Final report “Evaluation of the number and fate of microbial resources isolated in the frame of DFG-funded projects”

Background

During 2008 and 2010, a project was initiated within the frame of the EU project EMbaRC (www.embarc.org) to screen the holdings of major international European Microbiological Resource Centers (MRC) in order to determine the

- Holdings of prokaryotes, fungi and yeast at all pylogenetic levels from Domain to strains,
- Overlap in holdings of prokaryotes at the genus and phylum level,
- Strength of national MRCs according to taxonomically defined holdings and taxonomic expertise.

This survey was performed to evaluate the basis for a future European MRC infrastructure, based on complementarity rather than on fragmentation and duplication. One observation was the obvious strength of all collections in holdings of the same few prokaryotic phyla. Most MRCs concentrated in addition on certain aspects of microbial biodiversity with only moderate overlap. Annex 1 shows the holdings of phyla in three major MRCs (DSMZ, Germany; CRBIP, France; LMG, Belgium).

During the same period (2009) a survey was performed on the number and taxonomic affiliation of prokaryotic strains used in selected scientific literature in order to determine

- Percentage of these strain deposited in public resource centers,
- Origin of type and reference material used for comparative studies,
- Research emphasis with respect to the taxonomic affiliation of strains.

This study was initiated to evaluate author’s ethical behavior concerning sharing authentic material to allow confirmation of results by peers. Of the 20.000 strains covered, the number of strains deposited into MRCs was extremely low (□1%), and the use of non-authenticated material (□30%) was obvious. Annex 2 is a Table of journals and strain numbers evaluated.

A third survey was initiated in 2012 after members of a 2011 workshop defined ‘key’ criteria for prokaryotic non-type strains worth depositing. During a period of 12 months a questionnaire was linked to the manuscript submission template of several international microbiology journals asking authors to

- Evaluate strains included in the respective manuscripts according to these ‘key’ criteria, and
- For their willingness to deposit these strains in public collections.

In contrast to the witnessed dismal deposition practice as indicated above, while 58% agreed to deposit and only 20 % of responder denied their willingness to deposit strains. Without knowing the actual strain number covered in these journals, authors agreed to deposit about 2.500 microbial strains. Annex 3 is a summary page of journal screened and the final surveymonkey report.

All these numbers cast only a glimpse on the actual situation. The extrapolation of data generated for some journals from a regional to a global level is associated with a high degree of ambiguity and no attempt was made to calculate either the number of strains included in a wider range of the microbiological literature or the fraction of strains deposited in public collections.

The proposal

This was the situation at the time when the proposal ‘Evaluation of the number and fate of microbial resources isolated in the frame of DFG-funded projects‘ (see Annex 4) was submitted to the German Research Council (DFG). Having received the grant STA 184/22-1, Annett Schumann, a Pharmazeut -im-Praktikum with knowledge in microbiology was employed to screen the GEPRIS database for projects with microbiological content (Annex 5), to compile mail addresses of addressees, organize the surveymonkey questionnaire (Annex 6) and to evaluate return questionnaires.

Responses

Of 203 addressees contacted, representing about 250 individual DFG grants, six mails returned undeliverable. Of the remaining ones, 31 scientists (15.2%) replied that no strains were isolated in the course of the granted proposal. Another fraction of 31 scientists (15.2%) responded anonymously to the questionnaire and their data are the basis of the following analysis.

Q2. How many strains are present in your collection

Table 1 is an overview of strains, ordered per taxonomic group and the number of responding scientists. When a scientist collected isolates from more than a single Domain the scientist is recorded under the Domain for which the majority of strains were indicated. Plasmids and Phages were recorded as isolates. The vast majority of the 63.188 isolates collected belong to the Prokaryotes (85.0%), followed by fungi and yeasts (13.4%). The other groups constitute minor components only (□2.6%).

If asked in which area of microbiology work has been performed (nomination of more than one area was permitted), 20 times the fields ‘ecology’ and ‘physiology’ were indicated (29% each), ‘phylogeny’ and ‘taxonomy’ were mentioned about 15 times (20% each), while the research area ‘genetics’ was only mentioned one time. Within the Domain Bacteria the taxa Proteobacteria (medical relevance, N_2 fixation, *Roseobacter*), Firmicutes /Tenericutes, Bacteroidetes and Actinobacteria were mentioned most, followed by individual callings of *Thermotogales* and *Aquificales* as well as candidate taxa. Rarely studies included members of more than a single Domain, such as *E. coli*, *Agrobacterium tumefaciens* and *Saccharomyces cervisiae.*

Table 1. Affiliation of isolates to Domains, summary of strains collected and number of responding scientists.

Domain	Resources collected	Percentage	Responding scientists
Bacteria	52.188	82.7	25
Cyanobacteria	407	0.6	4
Archaea	1.045	1.7	6
Fungi	6.000	9.8	7
Yeasts	2.233	3.6	6
Protozoa	50	0.08	3
Plasmids	992	1.6	6
Phages	148	0.24	6
Microalgae	25	0.04	1

- Prokaryotic isolates are present in 16 large research collections with up to several thousand representatives, including one with about 1.000 Archaea. Seven collections maintain 20 to 100 isolates, while the remaining collections harbor few strains only (5 to 20 isolates).

  During the reporting period, only 26 strains (0.04%) were deposited in the DSMZ and in the SAG, Göttingen.

- Mycological isolates are stored in 6 large collections, maintaining between 200 and 4.000 strains. Three collections have 20 to 1.000 strains while the other four scientists reported the presence of 1 to 20 strains.

  During the reporting period, a single strain only (0.01%) has been deposited in the DSMZ.

- Protozoa are present in 3 collections. While two collections contain respectively 10 and 40 strains, a third scientist did not indicate the number of isolates
- Plasmids are represented in three larger collections, embracing between 100 and 500 resources, two smaller collections harbor 40 and 50 plasmids, while one collection maintain two plasmids only.
- Phages are collected by six research groups maintaining between 3 and 100 strains.
- Only one researcher reported the presence of a small collection of 25 microalgae.

Q3. Estimating the percentage of taxonomically characterized resources

In order to facilitate the future transfer of resources into public service collections, researchers were asked for the taxonomic affiliation of the collected resources, down to the rank of genus or even species. As not all researchers responded to this question the percentage can only be given to the relevant fraction of resources for which answers have been received.

Table 2. Fraction of well characterized strains as percentage of strains per Domain

Domain	Resources collected*	Responding scientists	Number and fraction % of taxonomically characterized resources	Responding scientists
Bacteria	52.188	26	45.831 (87.8)	24
Cyanobacteria	407	4	402 (98.8)	3
Archaea	1.045	7	540 (51,6)	6
Fungi	6.000	7	1.468 (24.4)	7
Yeasts	2.357	6	2.128 (90.2)	4
Plasmids	992	6	840 (84.7)	3
Phagen	148	6	20 (13.5)	3
Protozoa	150	3	105 (70)	3
Microalgae	25	1	25 (100)	1

*Some scientists indicated the presence of members of Domains without telling the number of resources

Prokaryotes

- Ten researchers indicate that 100% of their collection of Bacteria was fully characterized, while 11 other researcher responded that 50% and less of their bacteria were characterized. Some of the well characterized collections harbor mutants or genetically modified bacteria which per se are well characterized at the species level. All cyanobacteria in three collections were fully characterized. Four small collections of Archaea (up to 20 strains) are fully characterized. No response has been received from the curator of the large Archaea collections.

Yeast and Fungi

- None of the seven responding mycologists responded to the actual number of characterized strains. 50% and more of the isolates were characterized in five collections while three yeast collections were characterized between 90 and 100%

Protozoen:

- The large protozoa collection has been fully characterized; the other two to a few percent only.

Microalgae, plasmids and phages

- All 25 microalgae strains of the one responding scientist have been fully characterized.
- Two researchers indicate that their phage collections are fully characterized, while 25% of a third collection was characterized. Three researchers did not respond to this question.
- The plasmids collections of four collections were fully characterized.

Q4. According to your assessment of well characterized strains, how many fulfill the requirement of 'key' strains, worth deposition in public collections?

The 'key' criteria have been defined 2011 during a workshop with journal editors, collection manager and authors and the criteria were made available in the questionnaire (See Table 3, left column [a manuscript has recently been submitted: Publication on 'key' criteria for strains to be deposited has been submitted (Stackebrandt et al. SpringerPlus)]. As these criteria were defined for Prokaryotes, scientists were asked use similar stringent criteria for other taxa. The responses are not straight forward as a single strain may match several, up to all, 'key' criteria. The 11.520 strains which match the stringent 'key' criteria are, as expected, lower than the total number strains in research collections (18.2%) and the number of well characterized strains (22.4%). The highest number of 'key criteria scores were received for 'metabolic uniqueness', followed by 'phylogenetic uniqueness' and 'presence of rare taxa'.

Table 3 Well characterized strains (based upon Table 2 - % of taxonomically characterized resources) matching the ‘key’ criteria. Multiple assessments were allowed.

‘Key‘ criteria	Number of assessment
Phylogenetic uniqueness (e.g. < 98% 16S rRNA gene sequence similarity to nearest neighbor, sequences of other genes and spacers)	2.393
Metabolic uniqueness (e.g. new or modification of known pathways, new metabolites)	5.471
Morphological/cytological uniqueness	367
Genomic uniqueness genome size, architecture, regulation)	623
Resources with ‘draft’ or completely sequences genomes	702
Second strain for a species for which only the type strain is available	134
Representatives of ‘rare‘ taxa	1.530
Others	300

Q5 Willingness to deposit

As the number of strains for which the researcher indicated an interest in deposition (44.790) exceeds the number of resources indicated in Q4 (11.520) it must be assumed that several researchers were referring to numbers indicated in Q2 (63.188) or Q3 (50.904) but not in Q4. Nevertheless, most researchers would like to see their strains maintained long-time in public collections to safeguard the investment made in isolation and scientific research.

SUMMERY

Research collections encompass a rich diversity of small to medium sized collections with restricted phylogenetic depth

- About 15 percent of grant recipients responded that no research collection has been established. Of the remaining addressees, another fraction of 15 percent agreed to respond to the questionnaire, which is an encouraging return.
- About 63.000 resources were maintained in the collections of the 31 researchers, containing about 81.1% characterized specimen.
- Collections of Bacteria, cyanobacteria, yeasts, plasmids, protozoa and microalgae are well characterized, but not so those of filamentous fungi and phages. A minor fraction of the well characterized strains (total of 27) have already been deposited in public culture collections
- Most researchers agreed to see their resources be maintained in public collections
- Not unexpected was the concentration of isolates affiliated to the phyla Proteobacteria, Bacteroidetes, Firmicutes and Actinobacteria as these
  - are the ones which can be isolated most easily from any environment on standard cultivation media,
  - embrace the highest number of named species,
  - are also well represented in the scientific literature and
  - constitute the major holdings of most public collections

The latter point is of special interest as for the majority of isolates belonging to the four main phyla expertise of their routine maintenance and long-term storage is already available in most public resource centers.

Nevertheless, the indication that a strain matches one or more of the ‘key’ criteria for their deposition are most likely biased by the researchers’ personal assessment and the researchers’ arguments for their deposition evaluated, discussed and negotiated with curators of public collections. Assuming that not all collection holders responded to the questionnaire it can be extrapolated on the basis of 15% returned questionnaire that at present research collections of DFG grant recipients hold more than 20.000 microbial resources worth depositing in public collections.

Recommendation

It is recommended that the DFG guidelines to grant recipients requests the disclosure of numbers and taxonomic affiliation of microbial isolates in research collections and requests for a strategy for long-term preservation (intra-or extramural) of microbial resources.