Russian Journal of Genetics, Vol. 38, No. 8, 2002, pp. 851–868. Translated from Genetika, Vol. 38, No. 8, 2002, pp. 1013–1033.

When we classify organisms, phenotypic classification is but the first step. The second step is an attempt at making a conclusion about the genotype, this evolutionarily formed genetic program, which is of a far higher cognitive and prognostic value than phenotype. Phenotypes may possess similarity that is irrelevant, and only an examination of the genotype can establish, which similar phenotypic traits are caused by convergence, and which express the ancestral genotype.

E. Mayr, 1968 [1]

Molecular DNA Markers in Phylogeny and Systematics

V. V. Grechko

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991 Russia; e-mail: grechko@eimb.ru

Received December 2, 2001

Abstract—The review considers data on the use of the main evolutionary markers (ribosomal, mitochondrial, and RAPD markers; dispersed and tandem repeats). Some circumstances impending analysis of these data are discussed.

INTRODUCTION

Recent years have witnessed increasingly rapid development of molecular phylogenetics and systematics. This is caused by the development of new diverse methods of analysis of molecular DNA markers. These methods allow researchers to assess genetic relationships among taxa at a new, more advanced level and obtain new evidence concerning their phylogeny and biodiversity. The advances in the use of molecular markers for solving general biological problems are reviewed in [2–6]. Apparently, applying molecular markers to the increasing range of species and other taxa indicates importance and actuality of this line of research.

At the same time, understanding and comparison of the results obtained on the same organism using different markers are often in conflict with the views of zoologists on the morphological (classical) systematics. The reasons of this conflict are diverse. On the one hand, they are explained by the great complexity of molecular data analysis, which is largely based on modeling (with introduction of various assumptions) of the evolutionary process that proceeds differently and at unequal, non-uniform rates in different taxa. On the other hand, the "molecular boom" in science, as everywhere else, has produced the situation when a universal enthusiasm for molecular markers among researchers who are only superficially acquainted with molecular biology promoted hundreds of studies, in which the use of some markers is unfounded and conclusions are based on a statistically nonsignificant number of characters and taxa. The most problematic feature of both molecular and morphological approaches is the fact that both of them cannot always discriminate between convergent and homologous characters and lack criteria for determining evolutionary significance of the character employed in phylogenetic or taxonomic analysis.

Some authors believe that the significance of molecular marker studies has been diminishing with the advent of sequencing of the human and other genomes. Indeed, the consequences of this breakthrough can hardly be overestimated, but also are hard to predict, since a direct comparison of huge DNA molecules is extremely difficult technically and methodologically and requires detecting of orthologous regions in different taxa. The latter is a separate task (see [4]), which reduces the problem of a comparison of genomes to that of analysis of genome parts.

In prokaryotes, the number of completely sequenced small (of the order of 10^6 bp) genomes amounts to several hundred (see [7] for review) and will reach several thousand in the near future but the number of taxa in this group numbers in hundreds of thousands. In eukaryotes, DNA sizes are greater by 3 to 6 orders of magnitude, and the costs of their sequencing are fantastically high. The efforts of hundreds of researchers from several countries were required to sequence the human genome. It is unwarranted to expect the same advances in sequencing the genomes of other taxa in the coming decades. Of course, sequencing the genomes of members of main classes and other major taxa will result in incomparable expansion of understanding pathways and principles of macroevolution, the evolutionary roles of horizontal transfer and genome instability. An example of such analysis is provided by comparison of the human genome with that of

1022-7954/02/3808-0851$27.00

a lower organism, nematode. However, microevolution and direct phylogenetic analysis of millions of taxa can hardly be based on these results although the principle of the choice of DNA regions as speciation markers will be better understood.

Molecular biologists who have dared to enter evolutionary studies with their methodology and tools are faced with the necessity to participate in the current discussion between the advocates of cladistical approach to systematization of living organisms (based only on phylogeny; Hennig 1965, cited from [8]) and their opponents who find significant and yet insurmountable shortcomings of this approach (see Emel'yanov 1989, cited from [9]) related exactly to its direct application to classification and systematics. In my opinion, the paradox of this situation lies in the fact that the proponents of cladistics, who attempted to circumvent the difficulties of the traditional, phenotype-based systematics by proposing to construct classification systems only on the basis of phylogenetic relationships among elementary evolutionary units, could not yet suggest instead a balanced, understandable, and consistent methodology for constructing such system [9].

Another paradox lies in the fact that an advocate of the Linnaean systematics in its pure form is unlikely to be found among biologists beginning from Darwin and Haeckel (see [1]), and any phylogenetic verification, be it paleontological, osteological, karyological, or molecular, is favorably received by taxonomists and accounted for in constructing classifications. Linnaeus himself understood the necessity of this (see the book by N.N. Vorontsov [10]). However, in real life it is hard to imagine a taxonomist who is waiting for the time when the systematic position of a newly found taxon will be phylogenetically substantiated.

This uncertainty results in the following situation: presenting new, molecular phylogenetic relationhips for a known groups, the critics of phenetic (in essence Linnaean) approach nevertheless compare their schemes to the phenetic ones. They are satisfied when the conclusions based on the both approaches coincide; if not, the preference is given to those more reasonably integrating the molecular and morphological results. The literature on molecular markers abounds with the attempts at finding such integration. This may indicate that researchers implicitly assume that genetic relatedness of groups based on their morphology obviously reflects the formation history of the taxon and its ancestors, i.e., the phylogenetic component. Another explanation is "unscrupulousness" of adherents of the phylogenetic approach who refute their results if they do not conform to the standing notions of the classic systematics.

This opinion of the author of the present study has developed during her work with taxonomists and was unexpectedly illustrated by a statement in a book by A.S. Antonov [6] on the situation in the field of botany, in the case when molecular and genetic concepts overlap. On the one hand, referring to important authorities (I.I. Schmalhausen, E. Mayr, A.N. Belozersky, and Th. Dobzhansky), Antonov admits that phenotypic constructions as such are not sufficiently substantiated in plant systematics and that "the system needs considerable revision"; he favorably presents the views of M. Donohue (see [6] for reference) on the necessity to construct a completely different, phylogenetic systematics of plants. On the other hand, Antonov believes that genosystematic data only "can or must be taken into account by classical taxonomists, especially when they help to choose the best among the phylogenetic or systematic hypotheses proposed by them." The author even thinks that "since the evolutionary trees of phenotypes and genotypes are only similar but not identical," the attempts to "harness together a horse and a graceful doe" are bound to be unsuccessful. However, the factual part of Antonov's work, which analyses the contribution of molecular-based phylogeny to the modern views on plant systematics, is in direct contradiction to the pessimistic closing sentence and suggests the necessity to consider molecular (and paleontological) phylogeny as the primary driving force in constructing the new and revising the available, Linnaean, systematics. Thus, the current situation in biology is characterized by the fact that molecular phylogenetic studies employ the traditional operational methods of phenetic systematics and are in delay (in relation to the traditional systematics) in their estimations.

There is also a counter movement, i.e., traditional taxonomists use cladistic terminology describing relationships within large taxa in concrete works. However, for instance, the cladistic notion of "sister taxon" implies that the authors using it accept the major cladistic postulate that the ancestral species cease to exist at speciation giving rise to two new taxa, each of which carries an autapomorphic character. So far, these contradictions remain unresolved. The most consistent adherents of cladistics, like, for example, President of the American Association of Systematics, M. Donohue (cited from [6]) maintains that the whole Linnaean (phenetic) systematics as not based on taxon phylogeny is incorrect, the modern systematics lacks biological meaning, and a completely new systematics must be developed.

I would like to see a compromise between these two alternatives. The main reason is that the similarity between and comparison of phenetic characters is not purely mechanistic and in most cases evidently reveals, albeit implicitly, the shared genetic roots of the groups examined. The occasional erroneous conclusions resulting from possible homoplasias (convergence and reversion of development) can occur in any systems of weakly substantiated notions, of which the concept of species is an example. This issue is discussed in the treatise "Molecular Systematics" [11] and in other monographs.

The second reason is that, as follows from the further discussion, traditional constructions of systematics in many cases do nor enter into a categorical and unsolvable controversy with the data of molecular phylogeny regarded from a cladistic viewpoint. The contradictions can be overcome, in case of lower taxa, by comparing phylogenies based on different markers and by searching for sources of possible homoplasias [12].

Resolving the complicated issues at the lever of higher taxa, the origin of life, multicellular organization, etc., is far more difficult, and the deficit of comparative morphological data and their difficult (and sometimes impossible) comparison in case of distant taxa (and often wrong selection of molecular markers) in many cases leaves grounds to doubt. Although the number of studies dealing with these issues has been increasing, the controversy in the views of the phylogeny of the kingdoms and major taxa still persists (McIntyre 1994; cited from [13]; [14–17]). This is largely related to the development and evaluation of mathematical methods of data treatment (Philippe *et al.* 1996, cited from [18]). However, even here molecular phylogeny becomes a decisive argument whereas unclear subjects and inconsistencies require only further studies. Thus, today we witness the development of a new paradigm of evaluation of biological diversity and its criteria. Molecular biologists should, in particular, cautiously use traditional systematics in attempts of its phylogenetic understanding in terms of molecular markers and employing tools of cladistics.

The hope that molecular markers can be a panacea for understanding evolution seems to be overly optimistic. The most general and main complication is the inconsistency of the assumption on the nonuniform rate and equal probability of mutations (Pauling and Zuckerkandle 1965; Britten 1986; cited from [19]), i.e., the "molecular clock" concept. The temporal changes of different genome parts and, as a consequence, genomic products are not uniform not only in different taxa but even within one taxon [19]. Moreover, such issues as mutation regularities, the contribution of neutral mutations, different frequencies of codon usage in different taxa, frequencies of transitions and transversions, ambiguity of splicing messenger RNA precursors leading to intragenic protein variability, and many other processes affecting molecular marker quality are still poorly understood and cannot be fully taken into account in constructing phylogenies based on molecular markers but have an effect on the results.

The observed discrepancies between morphological systematics and the data on different molecular markers in a taxon may be related to the fact that each morphological or behavioral trait on the molecular level is controlled by several genes and depends on random mutation in any of them. The formation of a morphological or behavioral trait involves not only structural genes of protein synthesis, but also regulatory genes determining switching on and off of mRNA synthesis on the structural genes or their blocks and the synthesis intensity. The phenomena of horizontal gene transfer and complete genome duplication [20] contribute to the complexity of this issue not only in lower, but also in higher eukaryotes. For instance, recently reported striking similarity of a dispersed repetitive sequence in bovine DNA and that of some reptile species [21] places these groups in one pseudo-monophyletic clade.

Thus, the huge amount of literature contains few unambiguous results while phylogenies based on different markers are often conflicting. It seems that we are in the formative period of the basic level of molecular evolution and in search of the simplest and controlled associations between biochemical and morphological characters. In what follows, I will focus on the main pro and contra of the use of the currently most popular molecular approaches.

MARKERS OF GENES FOR RIBOSOMAL RNAs

Ribosomal RNAs and their genes have been used as a phylogenetic tool for more than two decades. Their use in this respect is determined not only by their functional significance but also by the amount of information on relatively different evolution intervals of different parts of the enormous and complex operon of these RNAs (Gerbi 1985, Rotschild *et al.* 1986, Humbi and Zummer 1992, cited from [18]; see also [5, 22]). Thus, in vertebrate operons having most complex organization, the structure of regions encoding ribosomal RNAs (18S, 28S, 5.8S, 5S) is most conserved and used for studying the most ancient stages of evolution of the living matter. Other transcribed but not coding RNA parts (spacers) are located within (Internal Transcribed Spacer, ITS) and at the ends (External Transcribed Spacer, ETS) of the operon; they evolve at faster rates and are used for studying more close genetic relationships. The variability of nontranscribed (or, more exactly, weakly transcribed) spacers (NonTranscribed Spacer, NTS) separating tandemly positioned ribosomal operons is even higher; in addition, lengths of NTSs containing internal repeats depend on their number. These regions can be used to study close relationships including within-population ones. More simply organized operons can lack ITSs or ETSs; the size of NTS varies producing operon sizes ranging from 6.8 kb in *Stilonichia* (Protozoa) to 42 kb in *Dictiostelium* and 44 kb in the rat [23]. Thus, the main field of use of these markers in higher eukaryotes is the relatedness of remote taxa, many of which have little morphological similarity, and in prokaryotes and lower eukaryotes, minimal similarity.

First let us consider some of the enormous number of studies using ribosomal markers that have marked an era in the development of molecular evolution. At the end of this section, I turn to the critical evaluation of these advances in the most recent years. I would like to emphasize, though, that the critical comments pertain only to methodological aspects of the analysis not con-

cerning the standing value of the accumulated experimental material, the possibilities of additional and repeated analysis of which are practically unlimited.

The discoveries of three superkingdoms of living organisms—Bacteria, Archaea, and Eucaria—by K. Woese and his associates (Woese *et al.* 1990, cited from [18]) and of the absence of intermediate forms between animals having two- and three-layer blastophylla were of key importance permitting a radical revision of the phylogenetic tree of mammals [17].

Conserved regions permit studying the origin of Metazoa and the evolution within Metazoa and Protozoa, the evolution of Bilateria and Radiata (Field *et al.* 1988, Hendriks *et al.* 1990; cited from [17]), the origin of Coelomata and Acoelomata, and other problems related to the search for the common ancestors of all classes and other major taxa in animals (see [11, 24, 25]; Cristen *et al.* 1991, cited from [17]).

Using small RNA genes, the monophyletic origin of Metazoa and early splitting of Animalia, which is also monophyletic, have been demonstrated [26]. According to this evidence, animals and fungi have the common evolutionary history and their last shared progenitor was flagellates like the modern Choanoflagellata. The similarity between animals and fungi is also indicated by the similarity of elongation factors (cited in [26]) and complex biosynthetic pathways including synthesis of oxyproline, chitin, cellulose, and ferritin. Thus, among variants of relationships of fungi, which were regarded as plants, protozoans, and even as an independent kingdom, their close relationship to animals is most likely.

The relatedness of major taxa was assessed in the group Protostomia whose monophyletic origin, as that of Deuterostomia, was shown independently [17, 27]. It turned out that the use of the trait of body segmentation to unite Annelida (Coelomata) and Arthropoda, which were regarded as descendants of Annelida, was erroneous. The classification of Protostomia based on ribosomal markers corresponds to their molting capability and unites the monophyletic group of animals lacking the molting stage (Annelida, Mollusca, Plathyhelmintes, and some others) and having this stage (Nematoda, Arthropoda, Tardigrada, Priapulida) [27]. The authors proposed new names for the first (Ecdisozoa, i.e., molting) and the second group (Lophotrochozoa, derived from "lophophorates," the former name of tentacled invertebrates).

In constructing phylogenies of individual groups of higher Metazoa on the basis of ribosomal genes, a search for synapomorphic characters of the molecule itself rRNA (e.g., insertions) may be promising [24]. RNA of the aphid *Acyrthosiphon pisum* (Homoptera, Aphidiidae) is unusually long surpassing the range of variation between, say, human and pea or Paramecium (1.7 and 1.8 kb). Screening for different sequence regions characteristic for the extra part of this RNA in other insect taxa—in members of Hemiptera and four orders of Homoptera—showed that all of them had a longer 18S rRNA molecule, each of which contained apomorphic inserts. By including in the analysis sequences of other insects, a nonconflicting tree was constructed which was better than the tree based on similarities/dissimilarities at individual nuceotide positions. Genes for 18S rRNA of modern insects retain as phylogenetic markers (various inserts) the traces of events having occurred in the Carbon when Hemiptera and Oligoptera diverged (by one insert) and in the early Permian where the clade Sternorrhincha was formed (by three inserts) [24].

The phylogeny within Arthropoda, which so far had different classification schemes, has been extensively studied. Their monophyletic or polyphyletic origin according to morphological and development/life cycle traits has long been under active debate, the opponents and proponents of each of these point of view using different variants of these traits (Turbeville *et al.* 1991, cited from [27]). Cladistic analysis favors the monophyletic origin of Arthropoda based on synapomorphies. The arbitration of these issues that involved examination of partial sequences of 18S rRNA established the monophilmy of this group based on a representative set of species. Within Arthropoda, the monophilmy of the subtype Chelicerata including Arachnea and Acarina was demonstrated. This coinciding with morphological phylogeny result is important as such since it has a double support. A modern phylogeny and taxonomy of microbes based on ribosomal RNA genes has been developed (see [7]), which led to important conclusions: for instance, sulfate-reducing bacteria were not included in one group and other examples. (On the phylogeny of primates, see the review by Tetushkin [28].)

The literature on "ribosomal" phylogeny in plants is enormous and involves both nuclear and chloroplast genes. In my review, I will not consider this special field referring the reader to the review by Antonov and Troitskii, 1995 (cited from [6]). In the context of critical evaluation of molecular marker methodology, note that the situation in plants is no less, if not more, complicated than in animals. Moreover, the danger of convergent similarity of taxa due to reticular evolution is likely to be substantially greater in the plant kingdom.

The list of example could be extended. Here, I should only like to note that at present there is significant criticism of interpretation of data on ribosomal markers, as well as on any other markers describing the early stages of evolution. This is caused by the following interrelated and associated circumstances that I will attempt to classify.

(1) The rejection of the idea of "molecular clock" after the appearance of numerous data that did not confirm it (Britten 1986, cited from [6]). The mutation rates in different regions of nuclear, mitochondrial, and chloroplast DNAs significantly differ in different taxa; sometimes DNA mutation rates are different even in lineages of the same taxon. For instance, a comparison

of ribosomal gene evolution and evolution of phoraminifer fossils yields the difference up to two orders of magnitude [29]. The rates of change of the ribosomal gene operon in some cases also differ. All these facts reduce the likelihood of phylogenetic constructions and trees developed by distance analysis. Efforts have been made (and amply documented) to construct models accounting for errors in comparisons of distant branches in the neighbor-joining (NJ) method [30] as well as replacements of different sites in the maximum likelihood (ML) procedure [31].

Constructing models of alternative evolutionary pathways as such is subject to criticism: in one of the articles by Z. Yang [32] titled "How often incorrect models produce correct phylogenies?" the author states that the complexity of the task of phylogenetic tree reconstruction is apparently not fully understood and requires extensive theoretical elaboration and substantiation.

(2) The second most important circumstance follows from the first: the differences in mutation accumulation rates in the compared taxa may influence phylogenetic tree topologies and evaluation of the relative time of branching off. Overlooking this can result in the so-called "Long Branch Attraction" (LBA), against which has cautioned Felsenstein in one of his early papers (Felsenstein 1978, cited from [18]) and which is associated, in particular, with the choice of an outgroup upon rooting the tree. If any of the outgroups is long-branched while one of the internal groups evolves at a higher rate than the others, the latter will appear at a deeper level, which looks as if it were attracted by the very long branch of the outgroup [18, 33]. The models of this phenomenon correlate with the observed data. These problems led some authors to proclaim a crisis of molecular phylogeny whose revision upon constructing, for instance, the universal "tree of life" leads to an assumption that this tree is based on eukaryotes rather than prokaryotes, which are derived from eukaryotes through loss of genes and nonorthologous substitutions [16].

In other words, the order of branching off of the "deep" branches mainly depends on the rate of evolution. The most rapidly evolving taxon branches off earlier, which does not always reflect its phylogeny as exemplified by thermophilous Eubacteria (Gupta 1998, cited from [16]). Given these circumstances, some authors cast doubt on, e.g., the phylogeny of arthropods based on nuclear ribosomal genes [13, 31]. The accumulation effect probably affects the structure of all phylogenies based on eukaryotic rRNA genes.

(3) During evolution, recurrent changes of a site can occur, which are very difficult to test. If this is so, the actual mutation rates may be far higher than their experimental estimates. Approaches to evaluating differences in evolutionary rates of the "saturated" genes have been developed using various improved methods of the replacement number estimation [18], paleontological data or comparisons on a smaller evolutionary scale. In most cases the evidence of heterogeneity of evolutionary rates is based on paleontological data (Sorhannus 1996, Philippe 1997, Ayala 1997; cited from [16], see also [29]). An introduction of the assumption on the lack of constant sites and high and different rates at the remaining sites considerably increases the distance values (Sullivan and Swofford 1997, cited from [18]). The issue of different mutation rates at different gene sites (e.g., three codon positions and others) is very important and actively debated. Not accounting for this may result in the incorrect estimation of transversion and transition rates in protein genes, mutation rates of genes and other DNA sites, and ultimately to erroneous phylogenetic reconstructions (Yang 1996, Sorhannus 1996; cited from [16]).

(4) The possibility of horizontal gene transfer became evident after the discovery of virus transduction, plasmid transfer inducing drug resistance in bacteria, and, later, transfer of mobile elements and multicellular organisms and the possibility of gene vaccination. This lead virtually to the change from the paradigm of conservative hereditary material to the paradigm of genome instability (see [34]) although some authors maintain that the evolutionary significance of the latter phenomenon is considerably overestimated [35]. Indeed, it seemed at first that the phenomenon of horizontal transfer, which is in principle important, yet does not play a substantial role in speciation because of its rare occurrence. However, at present we should not overlook the fact that with accumulation of the complete DNA structures in a number of both pro- and eukaryotes taxa, the events of horizontal exchange of genome parts or of their horizontal transfer recurred in evolution between nuclear DNA and DNA of organelles, between parasites and hosts, among bacteria and between bacteria and eukaryotes (see [22]). As mentioned above, the cases of the appearance of long DNA stretches (LINE-like repeats) of reptilians in mammalian genomes [21] are known.

The acknowledgment of this fact as well as the phenomenon of reticulous evolution (speciation via interspecific crosses which has similar consequences) impose restrictions on the exclusively cladistic view of evolution (i.e., an idea of evolutionary process only as a dichotomy) and can lead to erroneous establishment of phylogenetic relationships based on sequences bearing foreign genes. The hazard of considering nonorthologous sequences in comparisons of small parts of the ribosomal operon also exists. Modeling of such events is currently in progress [36].

(5) In some cases, ribosomal operons exhibit variability (mainly in NTS quality and length) within an individual and among individuals within a population (e.g., in the newt [37]). It is known that mutation rates and 18S rRNA gene sequences are different at different developmental stages in the same individual, as in *Plasmodium vivax* [38]. There are a few operons (orphons)

not connected with a tandem family and usually having structural defects. All this warrants caution in establishing phylogenies on the basis of a few clones. Of course, the situation is different in different taxa. For instance, *Drosophila* and *Xenopus* each have one ribosomal locus per chromosome, while more highly organized organisms have several or many of them on different chromosomes or in different regions of the same chromosome. Operons of different chromosomes may differ in function. For example, in *Drosophila*, insertion of the X-chromosome rDNA region in the X chromosome that exhibits incomplete pairing with the Y chromosome restores pairing at the site of the insertion. Insertion of the autosomal rDNA does not affect X–Y chromosome pairing. Thus, the ribosomal operon has a cellular function except the direct function of rRNA synthesis (see also a review in [5]).

In most studies using rDNAs as evolutionary markers, the authors compare only relatively small operon sequences. However, to exclude faulty conclusions, the number of characters in most commonly used programs should be at least 1000 (Nei *et al.* 1988b, cited from [39]). These conclusions are confirmed in the study that presents even higher estimates of the required character number (of the order of 3000) [39]. This may be the reason why phylogenies of the same taxon based on ribosomal operons and other markers are often incongruent.

In spite of the tremendous advances in computer simulation and phylogenetic tree construction, these methods are being constantly improved taking into account various assumptions and conditions underlying the actual evolutionary process [30]. Analysis of the situation in this field is beyond the scopes of the present study. I should only like to note ongoing debates and constantly arising criticisms of the most commonly used algorithms of phylogenetic construction such as maximum likelihood (ML) and maximum parsimony (MP) procedures (Stewart 1993, Hillis *et al.* 1994; cited from [31]).

MITOCHONDRIAL GENE MARKERS

In the early 1980s, mitochondria gave a new impetus to evolutionary genetics, which was based on their maternal clonal inheritance, rapid evolutionary rate, properties of the global molecular clock, and high transition/transversion imbalance. These properties underlie the standing dogma on high value of mitochondrial markers. However, today this dogma has been questioned, which is discussed below (see [40]).

Let us consider the most impressive advances in the use of mitochondrial markers. Recall that the circular eukaryotic mtDNA (except in plants) is relatively small; it size ranges from 15 700 to 19 500 kb, which is 2500 times smaller than the size of the smallest genomic DNA in animals. It contains two genes for rRNAs (12 and 16S), 22 genes for mitochondrial tRNAs, 13 genes for protein subunits of enzymes of electron transport and ATP synthesis, and a regulatory part containing start sites of DNA replication and transcriptions. As with ribosomal genes, different stages of evolution can be marked by sequences of various DNA regions as well as by gene arrangement and the presence of indels.

The advances in using mitochondrial markers apparently depend on the degree of concerted evolution of cellular and mitochondrial DNA in the given taxon (see [41]). In many cases, this cannot be established *a priori*. However, in some instances this seems to be the case as, for example, in whales, in which the phylogenies of the major taxa (Mysticeti and Odontoceti) in relation to the closest to them mammals from the order Artiodactyla [42] inferred from the complete mitochondrial genome sequences corresponds to the other data. The closest relatedness to whales of the member of Suiformes, hippopotamus, not only confirms the previous results on mitochondrial markers but permits to establish a reference paleontological "point" for estimating mutation rate, since the time of divergence of Mysticeti and Odontoceti (32–34 Myr ago) is known. According to this reference point, the hippopotamus and whales diverged about 55 Myr ago.

The results concerning some species of fish (Slobodyanuyk *et al.* 1995, cited from [43]), birds, crocodiles, and mammals seem quite plausible. According to them, birds and crocodiles are more closely related [44] (which corresponds to osseous fossil data) and in the family of snakes, position 13 (of 16 shown by morphologists) corresponds to the most congruent views of systematics [45] and the hypothesis stating that the earliest snake progenitors are Scolecophidia and all snakes may have passed a subterranean stage in early evolution.

Informative results were obtained from comparisons of long mtDNA sequences (over 7500 bp) in order to establish relationships among three classes—Polychaeta, Hirudinea, and Oligochaeta—within the type Annelida as well as between the types Annelida and Pogonophora. Hirudinea, Oligochaeta and Pogonophora cannot be distinguished by gene arrangements whereas Polychaeta differ from them by the position of several tRNA genes, the presence of two additional tRNA genes and of a large noncoding sequence in this region. A comparison of gene arrangements in other taxa confirms the existence of the common clade Mollusca-Annelida which does not include Arthropoda. This suggests that Pogonophora do not form a separate type as was proposed upon their discovery but can be included in Annelida. These results are in agreement with some other evidence [46].

Macey *et al.* (1997, 1997a; cited from [47]) compared about 1500-bp mtDNA region including sequences for eight tRNAs and fragments of three enzymes. These authors found a shift in the order and the secondary structure of mitochondrial tRNA in agama *Uromastics acanthinarus*, which they interpret as caused by DNA replication errors. These phenom-

ena, as well as the replication start site of the light chain (Q_L) are taxon-specific in iguana lizards, which allows studying their phylogeny. Agama has a unique gene arrangement, in which the $tRNA^{Ile}$ and $tRNA^{Gln}$ genes are positioned in the order typical for vertebrates, but Q_L has a different position, and $tRNA^{Cys}$ is devoid of dihydrouridine stem. Iguana lizards (*Scelopus occidentalis*) lack these properties. Using these markers, the authors showed distinct monophyly of two major groups of iguana lizards (Acrodonta and Iguanidae), which had been a stumbling block that could not be removed by examining many dozens of morphological traits; these data agree with one of the two proposed and very dissimilar systems of these reptiles. The within-clade relationships of genera are essentially different from those accepted in systematics and are by no means ultimately resolved.

However, investigation of parasitic flat worms of the genus Schistosoma (Plathyhelminthes) showed that two closely related species of this genus differ in positions of mitochondrial genes in the genome. This implies the necessity of preliminary examination of within-taxon variability for this character which was seemingly "immune" for homoplasias [48]. However, the issue of low probability of homoplasias in the character of gene arrangement is not obvious and have been subjected to doubt (Dowtin and Austin 1999, cited from [49]), although phylogenetic congruence of data for this character and other molecular markers confirms the orthologous structure of the genome (Carole and Kocher 1999, cited from [49]).

In view of this, consider the value of data on mtDNA gene arrangement. The mitochondrial gene order was regarded as a powerful phylogenetic character (Boore 1999, cited from 46). Earlier, it was found that it can vary within a taxon, and the genes can have different orientation (Brown 1985, Wolstenholme *et al.* 1985; cited from [49]). This character was mainly used as specific for vertebrates (fish, amphibians, and mammals). However, duplications of some genes, their rearrangement, and loss of evident recognition sites for light DNA chain replication start between two tRNA genes were later reported (see [47] and references therein). In other phyla (insects, worms, urchins, and others) the order of genes is changed due to transpositions and inversions (Dowling *et al.* 1996, see [40]) or, as we can suggest now, due to recombination. In some cases, mtDNA evolves 5 to 10 times faster than the one-copy nuclear DNA, at the rate varying according to the taxon; different regions of mtDNA evolve at a different rate (Brown 1985, cited from [49]). However, for example, in urchins and some fishes these rates are approximately equal. These estimates are relative since they are based on times of divergences that are measured with different accuracy depending on the taxon. Some examples cast doubt on the principal possibility of using mtDNA mutation dynamics as a character for species phylogeny. For instance, in the case of particularly rapidly evolving Drosophila mtDNA, even two strains (!) of a Hawaiian Drosophila species had threefold different mutation rates (see [50] for references).

These circumstances may explain the cases when the use of mitochondrial markers did not produce positive results. The conclusions of these studies are often controversial demonstrating the necessity of a complex approach combining different molecular markers, especially using nuclear markers for the same taxon. An example is provided by investigation of phylogeny of the order of turtles (Testudines) which is regarded as a sister taxon of archeosaurs (birds, crocodiles, and dinosaurs), lepidosaurs (lizards, snakes, tuataras), or as a sister clade of the common clade for the former two ones. Data obtained with different molecular markers testify for different viewpoints (see [51]). Unfortunately, the mitochondrial marker mutation rates in snakes are extremely high and incomparable to other lepidosaurs, which raises the issue of attributing snakes to the same clade with lizards. A comparison of the complete mtDNA sequences of a skink species and the green turtle confirms the least popular idea that turtles are a sister group in relation to archeosaurs which share a common ancestor with lepidosaurs [51]. It is clear, though, that this conclusion will be repeatedly challenged and verified.

A comparison of nucleotide and amino-acid sequences for all 13 mitochondrial proteins in 19 taxa (from lancelet to human) yielded data that are at variance with the accepted phylogeny of chordates and, within this group, vertebrates (Naylor 1998, cited from [51]). The data on the assignment of Logomorpha to Rodentia are controversial (see [52]). The topology of three trees on the mtDNA COII gene constructed for mammals is ambiguous: among eight orders examined, only human and ape were similarly close on all of the trees. The positions of the other taxa were unstable; the orders grouped differently on different trees, which may also depend on the relatedness of the taxon taken as outgroup.

The molecular systematics of higher primates cast doubt on the suitability of mitochondrial markers for constructing phylogenies. The application of the maximum likelihood analysis to mtDNA data for four primate species (taking into account possible losses and acquirements of restriction sites and different probabilities of transversions and transitions as well as using model situations) resulted in combining gorilla and chimpanzee into a clade separate from human. This is in contradiction with a set of other, both molecular and paleontological, data that place human and chimpanzee closer to one another than each of them to gorilla (see [28]). However, according to other authors (Ruvolo *et al.* 1991, cited from [28]), a comparison of a fragment of mitochondrial cytochrome oxidase only 700 bp in size testifies to closer relatedness between human and chimpanzee.

In some cases, attempts to solve complicated problems of phylogeny and taxonomy at the family level did

not assist to their resolution. For instance, the attempts to infer relationships among three species of Baikalian fish from the cottoid complex with an unclear status, comprising more than 20 formal families, from the data on ATPase and cytochrome c genes did not correspond, first, to each other and, second, to the accepted generic classification within families (Grachev *et al.* 1992, cited from [53]). In this study, small (~607 bp) mtDNA regions were compared. Increasing the species number permits to make comparisons with morphological data, which show significant similarity between some taxa [53]. However, the number of species examined clearly was not sufficient for establishing phylogeny of this group.

The mtDNA-based phylogenetic constructions for lizard genera from the family Lacertidae substantially diverge. Using partial sequences (12S and 16S rDNAs, 954 bp) and different algorithms of tree construction yields incongruent trees [54]. The results of estimating within-genus relationships of lizards from the genus *Lacerta* based on mtDNA markers (cyt b, ~650 bp; ATPase, ~338 bp) and on allozyme markers reported by the same authors (Fu and Murphy 1997, cited from [54]) do not coincide. Averaging over the two mtDNA markers, and then colliding with the allozyme data, surprisingly yields reasonable conclusions that allow to identify three major clades with the more or less consistent species composition.

The repeated attempts to revise and test the existing systems in identification of Artiodactyla comprising the suborders of Ruminantia and Nonruminantia, or Suiformes, so far produced only partially successful results because of the extreme complexity of the task. Only about 20 out of 5 modern families and 225 genera are examined using dispersed and satellite repeats as markers [21, 55]. It is problematic to draw definite conclusions on the basis of these data since different authors examine different sets of species.

The attempts to infer relationships among Tetrapoda, Actinopterigii, Crossopterigii (Coelocantes, *Latimeria*), and Dipnoi (Sarcopterigii) from mitochondrial markers (cyt b and 12S rDNA) has been so far unsuccessful. Three scenarios are possible, each of which is partially confirmed by morphological and osteological data. Mitochondrial markers support the tree in which Crossopterigii and Sarcopterigii split from the Tetrapoda branch after branching off of Actinopterigii. However, this is in disagreement with the globin data which trade places of Crossopterigii and Sarcopterigii; in other studies, also based on mitochondrial markers, the Crossopterigii–Tetrapoda clade is questioned (see [22]).

What are the reasons for such great discrepancies in the results based on mitochondrial markers? Apparently, assumptions underlying their use are only partially (or only in some cases) valid, and some circumstances are completely overlooked in the incredibly numerous publications on this topic. Below, I attempt to summarize the disadvantages of using mitochondrial markers.

(1) The assumption on maternal inheritance and the absence of recombination in mitochondrial propagation is not totally correct: about 50 to 70 paternal mitochondria are introduced into oocyte upon fertilization (see [40]) and recombination events were found in reproduction of mitochondria in plants, fungi, protozoans, nematodes, and human [56]. If this is so, then the possibility of within-species variation cast doubt practically on all results not accounting for this fact. It may be that discrete positions of the group compared on the basis of these markers and often the congruence of the trees based on morphological and molecular characters indicates that the situation is not so catastrophic. Apparently, it becomes complicated only in the case of past hybridization or transfer events, particularly when these occurred in remote past, when mtDNA recombination could take place. This issue is analyzed in the monograph by R.B. Khesin [34]. Note also that in some cases (e.g., in nine *Drosophila* species) recombination events in mtDNA were not found [57].

(2) The symbiotic origin of mitochondria from rickettsia-like bacteria poses a question whether the nuclear and mitochondrial genomes evolve in concert and whether the evolution of the latter reflects the evolution of the organism or is associated with it. The same question is valid for studies using chloroplast genome markers. It is unclear whether the symbiont formation occurred only once, at the most ancient cell stages, or repeatedly, after the beginning of the divergence of cellular organisms.

(3) Various and diverse (in different taxa) mutation rates of nuclear, chloroplast, and mitochondrial genomes create additional difficulties for interpretation of the phylogenetic data [58]. For instance, it is unclear how to interpret the results that mtDNA of lemurs diverge at a slower rate than that of other primates. The same is typical of the strains of Hawaiian *Drosophila* species. As noted in the previous section, this problem is also characteristic of ribosomal markers. It is of particular importance in constructing phylogenies on the basis of mitochondrial protein genes, because not accounting for different substitution rates in different taxon lineages, imbalance in nucleotide composition, transitions, and transversions, different mutation rates in different codons and at different positions of the same codon can lead to incorrect phylogenies (Honeycutt *et al.* 1995, cited from [58]).

The above circumstances can lead to the artifact of "long branch attraction" (LBA) that distorts the tree topology as noted in the previous section. The within-taxon mtDNA mutation rates probably depend on body size in various groups: the tree position of species with smaller body size (e.g., some birds [59]) tends to shift toward the base of the tree. In higher mammals, this trend is absent. By contrast, a comparison of mtDNA in three human populations and three species of humanoid

apes showed that judging from the paleontological fossil dating, nonsynonymous substitutions in protein-coding and RNA genes accumulate at approximately equal rates [60].

The second problem mentioned in the previous section is the substitution accumulation in sites at different accumulation levels in different taxa and different transversion and transition rates in different taxa. This problem is also relevant to mitochondrial markers (Brown 1985, cited from [49]). Some authors think that these differences are not fundamental and do not lessen the value of mitochondrial markers in phylogenetic studies [61].

(4) Heteroplasmy, i.e., (mtDNA polymorphism in different (including reproductive) tissues of the organism and even in the same cell, is another phenomenon rarely taken into consideration but hazardous for phylogenetical use of mitochondrial markers.

(5) Horizontal gene transfer among nuclear and mitochondrial (and chloroplast) genomes has been conclusively demonstrated (see [62]) although its mechanism and functional consequences are still unclear. Mundy *et al.* [62] found in human nuclear DNA a huge (5842-bp) untranslated mtDNA region bearing several protein and tRNA genes which seems not to be derived from modern mitochondria.

(6) The wide use of the order and polarity of gene arrangement in the total mtDNA molecule as a phylogenetic marker and the presence of indels (see the beginning of this section for examples) are also open to doubt (Boore *et al.* 1998, cited from [46]). The mitochondrial gene arrangement and orientation exhibit variation within a taxon, in the number of small repeats, and in the length of the control regions. Nevertheless, as substantiated by Macey *et al.* [47], the gene order and even its fine differences, as well as some other properties of mitochondrial genome organization, can serve as a criterion of the phylogenetic taxon position. Further research and accumulation of data may be very helpful for evaluating of this important evolutionary character.

(7) The overwhelming majority of studies using mitochondrial markers consider as a phylogenetic character small (of the order of several hundred bp) DNA regions. The necessity of taking at least 1000 characters noted in the previous section is confirmed by many facts when partial fragments yielded conclusions contradicting other evidence. For instance, the phylogeny of a crustacean based on one mitochondria protein gene was different from the phylogeny based on seven genes [63]. The trees of the lizard family Lacertidae constructed from the data on small mitochondrial genome regions were different from the trees based on allozyme and satellite DNA analysis [54, 64].

Here, I do not discuss the ample literature on using hypervariable mtDNA regions for within- and among-population studies. The objective of the presented material is to note the most likely stumbling blocks in the employment of mitochondrial markers exactly as phylogenetic markers for major taxa.

MOLECULAR RAPD MARKERS

The method of randomly amplified polymorphic DNA (RAPD; Williams *et al.* 1990) is also known as AP–PCR (Arbitrary Primers–PCR, Welsh *et al.* 1990) and DAF (DNA Amplified Fragments, Caetano-Anolles *et al.* 1991). In this method, electrophoresis of a set of products translated from different DNA regions randomly chosen by the primer produces a series of bands (pattern) of different intensity (see [65] for references to the papers mentioned in this paragraph). The change rates of RAPD patterns in various major taxa (primates, *Drosophila*, horned ungulates) are significantly different. However, within a taxon, the pattern divergence level is related to the time of the taxon divergence [66]. The pattern polymorphism depends on the specific DNA and primer, primer length, the amplification and annealing conditions, and the enzyme quality. Some of these conditions are objective while others depend on the strict reproduction and protocol of the experiment.

The RAPD method can be used for studying genetic intrapopulation polymorphism and segregation; population history and geography, hybridization at the range boundaries of allopatric populations and species, the mechanism of formation of hybrid parthenogenetic forms, parental analysis and paternity identification in polyandrous communities; searching for relatedness of wild and domesticated forms; identifying criteria for isolation of the notions of species–subspecies–population; and for related problems (Hadris *et al.* 1992, cited from [65]).

Here, I consider mainly the problems falling outside this scope into the realm of phylogeny, where the results depend on polymorphism of a particular taxon and the choice of a primer that would transcribe more conserved DNA sequences. Some of the positive examples are as follows. In our experiments (Mel'nikova *et al.* 1993, cited from [67]), domesticated and wild sheep forms exhibited strikingly similar RAPD patterns obtained with different primers, which testifies to their close relatedness within the species and corresponds to our other data on taxonoprint analysis [68]. For instance, the RAPD tree of species from the families Bovidae and Cervidae (in three subfamilies: Cervinae, Odocoileinae, and Muntjaconae) was in good agreement with these systematic gradations according to cluster analysis and genetic distances, when both families were systematically remote, and subfamilies clustered separately from one another (Comincini *et al.* 1996, cited from [65]), which, however, was at variance with the corresponding mtDNA data (see the relevant section). This method was employed for examining some ungulates (Glazko *et al.* 1997), serpents (Wang *et al.* 1996), fishes (Borowsky *et al.* 1995, Sultman *et al.* 1995) (see [65, 66] for references in this paragraph).

The RAPD method has been recently subjected to substantial criticism. The reports casting doubt on its correctness because of the appearance of artifact bands [69] warrant caution in treating the previously published results. Note the following RAPD aspects that require careful consideration upon analysis of RAPD data. First, same-size fragments can be transcribed from different loci or, conversely, bands of different sizes could contain homologous loci. Second, orthologous fragments which can be detected by hybridization nevertheless are not checked in most works. Poor band resolution in short agarose gels can also lead to the location of nonorthologous fragments in one band. Third, we cannot exclude the possibility of the appearance of artifact bands due to the formation of secondary structures in the primer or the products. These concerns were confirmed in the study showing that a band contained fragments of equal length having different sequences (see [65]).

The situation is somewhat improved by the following facts. First, the critically-minded authors themselves admit that the appearance of artifacts depends on the properties of the DNA and primer and thus can be controlled by using appropriate procedures. Second, the appearance of artifact fragments due to the formation of secondary, only partially transcribed sequences of the copied DNA, random interchain interactions of single DNA chains during denaturation/renaturation, and the presence of nested sites for the primer within one region can be stochastic and, hence, make the experiment not reproducible. By obtaining reproducible results via selecting an optimal primer, the researcher overcomes the false band count. If, on the other hand, these processes are not stochastic, then they reflect specific properties of the given DNA and can consequently be regarded as meaningful characters. In some cases, the reading is not stochastic: for example, samples of clonal individuals of parthenogenetic insects [70] and lizards [67] are totally identical with regard to these RAPD markers.

I think that RAPD markers remain a useful tool of molecular genetics, but the early euphoric enthusiasm considering the possibilities of this approach is being replaced by the realization that RAPD data should be treated and interpreted with more caution. The RAPD method still has possibilities. It may prove useful for searching for chromosome-specific amplification products. In birds, the sequenced product of a W-chromosome-specific band turned out to be a conservative regulatory protein of DNA structuring (CHD2; Bello and Sanchez 1999). Specific Y-chromosomal markers were found in the sheep (Gutierrez-Adan *et al.* 1997), pig (Castellanos *et al.* 1996), cattle (Antoniou and Skidmore 1995); two species of cattle differed with regard to the presence of this fragment (Teale *et al.* 1995; see [71] for references). The scanning of the products obtained with 700 primers in males and females of the wild (*Bombyx mandarina*) and domestic *(B. mori*) silkworm species (whose sex chromosomes are ZZ in males and ZW in females) probably yielded W-specific fractions. Such data can be helpful for studying the mechanisms of sex determination and parthenogenesis.

The presence of species-specific RAPD characters makes it possible to use these markers in studying hybridization in the context of speciation. The data on individual variation of parental populations are required for experiments on interspecific hybridization upon investigating, e.g., reticulate evolution. The probability of speciation through cross evolution is particularly high for plants.

In our works, we have proven the hybrid origin of lizard parthenogenetic species and identified bisexual lizard species that had been parental in the past (Grechko *et al.* 1998, Kan *et al.* 1998, Ryabinina *et al.* 1999; see [72]). This line of studies is facilitated by monomorphism and stability of RAPD pattern in parthenogenetic reptilian populations, on the one hand, and by low parental species heterogeneity, on the other.

The studies on RAPD product identification are currently in progress, which is promoted by combining this method with Southern transfer followed by hybridization with the selected primer. The RAPD markers stop being anonymous. Apparently, all DNA types (unique genes, middle and high repeats) are present among the amplification products; for instance, in soybean DNA six out of eleven fragments examined were transcribed from unique genes, while three and two, from the above repeats, respectively (Williams *et al.* 1990, cited from [65]). In another plant study, the amplified fragments were transcribed mainly from repetitive sequences (Kazan *et al.* 1993, cited from [65]). The sequenced RAPD DNA fragments of parasitic nematodes from the genus *Trichinella* mainly contain unique and low-repetitive sequences including genes for chitinase, maturase, and numerous sequences homologous to genes of nematode Caenorhabditis elegans [73]. Apparently, one can expect to find any locus among RAPD fragments depending largely on the primer specificity.

Thus, if adequately used and interpreted, the RAPD method is an informative modern tool for studying genetic diversity of species in nature.

INTERSPERSED REPEATS AS EVOLUTIONARY MARKERS

The history of investigating the structure, function, and distribution of dispersed repeats is in essence the history of investigating the genome evolution (Cook and Tristem 1997, Buntjer 1997; cited from [2] and [74]). Since the very discovery of the intermediate repeat fraction it became evident that they are taxon-specific [74]. After the discovery of the short and long interspersed repeat families (SINEs and LINEs) it was demonstrated that these repeats are indeed characteristic (sometimes markedly) of the species and higher-rank taxa [75]. However, the significance of the func-

tion of these repeats was long doubted due to hypnotic but probably unproductive ideas on the existence of "junk" or "selfish" DNA in the genome [76, 77].

At present the importance of the repeat functions has been becoming clearer. These functions include regulation of gene transcription via repeat retroposition (Dobrowski *et al.* 1991, Britten 1994; cited from [2]); protein synthesis regulation upon cellular stress by two-strand RNA products that interact, e.g., with the kinase initiation factor of retroposons eIF2 (RKR) (Chu *et al.* 1989, cited from [2]); the pathogenic effect of newly formed *Alu* repeats associated with a hereditary disease (Labuda *et al.* 1995, Schmid 1996; cited from [2]); an obvious association with the development regulation due to different methylation of, e.g., *Alu* repeats of the male and female germline cells (see [2]).

Moreover, retropositions of dispersed repeats clearly are not random and have preferential insertion sites (Tatout *et al.* 1998, cited from [74]). Thus, it has been becoming increasingly likely that evolutionary "survival" of repetitive sequences, including dispersed repeats, is explained by selection favoring them due to their participation in cell functioning. A comparison of SINEs of remote taxa revealed a very conserved consensus sequence CORE-SINE which is present in taxa of various organisms from plants to mammals [74]. Hence, the sequences of any type are likely to play a significant role in speciation or, at least, to be directly associated with them.

D.A. Kramerov [78] suggested to use SINEs directly as phylogenetic markers. This suggestion was in agreement with earlier considerations, e.g., in a review by Singer [75]. Singer noted that dispersed sequences could make a contribution to the traits traditionally distinguishing species and genera. This opinion supported by other authors (Ryan and Dugaiczyk 1989, Buntjer 1997; cited from [79]) is currently universally accepted [79].

Thus, the presence or absence of a particular SINE group is in itself a marker of the clade, not mentioning sequence specificity and the properties of flanking sequences which were shown to be nonrandom upon retroposition of a new monomer (Tatout *et al.* 1998, cited from [2]; see also [74]). Since SINEs do not transpose, their independent (convergent) origin is extremely unlikely. A taxon lacking this repeat probably branches out before a taxon having it though the losses are not excluded. The variants of orthologous sequences and their different homology also permit determination of relative times of their origin and classifying on the basis of similarity of repeat groups.

Using SINEs as molecular markers is productive in all variants: in simple comparisons of homologies of the repeats themselves, in more complex examination of their localization and flanking sequences, and in in-depth analysis of consensus sequences. This approach has a great phylogenetic value as it ensures the orthologous character of the compared DNA sequences and the uniqueness of the sequences at the specific and generic levels. Its convenience is due to the fact that most eukaryotes, as noted above, probably contain SINE-like repeats (Okada 1991, cited from [2]).

Note that, strictly speaking, the problem of convergence in comparisons of dispersed sequences is still standing, which is also true for other marker types. For instance, since the common ancient precursors of each from the superfamilies of the B repeats have not been found yet and may not be recognizable because of numerous accumulated mutations, convergent similarity of, say, primates and rodents by these repeats cannot be excluded [2]. The high similarity of LINE Bov-A family in ruminants and the LINE family in some reptiles (snakes and lizards) [21] casts doubt on the absolute phylogenetic comparisons based exclusively on the dispersed repeats.

The discovery of the dispersed repeat MIR (Mammalian Interspersed Repeat) (see [80]), like formerly the discovery of families B1 and B2, made a significant contribution to the phylogeny of rodents and related taxa [81, 82]. Investigation of *Alu* repeats has provided valuable taxonomic data showing that primates are more closely related to tarsiers than to other monkeys [83]. The nonrandom localization of *Alu* repeat in the human genome is evidenced by computer-aided analysis of DNA sequences of human chromosomes 21 and 22. This analysis revealed positive correlation between the *Alu* repeat distribution and protein exons on the chromosome and the same number of repeat units in the direct and reverse orientations [84]. This testifies to a significant role of interspersed repeats in the evolution of mammals.

To sum up this short section which but outlines the extensive area of using dispersed repeats and is far from comprehensive, I would like to once again emphasize the main point of it. The diversity and different levels of homology of these repeats permits studying the relatedness of major taxa at the level of classes, orders, and families and in some cases, even at the level of species and populations.

TANDEM REPEATS AS MOLECULAR MARKERS OF EVOLUTION

Tandem (satellite) DNA repeats which vary in structure and sequences form an enormous group. These repeats are organized in continuous clusters with the "head-to-tail" orientation of monomer units. The length of monomers in a cluster ranges from several to several thousand base pairs; their number in various families vary from several dozens to several million. Strict classification of tandem repeats on the basis of these features is problematic and arbitrary. The team of Jeffreys has proposed to classify these repeats as microsatellites (2–6 bp per unit), minisatellites (up to 100 bp), midisatellites (about 100–400 bp), and macrosatellites (up to several thousand) [85]. However, other athors do not

strictly adhere to this scheme. The term "satellites" is often used to denote any specific and clearly distinct family of tandem repeats except microsatellites.

Any organism probably contains sets of particular tandem repeat families many of which are polymorphic for the series length and intrinsic monomer arrangement slightly differing in sequences. Discussing the mechanism of mutation and polymorphism formation is beyond the scope of the present study and can be found in [85] and another review [3]. Here I would only like to note that individual polymorphism revealed by fingerpringting, due to the specificity of this method, mainly reflects length polymorphism of some tandem series and DNA fragments containing these series. The level of sequence polymorphism can be higher or lower but generally conforms to the notion of concerted evolution of tandem repeats which implies relatively rapid spreading of adaptive mutation throughout the series. A monomer sequence contains strictly conserved regions (i.e., mutation involves only particular monomer regions). Because of this, some tandem repeat families are preserved during million years of evolution and can serve as markers at different time periods. The objective of the last section of the present review is to show these facts and examine the relationship between satellite specificity and speciation/taxon formation.

Microsatellites deserve special consideration beyond this review. High variability makes them a tool of population genetics, but some facts testify that some of them (e.g., telometic [86] or core in pericentric chromatin satellites [87]) are conserved. Conservation of some microsatellite localization at orthologous positions was recorded in various taxa. Based on this, this character can be used as phylogenetically significant. The involvement of microsatellites in hereditary pathologies also may reflect their important role in speciation.

Satellites of various sizes will be further considered as phylogenetic markers, regardless of their size and structure (see also [3, 88–90 for review]. As early as in 1970s, data have started to accumulate showing that repeated DNA components can contain rather conserved (even in a series of relatively remote species) sequences (Dover 1980, cited from [88]). This contradicted the notion on random mutation accumulation in repeats and on the secondary role of this DNA in organism functioning (the "junk DNA" hypothesis). Pioneering works of Dover's team showed that the distribution of 15 different satellite DNA fractions in seven very similar *Drosophila* species is not random: some of them contain fractions that are species-specific or shared by two or three species. The relationship scheme constructed as enkapsis (a system of circles of different diameters enclosing each other, the smallest of which shows the closest relatedness) revealed two species clusters (*melanogaster–simulans–mauritiana* and *erecta–jakuba–tessieri*) although the author specified statistical nonsignificance of these results and the limitations of the method. The first cluster include species that are very close to one another by other criteria (morphology, hybridization, and karyology). The species grouping in the second cluster does not significantly contradict data of morphology and ecology; moreover, the satellite pattern is paralleled by that of polytene inversions.

The results obtained in the 1980s indicated that the prevailing for two decades opinion that "highly repetitive DNAs lack function and are selectively neutral, is wrong and the new data of 1990s evidently mark a revival for interest for microsatellites showing that it is at least hazardous to consider the hypothesis on satellite neutrality as plausible" [91]. Let us consider some examples.

The only satellite family of darkling beetles (*Coleoptera*, family Tenebrionidae) is genus- and species-specific (see [92] for references). In the genera of burying beetles and leaf beetles, species-specific satellites were found (King and Cummings 1997, cited from [92]). Satellites were examined in Hymenoptera (Palomeque *et al*. 1999, see [92] for reference) and Decapoda (see [88]). A satellite family of an ancient origin was described in a nematode species [93].

Among vertebrates, fishes were studied. For this class, satellite markers may serve as a pilot for navigating in extreme variability and morphological diversity of its members (see, e.g., [94, 95]). In Atlantic salmon (*Salmo salar*), several tandem loci were found showing different degrees of relatedness to DNA of the family, subfamily, and genera of this group [96].

In caudate amphibians, the families Salamandridae and Plethodontidae differ with regard to satellite markers [91]; all examined *newt* species contain satellites of different specificity in different combinations [97]. It has been suggested that some satellites are present as precursors for all modern newts and persisted for over 20 Myr since the paleonthological divergence of the genus (Cremisi *et al.* 1988, cited from [97]).

Although reptiles are very poorly studied, the available data are in good agreement. Complete species-specificity was demonstrated for lizards from the family Lacertidae (*Podarcis* [98], *Lacerta, Eremias,* and *Ophisops* [64]); species-specificity with distinct distribution of one minisatellite among species was found in more than 20 lacertid species, one of groups of which was isolated in a separate genus based on a unique minisatellite [72].

In mammals, rodents are studies in most detail. This family is the first among higher animals that showed species- and genus-specific properties of satellites (Southern 1975, Horz and Zachau 1977, Brown and Dover 1980a, Horz and Altenburger 1981; see [88] for references). Based on restriction analysis of DNA satellites in mouse species from the genus *Apodemus*, phylogeny of its five species was constructed and molecular organization of satellites was studied. A substantially altered satellite of the class described in mice was found in other genera (*Nannomys, Pyromys, Rattus,*

Coelomys, *Mastomys*; see [90, 99]). Specificity was found within these genera (Fry, Salser 1977; Witney and Furano 1984; Epstein *et al.* 1984; Rossi *et al.* 1990; see [88] for references). Important conclusions on the phylogeny of rodents were made from the level of satellite genus-specificity (Rossi *et al.* 1990, cited from [88]). Rossi *et al.* note that these satellites persist for far longer time that it is thought possible for ostensibly excessively heterogenizing satellites. These authors believe that these genes could be highly selectively advantageous in some situations which may explain rapid species formation in a short time period. The data on the family Cricetidae (hamsters) are in principle similar (Chelomina *et al.* 1990, Fatyol *et al.* 1994, Ivanova and Modi 1996, Khrapov *et al.* 1998, for references, see the abstract of the dissertation by S.G. Potapov [100] and also [101]).

The order Artiodactyla (even-toed ungulates), in particular the family Bovidae (cattle) and Cervidae (deer) ranks second in the level of being studied. I will not discuss this evidence here because of its incompleteness and contradictiory nature. A number of several complexly organized satellites of seven (out of eight present) Artiodactyla families are considered in the review [102] as well as in other studies (Queresci and Blake 1995; Lee and Lin 1996; Lee 1997; see [103] for references). The authors do not doubt the correlation between taxonomy and satellite organization and resolve many disputable issues in this context.

Arnason *et al.* (1993) used conserved satellites to study relationships in the group Cetacea (whales). In this group, each taxon is characterized by a specific satellite or satellite variant depending on the degree of genetic relatedness (Arnason *et al.* 1993, cited from [104]). Generally similar results were obtained in inferring phylogeny of carnivores (the order Carnivora) from satellite markers (Fanning *et al.* 1987, 1988; Modi *et al.* 1988; Potapov *et al.* 1991; Ivanov *et al.* 1991a; see [22] for references). Giant panda and spectacled bear originated from the modern bears, and lesser panda is derived from another lineage leading to the modern raccoons (see [22]).

In the studies of the incomplete taxonomy of kangaroos (Marsupialia), satellite repeats assisted in solving the problem of close relationship between giant and red kangaroo and other issues (Elizur *et al.* 1982, cited from [105]). The examination of primate satellites indicates a closer relationship between human and chimpanzee and their clustering with gorilla rather than orangutan (Jorensen *et al.* 1992, Laursen *et al.* 1992; see [5] for references). Primates have specific satellite fractions [28, 106]. In three lemur genera (family Lemuridae), lemur and semi-lemur were shown to be more closely related than lemur and eulemur (Montagnon *et al.* 1993, cited from [28]). Each of the general from the monkey tribe Pitheciini (hairy saki, black saki, and uakari) has the taxon-specific satellite derived from a very early ancestor [107].

These and numerous other results indicate that the condition and structure of tandem DNA repeats are associated with evolution. The evidence showing this association suggest its causal nature. If this is true, then tandem DNA repeats must have functional significance underlying their selective persistence. The vast part of chromosomes, their heterochromatin regions, are indeed composed of satellites but, in spite of this, satellites play no apparent role in mitosis and meiosis [88]. This facts bred skepticism with regard to the evolutionary significance of satellites. Probably, the functional role is mediated by other properties of these repeats. The data on the functional role of satellites are summarized in the review by E.N. Trifonov [108], who hypothesizes on the function of tandem repeats as regulators in tuning the concerted operation of genes ensuring the adaptive potential of an organism. The important point is that the notion of repeats as "selfish" or "junk" DNA has been rejected (see [109]).

Some examples. The properties of long polymers, particularly tandem repeats, enable them to form higher-order structures, e.g., those described by curvature of the molecule (Israelewski 1983, cited from [110]). This curvature is supposed to participate in the positioning of nucleosome core particles and in forming compact satellite heterochromatin (Israelewski 1983; Radic *et al.* 1987, Constanzo *et al.* 1990; see [111] for references) as well as in binding of particular proteins [112]. A specific role of, say, alphoid DNA in compaction of centromeric nucleosomes cannot be excluded [113]. This is supported by the evidence of [114] showing that upon introducing in human or hamster cells, YAC constructs containing alphoid DNA from centromeric regions form structures with cytologically observed centromeric properties. They are destroyed when the cell enters anaphase and interact with specific antiserum to proteins CENP-A, -B, and -C. The satellite found in the genome of lizard *Lacerta graeca*, is found to be similar to the CENP-B box [115]. Investigation of the chromosome evolution of South American rodent tuco-tuco (*Ctenomys*) revealed correlation between chromosome rearrangements and a deletion of a major satellite of this animal [116]. Human megasatellite RS447 encodes a deubiquitinating enzyme, and transcription of the same region of the opposite chain producing antisense RNA can modulate the enzyme expression level in the brain [117].

Other structural elements found in satellite sequences include reverse repeats capable of forming doubly symmetrical structures (Bigot *et al.* 1990, cited from [118]). This is supported by Plohl *et al.* [118] who examined satellites in beetles (Coleoptera, Tenebrionidae). These authors advances a hypothesis that evolutionary similarity may concern the higher, rather than only the primary, satellite structure which is thus maintained in evolution. For instance, satellite sequences in beetles *Tenebrio molitor* and *T. obscurus*, on the one hand, and *Palorus ratzenbergii*, on the other, are similar in the tertiary, but not the primary, structure (Plohl and

Ugarkovic 1994, cited from [118]). The complete analysis of the problem can be found in [119].

Using the properties of dispersed and tandem repeats as phylogenetic markers, the author of the present review and coworkers have proposed a simple method of general characterization of all DNA repeats—taxonoprint (Fedorov *et al.* 1992, Grechko *et al.* 1997; see [120] for references). We proceeded from the assumption that the amounts of mutation variability and distribution of mutations in any sequence is indirectly reflected in the number of restriction enzymes recognition sites. Splitting a repeat family in each monomer with a restriction endonuclease yields fragments of equal length from all monomers with the same sequence. Mutations at restriction sites of this of related families will change the set of resultant fragments. Obviously, the DNA restriction products will be rich in fragments from the repeat sequences. They could be detected after electrophoretic separation and staining (or radioactively labeling the ends of the produced fragments) against the homogeneous background of many nonrepetitive DNA sequences.

This approach enables to isolate monomers of different lengths for all series of repeats having a restriction site for the given restriction endonucleases, while the use of frequently cutting restrictases with different specificity covers virtually all the range of the available repeats. Upon electrophoretic separation the repeat fragments produce a taxon-specific pattern (taxonoprint); the individual pattern specificity is absent. Thus, all members of the same population have identical taxonoprints that exhibit species-specific characters. Some of the taxonoprint bands (characters) are autapomorphic for the species and synapomorphic for the species of the genus and the genera of the family. Hence, these characters can serve as phylogenetic markers for lower taxa of the taxonomic system, which was shown by us for reptiles (lizards of the family Lacertidae) and primates (human and apes) [120], and by other authors for even-toed ungulates and rodents (Potapov *et al.* 1993, Ryskov *et al.* 1994; see [68] for references), insectivores (Bannikova *et al.* 1995, cited from [68]), and fishes [94]. Thus, taxonoprinting may prove useful for isolating a group of populations with identical taxonoprints, i.e., taxonoprint identity may be a criterion of the species.

In closing the review, I would like to emphasize that, regardless of the technical difficulties, molecular DNA markers are an efficient tool for phylogenetic studies. Data on ribosomal and mitochondrial DNA markers probably should be treated with caution until independent results confirming the relevant conclusions are received. However, this reasoning also applies to the other methods. A combined study of a taxon by various markers seems a promising line of research which might enable to avoid errors. As one of the prominent researchers in the field said, "Because the organism under study has a single story, systemic study of any set of genetically determined characters should be congruent with other such studies based on different sets of the characters of the same organism. Congruence between studies is strong evidence that the underlying historical pattern has been discovered; conflict may indicate theoretical or procedural problems in one or both analyses. Or it may indicate that additional data are needed to resolve the phylogenetic relationship in question" (D.M. Hillis, "Molecular Versus Morphological Approach to Systematics" [12]).

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, grant nos. 00-04-48245 and 99-04-49334.

REFERENCES

1. Mayr, E., *Animal Species and Evolution*, Cambridge: Belknap, 1963.
2. Schmid, C.W., Does SINE Evolution Preclude *Alu* Function?, *Nucleic Acids Res.*, 1998, vol. 26, pp. 4541–4550.
3. Ryskov, A.P., Multilocus DNA Fingerprinting in Population Genetic Studies of Biodiversity, *Mol. Biol.* (Moscow), 1999, vol. 33, no. 6, pp. 997–1011.
4. Sverdlov, E.D., The Microcosm of the Genome, *Mol. Biol.* (Moscow), 1999, vol. 33, no. 6, pp. 917–940.
5. Kupriyanova, N.S., Conservation and Variation of Eukaryotic Ribosomal DNA, *Mol. Biol.* (Moscow), 2000, vol. 34, no. 5, pp. 753–767.
6. Antonov, A.S., *Osnovy genosistematiki rastenii* (Basics of Plant Genosystematics), Moscow: MAIK "Nauka/Interperiodica," 2000.
7. Debabov, V.G., Bacterial Life Outside Laboratories, *Mol. Biol.* (Moscow), 1999, vol. 33, no. 6, pp. 1074–1084.
8. Pesenko, Yu.A., Methodical Analysis of the Taxonomy: I. Formulation of the Problem, Major Taxonomic Schools, *Printsipy i metody zoologicheskoi sistematiki* (Principles and Methods of Zoological Systematics), Borkin, L.Ya., Ed., Akad. Nauk SSSR, 1989, vol. 206, pp. 8–119.
9. Emel'yanov, A.F. and Rasnitsin, A.P., Systematics, Phylogeny, Cladistics, *Priroda*, 1991, no. 7, pp. 26–37.
10. Vorontsov, N.N., *Razvitie evolyutsionnykh idei v biologii* (Development of Evolutionary Ideas in Biology), Moscow: Progress-Traditsiya, 1999.
11. *Molecular Systematics*, Hillis, D.M., Moritz, C., and Mable, B.K., Eds., Sunderland, Mass.: Sinauer, 1996.
12. Hillis, D.M., Molecular versus Morphological Approach to Systematics, *Annu. Rev. Ecol. Syst.*, 1987, vol. 18, pp. 23–42.
13. Huelsenbeck, J.P. and Nielsen, R., Variation in the Pattern of Nucleotide Substitution across Sites, *J. Mol. Evol.*, 1999, vol. 48, pp. 86–93.
14. Purvis, A. and Quicke, D.L.J., Building Phylogenies: Are the Big Easy?, *Trends Ecol. Evol.*, 1997, vol. 12, pp. 49–50.

15. DeJong, W.W., Molecular Remodel the Mammalian Tree, *Trends Ecol. Evol.*, 1998, vol. 13, pp. 203–207.
16. Forterre, P. and Phillipe, H., Where Is the Root for the Universal Tree of Life, *BioEssays*, 1999, vol. 21, pp. 871–879.
17. Adoutte, A., Balavoine, G., Lartillot, N., and de Rosa, R., The End of Intermedia Taxa?, *Anim. Evol.*, 1999, vol. 15, pp. 104–108.
18. Philippe, H. and Laurent, J., How Good Are Phylogenetic Trees? *Curr. Opin. Genet. Dev.*, 1998, vol. 8, pp. 616–623.
19. Ayala, F.J., Vagaries of the Molecular Clock, *Proc. Natl. Acad. Sci. USA*, 1997, vol. 94, pp. 7776–7783.
20. Wolfe, K.H. and Shields, D.C., Molecular Evidence for an Ancient Duplication of the Entire Yeast Genome, *Nature*, 1997, vol. 387, pp. 708–713.
21. Kordis, D. and Gubensek, F., Unusual Horizontal Transfer of a Long Interspersed Nuclear Element between Distant Vertebrate Classes, *Proc. Natl. Acad. Sci. USA*, 1998, vol. 95, pp. 10 704–10 709.
22. Avise, J.C., *Molecular Markers, Natural History, and Evolution*, New York: Chapman & Hall, 1994.
23. Gerbi, S.A., Evolution of Ribosomal DNA, *Molecular Evolution and Genetics*, MacIntyre, R.J., Ed., New York: Plenum, 1985, pp. 419–517.
24. Aleshin, V.V., Vladycheskaya, N.S., Kedrova, O.S., *et al.*, Comparison of the 18S rRNA Genes in the Invertebrate Phylogeny, *Mol. Biol.* (Moscow), 1995, vol. 29, no. 6, pp. 1408–1426.
25. Pennisi, E., Genome Data Shake Tree of Life, *Science*, 1998, vol. 280, pp. 672–674.
26. Wainwright, P.O., Hinkle, G., Sogin, M.L., and Stickel, S.K., Monophyletic Origins of the Metazoa: An Evolutionary Link with Fungi, *Science*, 1993, vol. 260, pp. 340–342.
27. Aquinaldo, A.M., Turbeville, J.M., Linford, L.S., *et al.*, Evidence for a Clade of Nematodes, Arthropods and Other Molting Animals, *Nature*, 1997, vol. 387, pp. 489–493.
28. Tetushkin, E.Ya., Molecular Paleogenetics of Primates, *Genetika* (Moscow), 1997, vol. 33, no. 3, pp. 293–307.
29. Pawlowski, J., Bolinar, I., Fahrni, J.F., *et al.*, Extreme Differences in Rates of Molecular Evolution of Foraminifera Revealed by Comparison of Ribosomal DNA Sequences and the Fossil Record, *Mol. Biol. Evol.*, 1997, vol. 14, pp. 498–505.
30. Bruno, W.J., Socci, N.D., and Halpern, A.L., Weighted Neighbor Joining: A Likelihood-Based Approach to Distance-Based Phylogeny Reconstruction, *Mol. Biol. Evol.*, 2000, vol. 17, pp. 189–197.
31. Hwang, U.W., Kim, W., Tautz, D., and Friedrich, M., Molecular Phylogenetics at the Felsenstein Zone: Approaching the Strepsiptera Problem Using 5.8S and 28S rDNA Sequences, *Mol. Phyl. Evol.*, 1998, vol. 9, pp. 470–480.
32. Yang, Z., How Often Do Wrong Models Produce Better Phylogenies? *Mol. Biol. Evol.*, 1997, vol. 144, pp. 105–108.
33. Stiller, J.W. and Hall, B.D., Long-Branch Attraction and the rDNA Model of Early Eukaryotic Evolution, *Mol. Biol. Evol.*, 1999, vol. 16, pp. 1270–1279.
34. Khesin, R.B., *Nepostoyanstvo genoma* (Genome Instability), Moscow: Nauka, 1984.
35. Kurland, C.G., Something for Everyone: Horizontal Gene Transfer in Evolution, *EMBO Rep.*, 2000, vol. 1, pp. 92–95.
36. Xu', S., Phylogenetic Analysis under Reticulate Evolution, *Mol. Biol. Evol.*, 2000, vol. 17, pp. 897–907.
37. Andronico, F., De Lucchini, S., Graziani, F., *et al.*, Molecular Organization of Ribosomal RNA Genes Clusters at Variable Chromosomal Sites in *Triturus vulgaris meridionalis* (Amphibia, Urodela), *J. Mol. Biol.*, 1985, vol. 186, pp. 219–229.
38. Wellems, T.E., Walliker, D., Smith, C.L., *et al.*, Histidine-Rich Protein Gene Marks a Linkage Group Favored Strongly in a Genetic Cross of *Plasmodium viviparum*, *Cell* (Cambridge, Mass.), 1987, vol. 49, pp. 633–642.
39. Cumming, M.P., Otto, S.P., and Wakeley, J., Sampling Properties of DNA Sequence Data in Phylogenetic Studies, *Mol. Biol. Evol.*, 1995, vol. 12, pp. 814–822.
40. Wallis, G.P., Do Animal Mitochondrial Genomes Recombine?, *Trends Ecol. Evol.*, 1999, vol. 14, pp. 209–210.
41. Sheldon, F.E., Jones, C.E., and McCracken, K.G., Relative Pattern and Rates of Evolution in Heron Nuclear and Mitochondrial DNA, *Mol. Biol. Evol.*, 2000, vol. 17, pp. 437–450.
42. Arnason, U., Gullberg, A., Gretarsdottir, S., *et al.*, The Mitochondrial Genome of the Sperm Whale and a New Molecular Reference for Estimating Eutherian Divergence Dates, *J. Mol. Evol.*, 2000, vol. 50, pp. 569–578.
43. Avise, J.C. and Walker, D., Species Realities and Numbers in Sexual Vertebrates: Perspectives from Asexually Transmitted Genome, *Proc. Natl. Acad. Sci. USA*, 1999, vol. 96, pp. 992–995.
44. Hedges, S.B., Molecular Evidence for the Origin of Birds, *Proc. Natl. Acad. Sci. USA*, 1994, vol. 91, pp. 2621–2624.
45. Heise, P.J., Maxson, L.R., Dowling, H.G., and Hedges, S.B., Higher-Level Snake Phylogeny Inferred from mtDNA Sequences of 12S rRNA and 16S rRNA Genes, *Mol. Biol. Evol.*, 1995, vol. 12, pp. 259–265.
46. Boor, J.L. and Brown, W.M., Mitochondrial Genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: Sequences and Gene Arrangement Comparisons Indicate That Pogonophora Is Not a Phylum and Annelida and Arthropoda Are Not Sister Taxa, *Mol. Biol. Evol.*, 2000, vol. 17, pp. 87–106.
47. Macey, J.R., Larson, A., Ananjeva, N.B., and Papenfus, T.J., Evolutionary Shifts in Three Major Structural Features of the Mitochondrial Genome among Iguanian Lizards, *J. Mol. Evol.*, 1997, vol. 44, pp. 660–674.
48. Le, T.H., Blair, D., Agatsuma, T., *et al.*, Phylogenies Inferred from Mitochondrial Gene Orders: A Cautionary Tale from the Parasitic Flatworms, *Mol. Biol. Evol.*, 2000, vol. 17, pp. 1123–1125.
49. Hickerson, M.J. and Canningham, C.W., Dramatic Mitochondrial Gene Rearrangement in the Hermit Crab *Pagurus longicarpus* (Crustacea, Anomura), *Mol. Biol. Evol.*, 2000, vol. 17, pp. 634–644.
50. Moritz, C., Dowling, T.E., and Brown, W.M., Evolution of Animal mtDNA: Relevance for Population Biology

and Systematics, *Annu. Rev. Ecol. Syst.*, 1987, vol. 18, pp. 269–292.

51. Kumasawa, Y. and Nishida, M., Complete Mitochondrial DNA Sequence of the Green Turtle and Blue-Tailed Mole Skink: Statistical Evidence for Archosaurian Affinity of Turtles, *Mol. Biol. Evol.*, 1999, vol. 16, pp. 784–792.
52. D'Erchia, A., Gissi, G., Persole, G., *et al.*, The Guinea Pig Is Not a Rodent, *Nature*, 1996, vol. 381, pp. 597–600.
53. Slobodyanyuk, S.Ya., Kirilchik, S.V., Pavlova, M.E., *et al.*, The Evolutionary Relationships of Two Families of Cottoid Fishes of Lake Baikal (East Siberia) as Suggested by Analysis of Mitochondrial DNA, *J. Mol. Evol.*, 1995, vol. 40, pp. 392–399.
54. Fu, J., Toward the Phylogeny of the Family Lacertidae: Implication from Mitochondrial DNA 12S and 16S Gene Sequences (Reptilia: Squamata), *Mol. Phylogenet. Evol.*, 1998, vol. 9, pp. 118–130.
55. Buntjer, J.B. and Lenstra, J.A., Phylogeny within Families Deduced from Interspersed Repeat PCR Fingerprints, in *DNA Repeats in Vertebrate Genome as Probes in Phylogeny and Species Identification*, Utrecht: Univ. of Utrecht, 1997, pp. 25–38.
56. Eyre-Walker, A., Evolutionary Genomics, *Trends Ecol. Evol.*, 1999, vol. 14, p. 176.
57. Rand, D., Mitochondrial Genomics Flies High, *Trends Ecol. Evol.*, 2001, vol. 16, pp. 2–4.
58. Sorhannus, U. and Fox, M., Synonymous and Nonsynonymous Substitution Rates in Diatoms: A Comparison between Chloroplast and Nuclear Genes, *J. Mol. Evol.*, 1999, vol. 48, pp. 209–212.
59. Nunn, G.B. and Stanley, S.E., Body Size Effects and Rates of Cytochrome B Evolution in Tube-Nosed Seabirds, *Mol. Biol. Evol.*, 1998, vol. 15, pp. 1360–1371.
60. Hurai, S., Hayasaka, K., Kondo, R., *et al.*, Recent African Origin of Modern Humans Revealed by Complete Sequences of Hominid mtDNA, *Proc. Natl. Acad. Sci. USA*, 1995, vol. 92, pp. 532–536.
61. Gissi, C., Reyes, A., Pesole, G., and Saccone, C., Lineage-Specific Evolutionary Rate in Mammalian mtDNA, *Mol. Biol. Evol.*, 2000, vol. 17, pp. 1022–1031.
62. Mundy, N.I., Pissinatti, A., and Woodruff, D.S., Multiple Nuclear Insertion of Mitochondrial Cytochrome B Sequence in Callitrichine Primates, *Mol. Biol. Evol.*, 2000, vol. 17, pp. 1075–1080.
63. Garsia-Machado, E., Pempera, M., Donnebony, N., *et al.*, Mitochondrial Genes Collectively Suggest the Paraphyly of Crustacea with Respect to Insecta, *J. Mol. Evol.*, 1999, vol. 49, pp. 142–149.
64. Grechko, V.V., Ryabinin, D.M., Fedorova, L.V., *et al.*, Molecular Genetic Classification and Phylogenetic Relationships of Several Lizard Species of the Family Lacertidae as Evident from the Analysis of Specific Restriction Site Distribution in DNA Repeats (Taxonoprinting), *Mol. Biol.* (Moscow), 1998, vol. 32, no. 1, pp. 172–183.
65. Schierwater, B., Arbitrarily Amplified DNA in Systematics and Phylogenetics, *Electrophoresis* (Weinheim, Fed. Repub. Ger.), 1995, vol. 16, pp. 1643–1647.
66. Espinasa, L. and Borowsky, R., Evolutionary Divergence of AP-PCR (RAPD) Patterns, *Mol. Biol. Evol.*, 1998, vol. 15, pp. 408–414.
67. Ryabinina, N.L., Grechko, V.V., and Darevsky, I.S., DNA Polymorphism in Lizard Populations of the Family Lacertidae, as Revealed by RAPD Analysis, *Genetika* (Moscow), 1998, vol. 34, no. 12, pp. 1661–1667.
68. Grechko, V.V., Fedorova, L.V., Fedorov, A.N., *et al.*, Restriction Mapping of Highly Repetitive DNA Sequences in Studies of the Genetic Relationships in Lower Animal Taxa, *Mol. Biol.* (Moscow), 1997, vol. 31, no. 2, pp. 256–264.
69. Rabouam, C., Comes, A.M., Bretagnolle, V., *et al.*, Futures of DNA Fragments Obtained by Random Amplified Polymorphic DNA (RAPD) Assays, *Mol. Ecol.*, 1999, vol. 8, pp. 493–503.
70. Fukatsu, T. and Ishikawa, H., Differentiation of Aphid Clones by Arbitrary Primed Polymerase Chain Reaction (AP-PAR) DNA Fingerprinting, *J. Mol. Evol.*, 1994, vol. 3, pp. 187–192.
71. Ellegren, H. and Sheldon, B., New Tools for Sex Identification and the Study of Sex Allocation in Birds, *Trends Ecol. Evol.*, 1997, vol. 12, pp. 255–259.
72. Chobanu, D.G., Rudykh, I.A., Ryabinina, N.L., *et al.*, Reticulate Evolution of Parthenogenetic Rock Lizard Species of the Family Lacertidae: Inheritance of CLSat Tandem Repeats and Anonymous RAPD Markers, *Mol. Biol.* (Moscow), 2002, vol. 36, no. 2, pp. 296–306.
73. Khrisanfova, G.G., Semenova, S.K., and Ryskov, A.P., Cloning and Characterization of RAPD Genome Markers in Parasitic Nematodes *Trichinella spiralis* and *T. pseudospiralis*, *Mol. Biol.* (Moscow), 2000, vol. 34, no. 5, pp. 828–833.
74. Gilbert, M. and Labuda, D., CORE-SINEs: Eukaryotic Short Interspersed Retroposing Elements with Common Sequence Motifs, *Proc. Natl. Acad. Sci. USA*, 1999, vol. 96, pp. 2869–2874.
75. Singer, M.F., SINEs and LINEs: Highly Repeated Short and Long Interspersed Sequences in Mammalian Genomes, *Cell* (Cambridge, Mass.), 1982, vol. 28, pp. 433–434.
76. Doolittle, W.F. and Sapienza, C., Selfish Genes, the Phenotype Paradigm and Genome Evolution, *Nature*, 1980, vol. 284, pp. 601–603.
77. Orgel, L.E. and Crick, F.H.C., Selfish DNA: The Ultimate Parasite, *Nature*, 1981, vol. 284, pp. 604–607.
78. Kramerov, D.A., Major Repetitive Elements of the Mouse Genome, *Doctoral (Biol.) Dissertation*, Moscow, 1987.
79. Hillis, D.M., SINE of the Perfect Character, *Proc. Natl. Acad. Sci. USA*, 1999, vol. 96, pp. 9979–9981.
80. Hughes, D.S., MIRs as of Mammalian Gene Evolution, *Trends Genet.*, 2000, vol. 16, pp. 60–62.
81. Serdobova, I.M. and Kramerov, D.A., Short Retroposons of the B2 Superfamily: Evolution and Application for the Study of Rodent Phylogeny, *J. Mol. Evol.*, 1998, vol. 46, pp. 202–214.
82. Kass, D.H., Raynas, M.E., and Williams, T.M., Evolutionary History of B1 Retroposons in the Genus *Mus*, *J. Mol. Evol.*, 2000, vol. 51, pp. 256–264.

83. Sietkiewicz, E., Richter, C., and Labuda, D., Phylogenetic Affinities of Tarsier in the Context of Primate *Alu* Repeats, *Mol. Phyl. Evol.*, 1998, vol. 11, pp. 77–83.

84. Blinov, V.M., Denisov, S.I., Saraev, D.V., *et al.*, Structural Organization of the Human Genome: Distributions of Nucleotides, *Alu* Repeats, and Exons in Chromosomes 21 and 22, *Mol. Biol.* (Moscow), 2001, vol. 35, no. 6, pp. 1032–1038.

85. Armour, J.A.L., Monckton, D.G., Neil, D.L., *et al.*, Mechanisms of Mutation at Human Minisatellite Loci, *Genome Analysis*, vol. 7: *Genome Arrangement and Stability*, Cold Spring Harbor, New York: Cold Spring Harbor Lab., 1993, pp. 43–57.

86. Moyzis, R.K., Buckingham, J.M., Cram, L.S., *et al.*, A Highly Conserved DNA Sequence Presented at the Telomeres of Human Chromosomes, *Proc. Natl. Acad. Sci. USA*, 1988, vol. 85, pp. 6622–6626.

87. Grady, D.L., Rateiff, R.L., Robinson, D.L., *et al.*, Highly Conserved Repetitive DNA Sequences Are Present at Human Centromeres, *Proc. Natl. Acad. Sci. USA*, 1992, vol. 89, pp. 1695–1696.

88. Miclos, G.L.G., Localized Highly Repetitive DNA Sequences in Vertebrate and Invertebrate Genome, *Molecular Evolutionary Genetics*, McIntyre, R.J., Ed., New York: Plenum, 1985, pp. 241–321.

89. Beridze, T.G., *Satellitnye DNK* (Satellite DNA), Moscow: Nauka, 1982.

90. Elder, J.F. and Turner, B.J., Concerted Evolution of the Population Level: Pupfish *Hin*dIII Satellite DNA Sequences, *Proc. Natl. Acad. Sci. USA*, 1994, vol. 91, pp. 994–998.

91. Varley, J.M., Macgregor, H.C., and Barnett, L., Characterization of a Short Highly Repeated and Centromerically Localized DNA Sequences in the Crested and Marbled Newts of the Genus *Triturus*, *Chromosoma*, 1990, vol. 100, pp. 15–31.

92. Lorite, P., Palomeque, T., Garneria, I., and Petitpierre, E., Characterization and Chromosome Location of Satellite DNA in the Leaf Beetle *Chrysolina americana* (Coleoptera, Chrysomelidae), *Genetica* (The Hague), 2001, vol. 110, pp. 143–150.

93. Castagnone-Sereno, P., Leroy, H., Semblat, J.-P., *et al.*, Unusual and Strongly Structured Sequence Variation in a Complex Satellite DNA Family from the Nematode *Meliodogine chitwoodi*, *J. Mol. Biol.*, 1998, vol. 46, pp. 225–233.

94. Mednikov, B.M., Bannikova, A.A., Lomov, A.A., *et al.*, Restriction Enzyme Analysis of Nuclear Repetitive DNA, the Criteria of Species, and the Mechanism of Speciation, *Mol. Biol.* (Moscow), 1995, vol. 29, pp. 1308–1319.

95. Garrido-Ramos, M.A., de Herran, R., Jamilena, M., *et al.*, Evolution of Centromeric Satellite DNA and Its Use in Phylogenetic Studies of the Sparide Family (Pisces, Perciformes), *Mol. Phyl. Evol.*, 1999, vol. 12, pp. 200–204.

96. Goodier, J.L. and Davidson, W.S., Characterization of Novel Minisatellite Repeat Loci in Atlantic Salmon (*Salmo salar*) and Their Phylogenetic Distribution, *J. Mol. Evol.*, 1998, vol. 46, pp. 245–255.

97. Vignali, R., Rijili, F.M., Batistoni, R., *et al.*, The Dispersed Highly Repeated DNA Families of *Truturus vulgaris meridionales* (Amphibia, Urodela) Are Widely Conserved among Salamandridae, *Chromosoma*, 1991, vol. 100, pp. 87–96.

98. Olmo, E., Odierna, G., and Capriglione, T., The Karyology of Mediterranean Lacertidae Lizards, *Lacertid of the Mediterranean Region*, Valakas, E.D., Bohme, W., Rerer-Mellado, V., and Maragou, P., Eds., Athens: Alicante, 1993, pp. 61–84.

99. Kozlova, S.V., Elisafenko, E.A., Mazurok, N.A., and Zakiyan, S.M., Genome Organization and Chromosomal Localization of New Repeat MS7 Specific to Sex-Chromosomal Heterochromatin in Vole of the *arvalis* Group of the Genus *Microtus*, *Mol. Biol.* (Moscow), 2001, vol. 35, no. 5, pp. 792–797.

100. Potapov, S.G., Diagnostic Potentialities of Molecular Genetic Markers in Mammalian Systematics, *Cand. Sci. (Biol.) Dissertation*, Moscow, 1998.

101. Shevchenko, A.N., Slobodyanyuk, S.Ya., and Zakiyan, S.M., Analysis of the Variation of Repetitive DNA Sequences in Four Common Vole Species, *Mol. Biol.* (Moscow), 1999, vol. 33, no. 4, pp. 700–705.

102. Modi, W.S., Gallagher, D.S., and Womack, J.E., Evolutionary Histories of Highly Repeated DNA Families among the Artiodactyla (Mammalia), *J. Mol. Evol.*, 1996, vol. 42, pp. 337–349.

103. Blake, R.D., Wang, J.Z., and Bauregard, L., Repetitive Sequence Families in *Alces alces americana*, *J. Mol. Evol.*, 1997, vol. 44, pp. 509–520.

104. Milinkovitch, M.C., Molecular Phylogeny of Cetaceans Prompts Revision of Morphological Transformation, *Trends Ecol. Evol.*, 1995, vol. 10, pp. 328–334.

105. Seymour, J., DNA Tells of Species Links by Repeating Itself, *Ecos*, 1982, no. 30, pp. 21–23.

106. Mullenbach, R., Lutz, S., Holzmann, K., *et al.*, A Non-Alphoid Repetitive DNA Sequence from Human Chromosome 21, *Hum. Genet.*, 1992, vol. 89, pp. 519–523.

107. Alves, G., Seuanez, H.M., and Fanning, T., A Clade of a New World Primates with Distinctive Alphoid Satellite DNAs, *Mol. Phyl. Evol.*, 1997, vol. 9, pp. 220–224.

108. Trifonov, E.N., Tuning Function of Tandem Repeating Sequences: A Molecular Device for Fast Adaptation, *Proc. Int. Conf. Evol. Genomics*, Costa Rica, 1999, p. 74.

109. Dimitri, P. and Junakovic, N., Revising the Selfish DNA Hypothesis: New Evidence on Accumulation of Transposable Elements in Heterochromatin, *Trends. Genet.*, 1999, vol. 15, pp. 123–124.

110. Martinez-Balbas, A., Rodriguez-Campos, A., Garcia-Ramirez, M., *et al.*, Satellite DNA Contain Sequences That Induce Curvature, *Biochemistry*, 1990, vol. 29, pp. 2342–2348.

111. Constanzo, G., De Muiro, E., Salina, G., and Negri, R., Attraction Phasing and Neighbor Effects of Histone Octamer on Curved DNA, *J. Mol. Biol.*, 1990, vol. 216, pp. 363–374.

112. Romanova, L.Y., Deriagin, G.V., and Mashkova, T.D., Evidence for Selection in Evolution of α-Satellite DNA: The Central Role of CENP-B/pJ Binding Region, *J. Mol. Biol.*, 1996, vol. 261, pp. 334–340.

113. Warburton, P.E., Waye, J.S., and Willard, H.F., Nonrandom Localization of Recombination Events in Human α-Satellite Repeat Unit Variants: Implications for

Higher Orders Structural Characteristics within Centromeric Heterochromatin, *Mol. Cell. Biol.*, 1993, vol. 13, pp. 6520–6529.

114. Larin, Z., Fricker, M.D., and Tyler-Smith, C., De Novo Formation of Several Features of a Centromere Following Induction of a Y Alphoid YAC into Mammalian Cells, *Hum. Mol. Genet.*, 1994, vol. 3, pp. 689–695.
115. Capriglione, T., DeSanto, M.G., Odierna, G., and Olmo, E., An Alphoid-like Satellite DNA Sequence Is Present in the Genome of a Lacertid Lizards, *J. Mol. Evol.*, 1998, vol. 46, pp. 240–244.
116. Slamovits, L.H., Cook, J.A., Lessa, E.P., and Rossi, M.S., Recurrent Amplifications and Deletions of Satellite DNA Accompanied Chromosomal Diversification in South American Tuco-Tucos (Genus *Ctenomys*, Rodentia: Octodontidae): A Phylogenetic Approach, *Mol. Biol. Evol.*, 2001, vol. 18, pp. 1708–1719.
117. Saitoh, Y., Miyamoto, N., Okada, T., *et al.*, The RS447 Human Megasatellite Tandem Repetitive Sequence Encodes a Novel Deubiquitinating Enzyme with Functional Promoter, *Genomics*, 2000, vol. 67, pp. 291–300.
118. Plohl, M., Mistrovic, N., Bruvo, B., and Ugarkovic, D., Similarity of Structural Features and Evolution of Satellite DNAs from *Palorus subdepressus* (Coleoptera) and Related Species, *J. Mol. Evol.*, 1998, vol. 46, pp. 234–239.
119. Kipling, D. and Warburton, P.E., Centromeres, CENP-B and Tigger Too, *Trends Genet.*, 1997, vol. 13, pp. 141–145.
120. Fedorov, A.N., Fedorova, L.V., Grechko, V.V., *et al.*, Variable and Invariable DNA Repeats Characters Revealed by Taxonoprint Approach Are Useful for Molecular Systematics, *J. Mol. Evol.*, 1999, vol. 48, pp. 69–76.