https://doi.org/10.15407/scine21.06.087

KLYUCHKO, O. M. [1] (https://orcid.org/0000-0003-4982-7490),
BELETSKY, A. Ya. [1] (https://orcid.org/0000-0002-3798-8150),
MELEZHYK, O. V. [2] (https://orcid.org/0000-0003-3882-7102),
and GONCHAR, O. O. [3] (https://orcid.org/0000-0003-4134-03544)

[1] National Aviation University,
1, L. Huzar Ave., Kyiv, 03058, Ukraine,
+380 44 406 7130, post@kai.edu.ua,
[2] Open International University of Human Development "Ukraine,"
23, Lvivska St., Kyiv, 03115, Ukraine
+380 67 328 2822, office@uu.edu.ua
[3] Bogomoletz Institute of Physiology National Academy of Sciences of Ukraine,
4, Bogomoletz St., Kyiv, 01024, Ukraine,
+380 44 256 2400, science@biph.kiev.ua

BIOLOGICAL DATABASES: THREE NOVEL METHODS OF KEYS FORMATION FROM GENETIC CODES

Introduction. *Creation of novel databases (DB) with biomedical information based on the latest technologies is important task.*

Problem Statement. *That is why it is an important issue to invent innovative approaches to the construction of DBs with biomedical information using keys with extended capabilities.*

Purpose. *Description and characteristics of the three newest author's methods of forming (encoding) keys for different groups of biological objects creating of innovative databases with biomedical information.*

Materials and Methods. *Analysis of images, comparative analysis, object-oriented system analysis for DBs construction; methods of DB design and ER-diagrams design were used.*

Results. *Using examples of relational database (DB) development with information about certain biological species, the application of object-oriented analysis methods for DB construction has been characterized, and an algorithm for their construction has been described. Particular attention has been paid to solving the problem of creating keys based on the genetic codes of these species: in alphanumeric expression, images, and fragments of alphanumeric images. The proposed methods for code formation are especially important as primary keys, ensuring links between individual DB tables, maintaining data integrity in the system, and providing reliable data access. The high level of data organization and retrieval when using keys based on genetic codes has been thoroughly substantiated.*

Citation: Klyuchko, O. M., Beletsky, A. Ya., Melezhyk, O. V., and Gonchar, O. O. (2025). Biological Databases: Three Novel Methods of Keys Formation from Genetic Codes. *Sci. innov.*, 21(6), 87–103. https://doi.org/10.15407/scine21.06.087

Conclusions. *The results can be applied in information systems with biomedical information, for the further development of DB construction methods, and for improving certain methods of protecting medical-biological and personal data. Databases with such keys can be effective for organizing large datasets about various groups of people, biological organisms, and materials, and are capable of ensuring the fastest access to relevant information and reliable protection of such information.*

Keywords: information technologies, biological databases, key, primary key, coding, biological species.

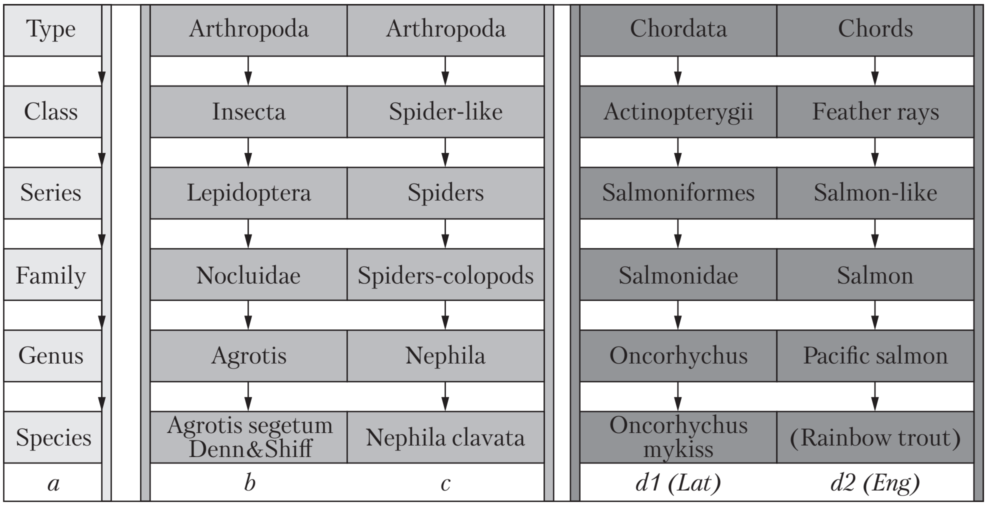
The creation of databases (DBs) with biological data, including data from taxonomy, systematics, genetics, etc. has become necessary and even urgent task in the last decades. The data organized in such DBs further can be used in medicine, ecology, nature conservation, agriculture, etc. — many practically important industries [1, 2]. Application of achievements of biology, biophysics in information technologies opens new possibilities for innovations, contributes to solving tasks at a new, higher level. In this article, we will examine briefly the technique of creating of relational DBs with information about biological organisms, tissues of organisms, their remains — that is, all natural objects that contain genetic material: DNA, RNA.

The creation of relational DBs with information about biological organisms, their fragments or remains, fragments that contain genetic material, DNA or RNA, some recent publications are devoted [1—4]. For such DBs and in process of their creating, the so-called “keys” play important role. The “keys” ensure fast and reliable information search in relation DBs, as well as in DBs with biological information (which can also be used in medicine). Three original ways of generating of 3 types of “keys” based on genetic codes for such DBs will be suggested below. Other words, there will be described three mechanisms for reliable and fast search of biological information in DBs.

But from other side, these 3 novel types of “keys” can also be used to protect personal information in such database (DB). Since the technique of relational DBs creating is abstract, not related to specific types of biological organisms, we will try to demonstrate its individual stages using the examples of various biological creatures — insects, fish, and so on. It is also necessary to emphasize the importance of developed methodology usage for DB for humans and their biological materials. The tatter is especially important during the war in Ukraine these days, when DBs with large volumes of the data are necessary for identification in military sphere, for ordering of information about wounded and affected people, police DBs of dead people, terrorists, etc. A lot of such DBs are being created now in Ukraine, they are really necessary, and their optimization can be really important, relevant and useful. As a result, there is another type of diversity had appeared — database diversity [1, 2]. Most often, such DBs are functionally related to biological or medical information systems (ISs), that is why the method of data recording and retrieving from a DB in which the data are logically ordered and structured is becoming more and more important.

Relational DBs are the most widely used for biological objects ordering now in contemporary science and practice [1—17]. Below the algorithm of construction of relational DBs for biological information ordering was described in details; it can be used also for the creation of individual DBs of scientists. The bases of DBs creation were given in some manuals too [18, 19], monography of the author [2], scientific publications [20—29], and patents [25—29]. The possibility to use invented specific keys, based on genetic material (genetic codes) will be demonstrated for DBs construction. Materials for DBs were taken from authors’ personal publications [20—29], as well as from other publications in open access [30—56].

The logical sequence of material suggested below can be given briefly in following order: 1) analysis of object area, defining of the main objects for created relational database; 2) creation of central “core” of the DB; 3) analysis of some limitations of the use of the keys formed in traditional manner;



Type	Arthropoda	Arthropoda	Chordata	Chords
Class	Insecta	Spider-like	Actinopterygii	Feather rays
Series	Lepidoptera	Spiders	Salmoniformes	Salmon-like
Family	Nocluidae	Spiders-colopods	Salmonidae	Salmon
Genus	Agrotis	Nephila	Oncorhychus	Pacific salmon
Species	Agrotis segetum Denn&Shiff	Nephila clavata	Oncorhychus mykiss	(Rainbow trout)
a	*b*	*c*	*d1 (Lat)*	*d2 (Eng)*

Fig. 1. The hierarchical structure of the classification of biological species corresponds very well to the concept of object-oriented analysis with the subsequent construction of DBs (details see in text)

4) description of three our invented original methods for keys formation on the base of genetic codes (including primary keys); 5) discussion of possibilities of novel types of keys use for innovative DBs construction with biological and medical data.

From a practical point of view, the algorithm for designing and creating biological DBs, as well as an algorithm for generating keys in them according to the three novel techniques was proposed by O.M. Klyuchko. According to this technique, keys for the relative DBs with biological (as well as medical information) can be based on the genetic codes of biological organisms. Five patents of Ukraine have been obtained for these methods.

The purpose of this research is to describe and characterize three newest author's methods of forming (encoding) keys for different groups of biological objects for the creation of innovative databases with biomedical information.

DESIGNING OF RELATIVE DATABASES WITH BIOLOGICAL AND MEDICAL DATA

Analyzing the characteristics of biological organisms that shall be placed in the DB for systematics, taxonomy, ecology etc., researchers can notice a set of regularities that determine the logical structure of the data in DBs (data representation model). Below we examined organisms-representatives of *Arthropod*s — night insects-moths (*Noctuidae, Lepidoptera*) [2, 20, 21, 35—40], fishes [22—26, 45—49, 54—56] and humans [51—53]. According to their characteristics, all biological objects form an ideal hierarchy. Figure 1 demonstrates the examples of hierarchical classification of phylogenetically distant organisms — representatives of insects (moths), spiders and fishes (see Fig. 1). It is possible to see that the so-called hierarchical model of DB construction, which was developed for engineering and natural sciences (like physics, mathematics) and has long been widely used there, is well suited to them. However, simple borrowing (acceptance followed by adaptation of this method) does not give successful experience due to the extreme individual variability of living organisms. Numerous deviations from standard characteristics and values do not allow creating successfully biological DBs without taking into account biological variability and other features of living things. Special techniques have to be developed when creating such DBs.

PROBLEMS OF DATABASE DESIGN FOR SYSTEMATICS AND TAXONOMY

Design of the information system (IS) which includes a DB is performed at the physical and logical levels. Solving of problems of design at physical level in many cases depends on the used DB management system (DBMS), often automated and hidden from the user. In a number of cases, the user is given the opportunity to adjust certain system parameters. In process of DB development for systematics and taxonomy, the main attention should first of all be directed to logical design. Logical design consists in determining the number and structure of tables, forming queries to the DB, determining the types of reporting documents, developing information processing algorithms, creating forms for entering and editing data in DB, and solving a number of other tasks. The nature of biological objects is the specificity that determines the solution of the problem of logical design. The most important here is the problem of data structuring. When designing data structures, it is necessary to distinguish three main approaches. Collection of information about the objects of the solved problem within one table (one relation) and its subsequent decomposition into several interconnected tables based on the procedure of normalization of relations. Further — formulation of knowledge about the system (definition of the types of initial data and relationships) and requirements for data processing, obtaining with the help of the automation system of the design and development of the DB, a ready-made DB scheme or even a ready-made application IS.

STRUCTURING INFORMATION FOR USE IN IS IN THE PROCESS OF SYSTEM ANALYSIS BASED ON A SET OF RULES AND RECOMMENDATIONS

The specificity of biological objects: their definition, standard data set for each species play a decisive role in the creation of biological DBs, as well as information collection in the field about such objects and its primary analysis. In our previous publications we had analyzed the characteristics of insects — moths (*Noctuidae, Lepidoptera*) [2, 20, 21, 35—40], and fishes [22—26, 45—49, 54—56] which should be included in the systematics and taxonomy DB. Such DBs can be attached to ecological ISs as their necessary component. Moths (*Noctuidae, Lepidoptera*) are nocturnal butterflies, many of which can be seen on summer nights. Their place in the hierarchy of living organisms according to Linnaeus (1761) [2] is shown in Fig. 1. According to this classification, in its modern form, the object we are considering — night insects is a subject to the hierarchical structure shown in Fig. 1. For comparison, the information about the fish species is given in parallel in Fig. 1 too — it is commonly organized in the same hierarchical manner. So, the information used for classification of living organisms in biology is structured very well, and this structure is hierarchical [2, 20—24, 35—40].

ANALYSIS OF THE OBJECT AREA IN PROCESS OF A DATABASE CONSTRUCTION, FINDING OF THE MAIN OBJECTS FOR THE DATABASE

The next step in DB design is the analysis of the so-called object area. Let's perform it and the following steps on the example of designing a DBs of fish. As mentioned above, we make such a change to the object of examination in order to demonstrate the generality of methods for DBs creation; they do not depend on the type of biological objects. Let's examine the methods of DBs construction only briefly for the following illustration of the invented methods.

Model task "Influence of the harmful substances-pollutants on fish fauna." Let's demonstrate invented methods on the model task about the influence of the harmful substances-pollutants on fish fauna" *and* using examples of the DB with 1) information about fishes, and 2) linked information about some substances-pollutants that can influence on fish fauna in the nature. Such ecological problems solutions are extremely important for the whole world, and especially for contemporary Ukraine, where the war occur great negative antropogenic pressure on the Nature. To solve the task, the main objects were selected for our analysis: “Species of fish” and “Chemical substance” of two types: a) some radioactive isotopes that can be environmental pollutants (for example, this happened as a result of the Chernobyl accident), and b) some organic substances (in our case with fragments — aromatic groups). Such substances flow to the water basins in industrial zones, regions where the war is continued, as a result of accidents, and all this has a negative effect on the fish population (Fig. 1). For example, only a few species of fish and a limited number of substances are taken, usually in the DB one can order great number of units of information. The third object on the scheme of Fig. 1 represents taxonomic characteristics of the main object “Fish species.”

OBJECT ANALYSIS: THE FIRST STAGE IN THE DATABASE CONSTRUCTION ALGORITHM

The first step in process of DBs construction is the analysis of objects, their characteristics and relations between them. In details this stage is described in [2, 20—24]. Interesting from the point of view of logic and perfect method for this is “ER-diagram” design (see Fig. 2). According to this method it is necessary to reveal all the most important objects for this DB (in our case — biomedical objects), to analyze their characteristics and relations between them. Basing on the results of this analysis, it is necessary to create a diagram called «ER-diagram» (Fig. 2). Moreover, some details of future DB become clear only in process of its ER-diagram complex analysis — so, without it the design of the DB of high quality is not possible.

The approach used for this purpose is known as Object Oriented Analysis (OOA). OOA approach is used here to help in design of optimal DB construction, and to do the process of DB construction in optimal way. Object (or “entity” is a basic notion of this approach. As example let’s demonstrate this approach creating the DB with the results of ecological monitoring of ichthyofauna. It is known that fish fauna is sensitive to various chemicals, both organic and inorganic compounds. A lot of such chemical substances, harmful to fauna, enter the natural environment, especially in the industrial regions of Ukraine, as well as during the current military operations, one of the consequences of which is the massive release of aggressive pollutants into the environment. In order to track the processes of the destructive impact on nature of such factors and help its restoration in the future, the creation of such DB is very important. The main objects that we highlight were the objects “Biological species” (in our example — fishes), its “Taxonomic category” and “Chemical substance” (a) radioactive isotopes, and b) harmful organic substance-pollutant). As a result of the analysis an ER-diagram can be created, a sample of which is shown in Fig. 2.

Further basing on this ER-diagram we will demonstrate author’s inventions — three methods how to use fragments of genetic codes of examined biological organisms as so-called “natural keys” for such DB design (Fig. 2). All other keys in Fig. 2 play roles of “alternate keys” (“GenCodeBioOrg*”). As it is understandable from above written, for the creation of functionally effective relation DB is necessary to do reliable connections between the tables (and they are transformed “objects”) in DB. To do this there was elaborated the method of keys formation from randomly selected symbols (letters or/and numbers), and this method is widely used today. Keys of this type in our case are called surrogate keys.

Necessary to mention that one of proposed objects “Biological species” has some identified characteristics (Fig. 2); they can be selected according to necessity and the task solved. Such characteristics were called “attributes” in the terms of DB. The same was done for other objects — “Chemical substance” and “Taxonomic category.” Any object can be characterized by their specific characteristics — “attributes.” Specific geometric symbols were traditionally used for all objects, attributes and relations (Fig. 2) [2, 20—24]. The special charac-

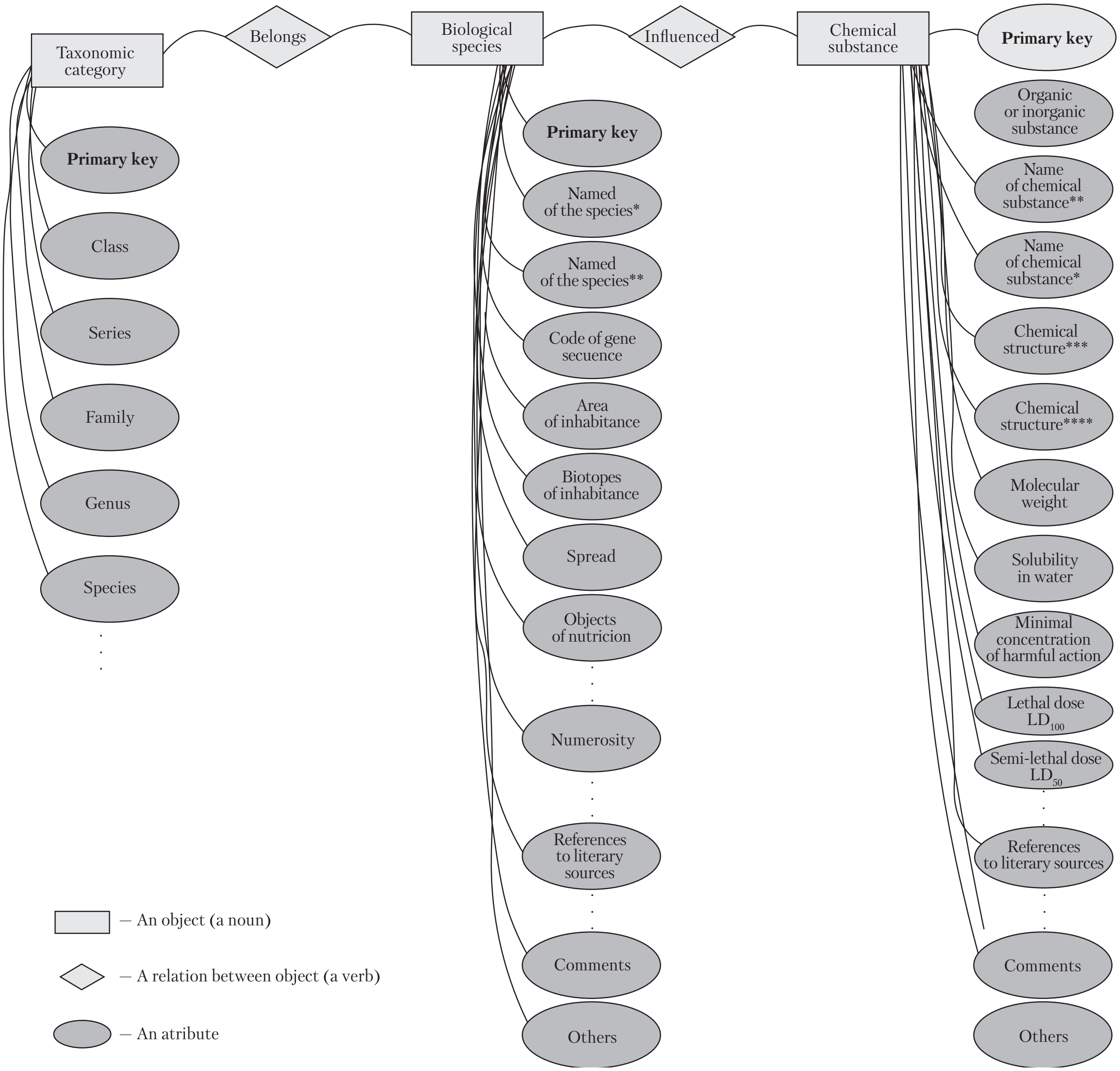


Fig. 2. Fragment of ER-diagram “Biological species” and “Chemical substance” that cause harmful influence on them: * Ukrainian; ** Latin; *** in alfanumeric form; **** in form of scheme

teristics which called “attributes” were used for ER-diagram in Fig. 2 and for the completion of corresponding table in Fig. 3. For proposed DB all attributes for *Noctuidae (Lepidoptera)* were taken from the “Red Book of Ukraine” [34] as well as from some publications about these biological species taxonomic hierarchy [2, 35—40]. For DBs with ecological information about chemicals (organic compounds) in environment of industrial regions of Ukraine the information was taken from [43—45]. These organic compounds with their derivatives were identified in wastewaters as environmental pollutants in industrial regions, near airports of Ukraine [43]. Further develop-

Object “Biological species”	Object “Chemical substance”
Attributes:	Attributes:
Primary key (based on the fragment of genetic sequence) Name of the species (Latin) Name of the species (Ukrsinian) **Code of genetic sequence** (if it is known) Area of inhabitance Biotopes of inhabitance Spread Location Numerosity Reasons of numerosity changes Features of biology Objects of nutricion Reproduction in captivity Protection measures .... References to literary sources .... Comments Others	**Key** Prganic or inorganic substance Name of chemical substance (Latin) Name of chemical substance (Ukraine) Chemical structure (alphanumeric form) Chemical structure (schematic image) Molecular weight Siolubility in water Minimal concentration of harmful action Lethal dose LD_{100} Semi-lethal dose LD_{50} .... References to literary sources .... Comments Others

Fig. 3. Characteristics of the objects “Biological species” and “Chemical substance” [2, 23, 24]

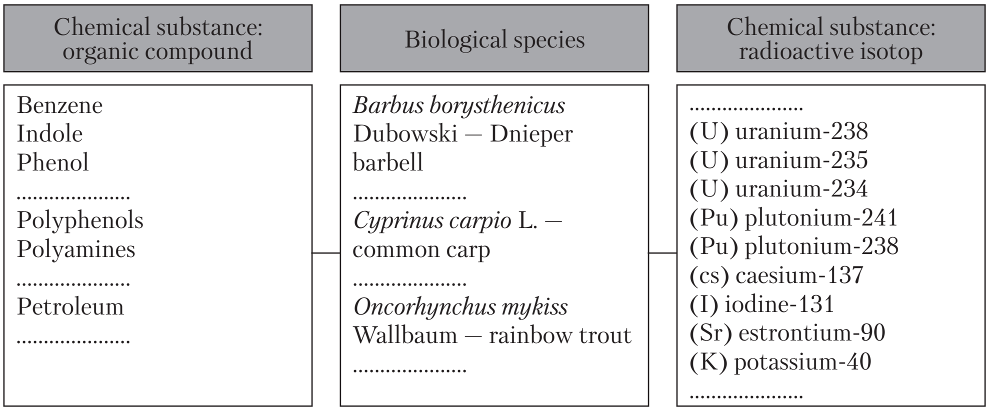


Fig. 4. Logical model of the DB with the links between some biological species (fishes) and different chemical substances influenced on them

ments of DBs were well described in [2, 20—24]. Power of relations was:

- “Fish species”:“Chemical substance” = 1 : N, and
- “Fish name” (Latin) : “Fish name” (Ukrainian) = = 1 : 1.

For the construction of the DB from the example with fish species, we had selected: *Barbus borysthenicus* Dybowski, *Cyprinus carpio* L., *Oncorhynchus mykiss* Walbaum [54—56]. Their connections with chemical substances a) radioactive isotopes, and b) organic substances-pollutants of the environment are presented in Fig. 4. Such logical models can be constructed for any biological object and associated characteristics.

A large-scale and deep analysis is necessary for the construction of biological DBs, because only on its basis the DB can be further created, from which all the necessary information about the studied objects can be obtained. The ER-diagram can be seen as abstraction that represents in general form all *objects* from the DB with all their *relations* and with all characteristics of the objects — *attributes*. In the process of DB developing the defined "*objects*" (Fig. 2) are transformed into tables (Fig. 4). "*Attributes*" (characteristics of objects are transformed into parts of the table — columns in which data about objects are written. At the same time, "*relations*" between objects (described by verbs) are transformed into connections between these related tables. For biological objects and materials (like remains after the organism death) following data (attributes) are important: species, genetic sequences, family, genus, other taxonomic characteristics and etc. ER-diagram helps to analyze better all objects. (entities) that we would like to order in the DB and examine them in their entirety. The ER-diagram helps to analyze better all objects (entities) that are arranged appropriately in the DB, and check the completeness of this information. Important requirement for the normal functioning of the DB and for its successful use is that the requested information has to be obtained from there reliably and as more complete as possible. As a result of such construction these requirements are satisfied.

Left column: some organic substances — environmental pollutants. Right column: some radioactive isotopes influencing on biological objects (explanation see in text). Logical model was done on the base of ER-diagram (see Fig. 2).

Description of invented methods of the links formation through the "keys" formed on the base of genetic codes. It is possible to notice how the objects can be transformed into the tables (details see in text).

It cannot happen that the information received at the request of the user is incomplete, that is, part of the records remained in the DB and could not be "retrieved» from there. If this happens, then it means that such a DB was designed and developed with errors. Such errors can be removed due to in-depth analysis, an example of which was described above. In addition, the connection between objects (tables) in the DB shall be reliable; its reliability is ensured using the "key" mechanism. Below we observed and examined such a "key" mechanism and explain the essence of our proposed innovations in creating a "key" mechanism to DBs with biological and medical information.

THE METHOD OF THE RELATIONS FORMATION BETWEEN THE OBJECTS (TABLES) USING THE "KEYS"

The method of "keys" was invented to ensure the data integrity in DBs, increase the connectivity of information and its level of individualization (this is quite important in biological and medical DBs), as well as to prevent the information losses in DB [2, 18—24].

Mechanism of "keys" was used for making relations between the objects ("entities"). When "keys" are professionally done, they provide accessible information in DB, it reliable retrieval. So, it cannot be lost and it can be retrieved from the DB. Inaccessible data cannot be retrieved! ""Keys" are fundamental elements of any relational database and of each table within it. If the information is retrieved from any table, it is consequently possible to output the information from another one linked, through a relation that unites their keys. Usually, as a "key" serves a code from random generated symbols (numbers, other symbols). In DBs with biomedical information the keys are generated traditionally in the same way. The links between such "key" codes from separated tables (transformed "objects") one can do easily in many contemporary software environments.

Figures 2—5 demonstrate how the "keys" serve for the relations' formation between different fields of various tables that have to be logically linked. All these 4 figures demonstrate consequently step-by-step the process of keys usage for relations' formation during DB construction. For example,

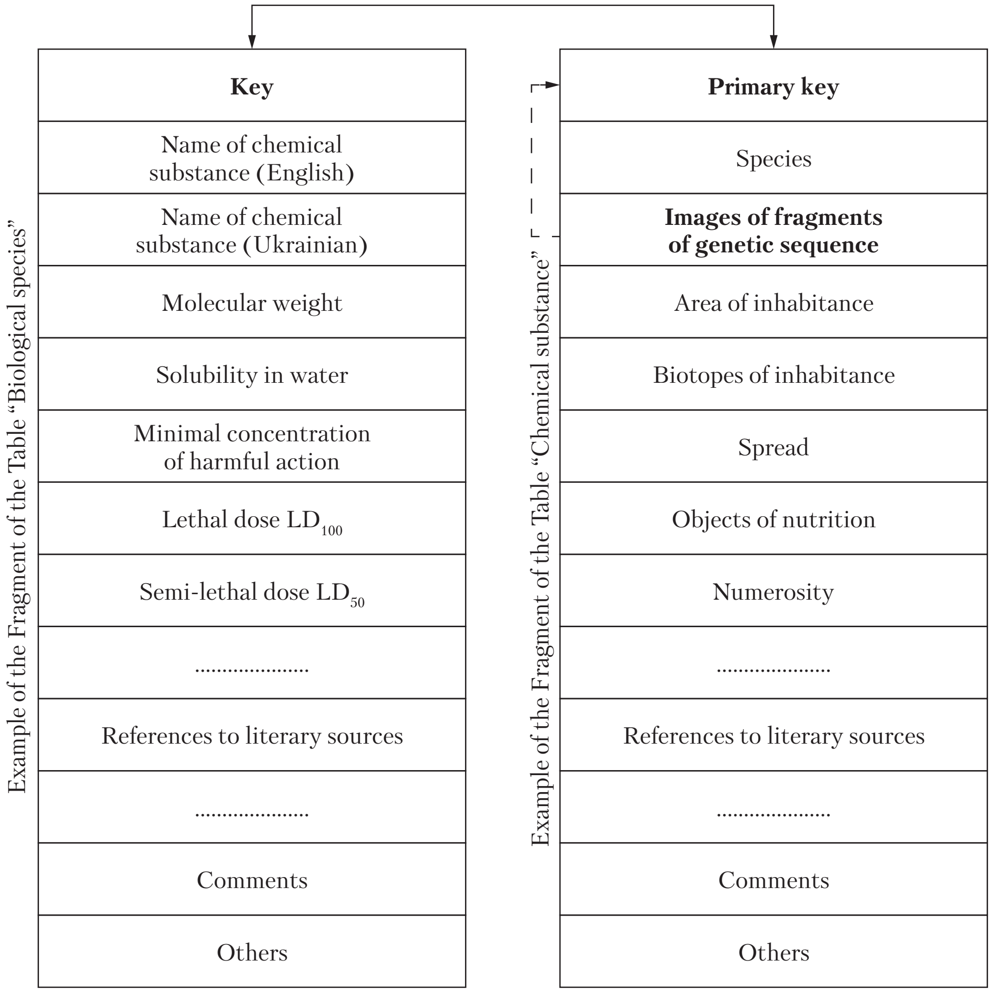


Fig. 5. Tables linked with two selected objects in our task

somebody is interested in how any defined chemical substance (f. e., any phenol derivative A) effects on any biological species (or this species population) [22, 23]. Let’s demonstrate this using the tables in Fig. 5. DBs developer knows that this is really true fact: examined “phenol derivative A” in the water really influences on certain biological organism. So, DBs developer has to make the link between EXACTLY this biological organism and EXACTLY the chemical substance (phenol derivative A) harmful for it. Let’s give a name “GenCodeBioOrg*” to the key that represent this biological species and which we would like to create. So, in program the key “GenCodeBioOrg*” from the table “Biological species” have to be linked with the key that is linked with “Phenol derivative” in the table “Chemical substance” (from water pollutants). And so on. From other side, the keys of chemical substances in Fig. 5 may be coded in trivial, commonly used manner — using the group of randomly generated numbers of any alphanumeric symbols (like the key for “Phenol derivative A,” for example). But for coding of the key “GenCodeBioOrg*” another way was suggested, and we realized it using three possible techniques. These three novel invented techniques based on the genetic codes of these species are described below.

THREE NOVEL INVENTED METHODS OF GENETIC CODES USE AS "KEYS" FOR THE CREATION OF DATABASES WITH BIOLOGICAL AND MEDICAL INFORMATION

So, all objects (entities) in our relative DBs have to be linked to each other through programmed links. In contrast to the creation of keys for the connection of DBs with biological and medical information based on random sequences of numbers and/or letters, Dr. O.M. Klyuchko proposed another mechanism for generating of keys. The prototype example of this invention is well known and traditionally used. During the development of DBs, for example, of bank customers, often the passport codes of these customers as keys are used or registration number of the taxpayer's registration card (NTPRC), which are unique. In addition, in such DBs, the columns with information about the passport code or NTPRC are mandatory. Therefore, it is very natural to use these data (codes) as keys.

Solving the problem of creating keys in a DB with biological and medical information, you can use a similar technique. After all, every biological organism in nature has its own individual code ("number") — genetic code. Moreover, many modern DBs include data on the genetic codes of organisms. The "genetic code" of organism can serve as a "key" in process of construction a DB of biological organisms, organ tissues, biological materials of human remains (animals, plants, etc.). Of course, here we are writing only about a fragment of the real genetic code that can be used because of limited dimensions, volumes, spaces in tables. Regarding the characteristics of this fragment (its length, sequence of nucleotides, etc.) they will be discussed below, in the discussion of suggested methods.

Genetics demonstrates nowadays great successes in genomes investigations. Genetic information is included into the biomedical DBs more and more from day to day. The structures of fragments of genetic sequences have been studied for numerous biological organisms, for some individual persons as well [51—52]. In many cases (for example in clinical practice) these data have been already organized in the biomedical DBs. Well known the importance of genetic information for each biological species, each individual person or animal as well. In each specific case, it is better to choose a really specific fragment of the sequence of nucleotides as a key. For example, such fragment can characterize exactly this species, individual person or patient [51—52] animal, tissue, and so on. Modern genetics suggest numerous results of genomes examinations — so, for many cases such keys with individual "characteristics" can be picked up and used as keys according to one of three our suggested forms for biomedical DBs creation.

Specific unique genetic code of each biological organism can be abstractly represented as the sequences of nucleotides. Such nucleotides can be coded in different manners — A) as sequences from the letters and numbers (alphanumeric codes) [24, 25, 28], or B) in the forms of images [26, 27]. Each sequence is highly specific and individual (for each species or for each organism). There is only a task for DB developer — to pick up preferable fragment of such genetic individual code, which would be suitable for the use as a "key" in each case. We suggest the term "natural key" for such keys — they can serve as "primary key." Using them the effect of "data integrity" can be realized very well.

Three novel invented by author methods of genetic codes use as "keys" can be explained using material represented in Fig. 6. Here, each figure demonstrates defined fishes' species in different forms of records. Accordingly, such different sequences of elements selected from the genetic codes of various organisms can serve as three different types of "key" codes. For example, Fig. 6, *a* demonstrates the fragment of the genetic sequence of one fish species [54, 55] in alphanumeric notation [24, 25, 28]. Figure 6, *b* shows fragments of genetic sequences of few fishes' species [54, 55] also in alphanumeric notation — however, when read, for example, with a scanner, such sequences are not recognized as separate symbols (as in Fig. 6, *a*); but they are perceived as complete image of a set of symbols (like similar to those outlined by rectangular frames) [29]. Two fragments of nucleotide sequences of fishes [54, 55] in the form of images of genetic codes are demonstrated in Fig. 6, *c*; they look like assets of multi-colored spots obtained during genome analysis in laboratory conditions. Actually, modified barcode scanners can be used for the work with them because they look like as a kind of barcode for such devices [26, 27]. Indeed, keys as images can be very convenient when scanning for quick recognition and identification with a modified barcode scanner. For human practice, such methods will be convenient for identification at borders, passing entrances to airports and banks, in medical institutions, identification of the remains of dead people, etc.

Concerning the characteristics of genetic code fragment that should be chosen to perform the functions of the "key" (its length, sequence of nucleotides, etc.) — the developer of specific DB makes a decision regarding their selection. His skill determines the success of the creation and application of each specific code. Consequently, it determines successful functioning of the created DB. The basis of the choice is the specifics of particular DB, characteristics of organism, the functions for which the DB is created and other factors.

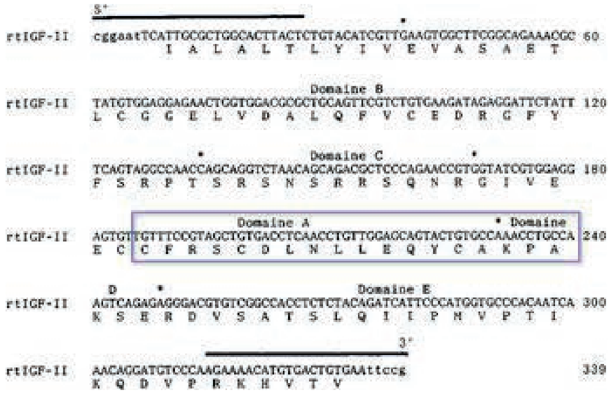
So, all code combinations represented in Fig. 6, can serve as keys in the process of DBs construction (see the text above). Of course, in DB with medical information about a person, such keys can perform the same role. Figure 7 demonstrates an example from real medical practice. This is a small fragment of a copy of the table with information about patients who were exposed to radiation during the accident at the Chernobyl accident (1986) and who obtained medical treatment or rehabilitation at the Elbrus Medical and Biological Station (EMBS) of the National Academy of Sciences of Ukraine. Patients' data were registered and processed in standard accepted records. However, during the processing of this data, developed methods of key formation were applied. Let us demonstrate how it works in such example. The method of generating the key for table 1 (personal data of patients) is usual — passport numbers are used there. Instead, in Table 2 with the patients' examination data, there is a "primary key" column, where a fragment of the genetic sequence can be inserted in the form of a code. But, according to the rules of DBs construction, the keys in both tables have to be connected by links [18, 19]. In this case, data about specific patient (or other person) will be obtained from this DB absolutely guaranteed. In addition, all the advantages of the methods we wrote about above (protection of information, high probability of searching for information in information systems, etc.) become real.

Generalized information on the construction, functioning, and application of relational databases (DBs) for biological objects, together with detailed information on innovative methods of building such DBs, has been presented. These innovations have been aimed at improving the overall quality of biomedical databases and making them more suitable and user-friendly for professionals in biology and medicine.

Conventional methods [18—19] rely on standard-type keys, typically generated as random numerical combinations, to ensure data integrity in technical systems. However, this approach has not always been adequate in modern practice. An unsuccessful choice of code may compromise data integrity, disrupt information flows, and reduce system reliability. For example, in a biomedical information system for environmental monitoring, integrity violations may lead to system malfunctions and, ultimately, a decline in the quality of monitoring chemical impacts on the environment.

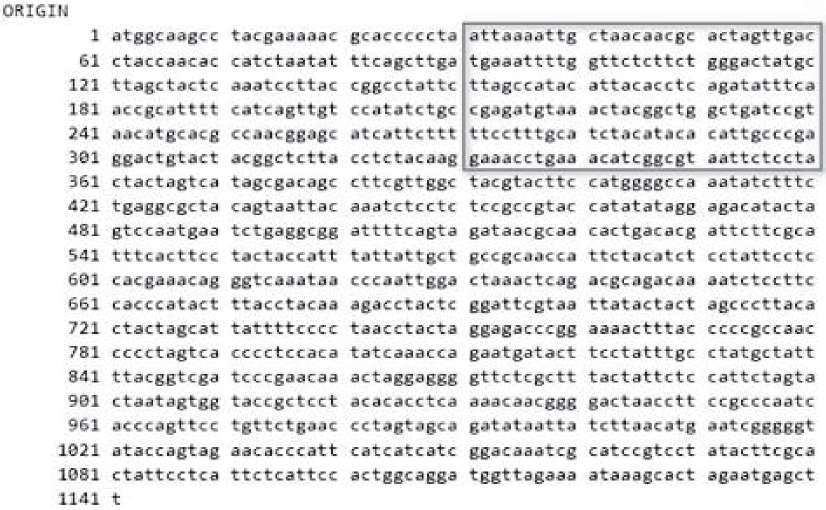
To address these challenges, we have invented and developed three fundamentally new methods of key encoding for biomedical databases, all based on the principles of genetic codes. Five patents of Ukraine have been obtained for these works [25—29]. These methods have improved the per-

```
ORIGIN
        1 tagatctttt gtgtctgcgt tctctggggt ttatcagagg tgctgggtga agtgggagag
       61 taggggagga gggtcagcag aggaacttca cagggcggtc acataaaaac tgcctggaga
      121 actggatgcg ggatcagacg gtgatgtcct ccgtgctcga ctattcctgg agctacagag
      181 atccgcgttc gcatgggact cagttggttt cttcaacgtg atgcggaact aatgatctaa
      241 aaattcgctt gaggattgga cttttgctgt gtgaggtgta ctcctttgaa aactccgttt
      301 taccgtcttc tcatccagga ataccgtgcg catctctacc aatggtaagc tggcatgttg
      361 tgtttaagtc agcaggcacg gtactttcat acgttcataa ccagaagtga tttatgaccg
      421 tatatgaggt ttgcgcacgt atttagatcc agtgacacgc gtttttgcca tttgtgcgta
      481 aactgattta ttttttattt ttttctctgc aggcggcgcg actaactaga acgctcccat
      541 tatattccac aaagctgtct tcacgtctac agacgcaatt gttataatca acgaagctgt
      601 acgtgtgcac cgaaaacggg agcgctctgc tttgtcgcgt agctaaagaa ttgcactggg
      661 acccggggtg ttgggaatgt tacacaggat tgaatatctt tgcttaccgg atagcgtagg
      721 ggaaaggttt tttttttttt tttttttttt tgtttttgca tgtttagaac caccaacgcg
      781 ctcaggcacc tgaccgcgcg ctcaccgggt ggatgaaagc tttcacatct gctcgtaacg
      841 tacgagtatg gcttttcatt caaacagaat tgacatttct gtttaaatga tgttttaatg
      901 tggtggctaa gtacagtttt acttgctgca tgtggtctaa ccacgaaggt ggtatattta
      961 ctttgtctag tatcaagcaa cgtatcaaga ttcactttta tatgagtttt tggaatctac
     1021 aagtgttccg aatcactggt tgcgaattca attctaacgc attccttaat caagcataca
     1081 ctggtttacg cagcgtttac gcatgacaaa tgttaagtta gcgtggaaaa caatagatgt
     1141 gtccttcatt tgatttaata aattaaggaa aaccggtctc cagtgtgcaa ctctagtgta
     1201 catttgacct ccacgaatgt tttgaagagt catactttgt cttaaatgag ttgtagtgaa
     1261 gagtgatggc gctcagtcga gcataagaac acgtctgaaa ggttttatga ggtgcattat
     1321 gacataaaga gacataaccc gagagtcact ttcgcaggga tgtttatctt tagttcaaaa
     1381 acgaaaaaaa atttttttaat tttttttat tacaaaatat attttttttag ttgttaaaaa
     1441 tttcaagagt tgtttttatt tagttagttg attgttgatt atagatagat gttttttttt
     1501 ggactagagc agaggtcttg agcctttttc aggccaagaa tgtggttctt gatggatgct
     1561 tatagtacta taaatctcta tttgcatgct ttgttatgac atttgaaagc tttaacaatc
```

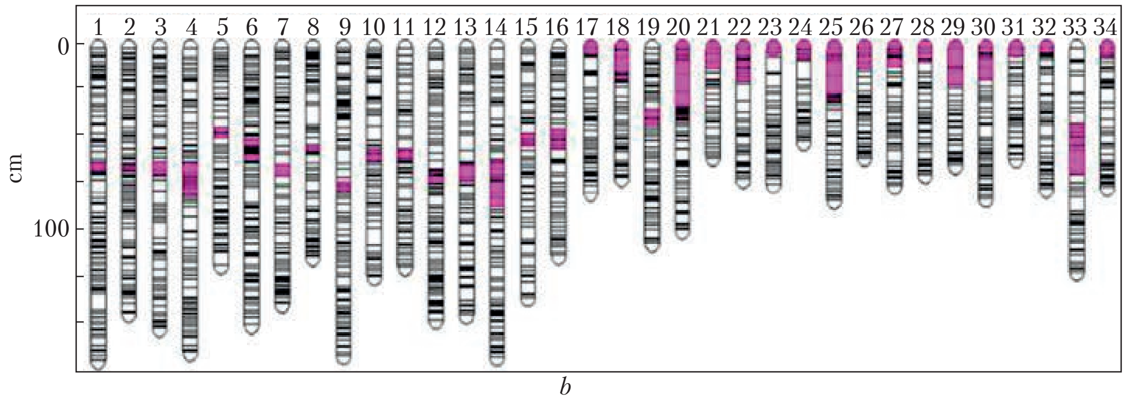


```
            5'
                                                    *
rtIGF-II    cggaatTCATTGCGCTGGCACTTACTCTGTACATCGTTGAAGTGGCTTCGGCAGAAACGC 60
                   I  A  L  A  L  T  L  Y  I  V  E  V  A  S  A  E  T

                                            Domaine B
rtIGF-II    TATGTGGAGGAGAACTGGTGGACGCGCTGCAGTTCGTCTGTGAAGATAGAGGATTCTATT 120
            L  C  G  G  E  L  V  D  A  L  Q  F  V  C  E  D  R  G  F  Y

                     *                      Domaine C           *
rtIGF-II    TCAGTAGGCCAACCAGCAGGTCTAACAGCAGACGCTCCCAGAACCGTGGTATCGTGGAGG 180
            F  S  R  P  T  S  R  S  N  S  R  R  S  Q  N  R  G  I  V  E

                                 Domaine A                    * Domaine
rtIGF-II    AGTGTTGTTTCCGTAGCTGTGACCTCAACCTGTTGGAGCAGTACTGTGCCAAACCTGCCA 240
            E  C  C  F  R  S  C  D  L  N  L  L  E  Q  Y  C  A  K  P  A

              D     *                           Domaine E
rtIGF-II    AGTCAGAGAGGGACGTGTCGGCCACCTCTCTACAGATCATTCCCATGGTGCCCACAATCA 300
            K  S  E  R  D  V  S  A  T  S  L  Q  I  I  P  M  V  P  T  I

                                                      3'
rtIGF-II    AACAGGATGTCCCAAGAAAACATGTGACTGTGAAttccg                      339
            K  Q  D  V  P  R  K  H  V  T  V
```

a



```
ORIGIN
        1 atggcaagcc tacgaaaaac gcacccccta attaaaattg ctaacaacgc actagttgac
       61 ctaccaacac catctaatat ttcagcttga tgaaattttg gttctcttct gggactatgc
      121 ttagctactc aaatccttac cggcctattc ttagccatac attacacctc agatatttca
      181 accgcatttt catcagttgt ccatatctgc cgagatgtaa actacggctg gctgatccgt
      241 aacatgcacg ccaacggagc atcattcttt ttcctttgca tctacataca cattgcccga
      301 ggactgtact acggctctta cctctacaag gaaacctgaa acatcggcgt aattctccta
      361 ctactagtca tagcgacagc cttcgttggc tacgtacttc catggggcca aatatctttc
      421 tgaggcgcta cagtaattac aaatctcctc tccgccgtac catatatagg agacatacta
      481 gtccaatgaa tctgaggcgg attttcagta gataacgcaa cactgacacg attcttcgca
      541 tttcacttcc tactaccatt tattattgct gccgcaacca ttctacatct cctattcctc
      601 cacgaaacag ggtcaaataa cccaattgga ctaaactcag acgcagacaa aatctccttc
      661 cacccatact ttacctacaa agacctactc ggattcgtaa ttatactact agcccttaca
      721 ctactagcat tattttcccc taacctacta ggagacccgg aaaactttac ccccgccaac
      781 cccctagtca cccctccaca tatcaaacca gaatgatact tcctatttgc ctatgctatt
      841 ttacggtcga tcccgaacaa actaggaggg gttctcgctt tactattctc cattctagta
      901 ctaatagtgg taccgctcct acacacctca aaacaacggg gactaacctt ccgcccaatc
      961 acccagttcc tgttctgaac cctagtagca gatataatta tcttaacatg aatcgggggt
     1021 ataccagtag aacacccatt catcatcatc ggacaaatcg catccgtcct atacttcgca
     1081 ctattcctca ttctcattcc actggcagga tggttagaaa ataaagcact agaatgagct
     1141 t
```



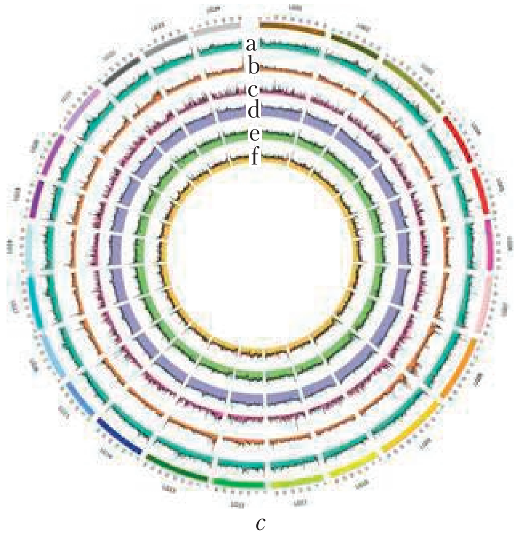


Fig. 6. Three methods of keys formation on the base of genetic codes and three examples of the fragments of genetic codes of various fishes' species that can serve as keys for such DBs construction (as well as primary keys), like "GenCodeBioOrg*" from our example [54–56]: *a* – Genetic code is represented in form of letter and number codes (fragment of genetic code *Cyprinus carpio* L.(common carp)); *b* – Genetic codes are represented in forms of images of the letter and number codes, locations of images are randomly picked up, they are surrounded by rectangular frames with dimensions (KxM): *Oncorhynchus mykiss* Walbaum (rainbow trout, above), and *Barbus borysthenicus* Dybowski (below) [54–56]; *c* – Two different images of the fragments of genetic codes that can be used as keys in the DBs [54–56].
Fragment of code of *Onchorincus mykiss* Walbaum (rainbow trout, above); and *Hypophthalmichthys molitrix,* Valenciennes (white carp, below)

formance characteristics of database keys (including primary keys), yielding superior results compared to prototype methods. Their application has ensured a higher level of data protection and greater individualization of data.

The three proposed methods generate keys in the following forms:

1. Alphanumeric sequences derived from fragments of genetic codes;
2. Nucleotide sequences expressed in alphanumeric form but processed as images (using a modified barcode scanner);
3. Images derived from laboratory studies of DNA and RNA structures.

An algorithm for creating novel DBs with biological material has been developed. It can be applied in biological and ecological research (e.g., databases of living organisms and identifiers), in medicine, and in other related fields. The novelty lies in employing genetic code–based keys for relational table linkages. These methods have demonstrated significant advantages and have shown both theoretical and practical value. In addition to biomedical and ecological contexts, they can also be applied in criminology, law enforcement, and military practice. Their importance has dramatically increased in wartime due to urgent needs in remains identification and other forms of biological identification.

Information protection. A crucial advantage of these methods is their contribution to information security. The protection of sensitive data in biomedical information systems has been realized at a qualitatively new level when any of the three

Table 1 Key 1

№	Passport number	Name	Surname	Date of birth	...
1	CH 158***	Sergiy	Osipenko	12.01.1997	...
2	AH 234***	Iryna	Koval	01.10.1978	...
...	...	...	...	...	...

Table 2 Key 2

№	Fragment of genetic code (fragment of nucleotide sequence)	Height (sm)	Weight (kg)	Trombocytes 10^9 /|	Leycocytes 10^9 /|	...
1	...	142	51	205	4.78	...
2	...	151	54	176	4.52	...
...	...	...	...	...	...	...

Fig. 7. Example of hypothetic DB with information about humans where keys were formed on the base of passport numbers (Table 1) and fragments of genetic codes (Table 2). Data for this DB were taken from archive publications of patients, who obtained medical treatment and rehabilitation after Chernobyl accident (1986) at Elbrus Medical and Biological Station (EMBS) — scientific base of the National Academy of Sciences of Ukraine in Caucasus Mountains [51]

methods are applied. Such codes have proven extremely stable and virtually impossible to compromise, thereby ensuring reliable protection of personal and biomedical data. They have also enabled more secure authorization and authentication processes. Importantly, these methods have provided new opportunities for encoding patient information and biological data, ensuring a high degree of individualization and substantially improving the efficiency and reliability of data retrieval in information systems.

Grounded in contemporary concepts from bioinformation systems theory, genetics, and molecular biophysics, it can be stated that the confidentiality and security levels of biomedical databases with such keys are exceptionally high, making unauthorized access practically impossible. Additionally, the methods have allowed for significant memory savings, a critical factor, given the vast storage requirements of biomedical databases.

These results have opened broad prospects for further research and practical application of the described methods.

Acknowledgements and funding sources. The study was funded under the themes according the State registration 0107U002666 and № 177-X04 (1.06.2004).

Conflict of interest. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethical standards and human rights. This article does not contain any studies with human subjects performed by any of the authors. Work was done in compliance with ethical standards

Animal studies. This article does not include animal studies. All biological data were taken from scientific literature and Internet.

REFERENCES

1. Schnase, J. L., Cushing, J., Frame, M., Frondorf A., Landis E., Maier D., Silberschatz A. (2003). Information technology challenges of biodiversity and ecosystems informatics. *Information Systems,* 28(4), 339—345. https://doi.org/10.1016/S0306-4379(02)00070-4
2. Klyuchko O. M. Information and computer technologies in biology and medicine. (2008). Kyiv. 252 p. [in Ukrainian].
3. Bor-Sen Chen, Shih-Kuang Yang, Chung-Yu Lan, Yung-Jen Chuang. (2008). A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. *Medical Genomics,* 1, 46. https://doi.org/10.1186/1755-8794-1-46
4. Pornputtapong, N., Wanichthanarak, K., Nilsson, A., Nookaew, I., Nielsen, J. (2014). A dedicated database system for handling multi-level data in systems biology. *Source Code for Biology and Medicine,* 9, 17. https://doi.org/10.1186/1751-0473-9-17.
5. Stobbe, M. D., Swertz, M. A., Thiele, I., Rengaw, T., van Kampen, A. H. C., Moerland, P. D. (2013). Consensus and conflict cards for metabolic pathway databases. *Systems Biology,* 7, 50. https://doi.org/10.1186/1752-0509-7-50
6. van Ommen, B., Bouwman, J., Dragsted, L. O., Drevon, C. A., Elliott, R., de Groot, P., Kaput, J., Mathers, J. C., ..., Wopereis, S. (2010). Challenges of molecular nutrition research 6: the nutritional phenotype database to store, share and evaluate nutritional systems biology studies. *Genes & Nutrition,* 5, 167. https://doi.org/10.1007/s12263-010-0167-9
7. Momin, A. A., James, B. P., Motter, T. C., Kadara, H. N., Powis, G., Wistuba, I. I. (2011). Integrating whole transcriptome sequence data and public databases for analysis of somatic mutations in tumors. *Genome Biology,* 12(1), 44. https://doi.org/10.1186/gb-2011-12-s1-p44
8. Stobbe, M. D., Houten, S. M., Jansen, G. A., van Kampen, A. H. C., Moerland, P. D. (2011). Critical assessment of human metabolic pathway databases: a stepping stone for future integration. *Systems Biology*, 5, 165. https://doi.org/10.1186/1752-0509-5-165
9. Chowbina, S. R., Wu, X., Zhang, F., Li, P. M., Pandey, R., Kasamsetty, H. N., Chen, J. Y. (2009). HPD: an online integrated human pathway database enabling systems biology studies. *Bioinformatics*, 10(11), S5. https://doi.org/10.1186/1471-2105-10-S11-S5
10. Goldstein, A. M. (2010). The NCBI Databases: an Evolutionist's Perspective. *Evolution: Education and Outreach*, 3, 258. https://doi.org/10.1007/s12052-010-0258-5.
11. Albà, M. (2000). Links to molecular biology databases. *Genome Biology*, 1—4 https://doi.org/10.1186/gb-2000-1-1-reports235
12. Walsh, J. R., Sen, T. Z., Dickerson, J. A. (2014). A computational platform to maintain and migrate manual functional annotations for BioCyc databases. *Systems Biology*, 8, 115.
13. Del Rio, A., Barbosa, A. J. M., Caporuscio, F. (2011). Use of large multiconformational databases with structure-based pharmacophore models for fast screening of commercial compound collections. *Journal of Cheminformatics,* 3(1), P27. https://doi.org/10.1186/1758-2946-3-S1-P27
14. Maier, C. V., Long, J. G., Hemminger, B. M., Giddings, M. C. (2009). Ultra-Structure database design methodology for managing systems biology data and analyses. *Bioinformatics,* 10, 254. https://doi.org/10.1186/1471-2105-10-254
15. Tan, T. W., Xie, C., De Silva, M., Kuan Siong, Lim, Patro, C. P. K., Lim, S. J., Govindarajan, K. R., Joo Chuan, Tong, Khar Heng, Choo, Ranganathan, S., Khan, A. M. (2013). Simple re-instantiation of small databases using cloud computing. *Genomics*, 14(5), 5—13. https://doi.org/doi: 10.1186/1471-2164-14-S5-S13
16. Bouzaglo, D., Chasida, I., Tsur, E. E. (2018). Distributed retrieval engine for the development of cloud-deployed biological databases. *BioData Mining*, 11, 26. https://doi.org/10.1186/s13040-018-0185-5
17. Eronen, L., Toivonen, H. (2012). Biomine: predicting links between biological entities using network models of heterogeneous databases. *Bioinformatics,* 13, 119. https://doi.org/10.1186/1471-2105-13-119
18. Microsoft Academy: Methods and means of software engineering. URL: https://www.intuit.ru/studies/courses/2190/237/lecture/6124 (Last accessed: 24.12.2021).
19. Harrington, Jan L. (2005). *Object-oriented database design clearly explained*. USA: Academic Press.
20. Klyuchko, O. M. (2019). Creation of biological databases: using object-oriented system analysis. *Biotechnologia Acta*, 12(3), 5—23. https://doi.org/10.15407/biotech12.03.005
21. Klyuchko, O. M., Klyuchko, Z. F. (2018). Electronic databases of *Arthropods*: methods and applications. *Biotechnologia Acta*, 11(4), 28—49 https://doi.org/10.15407/biotech11.04.028
22. Klyuchko, O. M., Buchatsky, L. P., Rud, Yu. P., Melezhyk, O. V. (2019). Creation of fish databases for electronic interactive map: tables and keys. *Ribogospod. nauka Ukr.*, 50(4), 37—57. https://doi.org/10.15407/fsu2019.04.037
23. Klyuchko, O. M., Buchatsky, L. P., Melezhyk, O. V. (2019). Fish information databases construction: data preparation and object-oriented system analysis. *Ribogospod. nauka Ukr.,* 49(3), 32—47. https://doi.org/10.15407/fsu2019.03.032

24. Klyuchko, O. M., Beletsky, A. Ya., Melezhyk, O. V., Gonchar, O. A. (2024). Innovative approaches to the construction of bioinformation systems with databases: keys based on genetic codes. *Science and Innovation*, 20(4), 33—48.
25. *Patent UA 143919 U*. Klyuchko, O. M., Biletsky, A. Ya., Lizunova, A. G. Method of application of monitoring system with databases and keys in symbolic records of genetic codes of fish and other aquatic organisms [in Ukrainian].
26. *Patent UA 143918 U*. Klyuchko, O. M., Biletsky, A. Ya., Lizunova, A. G. Method of applying monitoring system with databases and keys in form of images of genetic codes of fish and other aquatic organisms [in Ukrainian].
27. *Patent UA 143926 U*. Klyuchko, O. M., Biletsky, A. Ya., Lizunova, A. G. Method of use of monitoring system with databases and keys in form of images of genetic codes of biological organisms [in Ukrainian].
28. *Patent UA 155203 U*. Klyuchko, O. M., Biletsky, A. Ya., Lizunova, A. G. Method of application of information monitoring system with databases and keys in symbolic records of genetic codes of biological organisms [in Ukrainian].
29. *Patent UA* 157032 *U*. Klyuchko, O. M., Biletsky, A. Ya., Lizunov, G. V., Lizunova, A. G. Method of application of monitoring system with databases and keys in the form of images of sets of symbols of biological organisms' genetic codes [in Ukrainian].
30. Van der Laan, R., Eschmeyer, W. N., Fricke, R. (2014). Family-group names of Recent fishes**.** *Zootaxa Monograph*., 3882(1), 1—230. https://doi.org/10.11646/zootaxa.3882.1.1
31. Movchan, Yu. V. (2009). Fishes of Ukraine (taxonomy, nomenclature, remarks). *Collection of works of Zoological Museum,* 40, 47—87.
32. Fricke, R., Eschmeyer, W. N., R. van der Laan. Eschmeyer's catalog of fishes: genera, species, 2019. URL: http://esearcharchive.calacademy.org/research/ichthyology/catalog (Last accessed: 27.12.2021).
33. Di Génova, A. D., Aravena, A., Zapata, L., González, M., Maass, A., Iturra, L. (2011). SalmonDB: a bioinformatics resource for *Salmo salar* and *Oncorhynchus mykiss*. *Database (Oxford),* bar050. https://doi.org/10.1093/database/bar050
34. *Red Book of Ukraine*. Fauna. (2009). Kyiv. [in Ukrainian].
35. Klyuchko, Z. F., Plyushch, I. G., Sheshurak, P. N. (2001). *Annotated catalogue of noctuids (Lepidoptera, Noctuidae) of the fauna of Ukraine*. Kyiv.
36. Klyuchko, Z. F. *Noctuidae of Ukraine*. (2006) Kyiv. 248 p. [in Ukrainian].
37. Klyuchko, Z. F. (2002). Survey of moths (Lepidoptera: Noctuidae) of steppe reserves of Ukraine. *Proceedings of Kharkov Entomological Society*, 9(1—2), 114—122.
38. Klyuchko, Z. F., Kullberg, J. (2006). To the study of the noctuid fauna (Lepidoptera: Noctuidae sl) of the Ukrainian Carpathians. *Eversmannia,* 7—8, 69—74. [in Russian].
39. Klyuchko, O. M., Melezhyk, O. V. (2024). Innovative databases in ecomonitoring information systems: images of genetic codes as keys. *Journal of Applied Interdisciplinary Research (JAIR),* Special Issue, 35—49. https://doi.org/10.25929/re84ry96
40. Klyuchko, Z. F., Budashkin, J. I., Gerasimov, R. P. (2004). New and little-known species of noctuids (Lepidoptera) of Ukraine's fauna. *Bulletin of Zoology*, 38(1), 94.
41. Yukalo, V. G., Storozh, L. A., Datsyshyn, K. Ye., Krupa, O. M. (2018). Electrophoretic systems for the preparative fractionation of protein precursors of bioactive peptides from cow's milk. *Food science and technology*, 12(2), 26—32. https://doi.org/10.15673/fst.v12i2.932
42. Yukalo, V., Datsyshyn, K., Krupa, O., Pavlistova, N. (2019). Obtaining of β-LG, α-LA and BSA protein fractions from milk whey. *Ukrainian Food Journal,* 8(4), 788—798. https://doi.org/10.24263/2304-974X-2019-8-4-10
43. Franchuk, G. M., Isaenko, V. M. (2005). *Ecology, aviation and cosmos*. Kyiv. [in Ukrainian].
44. *Patent UA 134575 U*. Klyuchko O.M. Method for monitoring of chemicals influence on bioorganisms in few time intervals [in Ukrainian].
45. Brazhko, O. A., Omelyanchyk, L. O., Zavhorodniy, M. P., Martynovskyi, O. O. (2013). *Chemistry and biological activity of 2(4)-thioquinolines and 9-thioacridines*. Zaporizhzhia. 241 p. [in Ukrainian].
46. Tkachuk, R., Tkachuk, A., Stadnik, D., Yanenko, O. (2021). Ensuring high-precision testing of implants in the regulation of intra-eye pressure. *Intl. Conf. Advanced Applied Energy and Information Technologies (15—17 December 2021, Ternopil, Ukraine),* 157—161. Ternopil.
47. Arnot, J. A., Mackay, D., Parkerton, T. F., Bonnell, M. (2008). A database of fish biotransformation rates for organic chemicals. *Environmental Toxicology and Chemistry*, 27(11), 2263—2270. https://doi.org/10.1897/08-058.1
48. Tedesco, P. A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., ..., Oberdorff, T. (2017). A global database on freshwater fish species occurrence in drainage basins. *Sci. Data*, 4, 170141. https://doi.org/10.1038/sdata.2017.141
49. Daoliang, Li, Zetian, Fu, Yanqing, Duan. (2002). Fish-Expert: a web-based expert system for fish disease diagnosis. *Expert Systems with Applications,* 23, 311—320. https://doi.org/ 10.1016/S0957-4174(02)00050-7
50. Froese, R., Pauli, D. (2000). *FishBase 2000: Concepts, designs and data sources*. ICLARM. Los Banos, Philippines.
51. Beloshitsky, P. V., Baraboy, V. A., Krasyuk, A. N., Korkach, V. I., Torbin, V. F. (1996). *Post-radiation rehabilitation in mountain conditions. ICAMER*. Kyiv.

52. Human gene databases. URL: https://www.genecards.org/ (Last accessed: 23.04.2024).
53. Human genetic research databases. URL: https://www.alrc.gov.au/publication/essentially-yours-the-protection-of-human-genetic-information-in-australia-alrc-report-96/18-human-genetic-research-databases/what-are-human-genetic-research-databases/ (Last accessed: 23.04.2024).
54. Cyprinus carpio isolate SPL01 chromosome A17, ASM1834038v1, whole genome shotgun sequence. URL: https://www.ncbi.nlm.nih.gov/nuccore/NC_056588.1?from=25191087&to=25198977&report=genbank&strand=true (Last accessed: 23.01.2021).
55. Barbus borysthenicus isolate PK-977 cytochrome b (cytb) gene, partial cds; mitochondrial gene for mitochondrial product. URL: https://www.ncbi.nlm.nih.gov/nuccore/AY331026.1 (Last accessed: 23.01.2021).
56. User Reference for Fisheries Improvement Projects Database (FIP-DB) and Query Viewer. URL: https://ru.scribd.com/document/385739269/Readme-File-for-FIP-DB#download (Last accessed: 27.12.2021).

Received 01.09.2024
Revised 12.06.2025
Accepted 22.07.2025

О.М. Ключко [1] (https://orcid.org/0000-0003-4982-7490),
А.Я. Білецький [1] (https://orcid.org/0000-0002-3798-8150),
О.В. Мележик [2] (https://orcid.org/0000-0003-3882-7102),
О.О. Гончар [3] (https://orcid.org/0000-0003-4134-0354)

[1] Національний авіаційний університет,
просп. Л. Гузара, 1, Київ, 03058, Україна,
+380 44 406 7130, post@kai.edu.ua
[2] Відкритий міжнародний університет розвитку людини «Україна»,
вул. Львівська, 23, Київ, 03115, Україна,
+380 67 328 2822, office@uu.edu.ua
[3] Інститут фізіології ім. Богомольця НАН України,
вул. Богомольця, 4, Київ, 01024, Україна,
+380 44 256 2400, science@biph.kiev.ua

БІОЛОГІЧНІ БАЗИ ДАНИХ: ТРИ НОВИХ МЕТОДИ ФОРМУВАННЯ КЛЮЧІВ З ГЕНЕТИЧНИХ КОДІВ

Вступ. Створення сучасних баз даних (БД) з біомедичною інформацією на основі новітніх технологій є важливим завданням.

Проблематика. Актуальною є розробка інноваційних підходів до конструювання БД біомедичної інформації з ключами з розширеними можливостями, зокрема й реляційних БД з ключами, в основу кодування яких покладено генетичні коди організмів, біологічних матеріалів.

Мета. Опис і характеристика трьох новітніх авторських методів формування (кодування) ключів для різних груп біологічних об'єктів при створенні інноваційних БД з біомедичною інформацією.

Матеріали і методи. Застосовано методи аналізу зображень, компаративного аналізу, об'єктно-орієнтованого системного аналізу для оптимальної побудови БД, методи конструювання БД і проєктування *ER*-діаграм.

Результати. На прикладах розробки реляційних БД з інформацією про деякі біологічні види охарактеризовано застосування методів об'єктно-орієнтованого аналізу для конструювання БД, а також описано алгоритм їх конструювання. Особливу увагу приділено вирішенню проблеми створення ключів на основі генетичних кодів цих видів: у буквено-цифровому вираженні, зображень та фрагментів зображень буквено-цифрового Пропоновані способи формування кодів особливо важливі як первинні ключі, що забезпечують зв'язки між окремими таблицями БД, забезпечують цілісність інформації у системі, надійність доступу до даних. Високий рівень організації даних, їх отримання при застосуванні ключів на основі генетичних кодів детально обґрунтовано.

Висновки. Результати можна застосувати у інформаційних системах з біомедичною інформацією, для подальшого розвитку методів побудови БД, вдосконалення деяких методів захисту медико-біологічних та персональних даних. БД з такими ключами можуть бути ефективними для упорядкування великих масивів даних про різні контингенти людей, біологічні організми і матеріали та здатні забезпечити найбільш швидкий доступ до відповідної інформації й надійний захист такої інформації.

Ключові слова: інформаційні технології, бази біологічних даних, ключ, первинний ключ, кодування, біологічний вид.