Predictability in Science: Accuracy and Limitations
Pontifical Academy of Sciences, Acta 19, Vatican City 2008
www.pas.va/content/dam/accademia/pdf/acta19/acta19-arber.pdf

STOCHASTIC GENETIC VARIATIONS AND THEIR ROLE IN BIOLOGICAL EVOLUTION

WERNER ARBER

1. Introduction

It is still a widespread view, particularly in the general public, that all genes carried in the genome determine their encoded functions accurately and fully predictably. Scientists, however, realize more and more that this does not correspond to the reality by far. There are many reasons for a higher complexity, such as epigenetic phenomena, protein modifications, regulatory circuits in the control of gene expression and environmental impacts. Studies on the generation of genetic variants, that represent the driving force of biological evolution, have revealed rather unexpected properties of specific gene products that can be called variation generators. These proteins generate genetic variants with a high degree of unpredictability with regard to the site of mutagenesis on the genome and with regard to the time when a variation occurs. We will illustrate this situation with a few examples that were obtained by experimentation in microbial genetics. The evolutionary role of these activities will be discussed in the context of a more general theory of molecular evolution.

2. From the Neodarwinian Theory to a Theory of Molecular Evolution

The so-called modern evolutionary synthesis resulting in the Neodarwinian theory of biological evolution united, around 1940, Darwin's concept of natural selection with the notion of spontaneous mutagenesis as the source of hereditary phenotypic variation (Mayr, 1982). Shortly thereafter, DNA was identified as the carrier of genetic information (Avery *et al.*, 1944) and the filamentous, double helical structure of DNA molecules was

described (Watson and Crick, 1953). This then gave rise to molecular genetics. In the meantime it has become known that changes of inheritable traits are linked to alterations of nucleotide sequences in the underlying genetic information. However, not all changes occurring in genomic sequences result in an altered phenotypic trait, for a number of understood reasons. Because of this situation, two different definitions for the synonymously used terms of genetic variation and genetic mutation are found in the genetic literature. In classical genetics, a mutation is an altered phenotype that becomes transmitted to the progeny. In contrast, in molecular, reverse genetics a mutation is an alteration in the nucleotide sequence of the DNA.

In the context of today's knowledge on molecular evolution the process of biological evolution stands on three pillars. One of these is genetic variation that can, to some degree, affect life activities either positively or negatively. The second pillar is natural selection, i.e. the result of the interaction of individual organisms with their encountered environmental constraints. These depend both on physico-chemical properties of the environment and on the activities of all other organisms living in the same ecological niche. The third pillar is isolation that can be reproductive or geographic.

The generation of genetic variants represents the driving force of biological evolution. Natural selection together with the available genetic variants determines the different directions that evolution takes. Isolation modulates the process of evolution. These interpretations are schematically represented in the upper part of Figure 1.

Thanks to microbial genetics and genomics it has become possible to investigate the molecular processes that are sources of genetic variation. On the one hand, bioinformatic comparison of nucleotide sequences from more or less closely related organisms can reveal the accumulated alterations within the genomes since their evolutionary separation, as well as, at lower levels of genome organization, within a functional domain, a single specific gene or a group of functionally related genes. This can allow the researcher to postulate the molecular mechanisms that are likely responsible for having generated the observed sequence alterations and the observed phenotypic changes. On the other hand, it is also possible to experimentally document single steps of nucleotide sequence alterations, particularly with small, microbial genomes. The results of such investigations are summarily represented in the lower part of Figure 1. Obviously, a relatively large number of specific molecular mechanisms are at the source of the overall genetic variation (Arber, 1997). These insights allow us to formulate a theory of molecular evolution (Arber, 2003, 2007).

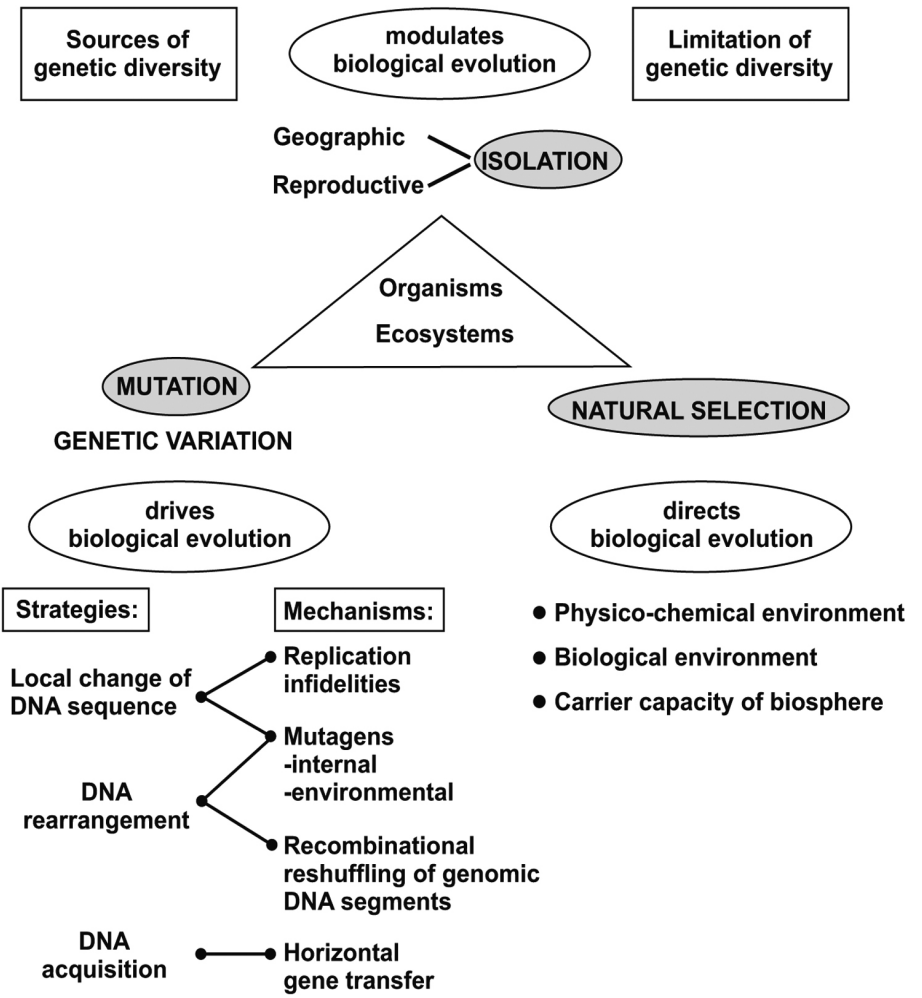


Figure 1. Schematic representation of elements involved in biological evolution and of the mechanisms and natural strategies of genetic variation.

3. Natural Reality Actively Takes Care of Biological Evolution

The specific molecular mechanisms of generation of genetic variants identified so far, some of which will be discussed in more detail below, can be classified into three natural strategies of genetic variation. Each of these strategies contributes qualitatively differently to the process of biological evolution.

3.1. *Local Sequence Change*

As it is shown in the lower left part of Figure 1, important contributions to the strategy of local change in the DNA sequence come from various types of replication infidelities that can, for example, result from tautomeric forms of nucleotides, from replication slippage or from a certain degree of chemical instability of nucleotides. Another source of local sequence change can be an impact of a chemical mutagen. It is well known that these kinds of mutational events would seriously threaten genetic stability if there were not could any enzymatic repair systems that could detect latent states of mutagenesis and reestablish the parental nucleotide sequence. Local sequence changes can occasionally give rise to the improvement of an available biological function or very rarely also to a new biological activity.

3.2. *Rearrangement of DNA Segments*

A second natural strategy of genetic variation is the intragenomic rearrangement of segments of DNA. In general, recombination enzymes catalyze the rearrangements that can result in the duplication, deletion, inversion or translocation of a DNA segment, or else in other kinds of reassortments. Occasionally, novel combinations can lead to a gene fusion or to the fusion of an open reading frame with an alternative sequence for the regulation of gene expression.

3.3. *DNA Acquisition*

The third natural strategy of genetic variation is DNA acquisition in which the genome takes up a segment of foreign DNA that had invaded the cell in question by horizontal/lateral gene transfer. This process is, in fact, at the basis of experimental microbial genetics and it is therefore well understood. More recently, horizontal gene transfer has also been identified

to occur in higher organisms. The quality of DNA acquisition for biological evolution can be seen in a sharing by the recipient organism in successful developments made by others. Various, mostly enzymatically mediated limitations of DNA acquisition keep the frequencies of DNA acquisition low, thus insuring sufficient genetic stability.

3.4. *Involvement of Products of Evolution Genes and of Non-Genetic Elements in the Generation of Genetic Variants*

The theory of molecular evolution postulates that in living organisms all three natural strategies of genetic variation contribute with their differing qualities to the steady process of biological evolution at the population level. Products of specific genes, which are called evolution genes, are involved in most mutagenesis events of all three strategies. However, non-genetic elements are also involved, such as a certain degree of chemical instability of nucleotides and an intrinsic structural flexibility of biologically active molecules. Other non-genetic elements with influence on mutagenesis are environmental mutagens and random encounter. In summary then, intrinsic properties of matter together with the products of evolution genes cause spontaneous genetic variations at fine-tuned frequencies. This insures both a relatively good genetic stability of the individual organisms and enough genetic variants in populations to allow for a steady evolutionary progress of the population.

The products of some of the evolution genes such as repair enzymes and restriction enzymes modulate the frequencies of genetic variation to tolerable rates. Other evolution gene activities act as variation generators; they do this also at tolerable rates. It is postulated that these two kinds of evolution gene activities have become fine-tuned in their own past evolution by second-order selection exerted at the population level (Weber, 1996).

3.5. *The Nearly-Stochastic Nature of Genetic Variation*

This leads us to raise the question whether genetic variation is brought about by directed processes. In other words, does a bacterium, for example, possess a sensory organ to identify what kind of genetic change in which gene could allow the cell to adapt to an encountered altered environment. As far as we know, generally this scenario does not correspond to reality. Rather, spontaneous genetic variation occurs largely at random and not directively. As a matter of fact, only a relatively small minority of genetic

variants turn out as favorable in comparison with their parental form. Many more spontaneous genetic variants provide to the concerned organism a selective disadvantage, and many other sequence alterations turn out to be neutral and silent under selective pressure.

4. Selected Examples of Molecular Mechanisms Generating Alterations in DNA

4.1. *Nucleotide Substitution Caused by Tautomeric Forms of Nucleotides*

Tautomery of nucleotides is known as one of the sources for nucleotide substitution, which is a local sequence change. In its standard form, adenine (A) pairs with thymine (T) in the double-stranded DNA molecules. Upon replication of DNA, the two daughter molecules will carry A-T pairs at the same sites as their parent, as long as the involved nucleotides are in their standard form. Occasionally, adenine can assume for a short moment its tautomeric imino form. Under these conditions, it cannot form a pair with thymine but it can do so with cytosine. In this latter case, as soon as adenine shifts back to its standard form, a mispairing becomes obvious. As we have already discussed, such mispairings can normally be rapidly repaired by appropriate enzymes. But occasionally, the mutation may become fixed as a nucleotide substitution that will later be transmitted to the progeny DNA molecules. In the scientific literature, this is often called a replication error. For conceptual reasons, I disagree with such interpretation. To my mind, natural reality rather makes use of the structural flexibility of tautomeric forms for the occasional stochastic production of nucleotide substitutions. The tolerable frequency of such local mutagenesis is thereby controlled by the efficiency of the enzymatic repair systems. In this case, the stochastic nature of mutagenesis is brought about by the non-genetic property of structural flexibility of the nucleotide, while the genetically encoded repair system modulates the frequency of such mutagenesis.

4.2. *Enzymatic Variation Generators*

Let us now direct our attention to two well-documented examples of enzymatic variation generators involved in the occasional rearrangement of intragenomic DNA segments.

4.2.1. *Transposition of Mobile Genetic Elements*

Mobile genetic elements are segments of DNA that carry genetic information which occasionally promotes either the translocation of the element or other kinds of DNA rearrangements (Shapiro, 1983). One widespread class of mobile genetic elements in bacteria is called IS (for inserted sequences) element. The transposition of an IS element into an alternative location in a genome can sometimes inactivate an important genomic function; in other cases, the process may have less drastic or even no observable effects. Quite rarely, as in most other mutagenesis events, the transposition activity may have a beneficial consequence.

Fig. 2 shows the result of an experiment in which lethal mutations in the genome of the bacterial virus P1 were screened for. The P1 genome can reside for prolonged times in the bacterial host as a provirus without producing viral particles. Only occasionally virus reproduction becomes spontaneously induced in one of the lysogenic cells. However, by experimental intervention, for example by UV-irradiation, one can induce virus reproduction in almost the entire bacterial population. A screening for bacterial subpopulations that do not produce viruses upon induction can reveal the presence of a lethal mutation in the resident viral genome. For the experiment reported in Figure 2, a population of bacteria that carried the P1 genome was propagated for several months under normal growth conditions and with periodical dilution into fresh medium (Sengstag and Arber, 1983). About 1% of the cells could then be shown by an appropriate screening not to produce active P1 virus any more. These cells still carried the P1 genome and about 95% of these independent isolates had suffered in the P1 genome an insertion of one of the IS elements that reside in the bacterial genome (most often it was the IS2 element). Obviously, in all these cases the IS insertion must have inactivated a viral function essential for viral reproduction. The P1 virus had suffered a lethal mutation. The upper part of Figure 2 shows the crude location of independent IS insertions within the 90,000 base pairs long P1 genome that is known to be densely populated with essential viral genes. Clearly, not all regions were used as insertion sites with comparable frequencies. Individual insertion sites of bacterial IS elements within the hottest region of IS insertion are shown in the lower part of Figure 2. Each of nine sequenced IS2 insertions had occurred at another location and these locations did not revel sequence homologies.

In a later experiment, either a stretch of the very hot region or a stretch of a cold region for IS2 insertion of the P1 genome was cloned in a plasmid

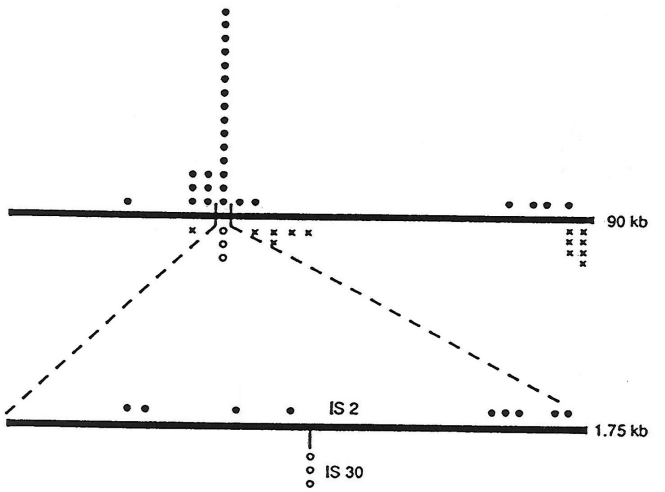


Figure 2. Location of independent IS insertions into the genome of bacteriophage P1 and resulting in mutants affected in the vegetative reproduction of the phage. The circular 90-kb genome of P1 is shown linearized in the upper part. Dots shown above the genome identify independent IS*2* insertions, crosses shown below the genome refer to insertions of IS*1*, IS*3*, IS*5*, Tn*1000* and circles to IS*30* insertions. As shown in the lower part, each of nine sequenced IS*2* insertions into the hot region for IS*2* transposition had occurred into a different sequence, while the three independent IS*30* insertions had occurred between the same base pairs and both orientations had been used. After Sengstag and Arber (1983).

gene vector (Sengstag and Arber, 1987). The resulting hybrid plasmids were then introduced into bacterial cells. Upon subsequent culture of these cells the hot region was still hot for IS2 insertion, while the cold region from the P1 genome was still cold at its plasmid location. In conclusion, there must be some still unknown regional characteristics in DNA molecules that can render this region attractive for IS2 insertion. Interestingly, however, each of the resulting transposition derivatives suffers the IS2 insertion at an individual location; the insertion event appears to be stochastic.

In contrast to IS2, the IS30 element shows a high degree of sequence specificity for its transposition. In the experiment reported in Figure 2, IS30 was found inserted into the P1 genome in three independent derivatives. All three insertions had occurred precisely at the same site of the P1 genome, twice in one direction and once in the opposite direction. IS30 is a good

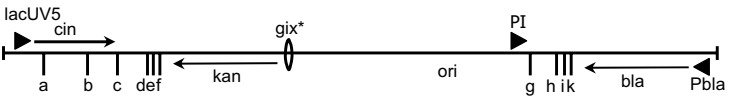
example for a mobile genetic element inserting with high probability into a specific sequence of nucleotides. With much lower frequencies, however, IS30 can also insert into other sites of DNA molecules.

With regard to its evolutionary relevance, IS transposition into many different genomic sites is of course more important than a nearly reproducible insertion into a specific and preferred DNA sequence.

4.2.2. *Site-Specific DNA Inversion Can Also Involve Secondary Sites of Inversion*

As the used terminology indicates, in site-specific DNA inversion a given segment of DNA can become inverted at its given location upon an enzymatic interaction (Glasgow *et al.*, 1989). The borders of this DNA segment are nearly-specific 'consensus' nucleotide sequences. In the well studied case of a DNA segment carried in the genome of the bacterial virus P1 the consensus sequence is 26 base pairs long and has a directedness in spite of a partial dyad symmetry. Two such consensus sequences carried in opposite directions flank a DNA segment that carries two different partial genes for determining the host range of the P1 virus. Just outside of one of the borders of the concerned DNA segment is another part of genetic information for the host range of the virus. In a so-called flip-flop process the DNA segment located between the two sites for inversion becomes periodically inverted. In both of the two possible orientations of the invertible DNA segment the host range gene product is active, but the host range of the virus differs for the two possible states of inversion (Iida, 1984). This is explained by the fact that a partial reading frame for the host range gene is constant and located outside of the invertible segment, while two alternative other parts of the reading frame are carried in opposite directions at the two ends of the invertible DNA segment. DNA inversion brings about an alternative fusion of the partial reading frames. Each of the two resulting fused genes determines its own specific host range of the P1 virus.

An interesting evolutionary contribution of such flip-flop systems of DNA inversion is brought about by the fact that, with much lower frequencies than the flip-flop reaction, the enzymatically mediated DNA inversion can also occur at so-called secondary sites of inversion. Figure 3 reports the results of an experiment in which an appropriate plasmid had been constructed with just one consensus inversion site (here called gix*). Proximate to the gix* site the plasmid carried an open reading frame for kanamycin resistance (kan), but without an expression promoter. As a matter of fact, the plasmid carried two promoters for gene expression (lacUV5 and PI) in the



	gix*	TTCctgTAAACC AA GGTTTTgGATAA	
a)	gix*Q391	Tatcacctctgg AA aGcaTacGgagt	(1)
b)	gix*Q390	aTagatgAttgC AA ccTgcTtacggg	(1)
c)	gix*Q389	cgCctgatAcag AA cGaaTTgctTgc	(2)
d)	gix*Q387	TTgggcTtcgga At cGTTTTccgggA	(5)
e)	gix*Q386	TTCTatgAAAgg tt GGgcTTcGgaAt	(1)
f)	gix*Q385	ccgcCtTctAtg AA aGgTTggGcTtc	(2)
g)	gix*Q384	actcttTttcCg AA GGTaactGgctt	(2)
h)	gix*Q388B	aTCTaggtgAag At ccTTTTtGATAA	(1)
i)	gix*Q393	aaaTtaaAAAtg AA GtTTTaaatcAA	(6)
k)	gix*Q392	aagTttTAAAtC AA tcTaaagtATAt	(1)

Figure 3. Nucleotide sequences used as secondary crossing over sites in Cin-mediated site-specific DNA inversion. On plasmid pSHI383 rare DNA inversion between the natural crossing over site *gix** and a secondary crossing over site brought the expression of the kanamycin resistance gene *kan* under the control of either promoter lacUV5 (sites a to f) or promoter PI (sites g and h). The plasmid also underwent unequal cointegration using *gix** and either site i or k and resulting in the fusion of *kan* with the operon under control of promoter P*bla*. Nucleotides corresponding to the *dix* consensus sequence of efficient crossing over sites are shown as underlined capital letters. Numbers in parenthesis refer to the number of independent isolates having used the crossing over site in question. The data were pooled from Iida and Hiestand-Nauer (1987), and the figure is reproduced from Arber (1995).

direction opposite to the kan reading frame, while a third promoter (Pbla) could not promote kan expression in view of its distant location. After a prolonged propagation of this plasmid in bacteria, kanamycin resistant derivatives were selected and the structures of their plasmids studied. Many derivatives had undergone DNA inversion or unequal cointegration between sister plasmids. In all of these cases analysed, recombination had occurred between the gix* consensus site and a secondary site (also called quasi site) on the plasmid (Iida and Hiestand-Nauer, 1987). Figure 3 shows the locations of ten different observed secondary sites and their nucleotide sequences (Arber, 1995). Capital letters identify nucleotides that still corre-

spond to the consensus sequence. Considerable deviations from the consensus sequence are obvious and they vary from case to case. In this experiment a total of 22 independent kanamycine resistant insertion derivatives carrying one of the promoters for gene expression in front of the kan reading frame were sequenced. Although all studied invertants were independent, the secondary site (i) was used six times and site (d) five times. Three other sites (c, f and g) were used twice, while the remaining five identified secondary sites were used only once. In conclusion, many DNA sequences can serve as secondary sites for DNA inversion and plasmid cointegration with some statistical reproducibility: DNA inversion can occur repeatedly at a given secondary site, although always with very low frequency.

As it is the case for the much more frequent flip-flop reaction described above, the use of secondary sites of inversion depends on the presence of the specific DNA invertase enzymes (in this case the product of gene cin) and of a system for recombinational enhancement. Therefore, in all cases the observed DNA inversions are mediated by enzyme interactions.

In subsequent experiments it could be shown that DNA inversion involving secondary sites of inversion and leading to novel sequence fusions occurs also within the bacterial genome (Rozsa *et al.*, 1995).

What could then be the evolutionary role of DNA inversion systems? In the discussed experimental set-up this cannot be the relatively efficient flip-flop reactions which alternatively activate one or the other of the two observed viral host ranges. The two possible host ranges provide to the viral population a widening of its host range, but the flip-flop process does not really represent an evolutionary progress. In contrast, the much rarer use of secondary sites for the DNA inversion can provide a large number of possibilities for novel gene fusions as well as of fusions of open reading frames with alternative promoters of gene expression. Occasionally, one such novel fusion may provide to the concerned organism a selective advantage, thus contributing to the evolutionary progress. Similar to DNA deletion, DNA inversion thus represents a natural way to fuse previously disconnected functional domains, and some of the resulting recombinants may, by chance, carry out a novel beneficial function. Compared with DNA deletion, DNA inversion has thereby the advantage that no DNA sequences are lost from the genome. For these reasons, I consider enzymatically mediated DNA inversion as an effective natural contribution to the evolutionary progress. The involved enzymes act actively, but at low frequencies, as variation generators. I consider the genes that promote these reactions as evolution genes.

4.3. *Several Fine-Tuned Systems of Horizontal Gene Transfer Contribute to Microbial Evolution*

Microbial genetics originated in the 1940s when several basic mechanisms for horizontal (also called lateral) gene transfer became known. These are the transformation with free DNA molecules (Avery *et al*., 1944), conjugation mediated by fertility plasmids (Lederberg, 1947) and virus-mediated transduction of bacterial DNA segments (Zinder and Lederberg, 1952). In the meantime, we know that products of specific genes are involved in all of these processes. Some of these genes can be classified as variation generators, others as modulators of the frequencies of genetic variation. For example, restriction-modification systems render horizontal gene transfer permissive within populations of bacteria belonging to the same restriction-modification system, while they largely reduce, but not completely suppress, the success of acquisition of DNA from more unrelated microorganisms (Arber, 1965, Arber and Linn, 1969). The stochastic nature of DNA acquisition can be seen at several levels of the process: in the mobilization of donor DNA for horizontal gene transfer, in the random encounter of the transferred DNA with a potential receptor cell and in the chance of the transferred foreign DNA to become part of the receptor genome.

Referring to section 4.2.1, it should be mentioned that bacterial IS elements do not only play an evolutionary role in intragenomic DNA rearrangements, they also contribute to horizontal gene transfer by their recombinational activities for the mobilization of chromosomal genes, as well as for the insertion of transferred genes into the receptor genome (Iida *et al*., 1983). This had first been documented for genes providing antibiotic resistance and it has later been shown to contribute also to the horizontal transfer of other genes.

5. Conclusions and Outlook

Biological evolution is a steady, long-term process. It is driven by spontaneous genetic variation and this is largely, although not entirely, stochastic. Many specific molecular mechanisms contribute with their specific ways and qualities to the generation of occasional genetic variants. Some of the driving forces of genetic variation are exerted by products of evolution genes, while other genetic variations can be assigned to effects caused by intrinsic natural properties of the non-living world. Overall, genetic vari-

ation is thus understood to result from a tight cooperation between non-genetic elements, on the one hand, and products of specific evolution genes, on the other hand. Evolution genes also keep the frequencies of genetic variation low; this insures a relatively high genetic stability in populations. It is assumed that in their long past history, the evolution genes of today's organisms have become fine-tuned for their evolutionary tasks by second-order selection at the level of populations.

5.1. *How Predictable is the Evolutionary Progress?*

A long-term prediction that one can make is that biological evolution will continue to go on as long as the organisms can find environmental niches in which life can persist. Even if some biodiversity will be lost, the intrinsic power for evolutionary progress will steadily replenish biodiversity as it has done to create the present degree of biodiversity. However, expansion of biodiversity becomes limited by the carrier capacity of the planet Earth.

Almost nothing else can be accurately predicted for biological evolution. On the basis of today's knowledge one can confess not to be able to make precise predictions on naturally upcoming events of genetic variation; in the best of cases only statistical predictions can be made. Changes in environmental conditions affecting natural selection cannot be predicted accurately either, nor is this possible with regard to isolation phenomena. Biological evolution is an open system.

In this context, it can be considered as good news that we can predict that all encountered living organisms must possess a certain degree of genetic stability. If this were not insured, organisms with unstable genomes might have already disappeared from our planet by natural selection.

In turning our attention to those evolution genes that act via their enzymatic products as variation generators, one can realize, with surprise, that in contrast to many other genes, such as housekeeping genes, variation generators neither act reproducibly from case to case, nor efficiently (Arber, 2005). Otherwise, they would not be effective generators of genetic variations. This is illustrated by the two described examples of bacterial IS elements and of site-specific DNA inversion systems using secondary sites of nucleotide sequences for inversion. The underlying gene products are bona fide enzymes that exert their evolutionary activities with very low frequencies and at a large number of different sites in the genome.

5.2. *World View Aspects and the Duality of the Genome*

In the past, many textbooks on genetics and biological evolution described genetic variation as due to errors, mistakes and illegitimate processes. In the light of an updated theory of molecular evolution this interpretation has to be replaced by a more pro-active attitude in the interpretation of available data. According to the views that I have presented in this article, biological evolution is seen as an active process, in which natural reality cares by a kind of self-organization for the steady preparation of novel genetic variants that are submitted to natural selection. Since the living conditions vary both spatially and temporally, a stochastic provision of genetic variants can make sense, since it can allow for beneficial adaptation at the population level to a number of different encountered environmental conditions.

One of the philosophical consequences of the postulate of evolution genes is the duality of the genome (Arber, 2005). We can realize that many, but not all, of the genes carried in a genome serve to satisfy the needs of the individual life of the carrier of the genome. They contribute to the fulfillment of the individual life. In contrast, the evolution genes contribute, at the population level, to a steady expansion of life, to the adaptation to novel living conditions and to a steady replenishment of biodiversity. This is welcome news of hope for us human beings, although we cannot precisely predict how the numerous forms of life on our planet will further develop in the long term.

REFERENCES

Arber, W. (1965), Host-controlled modification of bacteriophage, *Annu. Rev. Microbiol.*, 19, 365-378.

Arber, W. (1995), Genetic bases for evolutionary development of microorganisms, In: van der Zeijst, W.P.M. *et al.* (eds.), *Ecology of Pathogenic Bacteria; Molecular and Evolutionary Aspects*, Royal Netherlands Academy of Arts and Sciences, pp. 3-13.

Arber, W. (1997), The influence of genetic and environmental factors on biological evolution, *Commentarii*, Pontifical Academy of Sciences, vol. IV, 81-100.

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

Arber, W. (2005), Dual nature of the genome: Genes for the individual life and genes for the evolutionary progress of the population. *IUBMB Life*, 57, 263-266.

Arber, W. (2007), Genetic variation and molecular evolution, In: Meyers, R.A. (ed.), *Genomics and Genetics*, Wiley-VCH, Weinheim, vol. 1, 385-406.

Arber, W. and Linn, S. (1969), DNA modification and restriction, *Annu. Rev. Biochem*., 38, 467-500.

Avery, O.T., MacLeod, C.M. and McCarty, M. (1944), Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med*., 79, 137-158.

Glasgow, A.C., Hughes, K.T. and Simon, M.I. (1989), Bacterial DNA inversion systems, In: Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*, American Society for Microbiology, Washington, DC, pp. 637-659.

Iida, S. (1984), Bacteriophage P1 carries two related sets of genes determining its host range in the invertible C segment of its genome, *Virology*, 134, 421-434.

Iida, S. and 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.

Iida, S., Meyer, J. and Arber, W. (1983), Prokaryotic IS elements, In: Shapiro, J.A. (ed.), *Mobile Genetic Elements*, Academic Press, Inc., New York, pp. 159-221.

Lederberg, J. (1947), Gene recombination and linked segregation in *E. coli*, *Genetics*, 32, 505-525.

Mayr, E. (1982), *The growth of biological thought: Diversity, evolution and inheritance*, Harvard University Press, Cambridge MA

Rozsa, F.W., Viollier, P., Fussenegger, M., Hiestand-Nauer, R. and Arber, W. (1995), Cin-mediated recombination at secondary crossover sites on the *Escherichia coli* chromosome, *J. Bacteriol*., 177, 1159-1168.

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

Sengstag, C. and Arber, W. (1987), A cloned DNA fragment from bacteriophage P1 enhances IS2 insertion, *Mol. Gen. Genet*., 206, 344-351.

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

Watson, J.D. and Crick, F.H.C. (1953), Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid, *Nature*, 171, 737-738.

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

Zinder, N. and Lederberg, J. (1952), Genetic exchange in Salmonella, *J. Bacteriol.*, 64, 679-699.