Electronic supplement

Supplement: 1. Introduction

Conservation of microbial organisms can take place either in the natural environment (in-situ conservation) or outside the natural environment through technologies for long term preservation of microbial materials in frozen form, such as in liquid nitrogen or freeze-dried (ex-situ conservation). Public service (ex-situ) microbial collections (PSMCs) certify the quality of microbes as research materials and this supports knowledge production since researchers can use the certified material, avoiding duplication of effort. The role of PSMCs is based on the acquisition, authentication, conservation and distribution of living microbes and their replicable parts (e.g., DNA, genomes, plasmids, viruses) along with important information about their properties. PSMCs' specific added value consists not only in identifying the taxonomic nature of microbes, but also in characterising their biological function, and increasingly, sequencing them to identify the genetic code. Such information is organised in databases with molecular and physiological information diffused on PSMCs' internet sites (Sigler, 2004; Stern, 2004; Arora et al., 2005).

Specifically, the use of certified materials from public culture collections diminishes the cost from mistakes in cumulative research (Furman and Stern, 2006) and decreases the cost of finding appropriate materials (Evenson and Kislev, 1976; Gollin et al., 2000; Visser et al., 2000). [1]

In their role PSMCs support public sector and private sector research across different fields, including human health, agriculture, energy, environmental

[1] Additionally, the availability of large amounts of original and/or derived biomaterials supports high throughput screening of the activity of small molecules against drug targets (Parry, 2004; Rai et al., 2008). Finally, as general research infrastructures, PSMCs increasingly hold microbial organisms that are used for regulatory purposes in biosecurity, health and food safety and biodiversity preservation.

sciences and technology, evolutionary science, and taxonomy. A well-documented example of their biotechnological and economic importance is the discovery of the enzym Taq polymerase, which was recovered from the micro-organism *Thermus aquaticus* collected in a thermal spring of Yellowstone Park in the US and deposited in the American Type Culture Collection (Staley et al., 2010). This enzyme was successfully applied in the polymerase chain reaction, which is now one of the key processes in contemporary biotechnology. Other examples illustrate the breadth of use, even of a single micro-organism, such as the widely used cultures of the yeast *Saccharomyces cervisiae*, of which approximately 50.000 samples have been deposited in specialized and general PSMCs and which plays a role in fermentation in bakery, beverages, dairy products, as well as protein, amino acid and vitamin source in animal feed (Daniel and Prashad, 2010).

A variety of microbial resources are conserved in PSMCs. These resources are used as authenticated reference material, which can be reliably cloned for use in cumulative research in microbiology, given the rapid mutation rates of microbial organisms which make cumulative research not possible otherwise. A first category of resources is given by the general research resources, related to scholarly publishing in microbial research. In conformity with the traditional role of the PSMCs, many microbial strains that are held in the collections were deposited upon the publication of new research in scholarly journals, with the view to conserve them for possible future uses and for the verification of the published research results. These general research resources are typically the bulk of the holding of the general research collections that will be described in section 2.2 below (cf. also table 2).

With the genomic revolution in the life sciences, a new set of activities have been developed in culture collections, increasingly preserving and making available new biological materials such as rDNA, plasmids or vectors for genetic engineering, to perform research in microbiology and genomics. Culture collections also provide services based on expertise regarding the identification and handling of microbial strains, and the use of related databases and bioinformatic tools (OECD, 2001).

Supplement: 2.2. Responses to increasing interest from industry

Explanation to table 1

DSMZ is one of the largest culture collections and a recipient of substantial public funding. Its orientation is mainly towards conservation, including collection and storage. DSMZ supplies its holdings to a large number of institutions worldwide, including industry and non industrial users such as academia, hospitals and other research organisations. Furthermore DSMZ offers safety and patent deposits of microorganisms (http://www.dsmz.de, accessed on 12th Mar. 2013). By contrast, Forintec Canada Corp. is a commercial company with a specialised industry oriented culture collection that holds a collection of micro-organisms causing decay in wood. In this case, the company uses its collection to provide a broad range of commercial research services to the timber industry, including identification of microorganism samples provided by its clients (http://www.forintek.ca, accessed on 15th Dec. 2009).

A third category of research collections is exemplified by LMG Ghent. This is a research and infrastructure collection. It forms part of the BCCM group of collections, consisting of one public and three university collections, and is a coordination induced by the Belgian Government. In this case, the collection LMG Ghent is largely funded by the government but is hosted by a University. Similarly to DSMZ it also holds a substantial collection, and like DSMZ and Forintec Canada Corp, it holds a substantial in-house research capacity. However, unlike DSMZ and Forintec Canada Corp, it is located in a university. LMG Ghent distributes microbes by material transfer agreements, and offers several types of deposit services for microbes, including safety and patent deposits. LMG Ghent provides research services to both public and private institutions, and conducts phylogenetic studies, the identification and the fine-typing of various bacterial groups, including bioprospecting and sustainable agriculture (http://bccm.belspo.be/, accessed on 5th December 2009).

Finally, some other university laboratories specialise in a sub-field of microbial research and have built important in house collections that support that research. Some materials from these collections are exchanged with collaborating

scientists working in other research materials, but these collections typically do not offer a systematic distribution service. As our focus in this paper is mainly on the institutional and organisation structure of PSMCs holding publicly available micro-organisms, we did not include this fourth category of collections in our systematic survey of culture collections. The survey mainly focuses on the PSMCs that are members of the World Federation of Culture Collections, which encompasses collections of the three first categories presented in table 1.

Supplement: 4. Materials and methods

A worldwide survey of culture collections

In order to assure a high response rate a pilot survey was first analysed on 12 microbial PSMCs. In order not to overload respondents, three separate questionnaires were designed together with representatives of WFCC and MIRCEN, and distributed electronically with a two-month interval between January and July 2006 to all the members of WFCC and MIRCEN networks from Europe, Africa, the American continent, Asia and Oceania.

The first questionnaire contained the most relevant information for the present analysis and is the focus here. Besides using an online surveying approach, the questionnaire was also sent by post to 170 randomly selected PSMCs stratified by OECD membership of the country of origin.

All the collections of our final sample are members of WFCC (whether of WFCC only or of both WFCC and MIRCEN). Therefore these collections all have a similar, internationally recognized, level of quality management, in spite of the differences in funding and conservation choices which will be analyzed below.

It should be noted that while 52% of PSMCs in the population frame are located in OECD countries, the surveyed sample over-represents OECD collections (67%).[2] Based on WFCC (2005) registered categories of collections, the

[2] Due to this over representation the results are somewhat more representative for collections located in OECD countries. However, the difference is not dramatic and the inclusion in the

majority were held by universities (42%) and by governments (41%); the remaining collections were catalogued by WFCC as being semi-governmental (8%), private (4%), industry (1%) and inter-governmental (1%). Table ES1 presents a general overview of the main features of PSMCs' holdings of micro-organisms, segregated according to the geographical location of the PSMC.

Table ES1: Holdings of microbial strains in the sampled PSMC (standard deviations in parenthesis)

PSMCs	Avg. number of strains	Avg. number of TS	Avg. number of non-TS
OECD[a]	5,877 (13,294)	527 (1,206)	5,349 (12,876)
Non-OECD	2,775 (3,562)	214 (450)	2,561 (3,545)
Total	4,853	424	4,429

N=103. [a] OECD denotes that the PSMC is hosted by an OECD country (excluding Turkey and Mexico, including Brazil, India and China). As opposed to the variables used in the econometric model this variable includes USA. Standard deviations in parenthesis.

Based on the data from the sampled collections it can be seen that most of them receive substantial public funding, with 54% of the collections being associated with more than 80% of public funds (Table ES2). It can also be observed that PSMCs with large stocks of micro-organisms also depend on commercial funding, as do collections with a high proportion of TS. This might indicate that private/public cost-sharing plays a significant role in maintaining a large level of stock of micro-organisms, necessary for being able to answer requests for large quantities of materials by the private sector, hence in investing in the economic option value of the collection, and also in specialising in type strain holdings.

The data also reveal that collections with a high proportion of TS in their holdings are heterogeneous regarding the size of their stocks, geographical location and scope of the collection, as well as regarding conservation focus and user groups. This group of collections includes some of the high profile collections within the category of 'research collections' (c.f. Table 2), that hold

model of the OECD variable controls for the locations in OECD countries. Hence we hold that the general results are not likely to be affected in any significant way.

more than 15,000 strains. But it also includes some medium sized and small research collections, with fewer than 250 strains in their holdings, which include incidental collections in both universities and medically oriented organisations. TS constitute approximately 10% of the microbe stock among all sampled PSMCs. The PSMCs with high TS ratios are located mainly in OECD countries, but also in Brazil, India, China, Senegal and Egypt. Lastly, it is observed that it is more common for PSMCs located in OECD countries to receive public funding as compared to collections located in non-OECD countries.

Table ES2: Profile of the PSMCs stratified by level of public funding (Means, with standard deviations in parenthesis).[a]

	0% (n=15)	1-40% (n=8)	41-60% (n=8)	61-80% (n=13)	81-100% (n=56)	Total (n=103)
Microbial holdings of PSMCs						
Log of size of stock of Micro-organisms	6.62 (1.65)	8.21 (0.74)	7.66 (1.39)	7.44 (1.31)	7.45 (1.63)	7.40 (1.54)
Share of TS from total stock	0.17 (0.33)	0.09 (0.14)	0.45 (0.43)	0.07 (0.09)	0.13 (0.22)	0.15 (0.25)
Location						
OECD	67%	88%	88%	85%	88%	82%
Flows from PSMCs						
*Industry specialisation (*Share of distributed microbes to the industry*)*	0.23 (0.35)	0.40 (0.19)	0.29 (0.20)	0.31 (0.22)	0.18 (0.25)	0.23 (0.26)
*Academia specialisation (*Share of distributed microbes to academia	0.61 (0.42)	0.54 (0.21)	0.67 (0.20)	0.48 (0.32)	0.61 (0.35)	0.59 (0.34)
*Inflow of microbes (*percentage of received strains that the PSMC sourced from other PSMCs*)*[b]	1.67 (1.29)	1.88 (1.73)	1.25 (0.89)	0.77 (0.60)	1.00 (1.21)	1.17 (1.2)
Transaction modes						
Percentage of PSMCs Charging fees	60%	88%	88%	92%	57%	67%
Share of PSMC receiving strains regulated by MTA	0.38 (022)	0.44 (0.25)	0.38 (0.26)	0.33 (0.20)	0.38 (0.27)	0.38 (0.24)

N=103. [a] Due to the low number of observations in the category 1-20% public funding, this category was merged with the 21-40% category. [b] Table 5 provides the definition.

Specialisation in providing microbes to the industry and academia is associated with low-to-medium and medium level of public funding, respectively. This indicates that also industry relies on the public research infrastructure. In addition, a low level of public funding is associated with higher levels of inflows of microbial resources, i.e., sourcing of microbes from other PSMCs instead of sourcing the material from non PSMC, possibly due to lacking access to the broad research infrastructure in academia and hospitals.

In terms of transaction modes, Table ES2 indicates that the fee modality is mostly used by collections associated with medium-level share of public funding. Interestingly, the use of MTAs is rather evenly distributed across different ranges of funding sources. This suggests that formal private sector like contracts have been considerably adopted by PSMCs in general for an important part of the exchanges.

Modelling the PSMCs' management of type strains

PSMCs face a choice of strategy regarding which microbes to focus on. Budgetary constraints may impose limits to investing in specialized personnel, storage space and maintenance. Likewise, increase in TS microbial material implies an opportunity cost in terms of forgone benefits from storing a greater stock of other types of microbes. Further, since TS are search tools and hence have public good characteristic associated with significant positive externalities, it is expected that collections that focus more on storing TS might be more dependent on public funding. The conservation focus can be proxied by the ratio of TS that facilitate knowledge accumulation to the total stock of microbes held in the collection.

The question regarding PSMCs' strategy of investing in TS for knowledge accumulation needs to be complemented by the question of the orientation of the distribution of the microbial material, e.g., to traditionally typically public but also to typically commercial organisations. The demand for TS by basic research and the industry is increasingly interlinked. There is dependence of the industry

on PSMCs in order to gain access to the TS and there is influence by the industry regarding the decision making of public collections regarding their strategy regarding the level of TS supply. Both the conservation strategy and distribution focus of PSMCs are interlinked and can be seen as a joint decision process shaped by the PSMC microeconomic institutional context (Smith, Pers. Comm.).

A bivariate Tobit model (Greene, 2003) is estimated. Compared to traditional ANOVA or parallel multiple regression models, the bivariate Tobit model is better suited in the case of censored dependent variables and to surmount the potential interdependency of PSMCs' conservation choice regarding the level of TS supply and the orientation of the flow of microbial material, e.g., the extent of industry orientation (cf. for a similar use of the bivariate Tobit see e.g. Almus and Nerlinger, 1999). The model is composed of two estimable equations, as specified below:

The first of the two equations (c.f. Eq. 1) refers to the conservation model where the dependent variable *TSSHARE* stands for the ratio of type strains to the stock (total number) of microbes held by the collection. This ratio is expressed as a function of the extent to which the collection depends on public funding, once controlled for other key factors associated with the institutional environment in which it operates, its size or scale in terms of the total stock of microbes, being part of the broader network of public service collections, whether they charge a fee for sharing their material, the geographical context of the collections and the specialisation of the PSMCs into various categories of microbial resources (c.f. table ES2).

The second submodel (c.f. Eq 2) is associated with the industrial orientation (flow) model where in this case the dependent variable, *FLOWIND,* stands for the proportion of a collection's distributed microbes going to the private industry, as opposed to public sector affiliated users such as academia or public hospitals.

It is assumed that the error terms (μ and ε) follow a bivariate normal distribution (Green, 2003). The joint model is estimated using the maximum likelihood method. The Summary statistics of each of the two dependent variables as well as the control and independent variables appear in table ES3.

In what follows we provide an interpretation and rationale for the choice of the variables included in the model.

Table ES3. Summary statistics of the variables used in the econometric model [a]

Variable	Description	Mean	Std. dev	Min – Max
Dependent variables				
TSSHARE	Share of type strains over total number of strains in the PSMC's holding	0.15	0.25	0-1
FLOWIND	Share of the distributed microbes provided (flow) to the industry	0.23	0.26	0-1
Main explanatory variables				
SUPPORTNO	The PSMC receives no funding from public bodies (no funding from public bodies= 1, 0 otherwise) (1)	0.15		0-1
SUPPORTLOW	The PSMC receives between 1 and 40% of its funding from public bodies (yes= 1, 0 otherwise)	0.11		0-1
SUPPORTMEDIUM	The PSMC receives between 41% and 60% of its funding from public bodies (yes = 1, 0 otherwise)	0.08		0-1
SUPPORTHIGH	The PSMC receives between 61% and 80% of its funding from public bodies (yes= 1, 0 otherwise)	0.13		0-1
Control variables				
FEE	The collection does charge a per unit fee for providing microbes (yes = 1, 0 otherwise)	0.67		0-1
INFLOW	Interval variable: percentage of received strains that the PSMC sourced from other PSMCs, as opposed to from for example academia and hospitals (0%, 1-20%, 21-40%, 41-60%, 61-80%, 81-100%)	1.17	1.20	0-5
FLOWACAD	Share of the distributed microbes that are provided to academia and hospitals	0.60	0.34	0-1
STOCK	Natural log of number of strains in the collection's stock (type strains and non-type strains)	7.40	1.54	3-11.3
PR	Latent variable representing whether the collection received any strains regulated by Material Transfer Agreement (MTA) or contract. The variable is constructed from predicted probabilities of three instruments.	0.38	0.24	0-1
OECD	Collection is hosted by an OECD country (excluding Turkey, Mexico and USA, including Brazil, India and China) (yes = 1, 0 otherwise)	0.82		0-1
USA	Collection is located in the USA (yes = 1, 0 otherwise)	0.10		0-1
FUNGI	The PSMC holds this category of micro-organisms (yes=1, 0 otherwise)	0.52		0-1
YEAST	The PSMC holds this category of micro-organisms (yes=1, 0 otherwise)	0.45		0-1
ALGAE	The PSMC holds this category of micro-organisms (yes=1, 0 otherwise)	0.19		0-1
BACTERIA	The PSMC holds this category of micro-organisms (yes=1, 0 otherwise)	0.67		0-1
OTHER	The PSMC holds this category of micro-organisms (yes=1, 0 otherwise)	027		0-1

N. observations: 103; [a] Values correspond to the year 2005. (1) The survey data distinguishes between ranges of public funding (0, 1-20, 21-40, 41-60, 61-80, 81-100%). (1) Due to the low number of observations in the category 1-20% public funding, this category was merged with the 21-40% category.

The conservation strategy submodel

In order to address the question of the public mandate effect on the collections' conservation strategy, explanatory variables are included in the conservation regression which proxy the public influence over the collection related to the level of public funding received by the PMSCs. The independent categorical variables *SUPPORTHIGH, SUPPORTMEDIUM, SUPPORTLOW* and *SUPPORTNO* denote that a PSMC receives between 61-80%, 41-60%, 1-40%, 0% of their funding from public bodies, respectively. This set of categorical variables serve as comparison to the effect of being funded by public budgets beyond 80%, which represent a fully or close to fully publicly funded collection.

It is expected that both the heavily and the intermediate public funded collections are more specialized in TS than the group that derives all its income from private sources (c.f. table ES2). If such expectation is met it would also imply that in the future PMSCs that receive mixed public-commercial funding will increase in number as well as in the quantity and quality of the microbes that they hold and distribute.

Moreover, PMSCs in general are diversifying their income sources, such as illustrated by the increasing use of fees. Another control variable reflects the degree to which the handling of acquired microbes was subject to a MTA or formal contract and reflects the institutional environment in which the PSMC operates (*PR*). It controls for the existence of a traditionally informal reciprocity-based tier or a more formal and legalistic environment. We assume that formal transactions reflect mostly the behaviour of the providing PSMC as it is in the interest of the provider of strains to protect the property rights. There is not *a priori* expected effect of formalisation of transactions on the conservation outcome.

Another characteristic of PSMCs relates to their scale of operation approximated by the total stock of type- and non-type strains (*STOCK*). While a collection may be more conservation oriented in absolute terms by having a large stock of type strains, it generally contains a significantly larger stock of non-type strains.

Controlling for *STOCK* in the model allows accounting for such dilution effect. We also control for inflow of strains by including the variable *INFLOW* and for the relative distribution of microbes to academia and hospitals over the total flow of microbes from PSMCs (*FLOWACAD*).

Since it is expected that OECD countries would on average have a higher proportion of privately owned research collections, as compared to more general-purpose taxonomic collections that tend to prioritise type strains, the location of a PSMC in an OECD country is controlled for.[3] In this regard, the US is non-representative due to the special characteristics of its research funding in the life sciences, a property rights regime with strong commercial attributes, and due to economies of scale which have led to the presence of a high degree of centralization of culture collection facilities. Hence we also control for whether a PSMC is located in the USA.

Regarding the transfer mode, whether a collection charges a fee when distributing microbes from its own collection is also considered through the variable '*FEE*' since it is expected to affect the industry orientation of the collections. Finally, we also include a set of variables to control for the various categories of microbes that are held in the collections and which were reported in the survey. Finally, as explained in table ES3, we also control for the specialisation of the PSMCs in various categories of microbial resources, mainly fungi, yeasts, algae and bacteria.

The industrial spillover submodel

The second part of the model focuses on the potential spillover effect of investment in public general-purpose collections. A censored dependent variable is used to approximate the industry orientation (*FLOWIND).* It is expected that heavily publicly funded collections are less likely to pursue an industry oriented

[3] It should be noted that in order to proxy the presence of demand from biotechnology industry in different countries, Mexico and Turkey are coded as non-OECD, while Brazil, China and India are coded OECD.

strategy, since the government, rather than industry, is assumed to prioritise basic research.

The covariates included in the model include the following: The variable *PR* denotes a formal approach to the distribution of microbes, with an expected positive effect on the provision of microbes towards the industry. The variable *STOCK* is included since the size of the collection may also affect industry orientation, although we have no prior expectation about the effect of the scale of the collection except for very small PSMCs which are expected to be less likely to have the capacity and competence to supply to the private industry. Being part of the broader network of public service collections is controlled for by the variable *INFLOW*.

The control variable FEE is included in this second submodel[4]. Fees are charged by PSMCs that provide material to the industry, but also to other users. In fact, fees tend to be relatively low and hence they generally do not constitute an access barrier for industry. Rather, a collection that charges a fee is expected to signal that the collection is more commercially oriented. As such, we expect fee status to signal a policy orientation by collections rather than a direct income generation strategy. In this case, an endogeneity bias is not expected to be a major problem in the model. In other words it is not likely that the supply of strains to industry alone would induce collections to decide to whether or not charge a fee. Instead we expect the reverse relationship: charging a fee signals an industry orientation.

Another control variable, '*FLOWACAD'*, is included in the model as it can be expected to be negatively associated with the industry orientation by representing recipients of PSMCs materials that traditionally have been associated with public research as opposed to private research, even if industry orientation and academic orientation are not automatically mutually exclusive.

[4] It should be noted though that charging a fee does not automatically signal commercialization or a *de facto* industry orientation, but rather whether the collection has decided on an industry orientation policy or not.

The variable *OECD* is expected to positively affect industry orientation, due to a perceived higher industry demand in such countries.

Supplement: 5. Results and discussion

The results from the estimated bivariate Tobit model appear in Table ES4. The left hand side of the table presents the estimated results for the conservation strategy regression and the right hand side shows the estimates for industry orientation regression. The Wald test suggests that taken together the variables explain the variability in the dependent variables in a satisfactory way (*Wald chi2(18) = 278.25; Prob > chi2 = 0.00*).

The correlation coefficient between the two error terms or covariance term across the two equations is statistically significant and positive implying that there is, as expected, a correlation between the conservation profile of TS and an overall provision of microbial strains (TS and non TS) to industry.