

A Systematic Approach to the Interrogation and Sharing of Standardised Biofilm Signatures

Anália Lourenço^{*}, Andreia Ferreira, Maria Olivia Pereira, and Nuno F. Azevedo

Abstract. The study of microorganism consortia, also known as biofilms, is associated to a number of applications in biotechnology, ecotechnology and clinical domains. A public repository on existing biofilm studies would aid in the design of new studies as well as promote collaborative and incremental work. However, bioinformatics approaches are hampered by the limited access to existing data. Scientific publications summarise the studies whilst results are kept in researchers' private *ad hoc* files.

Since the collection and ability to compare existing data is imperative to move forward in biofilm analysis, the present work has addressed the development of a systematic computer-amenable approach to biofilm data organisation and standardisation. A set of in-house studies involving pathogens and employing different state-of-the-art devices and methods of analysis was used to validate the approach. The approach is now supporting the activities of BiofOmics, a public repository on biofilm signatures (<http://biofomics.org>).

Keywords: infection-causing microorganisms, biofilms, legacy data, data standardisation.

Anália Lourenço · Andreia Ferreira · Maria Olivia Pereira
IBB - Institute for Biotechnology and Bioengineering, Centre of Biological Engineering,
University of Minho, Campus de Gualtar, 4710-057 Braga – Portugal
e-mail: {analia,mopereira}@deb.uminho.pt, af18048@gmail.com

Nuno F. Azevedo
LEPAE, Department of Chemical Engineering, Faculty of Engineering, University of Porto,
Porto – Portugal
e-mail: nazevedo@fe.up.pt

^{*} Corresponding author.

1 Introduction

Microorganisms have evolved various strategies to survive and adapt to the ever changing environmental conditions. The formation of biofilms is an example of such adaptation strategies. Biofilms are structured communities of microorganisms that are able to survive virtually everywhere in Nature because of their ability to adhere to a surface and embed in a protecting, self-produced matrix of extracellular polymeric substances [1-2].

Due to their persistence and resistance to antimicrobial agents, biofilms are at the basis of a range of problems in areas of great importance to human development, such as hygiene and food safety in the food industry [3], nosocomial infections [4-7], acute and chronic infections [8-10], and clogging and contaminations in drinking water systems [11-12]. The interest of the scientific community for these problems is obvious and much research has been devoted to the understanding of initial cell adhesion and biofilm formation phenomena in the last decades.

Similarly to other domains, biofilm research has benefited from the technological evolution occurred in the last decades [13-14]. The development of high-throughput biofilm-forming devices (e.g. the 96-well plate, the microtiter plate with coupons and the Calgary device) has enabled the simultaneous testing of large sets of conditions. The implementation of automated spectrophotometry and microscopy systems (e.g. scanning electron, atomic force, and confocal laser scanning microscopy) has empowered the large scale analysis of biofilm features, such as biofilm biomass, biofilm activity and microbial composition. The “omics” platforms are supporting the study of the transcriptome [15-16], proteome [17-19] and metabolome [20-21] of biofilms.

Therefore, biofilm research is becoming data-intensive and thus, the need for suitable bioinformatics tools, namely databases on existing studies, is compelling [22]. However, scientific publications only summarise the obtained results and data files are not submitted to any public location, remaining on researchers’ private archives. Besides limited access, no protocol exists on how to document biofilm studies, i.e. the minimum information required to guarantee self-contained and explanatory documentation. Data files vary widely from laboratory to laboratory, from researcher to researcher and even from a researcher’s experiment to the next (Fig.1). Files lack comprehensive documentation on the experimental conditions evaluated (often they are only mentioned by abbreviations) and, even more important, on data quantification terms, such as the units of measure used.

The contribution of the present work is a novel computer-amenable approach to the systematic collection and storage of experimental data related to biofilm studies. Aiming to enable the large-scale classification, retrieval, and dissemination of biofilm signatures, the goal of this approach is to help incorporate existing studies.

The remainder of the paper is organised as follows. Section 2 describes the minimum set of information required to describe a biofilm study unambiguously, and the algorithm of the proposed approach. The proof of concept and the implementation are discussed in Section 3. Finally, some conclusions and future work are summarised.

2 Characterisation of Biofilm Signatures

Usually, biomedical databases derive their contents from literature and previously existing databases, or electronic records of some sort, and manual curation plays a decisive role in quality assessment. Biofilms differ from this scenario in that the data is not deposited in any location, but in the private *ad hoc* files of their owners. Therefore, biofilm data collection is dependent on the willingness of researchers to deposit their data and help in their correct interpretation and standardised structuring.

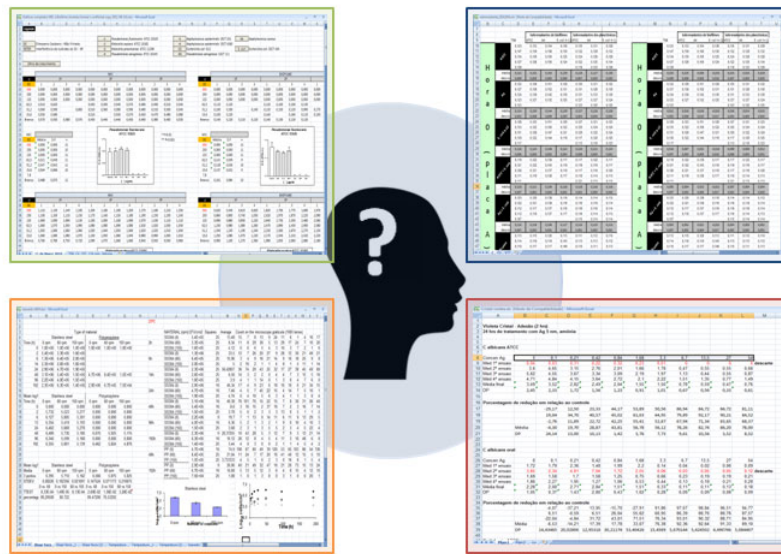


Fig. 1 Examples of biofilm data files from published biofilm studies.

2.1 Modelling Requirements

Biofilm signatures should be self-contained, i.e. there should be enough information about the purpose and execution of the experiment as well as the analysis of results. The signature should include information on the operating procedures (i.e. devices and associated settings that emulate the environment under study), the statistical validation (i.e. the number of replicates and reproductions performed), the methods of analysis and associated units of measure, and, of course, the actual result set (Fig.2). It is important to emphasise that results are fully comparable only for similar methods under identical conditions.

Biofilm signatures should also be self-explanatory, i.e. any researcher should be able to interpret their details. However, controlled vocabulary is scarce. Microorganism characterisation can be obtained from the NCBI Taxonomy Browser [21]

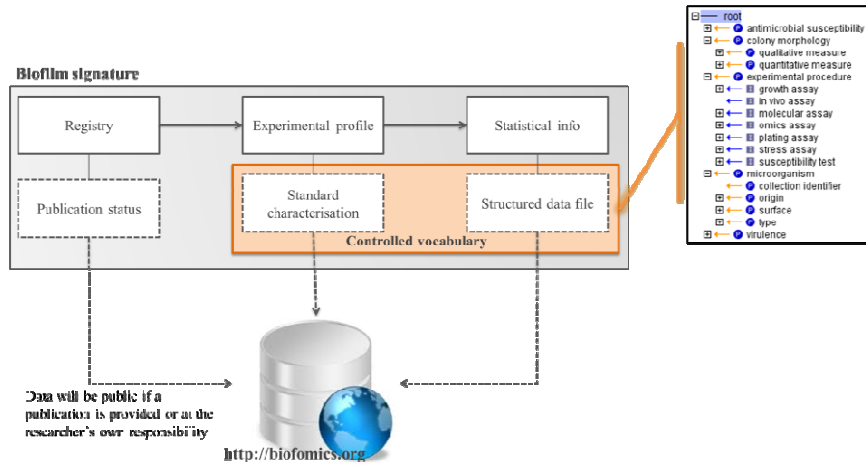


Fig. 2 Conceptual view of biofilm signature characterisation.

and terminology on antimicrobial agents can be retrieved from databases such as CAMP [22], but this is by no means sufficient.

In the scope of our bioinformatics initiative, a novel ontology on biofilm studies is being prepared to describe the overall concept of biofilm signature, i.e. a set of information on the microorganisms, the experimental procedures, the antimicrobial susceptibility and virulence profiles.

2.2 Systematic Organisation

Since most researchers already have their data in this format, data submission is for now simplified to data re-organisation and nomenclature standardisation on the familiar format of Microsoft Excel worksheets. Our approach aims to be flexible in terms of the type and number of conditions tested as well as the data series that are to be associated, accounting for the variability inherent to biofilm studies.

The modelling of biofilm signatures is challenged by the increasing complexity and variability of the studies. Therefore, the Excel file is customised according to researcher's specifications. As illustrated in the pseudo-code in Fig.3, experimental conditions are distributed into worksheets such that single value conditions (i.e. the conditions that are constant to the whole experiment) are at the first worksheet and the multi-value conditions (i.e. the conditions that are tested in the experiment) are organised hierarchically in descendent order. This hierarchical structure grows vertically – as many sub-levels as condition tests and experiment reproductions there are – and no restriction is imposed to the length of the data series of each-level (i.e. the number of test replicates).

```

Input:  $E_{cond}$ : list of experimental conditions;  $M_a$ : list of
methods of analysis
Output:  $F_{excel}$  customised file

 $E_{cond\_single} \leftarrow singleValueConditions(E_{cond})$ 
 $E_{cond\_multi} \leftarrow E_{cond} - E_{cond\_single}$ 
 $w_s \leftarrow initialiseWorksheet('experiment constants')$ 
for  $c \in sortByType(E_{cond\_single})$ 
     $w_s \leftarrow fillIn(c.name, c.value)$ 
end
 $F_{excel} \leftarrow addWorksheet(w_s)$ 

for  $m \in M_a$ 
     $w_m \leftarrow initialiseWorksheet(m.name)$ 
    for  $c \in sortDescendNumValues(E_{cond\_multi})$ 
         $w_m \leftarrow fillInHierarchy(c.name, c.values)$ 
    end
     $F_{excel} \leftarrow addWorksheet(w_m)$ 
end

return  $F_{excel}$ 

```

Fig. 3 Pseudo-code of the systematic construction of biofilm structured data files.

3 Proof of Concept and Implementation

A dozen of in-house published studies validated the proposed approach. These studies are quite diverse in nature, addressing problems such as microbial susceptibility to clinically relevant antibiotics, biofilm adhesion to abiotic surfaces and antimicrobial synergisms. They present different levels of complexity and include data quantifications based on various state-of-the-art analytical methods (e. g. crystal violet (CV), 4',6-diamidino-2-phenylindole (DAPI) and fluorescence *in situ* hybridization (FISH)).

The approach was applied successfully and, as a whole, it enabled the standard characterisation of over 10000 data points. As an example, Fig. 4 illustrates a small excerpt of the one of such signatures related to the study on the adhesion of water stressed *Helicobacter pylori* to abiotic surfaces (namely, polypropylene, glass, copper and stainless steel) by Azevedo et al. [25].

Currently, the approach is supporting the activity of BiofOmics database, a general scope biofilms database which is freely accessible to the community (<http://biofomics.org>). Its implementation is based on open-source, platform-independent software, namely MySQL server (version 14.14, distribution 5.1.52) and PHP 5.1.6.

One of the long term aims of the approach is to be able to perform data mining on the data sets collected. To achieve this goal, it is essential that future experiments follow very strict guidelines, presenting comprehensive and unambiguous

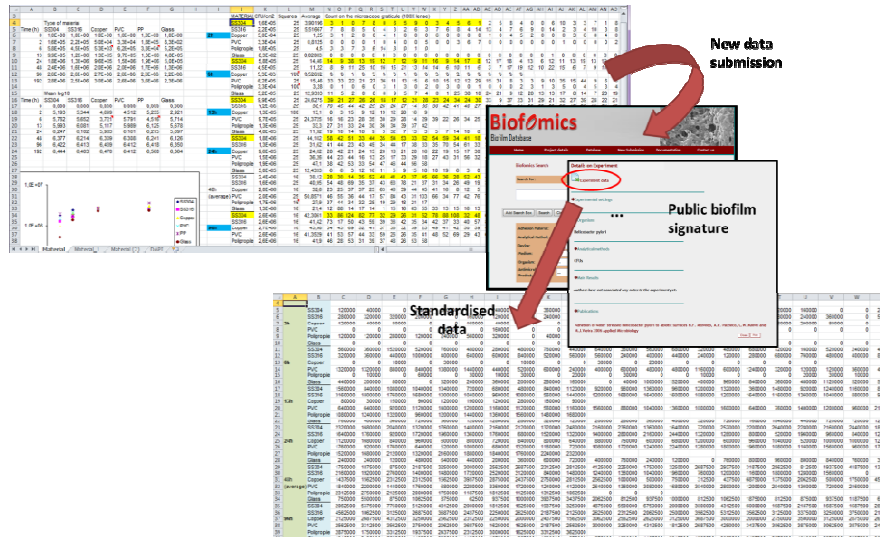


Fig. 4 Different stages of the data presentation of a study on the adhesion of water-stressed *Helicobacter pylori* to abiotic surfaces by Azevedo et al. [25].

information on how the experiment was performed and its outcomes. Namely, efforts should address the standardization of nomenclature in the biofilm area and the establishment of minimum information guidelines for reporting biofilm experiments. Once this has been achieved one should be able to query for patterns in biofilm signatures and thus empowering analysis abilities.

4 Conclusions

Biofilm research relies increasingly on large collections of data sets. This “big data” dimension calls for the development of novel computational tools, specialised in biofilm data management, interchange and analysis. However, we are unaware of any efforts to standardise and disseminate biofilm data at large scale. Currently, sharing of biofilm data among researchers is poor at best, in detriment of research and community at large.

Data standardisation is an urge to make knowledge more explicit, help detect errors, ensure data reliability, and promote data interchange. In here, we propose a systematic approach to the standardised organisation of biofilm data. The approach has been validated with a number of highly variable, already published experiments and it is already in practice, supporting the operation of the BiofOmics database, a database that facilitates data search and comparison as well as data interchange between laboratories (publicly accessible at <http://biofomics.org>).

Acknowledgments. The authors thank, among others, Rosário Oliveira, Maria João Vieira, Idalina Machado, Nuno Cerca, Mariana Henriques, Pilar Teixeira, Douglas Monteiro, Melissa Negri, Susana Lopes, Carina Almeida and Hélder Lopes, for submitting their data. The financial support from IBB-CEB, Fundação para a Ciência e Tecnologia (FCT) and European Community fund FEDER (Program COMPETE), project PTDC/SAU-ESA/646091/2006/FCOMP-01-0124-FEDER-007480, are also gratefully acknowledged.

References

1. McBain, A.J.: Chapter 4: In vitro biofilm models: an overview. *Adv. Appl. Microbiol.* 69, 99–132 (2009)
2. Jain, A., Gupta, Y., Agrawal, R., Khare, P., Jain, S.K.: Biofilms-a microbial life perspective: a critical review. *Crit. Rev. Ther. Drug Carrier Syst.* 24(5), 393–443 (2007)
3. Van Houdt, R., Michiels, C.W.: Biofilm formation and the food industry, a focus on the bacterial outer surface. *J. Appl. Microbiol.* 109(4), 1117–1131 (2010)
4. Machado, I., Lopes, S.P., Sousa, A.M., Pereira, M.O.: Adaptive response of single and binary *Pseudomonas aeruginosa* and *Escherichia coli* biofilms to benzalkonium chloride. *Journal of Basic Microbiology* 51, 1–10 (2011)
5. Frei, E., Hodgkiss-Harlow, K., Rossi, P.J., Edmiston Jr., C.E., Bandyk, D.F.: Microbial Pathogenesis of Bacterial Biofilms: A Causative Factor of Vascular Surgical Site Infection. *Vasc. Endovascular Surg* (2011)
6. Rodrigues, L.R.: Inhibition of bacterial adhesion on medical devices. *Adv. Exp. Med. Biol.* 715, 351–367 (2011)
7. Høiby, N., Bjarnsholt, T., Givskov, M., Molin, S., Ciofu, O.: Antibiotic resistance of bacterial biofilms. *Int. J. Antimicrob. Agents* 35(4), 322–332 (2010)
8. Bowler, P.G., Duerden, B.I., Armstrong, D.G.: Wound microbiology and associated approaches to wound management. *Clin. Microbiol. Rev.* 14(2), 244–269 (2001)
9. Simões, M.: Antimicrobial strategies effective against infectious bacterial biofilms. *Curr. Med. Chem.* 18(14), 2129–2145 (2011)
10. Smith, A., Buchinsky, F.J., Post, J.C.: Eradicating chronic ear, nose, and throat infections: a systematically conducted literature review of advances in biofilm treatment. *Otolaryngol. Head Neck Surg.* 144(3), 338–347 (2011)
11. Gião, M.S., Azevedo, N.F., Wilks, S.A., Vieira, M.J., Keevil, C.W.: Persistence of *Helicobacter pylori* in Heterotrophic Drinking-Water Biofilms. *Applied and Environmental Microbiology* 74(19), 5898–5904 (2008)
12. Jang, H.J., Choi, Y.J., Ka, J.O.: Effects of diverse water pipe materials on bacterial communities and water quality in the annular reactor. *J. Microbiol. Biotechnol.* 21(2), 115–123 (2011)
13. Ramage, G., Culshaw, S., Jones, B., Williams, C.: Are we any closer to beating the biofilm: novel methods of biofilm control. *Curr. Opin. Infect. Dis.* 23(6), 560–566 (2010)
14. Cos, P., Toté, K., Horemans, T., Maes, L.: Biofilms: an extra hurdle for effective antimicrobial therapy. *Curr. Pharm. Des.* 16(20), 2279–2295 (2010)
15. Wood, T.K.: Insights on *Escherichia coli* biofilm formation and inhibition from whole-transcriptome profiling. *Environ. Microbiol.* 11(1), 1–15 (2009)
16. Coenye, T.: Response of sessile cells to stress: from changes in gene expression to phenotypic adaptation. *FEMS Immunol. Med. Microbiol.* 59(3), 239–252 (2010)

17. Silva, M.S., De Souza, A.A., Takita, M.A., Labate, C.A., Machado, M.A.: Analysis of the biofilm proteome of *Xylella fastidiosa*. *Proteome Sci.* 9, 58 (2011)
18. Jiao, Y., D'haeseleer, P., Dill, B.D., Shah, M., Verberkmoes, N.C., Hettich, R.L., Banfield, J.F., Thelen, M.P.: Identification of biofilm matrix-associated proteins from an acid mine drainage microbial community. *Appl. Environ. Microbiol.* 77(15), 5230–5237 (2011)
19. Cabral, M.P., Soares, N.C., Aranda, J., Parreira, J.R., Rumbo, C., Poza, M., Valle, J., Calamia, V., Lasa, I., Bou, G.: Proteomic and functional analyses reveal a unique lifestyle for *Acinetobacter baumannii* biofilms and a key role for histidine metabolism. *J. Proteome Res.* 10(8), 3399–3417 (2011)
20. Xu, H., Lee, H.Y., Ahn, J.: Growth and virulence properties of biofilm-forming *Salmonella enterica* serovar typhimurium under different acidic conditions. *Appl. Environ. Microbiol.* 76(24), 7910–7917 (2010)
21. Workentine, M.L., Harrison, J.J., Weljie, A.M., Tran, V.A., Stenroos, P.U., Tremaroli, V., Vogel, H.J., Ceri, H., Turner, R.J.: Phenotypic and metabolic profiling of colony morphology variants evolved from *Pseudomonas fluorescens* biofilms. *Environ. Microbiol.* 12(6), 1565–1577 (2010)
22. Azevedo, N.F., Lopes, S.P., Keevil, C.W., Pereira, M.O., Vieira, M.J.: Time to “go large” on biofilm research: advantages of an omics approach. *Biotechnol. Lett.* 31(4), 477–485 (2009)
23. Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Federhen, S., Feolo, M., Fingerman, I.M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Mizrahi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y., Wilbur, W.J., Yaschenko, E., Ye, J.: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 39, D38–D51 (2011)
24. Thomas, S., Karnik, S., Barai, R.S., Jayaraman, V.K., Idicula-Thomas, S.: CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Res.* 38, D774–D780 (2010)
25. Azevedo, N.F., Pacheco, A.P., Keevil, C.W., Vieira, M.J.: Adhesion of water stressed *Helicobacter pylori* to abiotic surfaces. *J. Appl. Microbiol.* 111, 718–724 (2006)