen as an extension of neo-Darwinism at the molecular level.

It should be clearly stated that the theory of molecular evolution does not explain the origin of life. It can, however, explain that biological evolution exerted in all living beings is a steady, dynamic process that is actively promoted not only by intrinsic properties of matter but also by the intervention of products of evolution genes or more generally of evolutionary functions of many different gene products. A recently published book devoted to these exciting insights is entitled *Darwin in the Genome*.

The high philosophical value of this extension of our worldview is obvious and merits to be widely discussed and evaluated in its various cultural dimensions. One interesting aspect is the implied duality of the genome. Indeed, evolution genes are located together with all the other genes on the genome and on accessory DNA molecules such as plasmids and viral genomes. While probably a major part of gene products carries out functions to benefit the cell and the individual, often a multicellular organism, probably a minority of gene products works for the biological evolution of the concerned population. Note, however, that the generation of a novel genetic variant obviously also occurs in an individual cell. But this act of creation has only a small chance to bring to the concerned cell and to its descendants a selective advantage. More often, the mutation is unfavorable and renders the life of the concerned organism more troublesome. As long as the spontaneous rate of mutagenesis remains low (mostly thanks to the intervention of evolution genes), unfavorable mutations are tolerable at the level of propagating populations.

Incidentally, the situation described here offers a possible explanation to the quite difficult theodicean question: why does God, despite His love for the human creature, admit that physically evil events such as a mutation causing an inheritable disease can occur to individuals? As a matter of fact, Genesis describes creation as a stepwise process, which implies the permanent evolutionary expansion of the diversity of life forms. Genesis also states that God evaluated this system as good.

Hence, biological evolution occurs according to God's intention to amplify diversity of life on our planet. Occasional unfavorable mutations affecting rare individuals in populations is a sacrifice brought to the creative force residing in the system of molecular evolution.

In brief, the genetic information contained in each genome – of bacteria as well as of all higher organisms – represents an internal duality. It serves individuals for the fulfillment of their individual lives and it serves populations for a slow but steady expansion of life forms and thus of biodiversity.

10.3 Aspects Relating to Practical Applications of Scientific Knowledge on Molecular Evolution

Living organisms today occupy an amazing variety of ecological niches on our planet Earth. These niches include extreme physicochemical conditions such as elevated temperatures, high pressure, and quite unusual compositions of chemical elements. However, despite the intrinsic potential of the living world to evolutionarily expand, one can estimate the carrying capacity of the planet for life to be in the order of 10^{30} living cells. Although this is a very large number, it seriously limits free expansion of life in its various forms.

The following reflections should help illustrate this statement. An adult human being carries in the order of 10^{13} cells. The human population today, thus occupies a share of about 10^{23} cells of the available 10^{30}. Incidentally, this happens to be close to the average available for each of the estimated 10^{7} different species of organisms on the planet. Bacteria propagate by cell division as outlined in Sect. 3. Extending the reflections made there, one can conclude that from an inoculum with one single bacterial cell, one can theoretically expect to obtain 10^{30} cells within only 50 h. In reality, growth will be stopped much earlier by lack of nutrition, but this reflection illustrates well the enormous internal forces for expansion of a given form of life.

Similarly, a high potentiality for evolutionary expansion toward more diversity resides in the mechanisms of molecular evolution that are described in this article. These mechanisms can serve us as a basis to better understand both the origin and the steady replenishment of biodiversity as well as the internal limits set to evolutionary expansion. This knowledge can and should increasingly be used as a background for any measures taken toward the protection of biodiversity and of habitats for diverse forms of life. Last but not the least, a better understanding of the evolutionary process can be of help to render the development of agricultural and related practices more sustainable.

Genetic engineering offers ample new possibilities for the sustainable production of medical drugs, to obtain food of higher quality, and to reduce the nocent impact of the human civilization on the environment. The serious reservations made by large parts of the human population impede many of the proposed biotechnological applications. A part of these concerns refer to unpredictable long-term effects of genetically modified organisms (GMO) that are released into the environment, such as in agricultural applications. Scientific assessments of long-term and, in particular, evolutionary effects of such applications are thus required. Knowledge of the natural strategies of molecular evolution can provide a good basis for such studies. As a matter of fact, in genetic engineering, DNA sequence alterations are

brought about within the genome by site-directed mutagenesis in studies of the biological functions of specific genes. In addition, well-defined segments of DNA are introduced into other organisms either in view of amplifying the particular DNA segment or in view of harvesting a specific gene product. GMO can also directly serve in applications as GM food and for bioremediation by microorganisms. A candid comparison of these practices involving genetic engineering with the natural strategies of generation of genetic variations reveals a high degree of similarities. The amounts of nucleotides or lengths of DNA sequences involved in these genetic modifications, both in genetic engineering and in the natural genetic variation, are of the same order of magnitude. Depending on the strategy involved, they may concern one to a few base pairs, or in other instances, a DNA segment containing a sequence domain or one to a few genes, both in intragenomic DNA rearrangements and in the horizontal transfer of DNA between two different organisms. Thus, one can principally expect that long-term evolutionary risks of GMO compare with the biohazard intrinsic to the natural process of biological evolution. Similar risks may also be inherent in classical breeding techniques.

These considerations ask for a more integral, holistic, and critical evaluation of the impact of past and present human activities on the natural process of biological evolution. Such assessments should address any human impact on genetic variation, natural selection, and isolation. As far as we know from the long-term history, the foundations of life and its evolutionary development on Earth are relatively robust. This is good news for us human beings, but it should not exempt us from a responsible and well-reflected use of our scientific knowledge in any attempt to render our own lives more easy and comfortable.

Bibliography

Books and Reviews

Caporale, L.H. (Ed.) (1999) *Molecular Strategies in Biological Evolution*, Vol. 870, Annals of the New York Academy of Sciences, New York, NY.

Caporale, L.H. (2003) *Darwin in the Genome: Molecular Strategies in Biological Evolution*, McGraw-Hill, New York.

Kucherlapati, R., Smith, G.R. (Eds.) (1988) *Genetic Recombination*, American Society for Microbiology, Washington, DC.

Moses, R.E., Summers, W.C. (Eds.) (1988) *DNA Replication and Mutagenesis*, American Society for Microbiology, Washington, DC.

Shapiro, J.A. (1983) *Mobile Genetic Elements*, Academic Press, New York.

Primary Literature

Arber, W. (1991) Elements in microbial evolution, *J. Mol. Evol.* **33**, 4–12.

Arber, W. (1993) Evolution of prokaryotic genomes, *Gene* **135**, 49–56.

Arber, W. (1995) The generation of variation in bacterial genomes, *J. Mol. Evol.* **40**, 7–12.

Arber, W. (2000) Genetic variation: molecular mechanisms and impact on microbial evolution, *FEMS Microbiol. Rev.* **24**, 1–7.

Arber, W. (2002) Evolution of prokaryotic genomes, *Curr. Top. Microbiol. Immunol.* **264/I**, 1–14.

Arber, W. (2002) Molecular evolution: comparison of natural and engineered variations, *The Pontifical Academy of Sciences. Scripta Varia* **103**, 90–101.

Arber, W. (2003) Elements for a theory of molecular evolution, *Gene*. **317**, 3–11.

Arber, W. (2003) Traditional Wisdom and Recently Acquired Knowledge in Biological Evolution, *Proceedings of UNESCO Conference on* ''Science and the Quest for Meaning''; in press.

Arber, W., Hümbelin, P., Caspers, P., Reif, H.J., Iida, S., Meyer, J. (1981) Spontaneous mutations in the *Escherichia coli* prophage P1

and IS-mediated processes, *Cold Spring Harbor Symp. Quant. Biol.* **45**, 38–40.

Arber, W., Naas, T., Blot, M. (1994) Generation of genetic diversity by DNA rearrangements in resting bacteria, *FEMS Microbiol. Ecol.* **15**, 5–14.

Bernardi, G. (2000) Isochores and the evolutionary genomics of vertebrates, *Gene* **241**, 3–17.

Bernardi, G. (2000) The compositional evolution of vertebrate genomes, *Gene* **259**, 31–43.

Drake, J.W. (1991) Spontaneous mutation, *Annu. Rev. Genet.* **25**, 125–146.

Echols, H., Goodman, M.F. (1991) Fidelity mechanisms in DNA replication, *Annu. Rev. Biochem.* **60**, 477–511.

Galas, D.J., Chandler, M. (1989) Bacterial Insertion Sequences, in: Berg, D.E., Howe, M.M. (Eds.) *Mobile DNA*, American Society for Microbiology, Washington, DC, pp. 109–162.

Glasgow, A.C., Hughes, K.T., Simon, M.I. (1989) Bacterial DNA Inversion Systems, in: Berg, D.E., Howe, M.M. (Eds.) *Mobile DNA*, American Society for Microbiology, Washington, DC, pp. 637–659.

Iida, S., Hiestand-Nauer, R. (1987) Role of the central dinucleotide at the crossover sites for the selection of quasi sites in DNA inversion mediated by the site-specific Cin recombinase of phage P1, *Mol. Gen. Genet.* **208**, 464–468.

Lorenz, M.G., Wackernagel, W. (1994) Bacterial gene transfer by natural genetic transformation in the environment, *Microbiol. Rev.* **58**, 563–602.

Naas, T., Blot, M., Fitch, W.M., Arber, W. (1994) Insertion sequence-related genetic variation in resting *Escherichia coli* K-12, *Genetics* **136**, 721–730.

Naas, T., Blot, M., Fitch, W.M., Arber, W. (1995) Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12, *Mol. Biol. Evol.* **12**, 198–207.

Sandmeier, H. (1994) Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibers, *Mol. Microbiol.* **12**, 343–350.

Sengstag, C., Arber, W. (1983) IS2 insertion is a major cause of spontaneous mutagenesis of the bacteriophage P1: non-random distribution of target sites, *EMBO J.* **2**, 67–71.

Weber, M. (1996) Evolutionary plasticity in prokaryotes: a panglossian view, *Biol. Philos.* **11**, 67–88.

West, S.C. (1992) Enzymes and molecular mechanisms of genetic recombination, *Annu. Rev. Biochem.* **61**, 603–640.

Woese, C.R. (1987) Bacterial evolution, *Microbiol. Rev.* **51**, 221–271.