**Systematic literature searches and analysis of recovered data**

We performed two systematic literature searches using the Scopus database to cover our research goals. Settings for each search and the analyses performed on the retrieved data are detailed below.

*1. Systematic literature search 1 (Search1)*

This search was performed to cover the use of flow cytometry on scientific research using the keywords “flow cytometry” restricted to the title of the articles (Table S1). Two main analyses were performed on the retrieved articles from this search.

* 1. Network analysis

Co-occurrence analysis was performed on a subset of 1,880 articles focused on the use of flow cytometry for microbial characterization. These articles were selected by reading the Title, Abstract and Keywords of the 12,960 articles recovered from Search1. This analysis was performed using the software Vosviewer (van Eck & Waltman, 2010). The following settings were used for the analysis: type of analysis: co-occurrence, unit of analysis: title and abstract, counting method: full counting, minimum number of occurrences of terms: 10. All other parameters were used with the default setting.

* 1. Frequency analysis

Frequency analysis was used to identify main soil research goals assessed by flow cytometry. This analysis was based on a subset of articles of the Search 1 including the term “*soil*” in Keyword, Title or Abstract (Table S2). Articles were filtered by pertinence (FCM use to characterized bacteria in soil samples). Then the research goal and the type of analysed cells, native (i.e. cells that naturally occur in soils) or spiked (i.e. cells added to soils), were recorded and considered in the frequency analysis.

*2. Systematic literature search 2 (Search2)*

This search was performed to recover articles developing or using protocols for the extraction of bacteria from soil. It encompassed three sets of search keywords and six steps (Figure S1, Table S1). The first two steps involved a search on Scopus database using different combination of keywords related to soil matrix, methods of soil bacterial extraction and soil microbes. Results of these two steps could be biased by the combination of keyword used; therefore, in the third step we used this results as input for litsearchr package (Grames et al., 2019). This R package facilitate a quasi-automatic search strategy development for systematic reviews. In the fourth step we compared and combined results from the previous steps, and in the next step we filtered retrieved articles considering pertinence to the search goal by reading title and abstract or the complete article when required. In the last step the database was enriched, recovering data for the following 14 variables: cell type extracted, method used for cell analysis after extraction, soil diluent and soil/diluent ratio used for extraction, soil treatment, number of steps and number of step repetitions performed in soil treatment, methods used for dispersion, separation and purification, reagents used for density gradient centrifugation, protocol validation, nature of extracted cells (native or spiked), and article cited as reference of the protocol used or adapted. Missing reference articles were added to generate the final database (Table S3). Additionally, information of recovery efficiency was extracted from all articles, recording type of extracted cells, type of protocol used, use of soil, soil conservation (fresh or dried) and soil physicochemical properties. Although recovery efficiency has been reported based on different methods (e.g. microscopy count, plate count, microbial activity) we decided to recover data based on microscopy count, as it is usually considered as a reference method and therefore more data were available for it. Four main analyses were performed on articles from this search.

2.1 Frequency analysis

Frequency analysis was used to characterize information recovered on the variables recorded in the database, and to identify their major trends. This analysis was also performed on a subset of articles in which flow cytometry was the method used to characterize bacterial cells.

2.2 Recovery efficiency analysis

This analysis was performed to identify potential factors affecting recovery efficiency. Spearman correlation test was used to evaluate the correlation between recovery efficiency and soil physicochemical properties, while Wilcoxon and Kruskal-Wallis test were used to evaluate the effect of categorical variables (e.g. soil use, soil protocol, soil texture) on recovery efficiency.

2.3 Network analysis

Network analysis was performed to visualize relationships among the retrieved articles. The network was based on article citations. A table of connections was produced based on articles cited as reference for the implemented extraction protocol. This table was imported, as a directed network, and analysed using Gephi software (Bastian et al., 2009). Plotting of the network was performed using the Fruchterman Reingold layout and scaling nodes (i.e. each article of the network) according to the number of connections (i.e. number of citations). Nodes were also coloured according to the microbial group analysed (e.g. bacteria, virus, fungi or their combinations) in order to highlight connections among protocols for different microbial groups. Visualization also included the identification of the protocols used in combination with flow cytometry.

2.4 Classification and correspondence factor analyses

These analyses were performed considering two different goals. The first goal was to classify text information provided in the abstracts of the articles and to identify terms and articles assigned to each class. The classification was performed using the Reinert’s method and the post-hoc correspondence factor analysis as implemented in Iramuteq software (http://www.iramuteq.org/). The classification was based on a corpus text including the abstracts of the articles and considering the ID of each article as an illustrative variable for the correspondence factor analysis. The second goal was to identify classes of categorical variables and protocols assigned to each class. This analysis was also performed using the Reinert’s method and the post-hoc correspondence factor analysis. The classification was based on a matrix of six categorical variables of soil extraction protocols (diluent, diluent/soil ratio, dispersion treatment, separation treatment, purification treatment, and nature of analysedcells) and considering the ID of each protocol as an illustrative variable for the post-hoc correspondence factor analysis. For each categorical variable, mutually exclusive categories were defined considering frequency analysis.

**References**

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media.*

Grames, E. M., Stillman, A. N., Tingley, M. W., & Elphick, C. S. (2019). An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods in Ecology and Evolution*, *10*(10), 1645–1654. doi:https://doi.org/10.1111/2041-210X.13268

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. doi:10.1007/s11192-009-0146-3