



GeneDBase – Genetic Database of Selected Species of Mammals in the Czech Republic

Luděk Skočovský, Tomáš Šikula, Martin Ernst, David Kouřil

► **To cite this version:**

Luděk Skočovský, Tomáš Šikula, Martin Ernst, David Kouřil. GeneDBase – Genetic Database of Selected Species of Mammals in the Czech Republic. Jiří Hřebíček; Gerald Schimak; Miroslav Kubásek; Andrea E. Rizzoli. 10th International Symposium on Environmental Software Systems (ISESS), Oct 2013, Neusiedl am See, Austria. Springer, IFIP Advances in Information and Communication Technology, AICT-413, pp.661-668, 2013, Environmental Software Systems. Fostering Information Sharing. <10.1007/978-3-642-41151-9_62>. <hal-01457530>

HAL Id: hal-01457530

<https://hal.inria.fr/hal-01457530>

Submitted on 6 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

GeneDBase – Genetic Database of Selected Species of Mammals in the Czech Republic

Luděk Skočovský¹, Tomáš Šikula², Martin Ernst³, and David Kouřil²

¹ freelance programme writer, Brno, Czech Republic
ludek@skocovsky.cz

² HBH Projekt spol. s r.o., Brno, Czech Republic
{t.sikula, d.kouril}@hbh.cz

³ Mendel University, Brno, Czech Republic
martin.ernst@mendelu.cz

Abstract. GeneDBase is a part of a research project titled “Creating a genetic database of selected species of mammals in the Czech Republic to be used for sustainable transport development”. The aim of the research project is to assess the impact of roads on the fragmentation of animal populations and their genetic variability. Further, the objective is to create a genetic database of mammals comprising the data from collected samples of selected wild animals which could be used widely, not only for road planning, but also in other fields (agriculture, environmental protection, land-use planning).

Keywords: GeneDBase, migration, genetic variability, landscape fragmentation, genetic isolation, genetic database, database application, sustainable road development

1 Introduction

Wildlife migration preserves balance among populations, enables exchange of genetic information and the use of food sources and maintains genetic variability. Due to migration, populations are able to adapt to changes in their environment. The development of transport infrastructure leads to landscape fragmentation and decreases the migration permeability. A grant project was established to assess the impact of landscape fragmentation onto genetic variability of populations. A part of the grant project aims to create a database application called GeneDBase.

2 About the project

GeneDBase is a name for a database application, which is one of the outcomes of a grant project named: “Creating a genetic database of selected species of mammals in the Czech Republic to be used for sustainable transport development”. The grant project is supported by the Technology Agency of the Czech Republic (within Alfa Programme). The contractor is a private road design company, HBH Projekt spol. s r. o.,

together with a project participant, Mendel University in Brno, Department of Forest Protection and Wildlife Management.

The aim of the grant project is to create and evaluate a genetic database, which will contribute to a more objective and more complex assessment of the barrier effect of the transport infrastructure on wildlife populations, especially on their genetic variability. To reach the aim, the following steps were defined:

- establishing a DNA-bank with genetic samples (especially samples of tissue, hair bulb) of selected mammals;
- creation of a database with records of collected samples (not only results of genetic analyses, but also other entries will be stored);
- creation of a database application to view and edit the database using the internet.

Expected outcomes of the grant projects are:

- Evaluation of a landscape fragmentation impact on genetic variability of populations in selected areas, also concerning an isolation level of populations.
- Universal database application which can be continuously edited and developed.
- Certified methodology to maintain and develop the database.
- Certified methodology for the landscape fragmentation assessment in relation to genetic factors, in order to supplement existing methodologies used to assess decreasing landscape fragmentation level.

The grant project was launched in January 2012 and will be completed in December 2015. The completion of subsequent issues is designed/planned as follows: The following works in designated years are expected:

2012 – analysis of the current status, literature research;

2013 – sample collection and processing, complementing research, database design;

2014 – data analysis, database application;

2015 – testing of the designed genetic database and completion of the methodology.

At the beginning of the grant project, areas for collecting samples were chosen. The key factors in the selection process included the presence of distinctive migration barriers (roads) where the migration routes disappeared. The selection of the area was also based on the previous studies, and on the knowledge of the area. The localities for the collection of the samples are defined by hunting districts. The sample collection is performed in cooperation with local game keepers. The sampling will be concentrated on following species: Eurasian elk (*Alces alces*), red deer (*Cervus elaphus*), sika deer (*Cervus nippon*), roe deer (*Capreolus capreolus*), wild boar (*Sus scrofa*), European hare (*Lepus europaeus*), European badger (*Meles meles*) and red fox (*Vulpes vulpes*). Biological samples (tissue, skin, hair bulb, droppings) are stored in a DNA-bank deep frozen within Mendel University in Brno. An analysis of selected microsatellites will be performed and genetic indicators will be calculated, based on the results of the analysis.

The indicators include: inbreeding coefficient (indicates the degree of genetic relatedness within the population, the higher the value is, the greater risk of population

destabilization may be), genetic variability, heterozygosity, polymorphic information content (the higher the values are, the greater is the presumption of stability, adaptability or reproduction of the population). The above mentioned genetic indicators refer to certain condition of the population. Isolation of populations of the same species can increase the inbreeding coefficient, decrease genetic variability and gradually deprive the gene pool together with vanishing of certain alleles. If the reliability of individual identification using the microsatellite analysis is 99,99 %, the results can be used for migration studies (droppings or hair bulb need to be used for sampling).

3 Software solution

A database application called GeneDBase is being prepared for storing and handling all the data gathered during the project. The author of the software is a freelance program writer Luděk Skočovský. The database application should meet the following requirements:

- the database will record selected genetic properties of selected species of mammals
- the database will be based on DNA-bank, i.e. collected genetic samples of individuals and their genetic analysis, i.e. the record of microsatellites and the calculated indicators of genetic variability
- accompanying visual material will be part of the database (not just individuals themselves, but also elements related to the registration of the individual and its properties that are important for the evaluation)
- the database will enable export/import of data to/from different formats for their processing by third-party programs (e.g. for the calculation of genetic variability), or for their use in mapping servers to visualize the movement of registered individuals, or for share data with other databases of similar nature
- the database must be generally usable, so that the user will be able to add another species and attributes associated with them
- the content of the database must be professionally guaranteed and at the same time accessible to the public

Based on set requirements, it is clear that a suitable database should mainly provide professionally guaranteed storing of described genetic sample records and their analyses.

Further, as stated above, another important parameter of the database is the flexibility and connectivity to external data sources. The database should enable to edit e.g. new species (flexibility in terms of data structures), but also to add new user interfaces defined on the basis of newly formulated requirements on the data viewing or handling).

With respect to the status and availability of current technologies, the suitable and functional database should be implemented so that it is visible on the Internet, while visibility means availability of all its functions including editing and data management. Using the Internet it should be possible to export and import the data or to share data using inter-database protocols.

3.1 Data model

The proposal of a data model is based on the current needs which have been defined in the previous chapter. It is the evidence of:

- selected species;
- individuals of selected species, and their relationships;
- collected samples for genetic analyses;
- results of genetic analyses (the evidence of microsatellites and the calculated genetic variability indicators);
- photographs and related documentation;
- database users and their access rights;
- editing history.

3.2 Conceptual schema

A conceptual schema of the data model is given in Fig. 1.

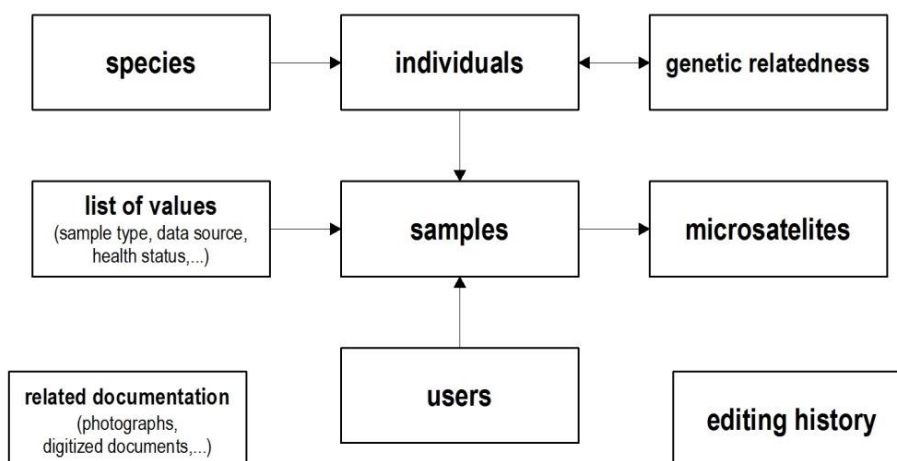


Fig. 1. GeneDBase – conceptual schema.

The basic platform of the database is **the table of samples**, which is shown in the center of the diagram. Each sample belongs to a particular individual (see Fig. 1). The table of samples and **the table of individuals** are separated and therefore it is possible to register several samples from one individual without redundancy of data. Separate tables are used to register the species and relatedness of each individual.

Samples, other than direct numerical or text data, also include given options, i.e. references to so called lists of values. For example, the list of value of a sample type includes the following options: (1. hair bulb, 2. tissue, 3. inner tissue, 4. bone, 5.dropping, 6. other)

The table of samples records all important data regarding collected sample so that the sample can be related to the specific individual that was at the time of sample collection present at a particular location and was in a particular status. Therefore, the table includes:

- the method of sample collection (reference to the list of values);
- sample type (reference to the list of values);
- location of sample collection (GPS coordinates);
- health status of the individual;
- weight of the individual;
- age of the individual;
- morphological and colour variations of the individual.

The table of samples is connected with the table of microsattelites so that each sample can be assigned with more microsattelites.

The table of microsattelites contains genetic information on the uniqueness of the individual and also the genetic variability indicators.

The table of individuals, the table of species, the table of samples and the table of microsattelites can be connected with the table of photographs and digitized documents in the way, that each record in these tables can be connected to any number of binary files.

The table of users is used for the purposes of access mechanisms to the database in a protected mode. It keeps a record of users that access the database in a protected mode using the login and password on the entrance portal of GeneDBase. According to user type, there are set levels with different access rights. User types are set as follows:

- ordinary editor;
- professional editor;
- administrator;
- database administrator;
- ordinary user.

An **ordinary editor** is a user with the permission to fill in the data or to edit the data that has been filled in. The data include information from the table of samples and individuals, or photographs and other digital documents.

A **professional editor** can not only fill in and edit own data, but also act as a competent authority confirming the data filled in by other users before these data become a part of the published database.

An **administrator** has additional rights to work with the table of microsattelites. Administrator can export the data and process them using external programs to calculate indicators of genetic variability, which can be then uploaded back to the database.

A **database administrator** can insert new records into the table of species and add new values to the lists of values. Database administrator has access to all data and can process the data using controlled algorithms according to the needs of the development of the database, can also use the history record of the development of the data-

base tables to get the insight into the data and, if it is necessary, to restore some of the original data.

An ordinary user has a permission to read data or export them to the selected format.

Access to the database is also possible in a public mode. In this mode, users do not have to log in and the access portal offers the display of the contents of the database through commonly available websites as it is common using the public Internet. Users entering a public mode can only read data and the data cannot be exported.

3.3 Implementation

GeneDBase will be implemented in open source technologies of current Internet. MySQL will be used as the database. Access algorithms will be programmed using PHP application server.

GeneDBase will be implemented as a typical network application of the client-server type, where the client (user) would use an Internet browser (Mozilla Firefox, Internet Explorer, Google Chrome etc.).

The output of the application server on the database side will be HTML and CSS code, the JavaScript client programming will be used too.

The running platform of the implementation is given in Fig. 2.

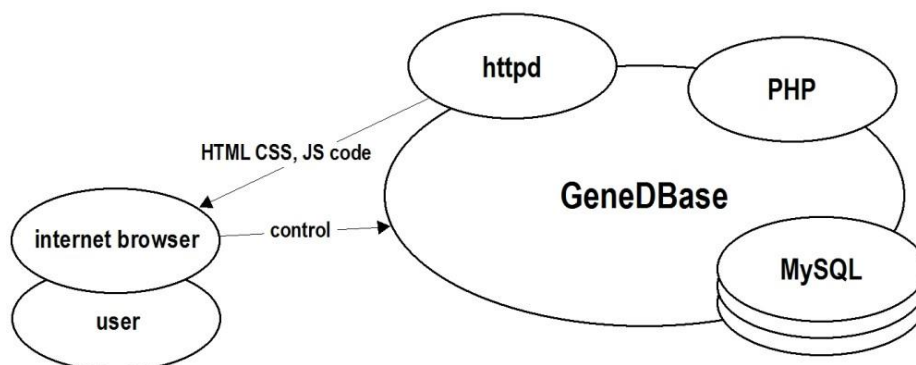


Fig. 2. The running platform of GeneDBase

No special hardware is needed for the implementation of GeneDBase, the commonly offered hosting of current Internet providers is sufficient in the initial phase. The domain name genedbase.eu (genedbase.cz) is already registered and used. Only in the case of a significant increase in the number of concurrent database accesses (app. 100 and above), special hardware should be considered. An increase in other parameters (number of records, total volume of the database etc.) will not require special hardware.

GeneDBase will be developed in strict compliance with all standard recommendations (especially ISO) and all technologies used (especially the database based on

general SQL, application language based on Java, and transmission and communication protocols based on XML). Therefore, any necessary transformation to another (commercially supported) platform will be possible without technological problems in the future.

3.4 Functionalities

The functionalities of the application, based on the prerequisites, are following:

- Data Collection Module – this is the basic functionality of GeneDBase. Data collection can be done by both ordinary and professional editor. Data can be loaded manually or by a batch file. Import batch should be in XML format.
- Verification Module, Microsatellite editing – all the loaded data must be verified by professional editor, who has access rights to all levels of records and is able to edit the data. Editing history is also recorded. The professional editor also has the rights to edit the table of microsatellites.
- Data Viewing Module – stored data can be listed and sorted by selected properties (date, individual...). As viewing is also considered sharing the database with other participants (mapservers, statistical applications)
- Data Sharing Module, Export-import – it is possible to share data with other users and with other partner projects. Data could be gained or received in different formats. Output formats are XLS, TXT, XML. Other possibilities are continuously checked and implemented into the database application.
- Data Administration Module – the module is specified for database administrator to run and maintain the application. New users can be added and their rights specified, the tables can be edited and lists of values can be added. Also to return changes is possible. Further development is expected.

3.5 List of Species (Taxalist)

A structure of Taxalist was designed by Ludek Skocovsky during the last decade. The structure has been implemented in several government information systems (ARROW, ENVIVoda, Salamander) and is guaranteed by the Ministry of Environment and the Ministry of Agriculture of the Czech Republic, it is also operated by the Czech Hydrometeorological Institute and basin management state enterprises. The same Taxalist is used for GeneDBase project in case of data interoperability.

4 GeneDBase application

Taking advantages and combining knowledge from different scientific branches contributes to finding connections and bringing new views into already known principles. Usage and future development of GeneDBase by different users can make the data accessible for other subjects. The database is accessible through www.genedbase.eu.

Acknowledgement. The grant project is supported by the Technology Agency of the Czech Republic, identification code TA02031259.

References

1. Ernst, M., Kuciel, J., Urban, T.: Analysis of genetic variation of eight candidate genes in two wild boar subspecies. *Czech Journal of Animal Science* 48, 533-539 (2003).
2. Feuereisel, J., Ernst, M.: Genetische Untersuchungen zum Einfluss langzeitiger Isolation einer Population weißer Hirsche (*Cervus elaphus* L.) in der Tschechischen Republik. In: Fragmentierung der Landschaft und andere anthropogene Einflüsse auf Wildtierpopulationen. 32. Strassfurt: Gesellschaft für Wildtier- und Jagdforschung, Prof.Dr.M.Stubbe. pp. 327-337 (2007)
3. Ernst, M., Kliment, J., Levý, E., Kourková, L., Stejskal, M.: Populace bílých jelenů – Využití mikrosatelitních analýz při šlechtění populace bílých jelenů u LČR, s.p. Edice Grantové služby LČR, Hradec Králové (2008)
4. Ernst, M., Matoušková, J., Křížek, L.: Využití mikrosatelitních analýz při monitoringu populace a přeshraniční migrace losa evropského v oblasti LS Vyšší Brod. Závěrečná zpráva projektu Grantové služby Lesy ČR, s.p. (2011)
5. Ernst, M., Putnová, L., Pokoradi, J., Matoušková, J., Židek, R., Maršálková, L., Lamka, J.: Microsatellite Analyses DNA for Controlled Breeding of Cervidae and Genetic Breeding of White Colored *Cervus elaphus* in Czech Republic. *Iranian Journal of Applied Animal Science* 2, 233-237 (2012)
6. Skočovský, L.: Informační systém SALAMANDER monitoringu ZVHS: Sběr biotických dat: Implementace. ZVHS Brno (2002)
7. Skočovský, L.: ARROW Supported Database: Sběrná podpůrná databáze informačního systému ARROW MŽP OOV pro zavedení potřebných opatření k implementaci Rámcové směrnice v oblasti vodní politiky. Ministerstvo životního prostředí ČR, Praha (2006)