ISSN 0006-2979, Biochemistry (Moscow), 2007, Vol. 72, No. 12, pp. 1284-1288. © Pleiades Publishing, Ltd., 2007.
Original Russian Text © A. S. Antonov, 2007, published in Biokhimiya, 2007, Vol. 72, No. 12, pp. 1576-1582.

From Birth to Christening

A. S. Antonov

Faculty of Biology, Lomonosov Moscow State University, 119992 Moscow, Russia;
fax: (495) 939-3181; E-mail: antonov@genebee.msu.su

Received July 26, 2007

Abstract—A brief history of comparative studies of nucleic acids for systematic purposes is given. These studies were initiated by a group of Moscow State University scientists headed by A. N. Belozersky. Based mostly on comparative DNA studies, some main dogmas of a new branch of systematics were gradually developed. In Russia, this new branch of systematics is called "genosystematics". Some of the main results obtained by genosystematics since its birth (1957) and up to its "christening" (1974) are described.

DOI: 10.1134/S0006297907120024

Key words: genosystematics, phenosystematics, DNA, systematics, evolution of organisms

In this paper I try to give a brief overview of the initial stages of the science that is in Russia called "genosystematics". This is very young science, it is just fifty years old, and it was named only seventeen years after birth. Such late "christening" may be explained by the need for clear determination of the subject for studies before definition as a particular branch of science. Representatives of "classical systematics" were opponents of this new science.

During the half-century history of genosystematics, it was influenced by willed or unwilled distortions. In this sense, my work is not an exception because I have written it using my own experience and observations; it is basically subjective. However, this subjectivism may be excused because I have been working in this new field of science from its birth up until now and have always tried not to distort history.

E. Chargaff, one of the "godfathers" of genosystematics, wrote about the willed distortions: "Quite recently it was relatively easy to discover new fields for activity and to treat these: nobody was concerned that he would be immediately robbed as it would almost inevitably happen now... Bibliographic references were compiled rather honestly, whereas now whole blocks of references are "dragged" via some transduction from one paper to another, so that if one paper is not cited any more this is forever. Such a gap in continuity of traditions is probably one of the most devastating consequences of the massive scientific community in which we live now".

In another work he wrote even sharper: "It is difficult to distinguish between hot-spirited search for truth and energetic campaign for making a career. Things initially started as an enterprise for the spirituals have been then transformed into survival for the most pushy and rapid ones" [1, 2].

Unfortunately, the history of genosystematics knows many examples validating the correctness of these words. However, I must say that at least a part of these willed mistakes originated from the existence of a semitransparent barrier between Soviet and foreign scientists, which basically did not know the Russian language.

Genosystematics began from careful chemical analysis of DNA preparations and based on the existing model of DNA molecules. In spite of evident achievements in molecular biology of DNA, the first method developed for discrimination of organisms by differences of their DNAs (e.g. the method of determination of nucleotide composition) attracted little interest (if any) in representatives of the classical school of systematics due to its low resolution capacity. Systematicians could not come to any serious conclusions on the basis of differences in composition of DNAs from calf, wheat germs, or *Bacillus tuberculosis*. They were just ready to take into consideration a hypothesis on tissue specificity of DNA structures. The correctness of this hypothesis has been proved many years later by genomics.

Predicting this situation many years ago zoologist E. Mayr wrote: "Organisms exhibit a unique property, which differs them from nonliving objects: they have genotype and phenotype... When we classify organisms, we classify them by phenotype; this is the first step. As the second step we try to make conclusion about genotype, the genetic program created during evolution, which has more cognitive and prognostic value than phenotype. Phenotypes may share various similarities that are not

related to the real problem and only analysis of logically evaluated genotype helps us to establish which similar features of phenotype are determined by convergence and which reflect ancestor genotypes" [3].

At that time comparative studies of one of the genotype components, DNA, began. This was a time when theory of molecular evolution just began to develop and zoologist-systematician E. Mayr considered data obtained during studies of genotypes and phenotypes as mutually complementary ones. This was a viewpoint widely shared among many scientists. Many years later, a distinguished botanist, A. Cronquist, evaluated importance of genotype studies. He believed that "one of the undoubted advantages of the method of nucleic acid sequencing consists in its relative independence on traditional methods and concepts... Firstly, it allows choosing the best of existing taxonomic schemes. Secondly, it allows proposing new possibilities that have been out of our attention" [4].

Gradually significant differences between the systems proposed by the classical school and the systems that could be built on the basis of results of studies of genotype evolution appeared. Finally, even the distinguished botanist A. L. Takhtajan (a main opponent of genosystematics in Russia at the first stage of its development) had to agree that conclusions made by this science coincide with botanists' viewpoints, studies of genotypes less often allow choosing one of alternative schemes, but unexplainable divergences are observed from time to time.

However, let us come back to the early stages of the development of genosystematics. How has the initial comparative study of evolution of genotypes and phenotypes of organisms been carried out?

In those distant years comparative works on chemical composition of the most important genotype component, DNA, were carried out by biochemists, who were interested in characteristic features of DNA composition in various organisms from viruses to man. At that time nobody knew how universal qualitative composition of DNA in various organisms was. It was found that there were interesting exceptions (e.g. DNA of some phages) from marked universality of chemical composition of canonical A, G, C, and T. Some nucleotides constituting DNA molecules were methylated. This stage of study has been described in the paper by B. F. Vanyushin in this issue.

The situation became interesting for systematicians only after a paper published by A. N. Belozersky's PhD student, A. S. Spirin, and his co-authors [5] in which (sometimes) significant differences in DNA composition were demonstrated even for organisms earlier considered as closely related and referred to the same lower taxon (e.g. enterobacteria). However, this paper was almost missed by the international scientific community due to a language barrier and certain concern of foreign scientists to studies by Soviet authors (the Western audience was familiar with Lysenko's "creative works"). Only after delivery of the plenary lecture by A. N. Belozersky at the Fourth International Biochemical Congress in which he demonstrated data of DNA composition in representatives of the most important groups of the living world from microorganisms to vertebrates and higher plants the works by the Belozersky's school were appreciated and highly evaluated.

Thus, let us to consider 1957 as the year of genosystematics birth.

Based on data accumulated during that period it was concluded that similarity in DNA composition does not say anything about the relation degree between various organisms, whereas differences in DNA composition definitely indicate that the compared organisms are not closely related to each other. Putative differences between DNAs of identical composition should be searched by comparing their nucleotide sequences.

During that period an interesting observation was made. We compared variation limits of DNA composition (the content of GC) in some large groups of animals separated by systematicians including protozoa, sponges, coelenterates, echinoderms, and chordates. The results demonstrated that earlier evolutionary formation of such groups corresponded to more pronounced differences in DNA composition among species forming animals. We explained this phenomenon of "molecular clocks" by duration of mutation process in the group, but other interpretations were also possible.

It was impossible to evaluate the accuracy of such "molecular clocks" because the chronicle of paleontology was too poor for elucidation of precise dates [6, 7]. However, these data obtained during studies of DNA evolution were in accordance with the hypothesis of the molecular clock proposed just before by Zuckerkandl and Pauling during comparison of structure of related proteins.

One of the most important consequences of early stages of the development of genosystematics was the conclusion made by A. Ravin [8, 9]. From his viewpoint, one should distinguish "phenospecies" and "genospecies" of microorganisms. Such contraposition of results obtained during studies of genotypes and phenotypes was very important. One can say that since that period systematics became to elaborate its own ideology of studies based on possible differences between phenotypes and genotypes.

Ravin also noted that all strains of some bacterial species might insignificantly differ by DNA structure whereas in other species such differences are well defined and are comparable with differences observed between "good" species. Thus, one phenospecies of microorganisms may represent a conglomerate of genospecies. The opposite situation is also possible when microbiologists refer a group of strains (belonging to one genospecies) indistinguishable at the DNA level to one species.

At first glance, this observation was the first indication that notions on the evolutional history of taxons and the systems of phenotypes and genotypes may differ from each other.

However, during deeper consideration Ravin's bacterial "genospecies" were old pals for the classical school systematicians who investigated eukaryotes. Many years before Ravin's discovery, they actively discussed a problem of so-called twin species, which were indistinguishable by phenotypic characters but behaved as "good species". The major reason consisted of insufficient knowledge on phenotypes, but nevertheless this approach had limited capacities. Thus, Ravin recognized similarity in some mechanisms of evolution of pro- and eukaryotes: in both cases there was divergence in the rates of evolution of genotypes and phenotypes.

So-called live fossil plants and animals give us the brightest examples of such divergences. Analysis of the Ginko genome tells us that it continues evolution since the dinosaurs' time, but morphology of plants remains unchanged during this period. Modern latimeria looks exactly the same as its distant ancestors, but its genotype has evolved. The same is true for many groups of ferns.

Many other examples could be also given for illustration.

However, let us come back to the 1960s, to bacteria. New methods attracted attention of systematicians-microbiologists and workers of the branches of food industry employing fermentation processes (fermented milk products, brewery, etc.). The number of organisms with determined DNA composition increased from year to year. During this period many observations directly related to systematics were made.

In 1966, analyzing results of studies on DNA composition in various microorganisms, Hill noted that in taxons of the same rank distinguished by microbiologists by comparison of complexes of phenotypic signs of microorganisms (e.g. within genus) DNA compositions might vary within different limits [10].

He concluded that some bacterial genera recognized by systematicians (e.g. *Lactobacillus*, a genus characterized by by very high variations in DNA composition) pooled together unrelated organisms and therefore this genus included unnatural, artificial species. Correctness of this conclusion was later confirmed by experiments on DNA hybridization. Thus, Hill proposed a principally new mode for determination of naturalness of the separated taxons. But rather earlier such mode was proposed by Moscow State University scientists studying enterobacteria.

If a study of DNA composition in species of some genus showed that one or a few species significantly differ in DNA composition from the main mass of the members of this genus, these species will have been definitely referred to this genus by mistake. N. Sueoka (USA) calculated that if content of GC differs by more than 10 mol% such organisms cannot have homologous DNA sequences, i.e. they diverged a long time ago. However, in reality the use of mathematical evaluation not always gives a correct result. Now in evolutionally stable genes (e.g. in rRNA encoding genes) homologous sequences have been found even in members of different living kingdoms; this indicates that life on the Earth originated from a common ancestor.

In some cases "good" bacterial genera differed however slightly by the degree of DNA variability of their species, but this variability was always significantly lower than in species of one bacterial family. Finally, use of statistical analysis allowed proposing quantitative criteria of the taxon rank in microorganisms by expressing it through variability of DNA composition in the taxon forming species.

Of course, such an approach to determination of equivalence of the taxon rank is not universal and is not free of shortcomings. Sizes of genomes may significantly differ, and so the rate of accumulation of differences in DNA composition will also vary. In taxons with small genomes it will appear faster than in taxons of higher forms. Nevertheless, this approach represented the basis for the first revision of the systems of microorganisms, which took place in 1970s and 1980s before the use of results of determinations of nucleotide sequences of their genomes in microbial systematics.

During that time studies of primary structure of proteins were actively developed. They created a basis for the hypothesis of "molecular clocks" [11]. Soon this hypothesis was theoretically substantiated by the theory of "non-Darwinian evolution"; according to this theory so-called "neutral mutations", which did not manifest in phenotypes, played an important role in evolution of genotypes [12, 13]. According to the hypothesis of molecular clocks, the earlier divergence of the compared organisms caused more pronounced differences in primary structures of their macromolecules—proteins and obviously nucleic acids. Based on this hypothesis certain attempts were undertaken to evaluate the time of divergence of compared taxons.

Soon reliability of such hypotheses was questioned [14, 15]. Large material on DNA composition in mammals and angiosperm plants was accumulated by that time, when it was believed that both groups originated in the Cretaceous period, i.e. they had similar evolutionary age. According to the hypothesis of molecular clocks, one may expect that DNA of both groups of organisms diverged to the same extent, but in reality DNAs of various angiosperm plants exhibit more pronounced differences than mammalian DNAs. In this case, researchers meet the problem of nonequivalence of taxons, having just one range (class) in the system. However, this did not seem to be unusual for systematicians because this is not the case of errors made by systematicians (as in the case of the microorganism system).

In his time E. Mayr wrote that considering genus a systematician of beetles understands a completely different community of organisms than a systematician of butterflies. For genosystematicians it was important that

these data indicated that the rate of accumulation of changes in DNA structure might be different in various phylogenetic lines and at various times.

Discoveries made by genosystematicians raised new problems for classical systematicians. How else could one explain different level of variability of DNA composition in taxons of the same evolutionary age? Explanation of these differences proposed at that time looked "heretical": for example, angiosperms could be a more ancient group than botanists suggested. In other words, more and more contradictions between classic and genosystematic notions on evolution of organisms accumulated, and this has been reflected in corresponding systems.

Evidences have been obtained during evolutionary studies of some proteins. At that time, genosystemicians already had first (still primitive) methods for phylogenetic tree construction [16]. Researchers from the Novosibirsk Institute of Cytology and Genetics actively participated in the development of problems of mathematical genetics [17, 18]. Subsequently they enormously helped Moscow State University scientists in treatment of results of RNA and DNA sequencing.

Finally, D. Boulter et al. based on comparison of plant cytochrome *c* structures and on the hypothesis of molecular clocks demonstrated that the angiosperms represented a more evolutionary ancient group, comparable with fished in this particular sign [19].

The arsenal of methods employed by genosystematicians constantly increased. In the 1960s and 1970s a series of new methods of determination of statistical characteristics of sequences of genetic texts was proposed. One of them was the method for determination of frequency of adjacent nucleotide pairs in DNA (AA, AG, AT, etc., 16 pairs in total). Using this method, it was demonstrated that prokaryotic and eukaryotic DNAs can be differentiated by frequency of some of these pairs. The methods for determinations of frequencies of mono- and oligonucleotide blocks, built only by pyrimidine or purine nucleotides (isopliths), have been developed. Finally, the method of specific enzymatic degradation of RNA and evaluation of quantities of oligonucleotides of various lengths and composition formed during this enzymatic degradation was proposed.

Using this method for analysis of rRNA of microorganisms, Woese and colleagues made one of the greatest discoveries in biology of the twentieth century. They discovered a new kingdom of living nature, archaebacteria. This was such a sensation that many researchers validated it using different methods for a rather long time. The paper summarizing the final results of these studies "Towards a natural system of organisms: proposal for domains *Archaea*, *Bacteria* and *Eukarya*" was published in 1980 [20]. At that time, the comparative analysis of oligonucleotides formed during hydrolysis of cyanobacterial rRNA and higher plant chloroplasts provided experimental substantiation of the hypothesis of symbiogenetic origin of chloroplasts. These studies gave convincing evidence for the correctness of notions according to which genosystematics actually has its own research object (genotype) and it represents one of two main sections of systematics as a science.

What is genotype in terms of molecular biology? Starting from the abstract notion "bank of genetic information" (given in the above-mentioned expression by E. Mayr) scientists gradually came to the conclusion that in terms of molecular biology the genotype is a complex multicomponent system including not only the "keeper" of genetic information (DNA), but also products of its realization at the level of RNA and proteins. This system is formed in accordance with environmental conditions, and that is why the same set of genotype genes may result in formation of several phenotypes accordingly differing from each other.

Efforts of genosystematicians to get a more comprehensive notion about homology of polynucleotide sequences resulted in the development of the method of nucleic acid hybridization. Initially it was thought that this method would be applicable for quantitative characterization of primary structures of DNA or RNA; however, soon disappointment appeared. During that time the complex structure of higher-organism DNA containing unique and repeated nucleotide sequences was recognized. The ratio of these unique and repeated sequences could significantly differ even in genomes of closely related forms and so quantitative evaluation of similarity in each fraction would be very insignificant. The first versions of this method were based mainly on hybridization of repeats.

In 1972, an attempt to pool together all the accumulated results on comparison of DNA was undertaken. A. N. Belozersky and his team collected a group of authors who wrote a series of review papers published under a common title "DNA Structure and Position of Organisms in the System" [21]. However, this and a subsequent publication of Moscow State University scientists [22] were missed abroad. However, several years later, when the method of nucleic acid hybridization became widely used, a similar edition was initiated by the American biochemist S. K. Datta. He ordered review papers from an international group of authors (including Russian scientists). At the time of publication of this book, I was in the USA where I met Prof. Datta. He showed me the book, which was ready for publication and which contained a short preface. Reading that preface I indicated some inaccuracies in presentation of history of "DNA systematics", which he had to accept in order to tell the truth. It was technically impossible to rewrite the preface, and S. K. Datta was able to add a few phrases to it. Below I give this S. Datta's preface (italics shows the added phrases): "Historically studies in the field of DNA systematics were initiated *more than twenty years ago* by Ellis Boulton, Roy Britten, Bill Haier, David Cohen,

Brian McCarty, and other scientists at the Carnegie Institution of Washington, and *also by other scientists in England*, *France*, *and in the USSR. These studies have been reviewed in the books published by Moscow State University Press in 1972 and 1980 and edited by A. N. Belozersky and A. S. Antonov*. For evaluation of similarity and differences of organisms, the authors mainly used the method of DNA–DNA hybridization, which is now considered as "the most suitable of existing methods" for building of the similarity tree as Levin writes. Using this method, C. G. Sibley and J. E. Ahlquist ... proposed the "DNA molecular clocks" used for phylogenetic tree building. *Earlier similar clocks were proposed by Antonov and his colleagues in the USSR*..." [23]. The result was a bit clumsy but it was definitely closer to the truth.

Disorder in names of the new direction in systematics appeared already at the first phase of studies (in the 60s and 70s). Initially "DNA systematics" was the most popular name, but later it became clear that it did not correspond to the scale of studies, because besides DNA these studies also employed RNA, proteins, and their complexes.

One of possible proposed variants for the name of new direction in systematics, the creation of molecular biology, was based on the commonly accepted subdivision "phenotype–genotype" [24]. The term "genosystematics" or "systematics of genotypes" has been proposed for comparison of genotypes. Thus, in Russia christening of a new direction in science has happened.

The other direction in systematics was proposed to be named as phenosystematics. The term "genosystematics" was proposed to underline the difference in research objects. Genosystematicians proposed use of the term "phenosystematics" as the homonymous to its meaning widely used by "classic" systematicians. For convenience of opposition, genosystematicians use this term to pool all directions of "classic" systematics (evolutionary systematics, cladism, etc.) based on results of studies of organism phenotypes.

It was originally suggested that the main task of genosystematics consists in determination of similarity degree of primary structures of genomes (cell DNA) and isolation of natural groups of genotypes from the biodiversity. Biologists define similar groups of organisms as taxons, whereas genosystematicians propose to name the groups of organisms with similar genotypes as genotaxons. The next task was to determine evolutionary links between them (this is the major goal for molecular phylogenetics) and to "award" certain rank to genotaxons in the hierarchy (i.e. creation of the system of genotaxons of organisms) if necessary.

In the English language literature such "uniting" term as genosystematics is basically not used, whereas such notions as DNA-systematics, molecular systematics as well as macromolecular phylogenetics are widely employed.

Molecular taxonomy partially solves the tasks of genosystematics. Construction of a system takes into consideration data obtained by molecular phylogenetics. However, alternative interpretation of data obtained by molecular phylogenetics is also possible. In forming species, it is proposed to distinguish clads of various levels of genetic relation. Possibility of use of so-called "phylocode" in systematics and many other innovations are also actively discussed. However, these are matters of modern but not past days.

REFERENCES

1. Chargaff, E. (1963) *Essays on Nucleic Acids*, Elsevier, Amsterdam-London-N.Y.
2. Chargaff, E. (1976) *Khimiya i Zhizn'*, **7**, 45-49.
3. Mayr, E. (1971) *Principles of Zoological Systematics* [Russian translation], Moscow, Mir.
4. Cronquist, A. (1988) *The Evolution and Classification of Flowering Plants*, N. Y. Botanical Garden, N. Y.
5. Spirin, A. S., Belozersky, A. N., Shugaeva, N. V., and Vanyushin, B. F. (1957) *Biokhimiya*, **22**, 744-754.
6. Antonov, A. S. (1966) in *Nucleic Acids* [in Russian], Meditsina, Moscow.
7. Antonov, A. S. (1966) *Abst. of the IIIrd FEBS Meet.*, Warsaw, No. 264.
8. Ravin, A. W. (1963) *Amer. Natur.*, **97**, 307-318.
9. Ravin, A. W. (1963) *Discussion*, in *Systematic Biology*, National Academy of Sciences of the USA, Washington DC, Publ. No. 1692, 275-278.
10. Hill, L. R. (1966) *J. Gen. Microbiol.*, **44**, 419-437.
11. Zuckerkandl, E., and Pauling, L. (1965) in *Evolving Gene and Proteins* (Bryson, V., and Vogel, H. J., eds.) Academic Press, N. Y., pp. 97-166.
12. Kimura, M. (1969) *Proc. Natl. Acad. Sci. USA*, **63**, 1181-1183.
13. Kimura, M. (1979) *Sc. Am.*, **241**, 94-104.
14. Marinova, E. I., Antonov, A. S., and Belozersky, A. N. (1969) *Dokl. Akad. Nauk SSSR*, **184**, 483-486.
15. Marinova, E. I., Vladychenskaya, N. S., and Antonov, A. S. (1969) *Dokl. Akad. Nauk SSSR*, **185**, 945-948.
16. Fitch, W. M., and Margoliash, E. (1967) *Science*, **165**, 279-284.
17. Ratner, V. A. (1972) in *Studies on Theoretical Genetics* [in Russian], Institute of Cytology and Genetics, Novosibirsk.
18. Ratner, V. A. (1974) in *Problems of Mathematical Genetics* [in Russian], Institute of Cytology and Genetics, Novosibirsk.
19. Boulter, D., Ramshaw, J. A. M., Thompson, E. W., et al. (1972) *Proc. Roy. Soc. London B.*, **181**, 441-455.
20. Woese, C., Randler, O., and Wheels, M. (1980) *Proc. Natl. Acad. Sci. USA*, **87**, 4576-4579.
21. Belozersky, A. N., and Antonov, A. S. (eds.) (1972) *DNA Structure and Position of Organisms in the System* [in Russian], Moscow University Press, Moscow.
22. Antonov, A. S. (ed.) (1980) *Molecular Bases of Genosystematics* [in Russian], Moscow University Press, Moscow.
23. Dutta, S. K. (ed.) (1986) *DNA Systematics*, Vol. 2, CRC Press, Boca Raton.
24. Antonov, A. S. (1974) *Usp. Sovr. Biol.*, **77**, 31-47.