Integrating different windows on reality: socio-economic and institutional challenges forculture collections

Heide-Marie Daniel (UCL/MBLA)
Uwe Himmelreich (Max-Planck-Institute for Neurological Research)
Tom Dedeurwaerdere (FRS-FNRS/UCL)

Bibliographical reference

Heide-Mairie D., Himmelreich , U., and Dedeurwaerdere, T., 2006, “Integrating different windows on reality: socio-economic and institutional challenges for culture collections”, International Social Science Journal, 2006, vol. 188: pp.368-380.

Integrating different windows on reality: socio-economic and institutional challenges for culture collections

Heide-Marie Daniel, Uwe Himmelreich, and Tom Dedeurwaerdere

The socio-economic function of culture collections

Microbial culture collections have grown in response to societies that need to address and solve problems associated with traditional food production, industry, and medicine and to develop modern microbial processes. Their general aim is to collect, authenticate, maintain, and distribute cultures of micro-organisms and associated information for high technology developments as well as for the day-to-day requirements of general health care, agriculture, food production, and teaching (Fig. 1). The current preference given to knowledge generation based on modern, high-technology approaches such as the determination of the genotype[1] – the inheritable information carried in the genetic code of the organism – starts to downgrade the existing knowledge of classically used phenotypic methods that analyse the observable structures and functions of a living organism. The resulting loss of expertise will reduce the possibility of relating modern and classical microbiological knowledge and the genotype to the phenotype.[2] It also hinders the continued use of technically less demanding classical methods in less developed countries. These countries most often hold the largest still existing biodiversity, as the destruction of the natural environment there is the least advanced.

Dr Ir H.-M. Daniel, Catholic University of Louvain, is scientist in a branch of the Belgian consortium of culture collections which concentrates on fungi of agro-industrial and environmental interest (MUCL/BCCM™). After an academic education in biotechnology, she specialised during her Ph.D. and postdoctoral studies in the reconstruction of genealogical relationships of yeasts by molecular methods. Her current research interests include the application of molecular, biochemical, and physiological methods to questions of yeast characterisation and identification.
Email: daniel@mbla.ucl.ac.be
Dr Himmelreich, currently working at the Max-Planck-Institute for Neurological Research, is a specialist in instrumental analytical methods and their application to microbiology and infectious diseases. The main focus of his research is the application of magnetic resonance spectroscopy to the biochemical and biophysical characterisation of micro-organisms. This research ranges from biochemical, phylogenetic to diagnostic applications. He is author of more than 70 scientific publications in peer-reviewed journals.
Email: himmelreich@nf.mpg.de
Tom Dedeurwaerdere is director of research at the Centre for Philosophy of Law and professor in the Faculty of Philosophy, both at the Catholic University of Louvain. Bibliographical information on his publications can be found on the website: http://www.cpdr.ucl.ac.be/perso/dedeurwaerdere.
Email: Dedeurwaerdere@cpdr.ucl.ac.be

The goal of microbiological culture collections is not only to maintain materials and expertise, but also to enhance the understanding of micro-organisms to promote their application. The large numbers of organisms aggregated in culture collections provide an excellent basis for this. To take advantage of these resources, research focusing on the properties of the organisms has to be accompanied by data repositories that facilitate data analysis in order to effectively allow the generation of knowledge.

More than 500 culture collections storing over one million microbials have

ISSJ 188 © UNESCO 2006. Published by Blackwell Publishing Ltd., 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

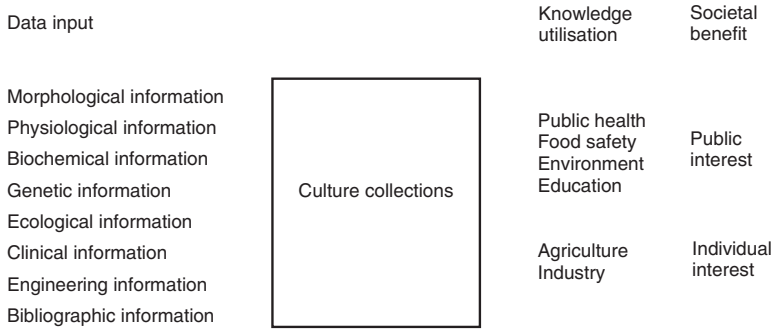


FIGURE 1. Information types generated and sectors of knowledge utilised by culture collections

been created in over 60 countries since the establishment of the first collection in Prague, Czech Republic in 1890 (World Collection for Culture Collections 2007; Sly *et al*. 1990). With the arrival of biotechnology and the industrial utilisation of living organisms which evolved through basic research in the 1970s and 1980s, the collections needed to become proficient in industrial relationships as well as remaining expert centres in microbial taxonomy,[3] conservation, and physiological and biochemical testing for research and educational purposes. Accordingly, collections catalogued and digitised their information, developed databases, coordinated their efforts for the benefit of the users, developed educational programmes to inform the public, promoted themselves as industrial partners, and, with the availability of the Internet, put all this information online for the benefit of international experts and the public. Entering the market economy was only possible with governmental support to add the new skills without loss of traditional expertise.

In 1992 the Convention on Biological Diversity was ratified by 150 nations (United Nations Environment Programme 1992). Following this, culture collections were required to develop new strategies to realise the conservation and sustainable use of biological diversity as well as the fair and equitable sharing of its benefits. Strategic changes that reflect this development must be accompanied by well-defined policies and by adequate funding. Current resources are not sufficient to face the challenge of securing access to materials and data in the best technological and socially most acceptable way. The debate continues as to what is needed, who should provide it and who should finance it. These basic questions must be resolved collectively among the international scientific community and the policymakers to establish a long-term strategy that will ensure continuous, stable, and progressive development of culture collections. Scientific and technological progress will continue to change the requirements of the end user and thus demand new services which must be developed and implemented simultaneously with the continuation of the traditional tasks of culture collections. To respond to this requirement, culture collections should be the place to conduct research that is unlikely to be carried out elsewhere and that relates to the primary goals of the collections (conservation and taxonomy). The skills for much of this research are embodied in the collection personnel by virtue of their daily tasks. It would be economic insanity not to utilise this resource for societal benefit.

Integrating different windows on reality

Human society is benefiting from our knowledge on micro-organisms. This knowledge is constantly increasing with improved technologies. In order to translate this accumulated knowledge into benefit for all and not just for experts in particular fields, it is important to share it and to correlate different findings with each other. For example the knowledge of the genetic code

of a micro-organism is useless without the knowledge of its translation into observable physical properties. Physical properties on the other hand have no meaning if they are not set into context with life. In addition, the physical properties of micro-organisms are not a constant but are known to change in response to the environment.

The basic argument of this article is that neglecting the importance of combining and sharing knowledge coming from these different levels of reality results in loss of expertise, knowledge, and social opportunities both for developed and developing countries, the latter as the reservoir of most of the remaining biodiversity on earth. In this section, we introduce the basic notions – in simple terms – that play a role in the different approaches to living organisms. Going beyond an atomistic and segmented approach to reality is the ultimate rationale for the development of appropriate institutional frameworks for sharing of data and resources.

Our world consists of observable objects. Apart from the small fraction of objects visible to the human eye, most of these objects, and also properties of the visible objects, are only detectable with the help of tools. Depending on the tools we use, we get different impressions of the objects as different tools explore different properties. Each of these properties is part of our reality. Although we have to deal with different properties, many of them are not independent of each other. For example, the property of taste of cheese is partly determined by the property of chemical composition of the cheese. In addition, properties may change with time, as can easily be imagined on the example of taste and chemical composition of the cheese. Only the complete properties of an object (including its temporal changes) would describe the object completely. However, the task of such a comprehensive description is impossible to complete, as human knowledge and with it the ability to detect more properties is constantly evolving.

Our natural environment is to a large extent determined by organic life forms. Most of them are hidden from the unaided eye. Many of these life forms, referred to as micro-organisms, are essential for the existence and the well-being of higher organisms, including humans. Examples of the role of micro-organisms are found in the natural recycling of organic material and their participation in food chains or symbiotic relationships, like in the digestive system of higher organisms. Even a large proportion of the human body mass consists of micro-organisms. Although we do not perceive micro-organisms as single individuals without the help of instruments like a microscope, they exist as individuals, are one of the earliest life forms on earth and determine the fate of our world.

Many micro-organisms show similar appearance when observed by microscopy. However, they often differ in their interactions with the environment. In contrast to other larger life forms, which are more easily distinguishable, the characterisation of micro-organisms requires the evaluation of as many as possible properties to differentiate between them. Their reliable identification is, for example, essential to distinguish pathogenic from benign and beneficial micro-organisms or to evaluate their suitability in industrial processes (for example, in cheese production). The sum of observable properties and characteristics of a micro-organism is regarded as the phenotype. This phenotype is determined both by the inheritable information carried by the organism, or the genotype, and the environment. The process of interaction between the molecules that represent the genotype and the environmental factors is complex and only partially understood. Hence the knowledge of the genotype, represented by the genetic code, does not always allow the prediction of the phenotype. For example, the taste of a particular cheese is not only determined by the genetic code of the cheese-producing micro-organism but also by its environment (such as temperature and milk composition). The phenotype of the micro-organism provides the most essential information for understanding an organism's position in the global balance of life and understanding it is essential for any potential exploitation. While the genetic material of cells can be decoded, digitised, and stored in standardised data formats on a large scale, the phenotypic information is highly diverse and is more difficult to transform into a digital format: standardised data formats need to be developed.

The ability of micro-organisms to multiply readily is the basis of their ancient role for the production, conservation, and spoilage of food and has made them famous as model organisms

in molecular genetics. Their importance in past and future applications leads to large amounts of accumulated information on the genotype and phenotype that has to be efficiently managed and utilised in its entirety to gain further benefit from it. However, many data management approaches pay declining attention to phenotypic data. This tendency may become critical to our knowledge of micro-organisms, combined with a loss of expertise in identifying it. On the other hand, modern analytical techniques allow the generation of highly detailed data that represent the total phenotype better than single biochemical tests.

The need to integrate phenotypic and genotypic properties

The recognition of the loss of taxonomic expertise when experienced specialists retire without being replaced by well-trained successors and the bias towards developing comprehensive classifications for attractive organisms such as mammals and flowering plants, while simultaneously neglecting difficult groups such as micro-organisms, has initiated discussions on an exclusively DNA-based taxonomy. It has been proposed that a DNA sample of an individual should be used as an exclusive reference and to generate one or several gene sequences as an identification tag for the species from which the individual was derived (Tautz *et al.* 2003). While such a system will improve the recognition and classification of organisms for which no other characteristics can be determined (such as non-cultivable organisms), all other organisms would be characterised very incompletely by a tiny portion of their genome while neglecting their phenotypic properties (Lipscomb *et al.* 2003). Although a DNA-based taxonomy is an efficient way to build an integrative database for all cultivable and non-cultivable organisms, it would disregard all other biological aspects that may contribute to our understanding of the organisms and their interactions.

Molecular methods are also excellent tools for reconstructing the natural relationships of organisms. However, further difficulties[4] arise, in addition to the issues mentioned above. from the continuum of individuals.[5] An unambiguous molecular definition of species would be possible only if the gene sequences used were constant among all members of one species and different from all other species. There is no evidence that most genes meet this criterion and any diagnostic character that would do so would work without the need to be molecular. Therefore, a definition of species from molecular data alone would be as subjective as it would be if based solely on phenotypic similarities. It is the combination of phenotypic and genotypic characteristics that has the potential to improve the recognition of microbial species effectively.

The definition of higher taxonomic groups would be even more difficult or impossible if based on molecular data alone. The current hierarchical system is based on the expertise of generations of taxonomists who decided which phenotypic characteristics are the most significant and informative for a higher order grouping (genera, families, kingdoms). These decisions were made according to the best knowledge of individual persons, therefore they are subjective and in some cases even inappropriate (such as in not following the natural pattern of ancestry). However, one would ignore centuries of knowledge accumulation if taxonomy was based solely on DNA sequences. Not only would science suffer from neglecting the phenotype, but more importantly, the valorisation of micro-organisms for biotechnology would potentially be limited, as different phenotypic abilities would not be recorded for newly discovered organisms.

It would be better to use the possibilities that are offered by genotypic data to carefully correct inappropriate groups than to invent a new system solely based on DNA data. As we currently do not fully understand the interactions of genotype and environment, data on all three parameters (genotype, phenotype, and the environment) need to be evaluated to describe living organisms in the most appropriate way.

Yeast as a model

Let us illustrate the importance of going beyond genetic information through an important and well-known model in the life sciences: the model of yeast. This model shows the importance of combing genotypes – the information encrypted

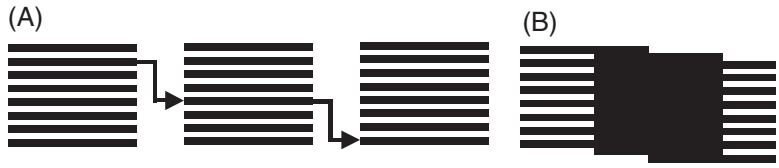


FIGURE 2. Not just the storage, linkage, and retrieval (A), but the processing, comparison, and analysis of data (B) is needed, effectively fusing data to new knowledge

in the genetic code – and phenotypes – the observable characteristics and properties of the organism. Here we focus on the scientific and socio-economic challenge to integrate both. This will set the stage for our discussion of the institutional challenges.

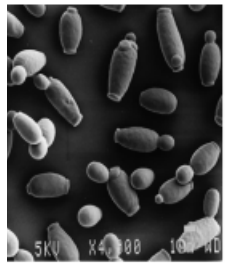
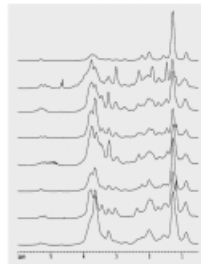
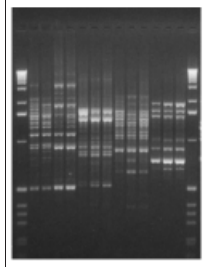
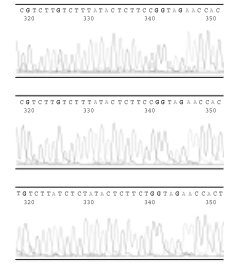
Yeasts are single celled fungi that are fast and easy to grow. Yeasts have been used in the production of food and beverages such as bread and beer since ancient times. The yeast species *Saccharomyces cerevisiae*, also known as baker's yeast, constitutes the best-developed eukaryotic system[6] to model physiological, biochemical, and many other processes. Consequently, it was the first eukaryote for which the complete genome[7] was decoded by sequence analysis of the entire genetic code (Goffeau *et al.* 1996). Complete genome sequences of about 20 yeast species are currently available. These are far fewer than for prokaryotes,[8] as the yeast genome is considerably more complex. In contrast to bacteria, yeasts are more similar to, and in some functions even equivalent to, higher organisms like humans. Using comparative approaches and models of interaction between the genome and cell functions allows us to draw general conclusions regarding simple biochemical processes.

Yeasts are also essential in many ecological networks, such as the recycling of biomass. They are far more specialised than bacteria regarding the nutrient sources that they are able to utilise and regarding the ecological niches they may thrive in, and therefore the associated environmental data are of greatest significance for the study and application of these micro-organisms. This is important for the evaluation of constantly shrinking, partly unexplored ecological systems, in particular those in developing countries. The data on the natural environment of yeasts are a primary source of information acquired directly in the process of collecting the organisms. They include information on the geographical location and the substrate from which they were recovered (such as flowers, soil, and animals). Ideally, a future database would link the organisms of different kingdoms (like bacteria, fungi, plants, and animals) that are found in close natural associations, so as to investigate the significance of possible interactions. The adaptation of many yeasts to well-defined ecological niches has led to a large diversity of particular physiological properties. This is currently not fully exploited as only a small fraction of the yeasts are utilised in industrial processes. These few species and very few strains of these species are often optimised by genetic engineering for various applications under considerable effort and cost. The natural potential of yeasts could be used more efficiently if their phenotypic properties were more accessible than they are currently.

Data management systems should not only facilitate the storage, linkage, and retrieval of the different data types but also facilitate the effective processing, comparison, and analysis of all available data so that conclusions can be drawn that extend knowledge beyond pure accumulation (Fig. 2). For example, the environmental survey of yeasts in a particular habitat generates data on the presence of particular yeast species. The analysis of the acquired data (description of yeasts by their morphology, physiology, genetic, and biochemical data, host organisms and geographic distribution) then allows predictions about species that have not been observed but are present in the habitat. The yeasts observed might allow conclusions to be drawn about their environmental adaptations and functions (Lachance 2006).

Museum of medicinal plants in the province of Yunnan, People's Republic of China. IRD / Michel Sanvin

TABLE 1. Phenotypic and genotypic properties commonly recorded for yeasts

Data type:	Phenotypic information		Genotypic information	
Data format:	Non-digitised	Digitised	Non-digitised	Digitised
Examples:	Culture appearance (Colour, odour, surface structure, texture) Shape and structure of colonies and cells (morphology) Mating behaviour Expression of growth inhibitors (mycocins) Isoenzyme profiles	Growth tests on various nutrient sources (physiology) Characterisation of cellular compounds (proteins, fatty acids, polysaccharides, etc.) and biochemical pathways by instrumental analytical chemistry (e.g. MS, HPLC, IR, NMR*)	Microsatellite primed polymerase chain reaction (MS-PCR) Randomly Amplified Polymorphic DNA analysis (RAPD) Restriction Fragment Length Analysis (RFLP) Hybridisation of total or partial genomic DNA Secondary structures of RNA	DNA sequences Amplified Fragment Length Analysis (AFLP)
				
	Cell morphology	NMR spectra	MS-PCR	DNA sequence data

*MS: Mass Spectrometry is an analytical tool used for measuring the molecular weight of molecules and their fragments.
HPLC: High Pressure Liquid Chromatography can be used to separate compounds that are dissolved in solution.
IR: Infrared spectroscopy is using the absorption of infrared light by substances to examine molecular structures.
NMR: Nuclear Magnetic Resonance spectroscopy is using different energetic states of molecules in a magnetic field to determine their molecular structure.

Yeasts have classically been grouped based on their phenotypic properties like the appearance of the colony, their cell morphology and their physiological properties (Table 1). This system of characteristics has served well up to the point at which genotypic characteristics allowed for a more precise distinction of organisms and enabled the reconstruction of natural relationships among them. Based on contradictions between genotypic and phenotypic classifications, it was recognised that the current phenotypic classification of yeasts is in large part artificial. This means that it groups organisms of phenotypic homogeneity but genetic heterogeneity, as it takes only that part of an organism's potential into account which is expressed under the given environmental conditions. The full phenotypic potential may involve additional, yet unrecognised properties.

A second problem is caused by the fact that phenotypic properties that might be variable within a group have been used to circumscribe this group. Genetically heterogeneous groups are not predictive of the full phenotypic potential of their members, as would be expected from a hypothetically natural classification. Such a natural classification would facilitate the valorisation of yeasts as it assists the selection of organisms that may possess a desired property. The awareness that (i) some crucial phenotypic properties were missed and (ii) that some less characteristic phenotypic properties were given undue weight necessitates the constant re-evaluation of the characteristics currently in

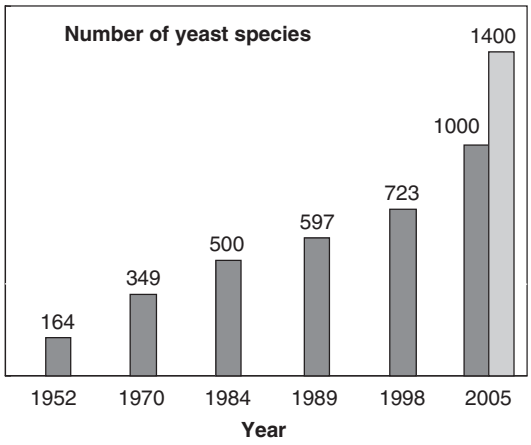


FIGURE 3. Quantitative development of taxonomically described yeast species (dark bars) and recognised yeast species (light bar). The numbers for the year 2005 are based on estimations of leading yeast taxonomists (Lachance pers. comm. 2005), the others taken from Barnett *et al*. (1990); Lodder (1970); Kreger-van Rij (1984); Kurtzman and Fell (1998).

use for yeast classification. The evaluation of increasing numbers of properties due to methodological and conceptual improvements has contributed to a continuing increase in the number of yeast species recognised (Fig. 3). The discrepancy between the estimates of recognised and described yeast species in 2005 demonstrates the urgent need for increased resources to deliver formal descriptions of the rapidly increasing number of recognised species. The continued evaluation of existing criteria and the search for new phenotypic and genotypic discriminative criteria is essential to classify the increasing numbers of new species in a realistic and meaningful scheme.

The utilisation of DNA-based molecular methods, namely the use of gene[9] sequences, is highly influential for the integration of new species and the approximation of natural relationships by the classification system. These natural relationships are essential for the development of a classification system that will predict the organism's full phenotypic potential. However, as explained before, DNA sequence data also show limitations. The recognition of species and species relationships may require sequences of different genes in different groups of organisms, making the approach of a single, all-purpose gene for identification impossible. It has also been recognised that one or two genes can often not resolve distinct species and therefore several gene sequences need to be determined for a reliable identification. Analyses of whole genome sequences have shown that the reliable reconstruction of natural relationships in a subgroup of yeasts required a minimum of 20 different gene sequences (Rokas *et al*. 2003).

The problems encountered with the use of either only classical taxonomic data or only DNA sequence data have led to the development of polyphasic approaches that utilise all available data from sources, phenotype, and genotype, to generate a consensus classification. In some cases the consensus classification is a compromise containing the minimum of contradictions. It is assumed that with more information the consensus will gain stability. Polyphasic taxonomy has been extensively developed in bacteria, as reviewed by Vandamme *et al*. (1996), who provided descriptions of the methods involved. However, equivalent principles are also applicable to yeasts. The evaluation and inclusion of many sources of information is essential to understand reality in a way that allows the effective valorisation of micro-organisms. The generation of phenotypic and genotypic types of data is currently achieved in a targeted way by determining characteristics that are a priori assumed to be informative.

The non-targeted search for informative characteristics has become feasible by recently developed methods that are screening the largest accessible parts of genotypes and phenotypes. New, rapid methods of instrumental analytical chemistry allow the simultaneous detection of the chemical cell composition, for example, the proteome[10] and the metabolome.[11] These methods provide an overview of all detectable chemical compounds in the cell. This simultaneous detection of hundreds of chemicals (and therefore potential characteristics) has the advantage of not having to select a particular compound or group of compounds as a target that is supposed to provide the information. Innovative methods for data analysis have been developed to extract valuable information for the characterisation of micro-organisms, and

their biochemical pathways and for the identification of potentially industrial useful products (Himmelreich *et al*. 2003; Raamsdonk *et al*. 2001; Wang *et al*. 2004).

Institutional challenges for culture collections

Information systems will be able to create increasingly realistic models of nature as more and more diverse and complex information will be fed into them. With this increasing knowledge, we not only face scientific and technical challenges, but have also the chance to achieve a new quality in utilising micro-organisms for the benefit of human society, allowing us to select the best-suited organisms for a particular application. This task can only be achieved if different parts of society (science, economy, and politics) work together in order to share costs and rewards, and to optimise resources.

The currently stagnating communication between increasingly specialised and therefore partitioned communities (industry, medicine, basic research) has restrictive consequences for the development of comprehensive knowledge. An example from the world of classification and taxonomy is a yeast known as *Pichia pastoris*, used commonly as a expression system for the production of heterologous proteins.[12] Recent biodiversity surveys have resulted in the discovery of similar yeasts leading to their reclassification, including *P. pastoris* in the new genus *Komagataella* (Kurtzman 2005). It now becomes apparent that many industrially used strains belong not to *K. (Pichia) pastoris*, but in fact to the newly described species *K. phaffii*. Knowledge of the distinct genotype-based classes within *P. pastoris* might have facilitated the selection of potent phenotypes as production strains. Further characterisation of *K. phaffii* might show subtle differences that have led to its empirical selection for industrial purposes. As biotechnologists and taxonomists have worked independently in this case, the discovery was made only by chance.

On one hand, the integration of many types of different data (genetic, ecological, biological, and chemical) into data and metadata repositories is technically demanding, as these enormous amounts of data require complex processing for digitisation (Table 1) and a high degree of data fusion for effective knowledge generation (Fig. 2). On the other hand, the design of a repository has to be as simple, intuitive, and user-friendly as possible to fulfil the social aspect of intellectual accessibility to all involved disciplines. The personnel that is managing and utilising the repository may be specialised in one or some types of data, but can never be an expert in all of them. To utilise all available data in polyphasic analyses (see above), the information should be presented in a format that is comprehensible for non-specialists. Software programs that fulfil the above requirements do not exist at present. This is mainly due to the diversity of data but also to the fact that some characteristics are difficult to digitise and model (for example, colour, odour, or shape) without over-simplification of their natural diversity. A large array of specialised software is necessary to manage and analyse the different types of data. To ensure the highest scientific quality of database and analysis software, the development demands a collaborative and multidisciplinary basis, including experts from the fields where the data originate.

This challenge has scientific, economic, and institutional dimensions. Comprehensive information systems should integrate traditional data (for example, morphology, physiology), genotypic data (for example, DNA sequences), novel phenotypic data (for example, spectroscopic data), and many other types of information. This information needs to be shared between different disciplines, as all can contribute to and gain from more specialised and detailed data that are not commonly available. Information systems will be essential for linking the genome as representative of the potential of an organism with the proteome, as the sum of expressed proteins in response to the environment, and the metabolome that indicates the current status of a cell by the totality of its small molecules, information on ecological factors, and pathogenesis. The establishment of such links will not only greatly contribute to our understanding of life but also have implications for the utilisation of micro-organisms. The challenge is to share this information between often contradictory interest groups without

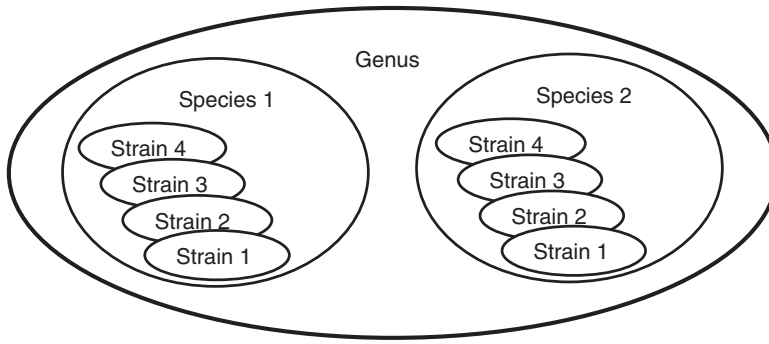


FIGURE 4. Hierarchy of individuals (strains) and the two lowest principal taxonomic ranks (species and genus) to group them

compromising on intellectual property rights. This not only applies to commercial entities but also to the protection of national resources. Developing countries are important partners for conserving biodiversity. They often possess very diverse, partly threatened, and poorly studied ecological systems. As they do not have the resources to implement the most recent technologies, it is crucial to generate inventories of their patrimony using basic microbiological methods.

Culture collections are situated at the intersection of societal requirements and multiple types of information. They have succeeded at this frontier and will continue to stimulate the establishment of a microbial commons by the way they manage biological, scientific, social, and ethical concerns.

Notes

1. The genotype is the inheritable information carried by all living organisms. This information is used as a set of instructions for building and maintaining an organism. These instructions are encrypted in the genetic code, copied at the time of cell division, and passed from one generation to the next. These instructions are intimately involved with all aspects of the life of an organism, controlling everything from the formation of protein macromolecules to the regulation of metabolism.

2. The phenotype includes anything that is part of the observable structures and functions of a living organism. These are physical parts, the sum of the atoms, molecules, macromolecules, cells, structures, metabolism, energy utilisation, tissues, organs, and behaviours.

3. Taxonomy is used here in its original sense as the science of finding, circumscribing, formally describing, and naming organisms. Taxonomic classification follows a hierarchical structure that creates groups of organisms with decreasing similarities in their properties and, more recently, in their natural genealogical relationships. The most similar individuals or strains are grouped in one species and similar species in one genus (Fig. 4). Thus, each strain can be assigned to a species; each species can be assigned to a genus, and so on.

4. Another important issue showing the limits of a ''genetic approach only'' is raised by comparative DNA analysis. These challenges include the difficulties to compare sequences of different lengths, distinguishing orthologs from paralogs and the selection of appropriate genes that are informative for a large range of diverse organisms. Orthologous genes are direct evolutionary counterparts derived from a common ancestor through vertical descent. As a

consequence, orthologs often, but do not not necessarily, assume the same function in different organisms. To compare the same gene from different species, those genes have to be orthologs. Paralogous genes originated from a common ancestor by duplication and then diverged from the ancestral copy by mutation and selection or drift. As a consequence, paralogs often, but do not necessarily, assume different functions in an organism (Koonin 2005).

5. No clear boundaries of distinct groups (for example, in species) exist. More and more sensitive techniques reveal a continuum of individuals with an increasing number of non-assignable individuals. Although it is difficult to develop models of such fuzzy groups, a future data management system has to be able to illustrate this reality without the current inevitability of reducing the true multidimensional reality.

6. Eukaryotes are organisms with a higher structural complexity of the cells than the more simple prokaryotes. Eukaryotes are characterised by having many functions segregated into semi-autonomous regions of the cells (organelles). The name of the eukaryotes origins from the most evident organelle, the nucleus (Greek, *eu* = true+*karyon* = nucleus). Eukaryotes include humans, other animals, plants, fungi, and a rich variety of micro-organisms.

7. The genome is the whole hereditary information of an organism that is encoded in the deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

8. Prokaryotes (Greek, *pro* = before+*karyon* = nucleus) are single celled organisms that lack the characteristic eukaryotic organelles. Neither their genome nor any other of their metabolic functions is restricted to an enclosed area of the cell. Instead everything is openly accessible within the cell. Prokaryotes include viruses, bacteria, and blue-green algae.

9. A gene constitutes a portion of the genome that encodes a single protein or another molecule of functional relevance. The genome of the yeast *Saccharomyces cerevisiae* contains about 6,000 genes.

10. The proteome is the totality of all proteins in a cell, produced under a given set of environmental conditions. While the genome remains constant (disregarding potential mutations) for the cells of an organism, the proteome varies with the activity and the environment of the cells.

11. The metabolome is the totality of all small molecules of a cell such as nucleotides, vitamins, and antioxidants. It mediates the information about environmental changes to the genome. While the genome is representative of what might be and the proteome is what is expressed, it is the metabolome that represents the current status of the cell (for example, nutrition, age, and effect of toxins).

12. Expression systems for heterologous proteins allow the production of proteins that are foreign to the producing cells which have been programmed by genetic engineering to express them. Such systems consist of the host cells, a DNA construct that contains the gene encoding the desired protein, and the appropriate environmental conditions for the expression. Expression systems are used for the large-scale production of high value proteins such as enzymes, vaccines, and various blood factors.

References

Barnett, J. A., Payne, R. W., and Yarrow, D. 1990. *Yeasts: characteristics and identification*, 2nd edn. Cambridge: Cambridge University Press.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldman, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Olicer, S. G. 1996. "Life with 6000 genes", *Science*, 274, 546–567.

Himmelreich, U., Somorjai, R. L., Dolenko, B., Lee, O. C., Daniel, H. -M., Murray, R., Mountford, C. E., and Sorrell, T. C. 2003. "Rapid identification of *Candida* species by using nuclear magnetic resonance spectroscopy and a statistical classification strategy", *Applied and Environmental Biology*, 69, 4566–4674.

Koonin, E.V 2005. "Orthologs, paralogs and evolutionary genomics", *Annual Review of Genetics*, 39, 309–338.

Kreger-van Rij, N. J. W. ed. 1984. *The yeasts: a taxonomic study*, 3rd edn. Amsterdam: Elsevier.

Kurtzman, C. P. 2005. "Description of *Komagataella phaffii* sp. nov. and the transfer of *Pichia pseudopastoris* to the methylotrophic yeast genus

Komagataella'', *International Journal of Systematic and Evolutionary Microbiology*, 55, 973–976.

Kurtzman, C. P., and Fell, J. W. eds. 1998. *The yeasts: a taxonomic study*, 4th edn. Amsterdam: Elsevier Science.

Lachance, M. A. 2006. ''Yeast biodiversity: how many and how much?'' *In*: Rosa, C. A. and Peter, G., eds *Yeast Handbook*, Berlin: Springer-Verlag, 1–9.

Lipscomb, D., Platnick, N., and Wheeler, Q. 2003. ''The intellectual content of taxonomy: a comment on DNA taxonomy'', *Trends in Ecology and Evolution*, 18, 65–66.

Lodder, J. ed. 1970. *The yeasts: a taxonomic study*. 2nd edn. Amsterdam: North-Holland.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N. S., Hayes, A., Walsh, M. C., Berden, J. A., Brindle, K. M., Kell, D. B., Rowland, J. J., Westerhoff, H. V., van, Dam K., and Oliver, S. G. 2001. ''A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations'', *Nature Biotechnology*, 19, 45–50.

Rokas, A., Williams, B. L., King, N., and Carroll, S. B. 2003. ''Genome-scale approaches to resolving incongruence in molecular phylogenies'', *Nature*, 425 (6960), 798–804.

Sly, L. I., Iijima, T., and Kirsop, B. 1990. ''100 years of culture collections'', *Proceedings of the Kral symposium to celebrate the centenary of the first recorded service collection*, 13 September 1990, International House, Osaka, WFCC, published by the Institute of Fermentation, Osaka, Japan.

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., and Vogler, A. P. 2003. ''A plea for DNA taxonomy'', *Trends in Ecology and Evolution*, 18, 70–74.

United Nations Environment Programme 1992. *Convention on Biodiversity*. Available from: http://www.biodiv.org/convention/articles.asp [Accessed 26 November 2006].

Vandamme, P., Pot, B., Gillis, M., De-Vos, P., Kersters, K., and Swings, J. 1996. ''Polyphasic taxonomy, a consensus approach to bacterial systematics'', *Microbiological Review*, 60, 407–438.

Wang, Y., Holmes, E., Nicholson, J. K., Cloarec, O., Chollet, J., Tanner, M., Singer, B. H., and Utzinger, J. 2004. ''Metabonomic investigations in mice infected with *Schistosoma mansoni*: an approach for biomarker identification'', *Proceedings of the National Academy of Sciences of the United States of America*, 101, 12676–12681.

World Federation of Culture Collections 2007. Available from: http://www.wfcc.info [Accessed 26 November 2006].