THE EVOLTREE REPOSITORY CENTRE
A CENTRAL ACCESS POINT FOR REFERENCE MATERIAL AND DATA OF FOREST GENETIC RESOURCES

Michael Stierschneider[1], Stephan Gaubitzer[1], Johanna Schmidt[1], Otto Weichselbaum[1], Dieter Kopecky[1], Antoine Kremer[2], Silvia Fluch[1] and Eva Maria Sehr[1].

1 AIT Austrian Institute of Technology GmbH, Health & Environment Dept., Konrad-Lorenz-Street 24, 3430 Tulln, Austria
2 INRA, UMR1202 BIOGECO, Cestas F-33610, France

The Repository Centre is a centralised storage facility that hosts biological and genetic resources and corresponding metadata collected by EVOLTREE partners, to be used for research upon request by any interested lab.

It deals with resource management ranging from whole organisms (plant and animal material, bacteria, and fungi) to DNA samples of whole genomes, organellar and sub-genomic clones and BACs or genes (ESTs).

Goal and objectives

For many decades, the importance of reference material for various types of plant based research from ecology to phylogenetic systematics has been acknowledged; whenever results need to be compared, the availability of reference material plays a critical role. Thus collections of reference material for plant breeding, biotechnology and biodiversity analyses have been established all around the world. These have traditionally been based on living plant material managed *ex situ* in seed banks, botanic gardens or *in situ* in nature reserves.

With the advent of DNA based analytics, a new and very powerful tool was uncovered to unlock the potential of the stored biological material. After all, genomic DNA samples represent the entire genetic information of an organism, from various resistance traits to phenotypic parameters, as well as records of their inherited characteristics and ancestral roots. Thus, since the discovery of the double helical structure of the DNA in the early 1950s (Watson and Crick, 1953), molecular data based on DNA sequences have become increasingly important for a variety of biological disciplines, including systematics, ecology, evolution and population genetics, genetic diversity assessment and data generation used as a basis for nature conservation.

Nowadays, the extraction of genomic DNA is easily done, even in high-throughput, and once purified, DNA can be stored for long periods of time. As a result, DNA based collections have become increasingly important, with biobanking in the human sector being the most prominent one; but natural history museums have also started engaging in uncovering the value of their historic collections by extracting DNA from various dried samples which have been collected and preserved for centuries (Droege *et al.* 2014).

Photos: AIT

However, the key to success is standardisation. Standardised tools and techniques are increasingly applied in diverse initiatives, such as the 'International Barcode of Life' (iBOL) which entirely relies on DNA information generated from reference samples collected throughout the world, aims to reconstruct the phylogenetic relationships in the Tree of Life This project can only succeed if all the partners apply the same techniques and have access to the same reference DNA, as well as collected data.

In order to meet the demands of future high-impact research, reference material and larger sets of DNA samples, as well as corresponding data, have to be made widely accessible. Particularly in the area of ecological research, where findings are based on the fact that population genetic patterns are being compared across borders and over large geographical distances and gradients, it is essential that researchers have access to reference material and to the respective data generated from this material. Such biological material, data and tools

are needed in a standardised and freely accessible way in order to guarantee comparability of research results across Europe. When dealing with forest trees, sample collection is a laborious and time-consuming task, involving long trips to remote areas, as well as dangerous or difficult situations when accessing the material (e.g., tree climbing, shooting down twigs); therefore, the sharing of this material and generated data leads to more economical and time saving research.

In view of this, one objective of the EVOLTREE network was to build up a centralised and standardised storage and management facility, known as the Repository Centre. By storing research material at one physical site, the aim was to generate high impact in forest research in the disciplines of ecology, genetics, genomics and evolution, not only in Europe and during the course of the project, but also as an international reference site for forest genetic resources.

Description of the facility

The Repository Centre gathers together all the available and dispersed research material in one reliable site and provides open-access to a continuously growing data-set. This high quality material is available to and provided by EVOLTREE partners and researchers outside the network.

Due to the huge number of samples which were already available at the EVOLTREE partner sites, special equipment for storing, managing and tracing material, together with a database for storage of all the corresponding (meta-) data, were prerequisites for the physical installation of the Repository Centre. The installation process for creating flexible and highly reliable workflows for DNA extraction, quality control and long-term storage at the Repository Centre laboratory was initiated in 2006. The goal was to have genomic DNA of populations, e.g., from the EVOLTREE Intensive Study Sites (Lefèvre *et al.* page 6, this book) as well as gene bank collections extracted and stored following the same extraction procedures. ESTs had to be available in the form of single clones, as well as spotted on micro arrays, in order to conduct large scale expression profiling in natural populations of non-model species. The most important features were guaranteed sample integrity, standardised quality of material and data, and sustainable and easy access to the material. To fulfil these requirements, a modular -20°C/-80°C fully automated storage system with a capacity for 11,230 microtiter plates at -20°C (Universal Store US-450; Nexus Biosystems) and 1,000 microtiter plates at -80°C was installed (BioBank™; Thermo Scientific).

The storage system comes with online monitoring and logging of the temperature status and includes an internal alarm system via SMS in case of any temperature failures. The re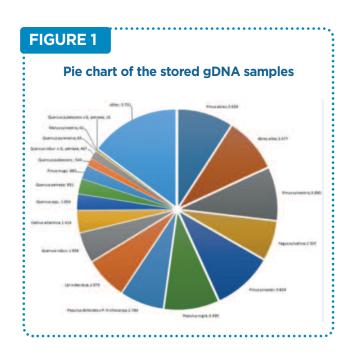dundant refrigeration systems are assembled outside the building; each system being able to hold the set temperature without the second one.

Barcoded, heat sealed microtiter plates are stored on blue barcoded trays (6 plates/tray) which are positioned randomly within a shelf system. A rail-based robotic arm stores or retrieves the plates on request and plates within the trays can be rearranged by a separate plate picking module under freezing conditions to prevent needless sample thawing. Plate picking and re-insertion is logged by a database system

Automation

To manage and distribute several thousand DNA resources, a high level of lab automation is required to guarantee failure-free sample and data handling, combined with high throughput and high quality. A newly introduced laboratory data management system, called *Material Administration and Preparation System* (MAPS™) serves as the backbone of the quality management system (Kopecky *et al.* 2012). It consists of a database which holds information about the lab processes and a web-based user interface for viewing and editing this information. Once the initial necessary Excel-based provider data has been imported into the system, the whole workflow being undertaken in the laboratory at the time (DNA extraction, electrophoresis images, PCR results, storage positions, DNA

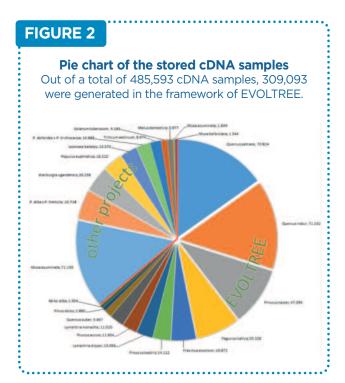

FIGURE 1

Pie chart of the stored gDNA samples

quantity/quality, and results of downstream analysis), together with all needed supply information (e.g., barcodes, volumes, well IDs), is logged within the system and can be retrieved accordingly.

Genomic DNA (gDNA) extraction protocols have been optimised for semi-automated (whereby the pipetting steps are run automatically and the centrifuging steps manually), high-throughput processing for 96-well format, using liquid handling platforms like *Tecan* or *Hamilton*. Supported by these infrastructures, the Repository Centre is able to execute up to 960 gDNA extractions per day.

Protocols for a wide range of raw plant material like leaves, needles, buds, roots, cambium, and wood have been established. In order to guarantee access to the reference material over time, lyophilised plant material is stored at room temperature alongside 2 x 100 µl of gDNA extracted from this material at concentrations of approximately 50 ng/µl. One of these 100 µl is located in the working copy deposited in the automated storage facility, whereas the second copy is used as a backup plate stored in another section of the building. This is due to the requirement of risk mitigation – so that in the case of any sort of accident (e.g., fire, system break down, etc.) which could destroy the working copy of a sample, the backup copy is still available for use.

As soon as the gDNA extraction is finished, an aliquot (mostly 1 µl) of each sample is verified and documented via agarose gel electrophoresis. Analysis of the gel images is carried out automatically using proprietary image analysis software (Bajla *et al.* 2005). This software enables the analysis of 96 individual samples (loaded on one gel) in parallel. Pixel intensities are measured and compared to a
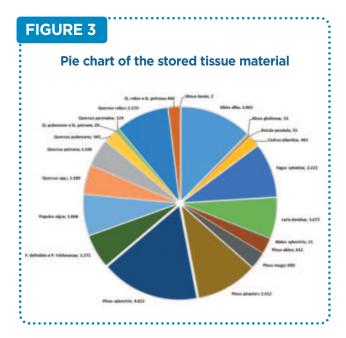

pre-defined standardised mass ladder, whereby the gDNA yield is calculated. The corresponding picture as well as the estimated gDNA concentration and size are saved back to MAPS™.

In the case of failed samples, or samples showing degraded gDNA, the sample state is set to «failed» within MAPS™, and these samples will not be available for downstream analysis. This way, the quality and amount of gDNA is documented for each and every sample so that in the case of any problems occurring in downstream analyses, the data quality data of a sample can be retrieved and checked by the user directly via a central search portal.

At the request of a customer, single samples can be selected out of 96- or 384-well microtiter plates, a process known as cherry-picking. The requested samples are often distributed over a larger number of microtiter plates and their picking necessitates an error free workflow. To fulfil these requirements, worklists are generated by MAPS™ which are generated within the software and sent to the liquid handling platforms directly. The pipetting process monitored by barcode tracking starts according to the worklists and the generated log files provide information about possible sample manipulation at any time.

Capacity

At the time of writing, 27,073 gDNA samples (Figure 1), 485,893 cDNA samples (Figure 2), 202,752 BACs, and 26,329 source samples (Figure 3) from various organisms of forest genetic resources and allied species (e.g., moths, caterpillars, fungi) are managed in the Repository Centre (Table 1). Of the capacity of 11,230 microtiter plates, currently 9,569 plates in both formats, 96- and 384- well, are stored. Most of the gDNA and tissue samples are stored in duplicates.

FIGURE 2

Pie chart of the stored cDNA samples
Out of a total of 485,593 cDNA samples, 309,093 were generated in the framework of EVOLTREE.

FIGURE 3

Pie chart of the stored tissue material

Data management and access

Databases are critical for efficient sample and data management, as well as for the efficient end utilisation of the DNA bank itself. To allow a quick and easy retrieval of a maximum amount of associated information (e.g., tissue type, collection conditions, date of collection, provider information, sample preparation, details on DNA quality and quantity), a tailor-made LIMS (Laboratory Information Management System) was developed in the framework of EVOLTREE to fulfil the demands of the Repository Centre. MAPS™ (Material Administration and Preparation System), a new concept of LIMS, meets these requirements with an innovative data model, as well as a modern service-oriented architecture, on top of state-of-the-art web technologies (Kopecky *et al.* 2012).

MAPS™ reflects all workflow steps in the laboratory and provides possibilities for recording these steps electronically. MAPS™ further communicates with the storage system via an application programming interface (API) based on the *Java Messaging Service* (JMS). This way the end user can access the organisation and reporting capabilities of the storage device seamlessly, without the need to work with two different applications in the laboratory. The user interface of MAPS™ has been implemented in a web-based way in *Java*. Access to the system therefore only requires a web-browser and no other specific software needs to be installed, so it is very easy to access MAPS™ from any computer available in the laboratory. The information in MAPS™ is recorded in a *PostgreSQL* database on a central server and regular backups of the data ensure its necessary integrity. Data access is only granted to laboratory workers and the administrator via user accounts with respective user rights. Should communication with external partners be required, MAPS™ offers an interface based on the *Simple Object Access Protocol* (SOAP), which provides standardised services for querying information about samples stored within the LIMS, as well as for adding new samples to the system. These services can also be made available across organisational borders, so that customers are able to query information about the stored samples via a central search portal, known as the eLab (Ehrenmann *et al.*, page 15, this book). Due to security restrictions, however, the MAPS™ service, especially the storage system interface, is located in an internal network not reachable from the outside. The eLab services cannot directly access the MAPS™ databases, since they are located at the same institution (although in a different network). The MAPS™ database is therefore treated as an external database resource by the eLab that needs to be queried at regular intervals. The queried information is then inserted into the eLab search system, where it can be accessed by the users (Kopecky *et al.* 2010).

Current and future use

Besides its central role as a storage and retrieval system for biological material, DNA, and data in the framework of the EVOLTREE network serving forest ecosystem research, the Repository Centre became part of the Trees4Future (http://www.trees4future.eu/) project. This was an Integrative European Research Infrastructure project that integrated forest tree breeding infrastructures to improve and enhance gains in the area of European forest tree breeding. In this framework, the repository centre and its links to the EVOLTREE data collections serves as an integrative hub for European forest ecosystem research and European tree breeding efforts.

TABLE 1

Overview of the sample types and number of samples stored in the Repository Centre

Sample type	Nr. of samples stored
Source material	26,329
gDNA	27,073
cDNA	485,593
BAC	202,752

This way, the Repository Centre is currently the largest data provider to the Global Genome Biodiversity Network (GGBN) – a network of repositories of genomic sample collections aiming at allowing open access to reference material from botanical gardens as well as natural history museums. Data of about 53,402 samples, divided into 27,073 gDNA and 26,329 tissue samples are accessible via the GGBN portal (http://www.ggbn.org/) following the ABCD standard used by the BioCASe provider software (Holetschek *et al.* 2012). In addition to this, all georeferenced EVOLTREE repository data sets can be found on GBIF, the Global Biodiversity Information System (GBIF, http://www.gbif.org/).

Being part of these networks enhances visibility and contributes to the ever growing idea of open-access resources and data in order to further the development of DNA based forest ecosystem research. As can be seen by all these activities, and as also requested by the European Commission, open data and open material initiatives will help future research to better integrate research results and to enhance forest ecological understanding on a large scale. By integrating genetic, as well as environmental data, we will be able to generate forest systems modelling approaches that will allow a more general understanding of the complex ecosystem forest and will help to mitigate the impact of climate change on one of the most important socioeconomic factors in Europe.

REFERENCES

Bajla, I., Holländer, I., Fluch, S., Burg, K., Kollár, M., 2005. An alternative method for electrophoretic gel image analysis in the GelMaster software. Comput. Methods Programs Biomed. 77, 209–231. doi:10.1016/j.cmpb.2004.09.007

Droege, G., Barker, K., Astrin, J.J., Bartels, P., Butler, C., Cantrill, D., Coddington, J., Forest, F., Gemeinholzer, B., Hobern, D., Mackenzie-Dodds, J., Ó Tuama, É., Petersen, G., Sanjur, O., Schindel, D., Seberg, O., 2014. The Global Genome Biodiversity Network (GGBN) Data Portal. Nucleic Acids Res. 42, D607–D612. doi:10.1093/nar/gkt928

Holetschek, J., Dröge, G., Güntsch, A., Berendsohn, W.G., 2012. The ABCD of primary biodiversity data access. Plant Biosyst. - Int. J. Deal. Asp. Plant Biol. 146, 771–779. doi:10.1080/11263504.2012.740085

Kopecky, D., Schmidt, J., Fluch, S., 2010. Large-scale integration of distributed heterogeneous data sources within Evoltree, in: Proceedings of the EVOLTREE Conference on Forest Ecosystem Genomics and Adaptation. San Lorenzo de El Escorial, Madrid, Spain.

Kopecky, D., Weichselbaum, O., Fluch, S., 2012. MAPS – A service-oriented, customizable, multi-purpose LIMS. Nat. Methods Appl. Notes.

Watson, J.D., Crick, F.H., 1953. Molecular structure of nucleic acids. Nature 171, 737–738.