BEW: Bioinformatics Workbench for Analysis of Biofilms Experimental Data

Gael Pérez Rodríguez[1], Daniel Glez-Peña[1], Nuno F. Azevedo[2], Maria Olívia Pereira[3], Florentino Fdez-Riverola[1], and Anália Lourenço[1,2]

[1] ESEI - Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas s/n, Universidad de Vigo, 32004 Ourense, Spain
[2] LEPABE – Dep. of Chemical Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
[3] IBB - Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
gprodriguez2@esei.uvigo.es, {dgpena,riverola,analia}@uvigo.es, nazevedo@fe.up.pt, mopereira@deb.uminho.pt, analia@ceb.uminho.pt

Abstract. Biofilms research has evolved considerably in the last decade and is now generating large volumes of heterogeneous data. MIABiE, the international initiative on Biofilms, is devising guidelines for data interchange, and some databases provide access to biofilms experiments. However, the field is lacking appropriate bioinformatics tools in support of increasing operational and analytical needs. This paper presents a flexible and extensible open-source workbench for the operation and analysis of biofilms experiments, as follows: (i) the creation of customised experiments, (ii) the collection of various analytical results, following community standardisation guidelines and (iii) on-demand reporting and statistical evaluation.

Keywords: Biofilms, standard operating procedures, data interchange, data analysis.

1 Introduction

Biofilms are surface-attached cellular agglomerates that are widespread in Nature and exhibit great abilities to adapt to environmental changes [1, 2]. Their actions can be seen as beneficial or detrimental to humans, depending on their ecological impact and our ability to act upon them [3, 4]. Notably, the resistance and resilience of biofilms attracts much attention from the biomedical research community due to the continuous emergence of multi-resistant strains in clinical settings, which are rendering conventional antibiotics ineffective [5, 6].

Biofilms research has witnessed a considerable development in the last decade. Conventional microbiological experimentation is giving place to large-scale multidisciplinary experimentation [7]. Cell viability, biomass formation, respiratory activity, morphological characterisation, and transcriptome and proteome profiling are just some examples of the methods of analysis now in use. Moreover, and due to the

J. Sáez-Rodríguez et al. (eds.), *8th International Conference on Practical Appl. of Comput. Biol. & Bioinform. (PACBB 2014)*, Advances in Intelligent Systems and Computing 294,
DOI: 10.1007/978-3-319-07581-5_6, © Springer International Publishing Switzerland 2014

diversity of settings where biofilms can be found, the environmental scenarios recreated in the lab can be quite different and sometimes challenging to implement. Repeatability, ruggedness and reproducibility tests are therefore conducted to ensure the quality of the acquired data and, in particular, the ability to compare results between laboratories [8, 9].

The MIABiE initiative (http://miabie.org), encompassing an international body of Biofilms experts, is working on the definition of guidelines to document biofilms experiments and the standardisation of the nomenclature in use. Biofilms databases such as BiofOmics (http://biofomics.org) [10] and MorphoCol (http://morphocol.org) [11] are endorsing these guidelines and making experimental data freely available. However, the community is missing tools to take the best advantage of these resources, and it is notable the inexistence of bioinformatics tools dedicated to biofilms data operation and analysis. In particular, unstructured data operation, using Excel or similar tools, compromises data standardisation and thus, any form of computer-aided processing and analysis.

This paper presents the first software tool dedicated to Biofilms: the Biofilms Experiment Workbench (BEW). The primary aim is to cover for primary intra-laboratory data collection and analysis necessities. Previous work on data standardisation and computerised data structuring [11], now complemented by MIABiE guidelines, established the starting point of development. A specialised markup language is defined now as grounds to document experiments, and to effectively promote data interchange across resources and software tools. Moreover, the application is developed with AIBench, an open-source Java desktop application framework [14], to ensure a flexible, cross-platform and interoperable development suitable to sustain future interactions with other Biofilms-related tools.

The next sections detail the markup language as well as the design and main functionalities of BEW.

2 Biofilms Markup Language

The current inability to exchange experiments between laboratories has its roots in the lack of a common format for describing Biofilms experiments. The modelling of Biofilm experiments is challenged by the complexity and variability of the studies. Studies may vary widely in aspects as important as: the conditions tested (related to the goals of analysis), the microorganisms studied (with implicit growth and other biological specifics), the methods of analysis used (data is only comparable for similar methods), specific data pre-processing (e.g. the calculation of dilution rates or log reductions), and the number of replicates and reproductions performed to ensure the validity of the study.

These issues were discussed with field experts - at the Eurobiofilms 2013 meeting in Ghent, Belgium (9-12 September 2013) - as means to accommodate for as much of this complexity and variability as possible, and anticipate new requirements resulting from the increasing use of high-throughput methods. Then, XML, the eXtensible Markup Language [12], was elected to formalise the data structure because of its portability and widespread acceptance as a standard data language for Bioinformatics [13].

At the end, we came by a very generic skeleton for the description of a biofilm experiment that considers four major conceptual elements:

- identification and authorship, including the name of the institution and the authors of the experiment, the title and a short description of the experiment, and any associated publications;
- method of analysis, which comprises a set of test conditions and a number of resulting data series;
- condition, i.e. a experimental condition tested by a given method of analysis (e.g. a value of temperature or a dose of an antimicrobial agent);
- data series that describe the data obtained by a method of analysis for a given set of conditions.

The definition of an experiment in XML simply consists of lists of one or more of these various components (Fig. 1).

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<ns0:bml xmlns:ns0="" xmlns:xsi="" xsi:schemaLocation="">
  <experiment contact="" date="" experimentName="" organization="" publication="">
    <authors><![CDATA[]]></authors>
    <notes><![CDATA[]]></notes>
    <methods>
      <method methodName="" methodUnits="">
        <conditions>
          <condition conditionName="" conditionUnits="">...</condition>
          ...
        </conditions>
        <dataSeriesSet>
          <dataSerie>
            <conditionValues>...#...</conditionValues>
            <measurements>...#...</measurements>
          </dataSerie>
          ...
        </dataSeriesSet>
      </method>
      ...
    </methods>
    <constantConditions>
      <constantCondition condition="" conditionUnits="" conditionValue=""/>
      ...
    </constantConditions>
  </experiment>
</ns0:bml>
```

Fig. 1. Skeleton of the definition of a biofilm experiment in XML format, showing all possible top-level elements

3 BEW: The Biofilms Workbench

BEW is a desktop-based application dedicated to the management and analysis of the data resulting from Biofilms experiments. The application was developed with AIBench, an open-source Java desktop application framework for scientific software development in the domain of translational biomedicine [14]. As illustrated in Fig. 2, BEW incorporates a plug-in for the R statistical computing tool (http://www.r-project.org/) and the libraries JFreeChart for data plotting (http://www.jfree.org/jfreechart/), JXL to read and write Excel sheets (http://jexcelapi.sourceforge.net/), and JSoup for working with

HTML documents (http://jsoup.org/). These third-parties support main data processing and data analysis operations whilst enable future adaptation or extension (notably regarding statistical analysis).

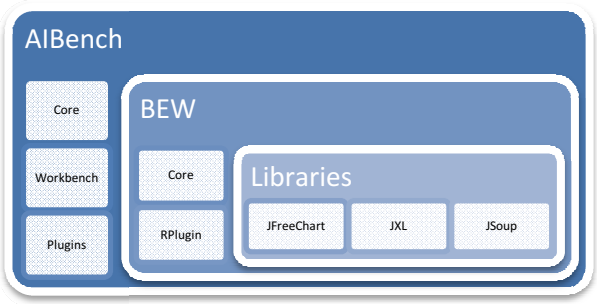

Fig. 2. General architecture of the software components in BEW

The next subsections describe the core management and analysis functionalities currently provided by the software.

3.1 Experiment Management

In accordance to standardisation guidelines and to cope with data legacy, BEW supports both .xml and .xls data representation formats. The XML is the native format (as described in section 2) and wrapping functionalities are provided to import/export experiments from/to the standardised and hierarchy-based Excel format [11].

The experiment should be comprehensively described in order to guarantee its community-wide unambiguous interpretation. First, the experiment is "identified" in terms of authorship (author and institution), author's summary and any associated publications. Then, the experimental setup is detailed and, more specifically, the overall settings or constant conditions, and the conditions tested by each of the methods of analysis. Given the wide range of analyses now conducted for biofilms, user interface was made as flexible and intuitive as possible to enable the construction of variable nature and extent data "worksheets" (Fig. 3). In particular, metadata standardisation and systematic data structuring were encapsulated in logical selections to detach the user from unnecessary computational details.

3.2 Statistical Analysis

BEW is equipped with a powerful analytical component that supports on-demand construction of data plots as well as statistical data testing.

Plots and statistical tests usually present in Biofilms publications served as references of expected functionalities, but the component was developed such to enable the use of a wider range of plots and tests. Two free third-party tools support BEW analytical component: the JFreeChart library provides for plotting abilities while the R plug-in grants access to broad statistical analysis.

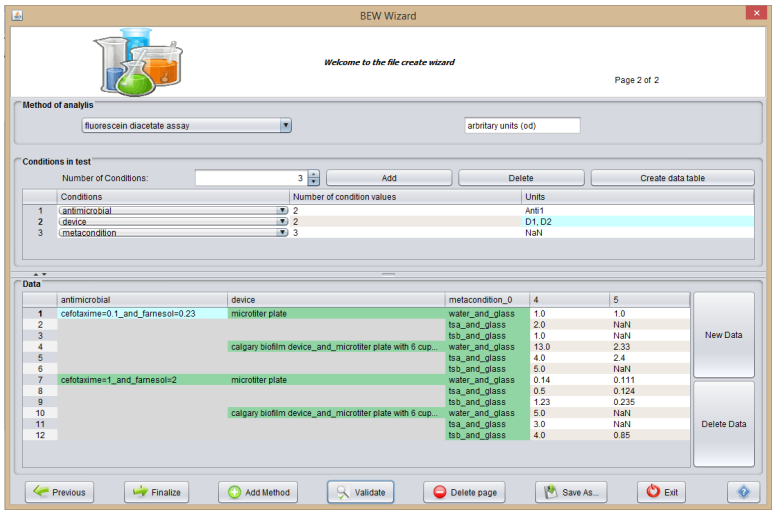

Fig. 3. Snapshot of the description of an experimental analysis

Data plotting was made as flexible as possible in order to accommodate the visualisation of the results produced by virtually any combination of test conditions (Fig. 4).

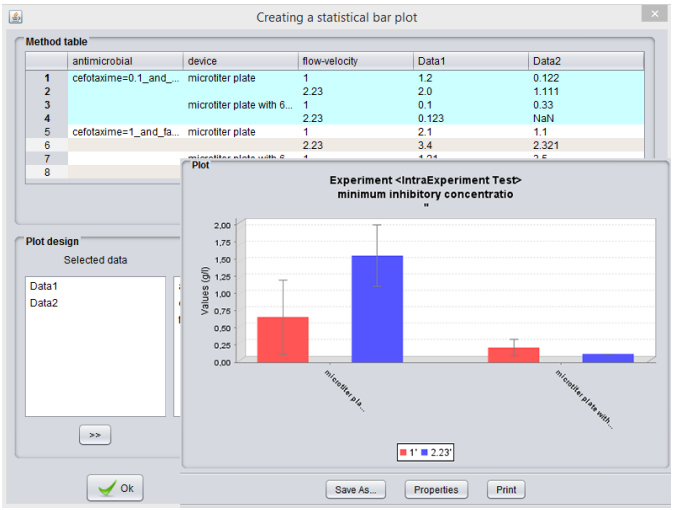

Fig. 4. Snapshot of a plot representing the test results for three conditions

Likewise, BEW has no pre-established pipeline for statistical analysis, i.e. the user decides which tests should be executed for his data. This is crucial given the heterogeneity of existing experimentation. For example, antimicrobial susceptibility testing usually aims to analyse the variance of results obtained under different sets of therapeutic conditions (Fig. 5). In turn, repeatability and reproducibility experiments

are quite unique in the goal of analysis – standard operation – and the interpretation of results. Typically, researchers are interested in conducting ANOVA Gauge R&R tests, which are not used in most of the other experiments.

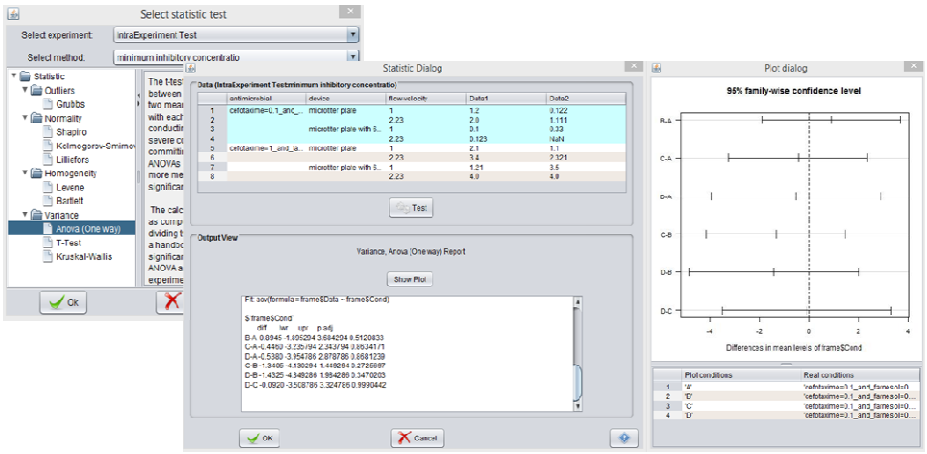

Fig. 5. Snapshot of the execution of an analysis of variance

3.3 Reports

Report capabilities are useful to automatically document the experiment, notably to describe the experiment in terms of methods of analysis and tested conditions, as well as to provide a summary of the resulting data (Fig. 6).

This functionality is meant to complement the XML file, being of practical use when depositing an experiment in a public database or submitting a manuscript to peer-review. It serves the purpose of introducing the work developed, prior to the actual inspection of results by means of the XML file. Also, it can be used to pre-visualise the experiments in Web resources and tools.

4 Conclusions

In this paper we present BEW, a desktop application devoted to the management of Biofilms experiments. It is the very first tool developed to meet the data processing and analysis requirements of the Biofilms domain. As such, its first contribution is a XML-based markup language for the general description of biofilm experiments.

Widespread use of Biofilms markup language in resources and software packages will benefit users as well as developers. With greater interaction between tools, and a common format for publications and databases, researchers will be able to perform systematic experiment comparison and data interchange.

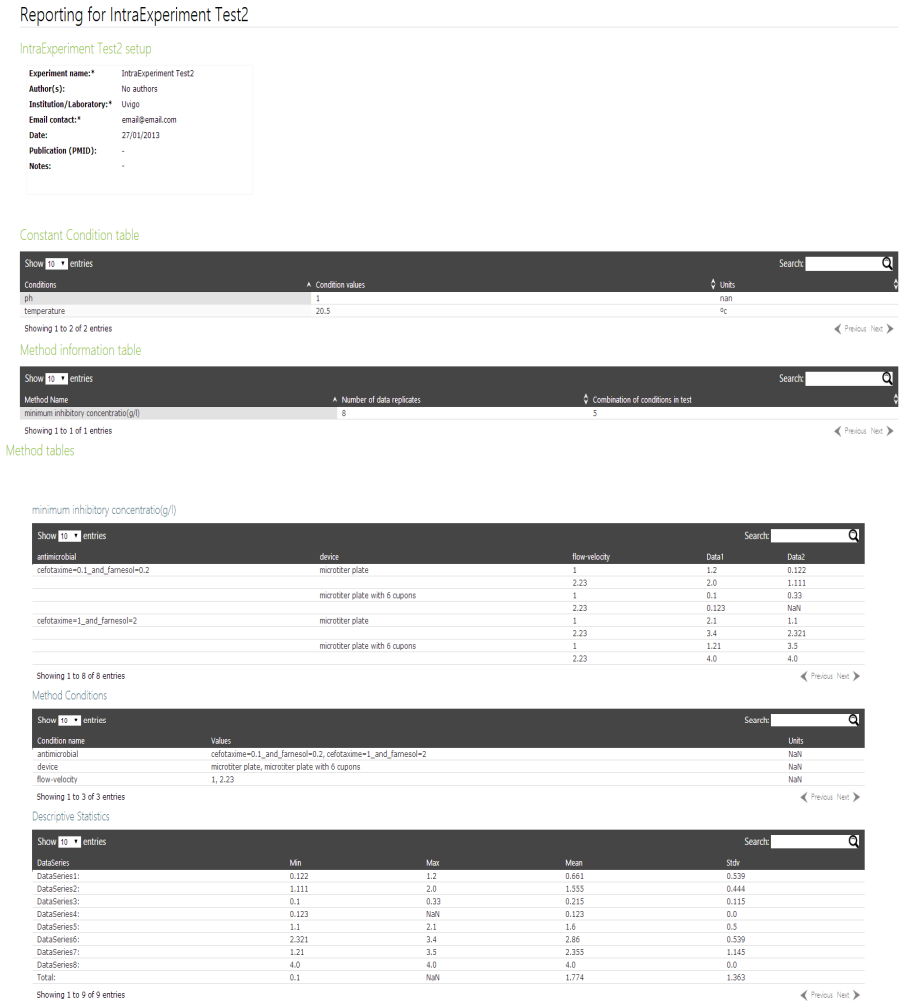

Fig. 6. Snapshot of the Web report of an experiment

Moreover, BEW architecture is designed to be flexible and easily extensible, in anticipation of the various requirements emerging from the growing multidisciplinary of these experiments and the diverse nature of studies. At first, BEW aims to support researcher's daily operation, both in dealing with the generated data and executing the necessary statistical discussion. In a near future, BEW is expected to go beyond intra-laboratory frontiers, notably by enabling the automatic deposition of standardised experimental data in community databases, such as BiofOmics, and enabling inter-laboratory experiment comparisons.

Acknowledgements. This work was supported by the [INOU13-07] project from the University of Vigo, the IBB-CEB, the Fundação para a Ciência e Tecnologia (FCT) and the European Community fund FEDER, through Program COMPETE [FCT Project number PTDC/SAU-SAP/113196/2009/FCOMP-01-0124-FEDER-016012], and the Agrupamento INBIOMED from DXPCTSUG-FEDER unha maneira de facer Europa (2012/273). The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7/REGPOT-2012-2013.1 under grant agreement n° 316265, BIOCAPS. This document reflects only the author's views and the European Union is not liable for any use that may be made of the information contained herein.

References

1. Hall-Stoodley, L., Costerton, J.W., Stoodley, P.: Bacterial biofilms: from the natural environment to infectious diseases. Nat. Rev. Microbiol. 2, 95–108 (2004)
2. Donlan, R.M.: Biofilms: microbial life on surfaces. Emerg. Infect. Dis. 8, 881–890 (2002)
3. Davey, M.E., O'toole, G.A.: Microbial biofilms: from ecology to molecular genetics. Microbiol. Mol. Biol. Rev. 64, 847–867 (2000)
4. 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, 393–443 (2007)
5. Römling, U., Balsalobre, C.: Biofilm infections, their resilience to therapy and innovative treatment strategies. J. Intern. Med. 272, 541–561 (2012)
6. Peters, B.M., Jabra-Rizk, M.A., O'May, G.A., Costerton, J.W., Shirtliff, M.E.: Polymicrobial interactions: impact on pathogenesis and human disease. Clin. Microbiol. Rev. 25, 193–213 (2012)
7. 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, 477–485 (2009)
8. Buckingham-Meyer, K., Goeres, D.M., Hamilton, M.A.: Comparative evaluation of biofilm disinfectant efficacy tests. J. Microbiol. Methods 70, 236–244 (2007)
9. Hamilton, M.A.: KSA-SM-03 - Desirable attributes of a standardized method. Stand. METHODS Test. Surf. Disinfect (2010)
10. Lourenço, A., Ferreira, A., Veiga, N., Machado, I., Pereira, M.O., Azevedo, N.F.: BiofOmics: A Web Platform for the Systematic and Standardized Collection of High-Throughput Biofilm Data. PLoS One 7, e39960 (2012)
11. Sousa, A.M., Ferreira, A., Azevedo, N.F., Pereira, M.O., Lourenço, A.: Computational approaches to standard-compliant biofilm data for reliable analysis and integration. J. Integr. Bioinform. 9, 203 (2012)
12. Bray, T., Paoli, J., Sperberg-McQueen, C.M.: Extensible markup language, XML (1998), http://www.w3.org/TR/1998/REC-xml-19980210
13. Sreenivasaiah, P.K., Kim, D.H.: Current trends and new challenges of databases and web applications for systems driven biological research. Front. Physiol. 1, 147 (2010)
14. Glez-Peña, D., Reboiro-Jato, M., Maia, P., Rocha, M., Díaz, F., Fdez-Riverola, F.: AIBench: a rapid application development framework for translational research in biomedicine. Comput. Methods Programs Biomed. 98, 191–203 (2010)