



**HAL**  
open science

## Rapid BOD assessment with a microbial array coupled to a neural machine learning system

Sullivan Jouanneau, Emilie Grangé, Marie-José Durand, Gerald Thouand

### ► To cite this version:

Sullivan Jouanneau, Emilie Grangé, Marie-José Durand, Gerald Thouand. Rapid BOD assessment with a microbial array coupled to a neural machine learning system. *Water Research*, 2019, 166, pp.115079. 10.1016/j.watres.2019.115079 . hal-02329921

**HAL Id: hal-02329921**

**<https://hal.science/hal-02329921>**

Submitted on 21 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Rapid BOD assessment with a microbial array**  
2 **coupled to a neural machine learning system.**

3 *Sullivan Jouanneau<sup>1</sup>, Emilie Grangé<sup>1</sup>, Marie-José Durand<sup>1</sup>, and Gérald Thouand<sup>1</sup>.*

4 *<sup>1</sup>: University of Nantes, UMR CNRS 6144 GEPEA, 18 Boulevard Gaston Defferre,*  
5 *85035 La Roche sur Yon, France*

6 Corresponding authors: Sullivan Jouanneau, UMR CNRS 6144 GEPEA, CBAC Laboratory,  
7 18 Boulevard Gaston Defferre, 85035 La Roche sur Yon, France, [sullivan.jouanneau@univ-](mailto:sullivan.jouanneau@univ-)  
8 [nantes.fr](mailto:sullivan.jouanneau@univ-nantes.fr)

9 Gérald THOUAND, UMR CNRS 6144 GEPEA, CBAC Laboratory, 18 Boulevard Gaston  
10 Defferre, 85035 La Roche sur Yon, France, [gerald.thouand@univ-nantes.fr](mailto:gerald.thouand@univ-nantes.fr)

11

12 ABSTRACT

13 The domestic usage of water generates approximately 310 km<sup>3</sup> of wastewater worldwide  
14 (2015, AQUASTAT, Food and Agriculture Organization of United Nations). This sewage  
15 contains an important organic load due to the use of this water; this organic load is  
16 characterized using a standard method, namely, the biological oxygen demand measurement  
17 (BOD<sub>5</sub>). The BOD<sub>5</sub> provides information about the biodegradable organic load (standard ISO  
18 5815). However, this measurement protocol is very time-consuming (5 days) and may  
19 produce variability in approximately 20% of results mainly due to variation in the  
20 environmental inocula.

21 To remedy these limitations, this work proposes an innovative concept relying on the  
22 implementation of a set of rigorously selected bacterial strains. This publication depicts the  
23 different steps used in this study, from bio-indicator selection to validation with real  
24 wastewater samples. The results obtained in the final step show a strong correlation between  
25 the developed approach and the reference method (ISO 5815) with a correlation rate of  
26 approximately 0.9. In addition, the optimization of the experimental conditions and the use of  
27 controlled strains (8 selected strains) allow significant reduction in the duration of the BOD<sub>5</sub>  
28 analysis, with only 3 hours required for the proposed method versus 5 days for the reference  
29 method. This technological breakthrough should simplify the monitoring of wastewater  
30 treatment plants and provide quicker, easier and more coherent control in terms of the  
31 treatment time.

32

33 KEYWORDS: BOD<sub>5</sub>, multivariate approach, bacterial biosensor, neural network, wastewater.

34

35

## 36 1. Introduction

37 The global production of domestic wastewater was assessed to be 315 km<sup>3</sup> in 2010 and will  
38 tend to increase in the future concomitant with population expansion (**Flörke et al., 2013;**  
39 **Mateo-Sagasta et al., 2015**). These wastewaters, which are mainly contaminated by organic  
40 materials (i.e., carbohydrates, fats, proteins, etc.) resulting from human body wastes (20% -  
41 faeces and urine) and domestic uses (80% - personal washing, laundry, food preparation, etc.)  
42 (**Duncan, 2004; Garcia et al., 2006**) present a real risk for the receiving environment and the  
43 exposed population. To limit the consequences of release of these wastewaters into the  
44 environment, they are treated to remove the organic load through bioprocesses, such as  
45 wastewater treatment plants (WWTPs). To measure the conformity of the released waters, the  
46 organic load in the treated water is tested weekly in average samples according to the  
47 reference method (*ISO 5815-1:2003, 2003; ISO 5815-2:2003, 2003*) based on BOD  
48 (biological oxygen demand).

49 This method dates from the beginning of the twentieth century (**Great Britain. Royal**  
50 **commission on sewage disposal, 1908; Jouanneau et al., 2014**), and its implementation is  
51 widespread worldwide. This method is based on assessment of the biochemical oxygen  
52 demand [i.e., the amount of oxygen consumed over 5 days by environmental microorganisms  
53 to biodegrade the organic matter under aerobic conditions (BOD<sub>5</sub> in mg.L<sup>-1</sup> of consumed O<sub>2</sub>)].  
54 However, this method has several drawbacks:

- 55 - The method exhibits important intrinsic variability (higher than 20%) that is mainly  
56 due to the broad diversity of inocula used to carry out the measurements (**Guyard,**  
57 **2010**).
- 58 - The 5-day period required for a measurement is longer than the retention time of water  
59 in WWTPs (close to 3 days).

60 - The measurement carries a significant risk of measurement error due to the available  
61 amount of oxygen in water. Indeed, the maximal load measured with this method  
62 cannot exceeded 6 mg/L BOD (*ISO 5815-1:2003, 2003; ISO 5815-2:2003, 2003*),  
63 whereas the BOD<sub>5</sub> value of wastewater is close to 300 mg/L and thus requires pre-  
64 dilution.

65 Over the last several decades, many research teams have focused on these issues to propose  
66 new measurement strategies with an aim of removing these drawbacks (**Jouanneau et al.,**  
67 **2014; Sun et al., 2015; Yang et al., 2015**). A specific focus on the biological aspect of the  
68 measurement highlights two strategies based on (i) the use of environmental inocula and (ii)  
69 the implementation of only one microbial strain. The first strategy has the advantage of  
70 ensuring great representativeness of the biological answers due to broad microbial diversity  
71 but induces significant variability (i.e., origin, density, and diversity) (**Liu et al., 2012;**  
72 **Rastogi et al., 2003; Rocher et al., 2011**) due to the intrinsic characteristics of the inocula.  
73 Conversely, the second strategy, based on the use of one known strain, is very reproducible  
74 but is limited by the range of metabolic abilities of the selected strain (**Arlyapov et al., 2012;**  
75 **Raud et al., 2012; Yoshida et al., 2001**), which consequently provides a truncated view of  
76 the biodegradable organic diversity.

77 In light of these issues, some researchers have proposed an alternative strategy based on the  
78 use of several known microbial strains. Two strategies were proposed based on either (i) an  
79 artificial mixture of 2-4 strains or (ii) a set of several strains used individually. The aims of  
80 these strategies were the same, namely, to enlarge the biodegradation spectra while ensuring a  
81 great level of reproducibility (**Catterall et al., 2003; Jia et al., 2003; Jiang et al., 2006; Lin**  
82 **et al., 2006; Xin et al., 2007**). In the case of multi-species mixtures (i), the stability of the  
83 initial properties of the consortia (i.e., cell density, diversity, ratio between strains,  
84 biodegradation abilities, etc.) was difficult to ensure long-term mostly due to competition for

85 the carbon substrate. The second strategy (ii) (**Pitman et al., 2015; Raud and Kikas, 2013**)  
86 also relies on several strains but is based on the use of several selected strains (not structured  
87 in the consortium). This strategy ensures good reproducibility while freeing the results from  
88 potential inocula drifts (this strategy was applied in this work). Nevertheless, this multivariate  
89 approach (based on several biological descriptors) increases the complexity of the data  
90 analysis (one value per strains). To overcome this limitation, Raud and Kikas (**2013**) analysed  
91 the data with a PLS (partial least squares) regression method to simultaneously take into  
92 consideration all the information provided by the different strains. The results were very  
93 encouraging, but the validation was performed only on synthetic wastewaters and not on real  
94 samples.

95 Moreover, bio-element selection is a crucial parameter in these metrological approaches based  
96 on controlled strains and directly conditions the potential performances of the future  
97 analytical method (i.e., biodegradation spectra and response kinetics) (**Jouanneau et al.,**  
98 **2014**). The selection step, when included, generally is based on either the intrinsic properties  
99 of the strains (i.e., biodegradation range and growth conditions) (**Kwok et al., 2005; Oota et**  
100 **al., 2010; Raud et al., 2012; Raud and Kikas, 2013**) or isolation of strains from a specific  
101 environment (**Chan et al., 2000; Kwok et al., 2005; Yoshida et al., 2000**). Nevertheless, no  
102 characterization or comparison has been carried out to determine the relevance of this choice.  
103 Moreover, questions remain concerning how to ensure the representativity of the proposed  
104 analytical approach under these conditions.

105 The aim of the proposed study was to design a relevant analytical method for the assessment  
106 of the BOD<sub>5</sub> to address these limitations. This strategy is based on the separate use (not  
107 structured in the consortium) of a set of 8 bacterial strains that have been rigorously selected  
108 using defined characteristics, such as the biodegradation potential, metabolic rate and  
109 toxicological robustness (Fig. 1). Coupled with an algorithm of data processing allowing to

110 analyses the multivariate information provided by the bacterial set, the proposed method is  
111 able to assess BOD<sub>5</sub> in only 3 hours with a high reliability ( $r^2=0.85$ ).

## 112 **2. Experimental section**

### 113 **2.1. Bacterial strains**

114 A total of 28 strains were preselected from the literature according to the following  
115 characteristics: (i) their representativity with regard to the microbial population of an  
116 activated sludge from a WWTP (Kong et al., 2007; Snaidr et al., 1997; Xia et al., 2010), (ii)  
117 their weak pathogenicity (mainly group 1 risk according to the 2000/54/EC classification  
118 proposed by the European Directive), (iii) their broad biodegradation abilities (Gao et al.,  
119 2010) and (iv) their resistance to freeze-drying conditions (information collected from  
120 microbial collections, such as ATCC, DSMZ or CIP) (specification relative to an eventual  
121 industrial transfer). The full list of pre-selected strains at this step is described in the  
122 supplementary data (Table SD-1).

### 123 **2.2. Growth medium of the strains and other solutions**

124 The Luria Bertani medium (LB medium) was prepared in distilled water and supplemented  
125 with 5 g.L<sup>-1</sup> NaCl (Merck, Germany), 5 g.L<sup>-1</sup> tryptone (Biokar Diagnostics, France) and 1 g.L<sup>-1</sup>  
126 yeast extract (Biokar Diagnostics, France). 15 g.L<sup>-1</sup> of agar type E (Biokar Diagnostics,  
127 France) was added to obtain the solid medium. The pH was adjusted to 7 using a solution of  
128 HCl (Sigma Aldrich, USA) or NaOH (Merck, Germany) before autoclaving at 120°C for 20  
129 min.

130 The synthetic wastewater (SWW) solution was prepared in accordance with the OECD 209  
131 standard<sup>28</sup> as follows: 1 L of distilled water was supplemented with 16 g of tryptone (Biokar  
132 Diagnostics, France), 11 g of meat extract (Biokar Diagnostics, France), 3 g of urea (Sigma-

133 Aldrich, USA), 0.7 g of NaCl (Merck, Germany), 0.4 g of CaCl<sub>2</sub> • 2H<sub>2</sub>O (Merck, Germany),  
134 0.2 g of MgCl<sub>2</sub> • 7H<sub>2</sub>O (Sigma-Aldrich, USA), and 2.8 g of K<sub>2</sub>HPO<sub>4</sub> (Merck, Germany).

135 The mineral medium [MM, derived from M9 medium (Atlas, 1997)] without a carbon source  
136 used in this study was prepared in two steps. The first step consisted of preparation of several  
137 concentrated solutions (S1, S2, S3, St and Sb) in distilled water. S1 consists of 60 g.L<sup>-1</sup>  
138 Na<sub>2</sub>HPO<sub>4</sub> (Fluka, Germany), 30 g.L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub> (Merck, Germany), 10 g.L<sup>-1</sup> NH<sub>4</sub>Cl (Merck,  
139 Germany) and 5 g.L<sup>-1</sup> NaCl (Merck, Germany). S2, S3, Sb and St are composed of 264.5 g.L<sup>-1</sup>  
140 MgSO<sub>4</sub> • 7H<sub>2</sub>O (Sigma-Aldrich, USA), 14.7 g.L<sup>-1</sup> CaCl<sub>2</sub> • 2H<sub>2</sub>O (Merck, Germany), 0.5 g.L<sup>-1</sup>  
141 biotin (Sigma-Aldrich, USA) and 0.15 g.L<sup>-1</sup> thiamine • HCl (Sigma-Aldrich, USA),  
142 respectively. S1, S2 and S3 were sterilized by autoclaving, and St and Sb were filtered  
143 through 0.2 µm cellulose acetate filters (146560, Dutscher, France) and stored at +4°C. The  
144 second step consisted of mixing these solutions according to the following ratio (for 1 L of  
145 MM): 100 mL of S1, 1 mL of S2/S3/St/Sb and if required 10 mL of a carbon source at the  
146 desired concentration. The mixture was adjusted to 1 L with sterile distilled water.

147 Glucose solutions (GSs) were prepared at several carbon concentrations in the MM medium  
148 and sterilized by filtration (cellulose acetate filters, 146560, Dutscher, France).

149 Glycerol solution (GlyS) at 50% (v/v) was prepared in distilled water and sterilized by  
150 autoclaving.

### 151 **2.3. Growth and preparation of bacterial cells**

152 All strains were stored and cultivated under the same conditions. The cells were stored at -  
153 80°C in a mixture (v/v) of LB medium (50%) and GlyS (50%). The cells were isolated on  
154 solid LB medium at +30°C during 12 to 24 hours and the obtained colonies were used to  
155 inoculate the first cultivations.

156 For the first cultures, the cells were grown overnight (14 hours) in LB medium at 30°C under  
157 shaking at 250 rpm. Then, the cells were diluted to an optical density of  $OD_{620nm} = 0.2$  in  
158 fresh LB medium and cultivated at 30°C under shaking at 250 rpm until an  $OD_{620nm} = 0.6$   
159 (exponential growth stage) was reached. This  $OD_{620nm}$  value (0.6) corresponds to a cell  
160 density varying between  $10^8$  and  $10^9$  CFU.ml<sup>-1</sup> for the tested strains (results not shown). For  
161 the sake of simplicity, we considered that these cell densities were not significantly different  
162 from strain to strain.

163 The cultivated cells ( $OD_{620nm} = 0.6$ ) were refreshed in an ethanol bath at -20°C for 1 minute.  
164 The aim of this step is not to freeze the cells but to decrease quickly the temperature in order  
165 to avoid any change of their physiological state. Then, the cells were centrifuged (6,400 x g, 5  
166 min, +4°C) to remove the growth medium and resuspended in a cold washing solution of  
167  $MgSO_4 \cdot 7H_2O$  ( $10^{-2}$  M). After three washes, the cells were resuspended in MM without a  
168 carbon source and adjusted to obtain the required cellular density  
169 ( $OD_{620nm} = 0.43/0.86/1/1.72/2.57/3.43/4.29$ ).

#### 170 **2.4. Optimization of the experimental conditions**

171 We were interested in improving some of the experimental conditions through two strategies:  
172 the replacement of oxygen (used as an activity marker of biodegradation in the reference  
173 method) with a fluorescent dye (resazurin) that was *a priori* better suited for this metrological  
174 context and the optimization of the cell density in the biodegradation assays. The aim of this  
175 second strategy was to define the condition that ensured both a reduced analysis duration and  
176 a measurement range that was as large as possible. The resazurin solution used in this work  
177 was a commercial solution (PrestoBlue®, Invitrogen™, USA). No information was available  
178 relative its concentration.

179 These steps were performed with the model strain *E. coli* K12 MG1655 in duplicate.

180 2.4.1. Comparison of two activity markers: oxygen versus resazurin (fluorescent dye)

181 **2.4.1.1. Monitoring the biodegradation activity via the fluorescent marker.**

182 The bacterial suspension was prepared as described above (see 2.3) at an  $OD_{620nm} = 0.43$ . The  
183 assay solution was prepared by mixing the washed bacterial suspension, MM supplemented  
184 with glucose and the commercial resazurin solution (PrestoBlue<sup>®</sup>, Invitrogen<sup>™</sup>, USA).

185 Several glucose concentrations (final concentration) ranging from  $0.1 \text{ mg.L}^{-1}$  to  $500 \text{ mg.L}^{-1}$   
186 were tested to define the limits of the measurement range. The mixtures were prepared  
187 directly in the microplate wells (Nunc<sup>™</sup> F96 MicroWell<sup>™</sup>, 236108) in following the ratio  
188 (final volume =  $150 \mu\text{L/well}$ ):  $35 \mu\text{L}$  of bacterial suspension (final  $OD_{620nm} = 0.1$ ),  $100 \mu\text{L}$  of  
189 MM with glucose and  $15 \mu\text{L}$  of dye (PrestoBlue<sup>®</sup>, Invitrogen<sup>™</sup>, USA, in accordance with the  
190 provider protocol: 10% v/v).

191 Monitoring of the biological activity via the fluorescent marker (**Cregut et al., 2013; Dudal**  
192 **et al., 2006; Tizzard et al., 2006**) was performed with the TriStar LB 941 Multimode  
193 Microplate Reader (Berthold Technologies GmbH, Germany) every ten minutes [ $\lambda_{excitation} =$   
194  $540 \text{ nm}$  (bandwidth =  $10 \text{ nm}$ ),  $\lambda_{emission} = 610 \text{ nm}$  (bandwidth =  $10 \text{ nm}$ ), counting time =  $0.1 \text{ s}$   
195 and incubation temperature =  $30^\circ\text{C}$ ].

196 **2.4.1.2. Monitoring of the biodegradation activity via oxygen consumption.**

197 For the assay based on the oxygen marker, the resazurin solution was replaced by distilled  
198 water in the assay solution. In this case, the biological activity was monitored in an Oxodish<sup>®</sup>  
199 OD24 microplate reader (24-well microplate with an integrated oxygen sensor, PreSens  
200 Precision Sensing, GmbH, Germany). Each well was filled with  $3 \text{ mL}$  of assay solution (same  
201 ratio as in 2.4.1.(a) –  $700 \mu\text{L}$  of bacterial suspension,  $2 \text{ mL}$  of MM with glucose but  $300 \mu\text{L}$  of  
202 distilled water) and sealed using an AB-0558 adhesive PCR film (Thermo, ABgene, France)  
203 to prevent atmospheric oxygen from entering the system (**Cregut et al., 2013**) during the test.  
204 Oxygen was monitored at  $30^\circ\text{C}$  with the SensorDish Reader (PreSens Precision Sensing).

205 2.4.2. Optimization of the cellular density.

206 The impact of the cell density was evaluated as described in the previous paragraph (2.4.1.).

207 The same protocol based on resazurin as the biodegradation marker was used except for the

208 final bacterial concentration, which ranged from  $OD_{620nm} = 0.1$  to  $OD_{620nm} = 1$ .

209 **2.5. Characterization of the strains**

210 The characterization steps described below were performed for the 28 selected strains in

211 parallel with the optimization phases. The cellular density after the washing step (see 2.3) was

212 adjusted to  $OD_{620nm} = 1$  for all strains.

213 2.5.1. Metabolic profile and stress robustness

214 The objective was to characterize the selected strains according to two criteria: their

215 biodegradation capabilities and their intrinsic robustness towards several environmental

216 stresses (i.e., pH, salinity, antibiotics, reference toxic compounds, etc.). These two parameters

217 were assessed simultaneously using GEN III microplates<sup>TM</sup> (Biolog Inc., USA – see Fig. SD-

218 1) with 71 different carbon sources and 23 stress conditions (test based on the growth

219 inhibition induced by specific conditions such as osmotic and pH stresses or several reference

220 toxics) supplemented with tetrazolium redox dyes (as a colorimetric respiratory marker)

221 following the protocol provided by the manufacturer.

222 Briefly, after the washing step (see 2.3), the bacterial suspension was diluted tenfold in an IF-

223 A solution (Biolog Inc., No. 72401, IF-A, USA) to obtain a cellular density between  $10^7$  and

224  $10^8$  cell.mL<sup>-1</sup> ( $OD_{620nm} = 0.1$ ). Then, 100  $\mu$ L of this diluted suspension was added to each

225 microplate well, and the plate was incubated for 40 hours at 30°C. The reading was performed

226 at a wavelength of 590 nm (**Preston-Mafham et al., 2002**) with the TriStar LB 941

227 multimode microplate reader (Berthold Technologies, Germany). Each strain was tested in

228 duplicate.

229 The results obtained from the 71 carbon sources were used to establish biodegradation  
230 profiles specific to each strain and then compared and clustered using statistical approaches  
231 (agglomerative hierarchical clustering, XLStat, Addinsoft).

232 For each strain, an index of robustness was calculated from the number of stress conditions,  
233 conditions inducing a growth inhibition (none observed cellular development), divided by the  
234 overall number of tested conditions (n=23). This index aims to compare the sensitivity of the  
235 strains to potential environmental stresses.

#### 236 2.5.2. Assessment of the biodegradation kinetics

237 The biodegradation kinetics of the 28 strains were compared in the SWW medium to evaluate  
238 their respective response times. First, 2.7 mL of SWW and 300  $\mu$ L of the washed suspension  
239 ( $OD_{620nm} = 1$ ) were added to the wells of an Oxodish<sup>®</sup> OD24 microplate (PreSens, Germany).  
240 An AB-0558 adhesive PCR film (Thermo, ABgene, France) was applied on the microplates to  
241 limit atmospheric oxygen diffusion in the reaction medium. Measurements were performed  
242 with the SensorDish Reader (PreSens Precision Sensing) each minute at 30°C for 10 hours in  
243 triplicate. The kinetics values were calculated from the maximal slope of the measured  
244 oxygen consumption.

#### 245 **2.6. Environmental samples**

246 The environmental samples (n = 104 - Table SD-2) had several origins, including raw or  
247 treated waters from several WWTPs in the Nantes countryside (France, n = 72), stormwaters  
248 from the Bordeaux Region (France, n = 8) and treated waters collected from non-collective  
249 treatment units (autonomous sanitation system dedicated to the isolated private dwellings)  
250 installed in the Scientific and Technical Centre for Building of Nantes (CSTB, France, n =  
251 24). The sampling procedure used for this study follows the conditions described by the ISO  
252 5667-2 and ISO 5667-3 standards. The samples were collected in 250 mL disposable plastic

253 bottles (PE, 030424, Dustcher, France) that were previously washed with distilled water. The  
254 samples were aliquoted into 2 mL tubes (PP, Biopur™ Safe-Lock™ tubes, Eppendorf®,  
255 Germany) and stored at -20°C prior to use.

256 Several parameters were measured to characterize the samples, including the BOD<sub>5</sub> (see 2.7),  
257 chemical oxygen demand (NF T90-101/NF EN 872), and toxicity (ISO 11348, ISO 6341 and  
258 ISO 20666) (data shown in Table SD-2).

## 259 **2.7. BOD<sub>5</sub> measurement**

### 260 **2.7.1. Reference method (ISO 5815)**

261 The reference analyses for BOD<sub>5</sub> were performed according to the standard method  
262 (ISO 5815) by the INOVALYS NANTES Lab (Nantes, France), which is certified by the  
263 French Accreditation Committee (COFRAC).

### 264 **2.7.2. Method developed in this work**

265 The measurement protocol for the proposed BOD<sub>5</sub> test was carried out in 96-well microplates  
266 (Nunc™ F96 MicroWell™, 236108) according to the following ratio: 35 µL of bacterial  
267 suspension at an OD<sub>620nm</sub> = 2.57 (final OD<sub>620nm</sub> in well = 0.6), 100 µL of sample and 15 µL of  
268 resazurin (PrestoBlue®, Invitrogen™, USA). The microplate was incubated for 3 hours at  
269 30°C. These test conditions (incubation period, cell density, oxygen marker, and used strains)  
270 were determined during the optimization step of the experimental conditions (see results in  
271 3.1 and 3.2). Subsequently, the fluorescence signal relative to the biological activity of the  
272 strains was individually read with the TriStar LB941 Multimode Microplate Reader as  
273 described above (see 2.4.1).

274 All the samples (see 2.6) were tested in duplicate, and the collected values were exported to  
275 an overall database. This data set was used to establish the correlation algorithm between the  
276 biological values provided by the bacterial set and the BOD<sub>5</sub> reference method (ISO 5815).

277 **2.8. Data processing: design of correlation models by a statistical approach**

278 The aim of the data processing was to design algorithms to correlate the biological data  
279 (fluorescence values) obtained from the selected strains with the BOD<sub>5</sub> reference values; the  
280 final goal was to propose a mathematical model to predict the BOD<sub>5</sub> value from only the data  
281 provided by these strains. All models were designed via a dedicated software (Neuro One  
282 6.13.0.5, Netral, France).

283 **2.8.1. Architecture of the model (number of biological descriptors and nature of the model)**

284 First, the data were analysed using a mono-parametric approach. The information  
285 (fluorescence values) from each strain was individually correlated with the reference values  
286 by linear regression. Because the results were insufficient (low correlation rates), the data  
287 processing focused on a multi-component strategy (based on several strains), which was also  
288 based on a linear regression algorithm. This step allowed the determination of the number of  
289 biological descriptors (selected strain) required to assess the BOD<sub>5</sub>. The predictive abilities of  
290 these models are limited.

291 To improve the performance of the models for the assessment of the BOD<sub>5</sub>, non-linear  
292 algorithms by neural networks were deployed. These models are defined according to  
293 architectural characteristics, such as the number of hidden neurons, the activation modalities  
294 of the neurons, and the data standardization conditions. Several architectures were compared  
295 for the overall database to identify *a priori* the most relevant algorithms.

296 **2.8.2. Model validation**

297 The last data processing step consisted of comparing the preselected architectures based on  
298 their predictive abilities with new samples. For this purpose, the overall database was  
299 randomly separated into two groups: a learning set (n= 148 assays) and a validation set (n= 60  
300 assays). The learning set was used to design new algorithms based on the preselected

301 architectures (number of hidden neurons, activation mode, and data standardization). To  
302 assess the robustness of these new models and choose the most relevant model, the estimated  
303 BOD<sub>5</sub> values obtained from the validation set were compared to the reference values. This  
304 step was also performed with Neuro One 6.13.0.5 (Netral, France).

### 305 **3. Results and discussion**

#### 306 **3.1. Improvement of biological activity monitoring**

307 The reference method used to assess BOD<sub>5</sub> is based on oxygen consumption as a marker of  
308 the biodegradation activity. This approach induces some technical constraints, such as a  
309 measurement range that is limited by the saturation rate of oxygen in water (**Cregut et al.,**  
310 **2013**) or the tightness of the system to air (contaminant oxygen). In this context, resazurin  
311 appears to be a relevant marker for monitoring biological activity (**Cregut et al., 2013**). This  
312 non-toxic non-fluorescent dye ( $\lambda_{\text{absorbance max}} = 601 \text{ nm}$ ) is reduced in the cell cytoplasm to a  
313 fluorescent compound, resorufin ( $\lambda_{\text{absorbance max}} = 571 \text{ nm}$ ,  $\lambda_{\text{excitation}} = 535\text{-}560 \text{ nm}$ ,  $\lambda_{\text{emission}} =$   
314  $590\text{-}615 \text{ nm}$  – see Fig. SD-2), by the enzymatic chain involved in cellular respiration (**Cregut**  
315 **et al., 2013; Fai and Grant, 2009; O'Brien et al., 2000**).

316 These two cell activity markers were compared during glucose biodegradation monitoring  
317 (MM supplemented with glucose as the sole carbon source) in *E. coli*. Several glucose  
318 concentrations (range from 0.1 to 500 mg.L<sup>-1</sup>) were tested to determine the highest  
319 measurable concentration according to the marker used (Fig. 2A). Comparative analysis of  
320 these results showed that use of resazurin as a marker could assess the glucose concentration  
321 up to 495 mg.L<sup>-1</sup> versus only 125 mg.L<sup>-1</sup> with oxygen. In light of these data, the fluorescent  
322 dye was selected for continuation of the study.

323 Subsequently, we were interested in the effect of the cell density of *E. coli* on the  
324 measurement duration with resazurin as cell activity marker. Figure 2B shows that an increase  
325 in the cell density substantially reduced the analysis duration; a cell density of 0.1 (OD<sub>620nm</sub>)  
326 resulted in a measurement duration of almost 12 hours versus 2 hours for an OD<sub>620nm</sub> = 1. The  
327 results also showed a reduction in the measurement range that was linked to the increasing  
328 cell density. This trend can be explained by an increase in the signal-to-noise ratio (see Fig.  
329 SD-3). An important noise (fluorescence baseline) measured with the high cell densities  
330 probably is induced by endogenous cellular respiration. At the same time, the saturation limit  
331 of the approach is reached at approximately 800,000 RFU. Consequently, the maximal  
332 measurable concentrations should not exceed 500 mg.L<sup>-1</sup> for a cell density of 0.1 (OD<sub>620nm</sub>)  
333 versus 170 mg.L<sup>-1</sup> for an OD<sub>620nm</sub> = 1.

334 To comply with the approach expectations (analysis limited in time but ensuring a relevant  
335 measurement range), the better compromise seemed to be a cellular density of 0.6  
336 corresponding to an analysis duration of approximately 3 hours.

### 337 **3.2. Selection of a representative set of strains**

338 The objective of this step was to constitute a set of controlled bacterial strains that were able  
339 to reliably assess the BOD<sub>5</sub> in a limited time and under reproducible conditions.

#### 340 **3.2.1. *Pre-selection of bacterial candidates***

341 The first step was dedicated to the preselection of bacterial strains that were likely to become  
342 a biological strain capable of assessing the BOD<sub>5</sub>. A literature review identified the major  
343 phyla present in the activated sludge of wastewater treatment plants (main targets of the  
344 proposed metrological approach) (Isazadeh et al., 2016; Xia et al., 2010; Zhang et al.,  
345 2018). More than 1200 operational taxonomic units (OTUs) were identified as potential  
346 candidates in activated sludge from WWTPs. Several of these OTUs were directly rejected

347 due to their strong pathogenicity (i.e., *Campylobacteraceae*, *Helicobacteraceae*,  
348 *Xanthomonadales*, some *Burkholderiales*, some *Bacteriodetes*, etc.) or their unsuitable  
349 metabolism (strictly anaerobic strains, such as *Desulfobacterales* and *Desulfovibrionales*, and  
350 marine strains, such as *Alteromonadales*), which were incompatible with the intended use of  
351 the technology (Fig. 3).

352 Among the selected families, we focused on the identification of bacterial species that met the  
353 criteria of the approach. The pre-selected strains should preferably be isolated from activated  
354 sludge of WWTPs, should be largely represented in these environments, should have wide  
355 and complementary *a priori* biodegradation abilities, should not be pathogenic and should be  
356 available in the current strain banks (i.e., DSMZ or ATCC). In accordance with these five  
357 criteria, 31 strains (among more than 1200 identified OTUs) were pre-selected before a  
358 second characterization step based on their ease of use (quick growth and ease of cultivation).  
359 From the obtained results, 3 strains were discarded (complexity of culture or too low a growth  
360 rate), which reduced the bacterial preselection to 28 strains (see Table SD-1). The overall pre-  
361 selection strategy is shown in Figure 3.

### 362 3.2.2. Selection of a representative set of strains

363 The simultaneous use of the 28 strains as individual descriptors of the BOD<sub>5</sub> was extremely  
364 technically complex and expensive. Consequently, we needed to reduce the number of strains  
365 used in the bioassay. First, we focused on the biodegradation capabilities of the selected  
366 strains. The biodegradation profile of each candidate was determined using 71 carbon sources  
367 (see Fig. SD-1), and hierarchical clustering was performed according to their similarities  
368 (agglomerative hierarchical clustering) (Fig. 4). Eight clusters with comparable  
369 biodegradation capabilities were identified among the 28 preselected strains (similar profiles  
370 at 95%). For example, group A, which consisted of 6 strains (*A. townneri*, *B. aquatica*, *P.*

371 *panthropus*, *Z. resiniphila*, *B. vertrisii* and *B. funiculus*), exhibited a modest range (only 6) of  
372 molecules that were likely to be biodegraded among the 71 proposed carbon sources.  
373 Conversely, group H offered a very wide set of biodegradation capabilities with ca. 85%  
374 biodegraded molecules (*R. radiobacter*, *B. subtilis*, *Staphylococcus sp.*, *S. yanoikuyae* and *A.*  
375 *aurescens*).

376 In parallel, all strains were characterized according to their robustness and biodegradation  
377 kinetics (Fig. 5). These criteria were taken into account to select a representative among the  
378 bacterial groups with similar biodegradation profiles. To limit the effect of the water toxicity  
379 on the metrological approach, strains with strong robustness were preferred because of their  
380 inherent capabilities to resist environmental stresses. Indeed, in the case of excessive water  
381 toxicity or too high sensitivity of strains, the bacterial cells will be not able to provide reliable  
382 information inducing erroneous results. However, strains with high biodegradation kinetics  
383 were *a priori* more likely to provide a biological signal quickly. Consequently, these strains  
384 were also preferred in view of the context of technological development of this study.

385 Figure 5A reports the specific robustness of each strain. These values were calculated based  
386 on 23 stress conditions and varied between 0.13 for *B. mycoides* (the most sensitive strain  
387 among the 28 tested) and 1 for *E. coli*. Figure 5B depicts the biodegradation kinetics of each  
388 strain for the synthetic wastewater (SWW). From these data that were calculated from the  
389 maximal measured slope of oxygen consumption, we noted significant differences between  
390 strains. For example, the biodegradation kinetics reached 0.33 mg BOD.L<sup>-1</sup>.h<sup>-1</sup> for *B. vertrisii*,  
391 whereas the kinetics achieved by *P. putida* (DSM 1868) were approximately 12 mg BOD.L<sup>-1</sup>.h<sup>-1</sup>.  
392 Consequently, the second strain should *a priori* provide a faster biological signal, as  
393 desired for our bioassay.

394 The final selection process of representative strains among each cluster was based on these  
395 last two criteria in addition to some criteria relative to the utilization of the strains, such as the  
396 flocculation of cells during growth (complexity of the preparation step). For example, group  
397 A consists of 6 strains (*B. funiculus*, *B. vertrisii*, *Z. resiniphila*, *P. pantrophus*, *B. aquatica*  
398 and *A. towneri*). Among these 6 strains, *B. aquatica* was selected to represent this group  
399 because of its capabilities (highest robustness and fastest biodegradation kinetics).

400 Finally, among the 28 candidates, a representative set of 8 strains was definitively selected,  
401 consisting of *B. aquatica*, *C. testosteroni*, *P. putida* (DSM 1868), *V. paradoxus*, *C.*  
402 *pseudodiphtheriticum*, *P. mirabilis*, *E. coli* and *B. subtilis*.

### 403 **3.3. BOD<sub>5</sub> assessment with the standard method and the bioassay.**

404 The BOD<sub>5</sub> values of the 104 real samples (collective or non-collective sanitation or  
405 stormwater) were determined according to the standard method (ISO 5815) (see Table SD-2).  
406 As expected, the samples taken before treatment by a WWTP showed significant pollution  
407 levels, with an average BOD<sub>5</sub> of nearly 120 mg.L<sup>-1</sup> (min. 66 mg.L<sup>-1</sup>; max. 190 mg.L<sup>-1</sup>). The  
408 levels measured in the treated waters varied strongly according to their origins (collective or  
409 non-collective sanitation). The average BOD<sub>5</sub> of stormwater was approximately 40 mg.L<sup>-1</sup>  
410 (min. 12 mg.L<sup>-1</sup> ; max. 54 mg.L<sup>-1</sup>); this pollution is probably due to runoff of water on roads  
411 and roofs, which transfers pollutants from support to water.

412 In parallel, each sample was subjected to the selected strains to determine the corresponding  
413 biological activity (fluorescence values obtained after 3 hours of exposure for the 104  
414 analysed samples in duplicate). The final database consisted of 208 assays. Each sample is  
415 defined by 8 fluorescence values (one per strain) and a BOD<sub>5</sub> reference value obtained using  
416 the standard method. The data analysis was performed from this database. The main objective  
417 of this supplementary step was to propose a reliable correlation model that allowed prediction

418 of the BOD<sub>5</sub> after only 3 hours. Nevertheless, we were focused on the number of strains  
419 required to assess the BOD<sub>5</sub> before improving the correlation algorithm a second time  
420 (numerical model allowing calculation of the BOD<sub>5</sub> from fluorescence values provided by the  
421 selected strains).

### 422 3.3.1. Determination of the number of biological descriptors needed to assess the BOD<sub>5</sub>

423 The data provided by each selected bacterial strain were individually correlated by linear  
424 regression with the BOD<sub>5</sub> reference values (Fig. 6). The average correlation rate was  
425 approximately 0.11, with a standard deviation of 0.08. Among the 8 selected strains, the best  
426 correlation rate obtained with *B. aquatica* ( $r^2=0.2638$ ) was too low to be used as a reliable  
427 bio-indicator of the BOD<sub>5</sub>.

428 Nevertheless, each strain provided a portion of the overall information. Consequently,  
429 increasing the number of strains used to assess the BOD<sub>5</sub> should improve the relevance of the  
430 approach. However, the number of strains required to ensure a reliable measurement is  
431 unknown.

432 Correlation models (linear regression) were designed based on the use of 1 to 8 strains. The  
433 obtained models are presented on Figure 7. The correlation rate between the predicted  
434 (calculated from fluorescence data provided by the selected strains) and reference values  
435 (standard method) tended to increase with the number of strains used. Indeed, the average rate  
436 obtained from a single strain was approximately 0.11, whereas this rate reached  
437 approximately 0.6 when the 8 strains were used.

438 Increasing the number of strains used to assess the BOD<sub>5</sub> was associated with a significant  
439 increase in the correlation rate of the models. However, the predictive capabilities of the best  
440 model based on the 8 strains was not sufficient ( $r^2=0.58$ ). In view of these results, a second  
441 step was required to improve the predictive capabilities of the data analysis model.

442 3.3.2. Improvement of the correlation model

443 To overcome this limitation and to propose a more reliable method for the assessment of the  
444 BOD<sub>5</sub>, the regression linear algorithms were compared to a neural network approach. The  
445 neural network is a framework for many different machine learning algorithms that can  
446 process complex data. Consequently, an exploratory step was performed to screen the  
447 architectures with the most potential. Several architectures were compared with the dedicated  
448 Neuro One software (Version 6.13.0.5, Netral, France). The differences between the tested  
449 architectures were based on the number of hidden neurons, the activation mode, and the data  
450 standardization. More than 800 models (80 architectures and 10 models by architecture  
451 obtained by iterations) were generated (see Table SD-3) by supervised learning from the  
452 overall database. The correlation rate obtained between the predicted and reference BOD<sub>5</sub>  
453 values varied from 0.36 (below the model based on linear regression – 0.58) to 0.93. Among  
454 the 80 tested architectures, several architectures had relevant predictive capacities with  
455 correlation rates superior to 0.8. The best model ( $r^2 = 0.93$ ) was based on an  
456 orthonormalization (Gram Schmidt process) of inputs, 4 hidden neurons activated by a  
457 hyperbolic tangent algorithm and post-treatment of the outputs.

458 These models are iteratively designed from existing data (i.e., they modify the different  
459 variables that control the variables during the learning process according to the data taken into  
460 account). Consequently, over-learning is a significant risk, which leads to questions about the  
461 relevance and the robustness of the predicted BOD<sub>5</sub> values obtained from new data. To  
462 remove this uncertainty, new models were designed using a fraction of the overall database.

463 3.3.3. Model validation

464 The global database was randomly separated into two sets: the learning (n=148) and the  
465 validation (n=60) sets (see 2.8.2). The learning set was used to design models with the 10 best  
466 *a priori* architectures determined during the previous step (see Table SD-3, conditions

467 overlined in grey). Ten models were generated based on the architecture. The robustness of  
468 these models was compared from unused data in the validation set based on the correlation  
469 rate between the predicted and reference values.

470 Among the 100 models, the most relevant was able to estimate the BOD<sub>5</sub> of the validation  
471 samples (data not included during the learning process) with strong reliability ( $r^2=0.851$ ) (Fig.  
472 8). This model consists of a hidden layer of three neurons activated by a hyperbolic tangent  
473 function. The input and output data are standardized with respect to a centred, reduced,  
474 normal distribution. From a global perspective, the BOD<sub>5</sub> values calculated from the bacterial  
475 set are relatively consistent with the reference values. The model seems to be particularly  
476 reliable between 25 and 300 mg/L<sup>-1</sup> BOD<sub>5</sub> (according to the reference method) for  
477 wastewaters (collective or non-collective sanitation). Nevertheless, its performance decreases  
478 for samples with low biodegradable organic loads (inferior to 25 mg.L<sup>-1</sup> according to the  
479 reference method) and for stormwaters. Two hypotheses can explain this result: (i) the current  
480 model is not sufficiently accurate to predict low BOD<sub>5</sub> values (a specific model must be  
481 designed to analyse specifically these low values) and/or (ii) an organic composition of  
482 stormwater different from that of wastewaters (organic compounds collected during water  
483 runoff on roads or roofs) induces specific findings from the bacterial descriptors.

484 Similar results were obtained by Raud and Kikas (2013). Their strategy was based on a  
485 microbial set consisting of 7 bacterial strains selected from their intrinsic biodegradation  
486 properties. The processing of multivariate data was performed using a PLS algorithm. The  
487 validation was only performed on artificial samples (synthetic wastewater spiked with phenol,  
488 milk, fat or cellulose) and not on real wastewater samples. Nevertheless, the correlation rate  
489 between the calculated and reference values was close to 1 ( $r^2=0.85$  in our study). Opposition  
490 to real samples seems to decrease the correlation rate (calculated values versus reference  
491 values), because mimicking a real sample with synthetic effluents is difficult.

492 The reference values own an inherent variability close to 20% (Jouanneau et al., 2014)  
493 notably due to the environmental variability of used inoculum. Conversely, the mean  
494 variability of data provided by the selected strains is 6.62% (results not showed). During the  
495 learning process, the model adjusts the model variables to match the data provided by the  
496 bacterial set with the reference data integrating at the same time its inherent variability (20%).  
497 To improve prediction reliability, it is required to know the “true” reference value. For this, it  
498 would be necessary to carry out several reference analysis by samples to determinate these  
499 "true" values (statistical analysis) in order to design new predictive models.

#### 500 3.3.4. *Post analysis*

501 In accordance with the selected architecture, a post analysis was performed to demonstrate the  
502 relevance of the use of several strains as bio-indicators associated with data processing by the  
503 neural network. Figure 9 shows the average correlation rates (linear least squares method) and  
504 the associated standard deviations depending on the number of strains used and according to  
505 the models used (linear regression or neural network). The results show a significant  
506 improvement in the reliability of the provided BOD<sub>5</sub> values compared with those obtained  
507 with the linear regression models.

508 With this post-analysis, we demonstrated firstly that the number of biological descriptors was  
509 a crucial factor in the development of a method to assess the biodegradable organic load. The  
510 use of a single bio-indicator seems to significantly limit the measurement performance of the  
511 strategy. The review of Jouanneau et al. (2014) tended to reinforce this observation. Indeed,  
512 among the many cited publications, strategies based on a unique biological descriptor show  
513 lower abilities to assess the BOD<sub>5</sub> in real samples (see Fig. SD-4) and wider variability  
514 between studies. From the publications cited in this review, the average correlation rate was  
515 close to 0.81 (SD=0.24) with a unique biological descriptor ( $r^2_{\min} = 0.09 / r^2_{\max} = 1$ ) versus  
516 0.95 (SD=0.07) with a consortium ( $r^2_{\min} = 0.81 / r^2_{\max} = 1$ ). Strategies based on consortium-

517 related systems use microbial mixtures from natural or artificial origins in most cases. These  
518 strategies are in agreement the protocol described in the reference method (ISO 5815) but  
519 induce wide inherent variability due to the descriptors used.

#### 520 **4. Conclusion**

521 In this work, we proposed an overall strategy to assess the BOD<sub>5</sub> from the selection of  
522 bacterial indicators to the validation of real wastewater samples. By acting on some levers,  
523 such as the measurement conditions or the selection of 8 bacterial bio-indicators, we  
524 developed a bioassay that allowed reliable and accurate BOD<sub>5</sub> measurement and required only  
525 3 hours versus 5 days for the reference method. The measurement range being enlarged, no  
526 dilution is required to assess BOD<sub>5</sub> in most wastewater, contrary to the reference method  
527 limited to 6 mg/L BOD. The obtained results are very encouraging and show a strong  
528 correlation with the reference BOD<sub>5</sub> values ( $r^2=0.85$ ). Nevertheless the method seems to reach  
529 its limits regarding to the assessing of low BOD<sub>5</sub> levels (inferior to 25 mg.L<sup>-1</sup> according to the  
530 reference method). Further work will have to be carried out to improve the model in these  
531 cases.

532 This significant technological improvement should increase the monitoring level of  
533 wastewaters. Indeed, by reducing the analysis duration, it is easier to control the process of  
534 water treatment and, if necessary, to adjust some parameters, such as the oxygen intake or the  
535 water flow, due to the time scale that is consistent with the treatment duration (2-3 days in  
536 WWTPs).

537 This metrological strategy was filed with the National Institute of Industrial Property to be  
538 patented (deposit number: 1851496).

539 **5. ACKNOWLEDGMENT**

540 The research was supported by the French National Research Agency (program BIOGUARD  
541 - ANR/Ecotech 2011-2014).

542 **Bibliography**

- 543 Arlyapov, V., Kamanin, S., Ponamoreva, O., Reshetilov, A., 2012. Biosensor analyzer for  
544 BOD index express control on the basis of the yeast microorganisms *Candida maltosa*,  
545 *Candida blankii*, and *Debaryomyces hansenii*. *Enzyme Microb. Technol.* 50, 215–220.  
546 <https://doi.org/10.1016/j.enzymictec.2012.01.002>
- 547 Atlas, R.M., 1997. *Handbook of Microbiological Media*. Second Edition, Lawrence C.Parks.  
548 ed. CRC press.
- 549 Catterall, K., Zhao, H., Pasco, N., John, R., 2003. Development of a Rapid Ferricyanide-  
550 Mediated Assay for Biochemical Oxygen Demand Using a Mixed Microbial  
551 Consortium. *Anal Chem* 75, 2584–2590. <https://doi.org/10.1021/ac0206420>
- 552 Chan, Chiyui, Lehmann, M., Chan, K., Chan, P., Chan, Chiwai, Gruendig, B., Kunze, G.,  
553 Renneberg, R., 2000. Designing an amperometric thick-film microbial BOD sensor.  
554 *Biosens. Bioelectron.* 15, 343–353. [https://doi.org/10.1016/S0956-5663\(00\)00090-7](https://doi.org/10.1016/S0956-5663(00)00090-7)
- 555 Cregut, M., Jouanneau, S., Brillet, F., Durand, M.-J., Sweetlove, C., Chenèble, J.-C.,  
556 L'Haridon, J., Thouand, G., 2013. High throughput and miniaturised systems for  
557 biodegradability assessments. *Environ. Sci. Pollut. Res.* 21, 9545–9552.  
558 <https://doi.org/10.1007/s11356-013-2236-2>
- 559 Dudal, Y., Holgado, R., Knoth, K., Debroux, M., 2006. A fluorescence-based microplate  
560 assay to quantify DOM-induced catabolic activity. *Anal. Bioanal. Chem.* 384, 175–  
561 179. <https://doi.org/10.1007/s00216-005-0140-4>
- 562 Duncan, M., 2004. *Domestic wastewater treatment in developing countries*. Earthscan, UK,  
563 USA.
- 564 Fai, P.B., Grant, A., 2009. A rapid resazurin bioassay for assessing the toxicity of fungicides.  
565 *Chemosphere* 74, 1165–1170. <https://doi.org/10.1016/j.chemosphere.2008.11.078>
- 566 Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., Alcamo, J., 2013. Domestic and  
567 industrial water uses of the past 60 years as a mirror of socio-economic development:  
568 A global simulation study. *Glob. Environ. Change* 23, 144–156.  
569 <https://doi.org/10.1016/j.gloenvcha.2012.10.018>
- 570 Gao, J., Ellis, L.B.M., Wackett, L.P., 2010. The University of Minnesota  
571 Biocatalysis/Biodegradation Database: improving public access. *Nucleic Acids Res.*  
572 38, D488–D491. <https://doi.org/10.1093/nar/gkp771>
- 573 Garcia, I.M., Rodriguez, J.R.B., Rodríguez, J.J.S., Suárez, B.P., Bocardo, J.R.P., Martín, N.S.,  
574 2006. *Guide for the Treatment of Urban Waste Water in Small Settlements (Main*  
575 *Text)*, ITC. ed.

576 Great Britain. Royal commission on sewage disposal, 1908. Final report of the commissioners  
577 appointed to inquire and report what methods of treating and disposing of sewage  
578 (including any liquid from any factory or manufacturing process) may properly be  
579 adopted. General summary of conclusions and recommendations ., Cornell University  
580 Library. ed.

581 Guyard, C., 2010. DBO5: un paramètre qui monte. *Eau Ind. Nuis.*

582 Isazadeh, S., Jauffur, S., Frigon, D., 2016. Bacterial community assembly in activated sludge:  
583 mapping beta diversity across environmental variables. *MicrobiologyOpen* 5, 1050–  
584 1060. <https://doi.org/10.1002/mbo3.388>

585 ISO 5815-1:2003, 2003. , Water quality -- Determination of biochemical oxygen demand  
586 after n days (BOD<sub>n</sub>) -- Part 1: Dilution and seeding method with allylthiourea  
587 addition.

588 ISO 5815-2:2003, 2003. , Water quality -- Determination of biochemical oxygen demand  
589 after n days (BOD<sub>n</sub>) -- Part 2: Method for undiluted samples.

590 Jia, J., Tang, M., Chen, X., Qi, L., Dong, S., 2003. Co-immobilized microbial biosensor for  
591 BOD estimation based on sol–gel derived composite material. *Biosens. Bioelectron.*  
592 18, 1023–1029. [https://doi.org/10.1016/S0956-5663\(02\)00225-7](https://doi.org/10.1016/S0956-5663(02)00225-7)

593 Jiang, Y., Xiao, L.-L., Zhao, L., Chen, X., Wang, X., Wong, K.-Y., 2006. Optical biosensor  
594 for the determination of BOD in seawater. *Talanta, China-Japan-Korea Environmental*  
595 *Analytical Chemistry Symposium*China-Japan-Korea Environmental Analytical  
596 *Chemistry Symposium*China-Japan-Korea Environmental Analytical Chemistry  
597 *Symposium* 70, 97–103. <https://doi.org/10.1016/j.talanta.2005.11.046>

598 Jouanneau, S., Recoules, L., Durand, M.J., Boukabache, A., Picot, V., Primault, Y., Lakel, A.,  
599 Sengelin, M., Barillon, B., Thouand, G., 2014. Methods for assessing biochemical  
600 oxygen demand (BOD): A review. *Water Res.* 49, 62–82.  
601 <https://doi.org/10.1016/j.watres.2013.10.066>

602 Kong, Y., Xia, Y., Nielsen, J.L., Nielsen, P.H., 2007. Structure and function of the microbial  
603 community in a full-scale enhanced biological phosphorus removal plant.  
604 *Microbiology* 153, 4061–4073.

605 Kwok, N.-Y., Dong, S., Lo, W., Wong, K.-Y., 2005. An optical biosensor for multi-sample  
606 determination of biochemical oxygen demand (BOD). *Sens. Actuators B Chem.* 110,  
607 289–298. <https://doi.org/10.1016/j.snb.2005.02.007>

608 Lin, L., Xiao, L.-L., Huang, S., Zhao, L., Cui, J.-S., Wang, X.-H., Chen, X., 2006. Novel  
609 BOD optical fiber biosensor based on co-immobilized microorganisms in ormosils  
610 matrix. *Biosens. Bioelectron.* 21, 1703–1709.  
611 <https://doi.org/10.1016/j.bios.2005.08.007>

612 Liu, Changyu, Zhao, H., Zhong, L., Liu, Chang, Jia, J., Xu, X., Liu, L., Dong, S., 2012. A  
613 biofilm reactor-based approach for rapid on-line determination of biodegradable  
614 organic pollutants. *Biosens. Bioelectron.* 34, 77–82.  
615 <https://doi.org/10.1016/j.bios.2012.01.020>

616 Mateo-Sagasta, J., Raschid-Sally, L., Thebo, A., 2015. Global Wastewater and Sludge  
617 Production, Treatment and Use, in: Drechsel, P., Qadir, M., Wichelns, D. (Eds.),  
618 Wastewater. Springer Netherlands, pp. 15–38. [https://doi.org/10.1007/978-94-017-9545-6\\_2](https://doi.org/10.1007/978-94-017-9545-6_2)  
619

- 620 O'Brien, J., Wilson, I., Orton, T., Pognan, F., 2000. Investigation of the Alamar Blue  
621 (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur. J.*  
622 *Biochem.* 267, 5421–5426. <https://doi.org/10.1046/j.1432-1327.2000.01606.x>
- 623 OECD, 2010. Test No. 209: Activated Sludge, Respiration Inhibition Test (Carbon and  
624 Ammonium Oxidation). Organisation for Economic Co-operation and Development,  
625 Paris.
- 626 Oota, S., Hatae, Y., Amada, K., Koya, H., Kawakami, M., 2010. Development of mediated  
627 BOD biosensor system of flow injection mode for shochu distillery wastewater.  
628 *Biosens. Bioelectron.* 26, 262–266. <https://doi.org/10.1016/j.bios.2010.06.040>
- 629 Pitman, K., Raud, M., Kikas, T., 2015. Biochemical oxygen demand sensor arrays. *Agron.*  
630 *Res.* 13, 382–395.
- 631 Preston-Mafham, J., Boddy, L., Randerson, P.F., 2002. Analysis of microbial community  
632 functional diversity using sole-carbon-source utilisation profiles – a critique. *FEMS*  
633 *Microbiol. Ecol.* 42, 1–14. <https://doi.org/10.1111/j.1574-6941.2002.tb00990.x>
- 634 Rastogi, S., Kumar, A., Mehra, N.K., Makhijani, S.D., Manoharan, A., Gangal, V., Kumar,  
635 R., 2003. Development and characterization of a novel immobilized microbial  
636 membrane for rapid determination of biochemical oxygen demand load in industrial  
637 waste-waters. *Biosens. Bioelectron.* 18, 23–29. [https://doi.org/10.1016/S0956-5663\(02\)00108-2](https://doi.org/10.1016/S0956-5663(02)00108-2)
- 639 Raud, M., Kikas, T., 2013. Bioelectronic tongue and multivariate analysis: A next step  
640 in BOD measurements. *Water Res.* 47, 2555–2562.  
641 <https://doi.org/10.1016/j.watres.2013.02.026>
- 642 Raud, M., Tenno, T., Jõgi, E., Kikas, T., 2012. Comparative study of semi-specific  
643 *Aeromonas hydrophila* and universal *Pseudomonas fluorescens* biosensors for BOD  
644 measurements in meat industry wastewaters. *Enzyme Microb. Technol.* 50, 221–226.  
645 <https://doi.org/10.1016/j.enzmictec.2012.01.003>
- 646 Rocher, V., Meche, P., Rechdaoui, S., Paffoni, C., Goncalves, A., Pichon, S., Pauss, A., 2011.  
647 New tools for the determination of PO<sub>4</sub><sup>3-</sup> and BOD in municipal wastewaters? *Eau*  
648 *Ind. Nuis.* 143–149.
- 649 Snaidr, J., Amann, R., Huber, I., Ludwig, W., Schleifer, K.H., 1997. Phylogenetic analysis  
650 and in situ identification of bacteria in activated sludge. *Appl Env. Microbiol.* 63,  
651 2884–96.
- 652 Sun, J.-Z., Peter Kingori, G., Si, R.-W., Zhai, D.-D., Liao, Z.-H., Sun, D.-Z., Zheng, T., Yong,  
653 Y.-C., 2015. Microbial fuel cell-based biosensors for environmental monitoring: a  
654 review. *Water Sci. Technol. J. Int. Assoc. Water Pollut. Res.* 71, 801–809.  
655 <https://doi.org/10.2166/wst.2015.035>
- 656 Tizzard, A.C., Bergsma, J.H., Lloyd-Jones, G., 2006. A resazurin-based biosensor for organic  
657 pollutants. *Biosens. Bioelectron.* 22, 759–763.  
658 <https://doi.org/10.1016/j.bios.2006.01.011>
- 659 Xia, S., Duan, L., Song, Y., Li, J., Piceno, Y.M., Andersen, G.L., Alvarez-Cohen, L.,  
660 Moreno-Andrade, I., Huang, C.L., Hermanowicz, S.W., 2010. Bacterial community  
661 structure in geographically distributed biological wastewater treatment reactors. *Env.*  
662 *Sci Technol.* 44, 7391–6.

- 663 Xin, L., Wang, Xudong, Guo, G., Wang, Xiaoru, Chen, X., 2007. An optical biosensing film  
664 for biochemical oxygen demand determination in seawater with an automatic flow  
665 sampling system. *Meas. Sci. Technol.* 18, 2878–2884. [https://doi.org/10.1088/0957-](https://doi.org/10.1088/0957-0233/18/9/017)  
666 0233/18/9/017
- 667 Yang, H., Zhou, M., Liu, M., Yang, W., Gu, T., 2015. Microbial fuel cells for biosensor  
668 applications. *Biotechnol. Lett.* 37, 2357–2364. [https://doi.org/10.1007/s10529-015-](https://doi.org/10.1007/s10529-015-1929-7)  
669 1929-7
- 670 Yoshida, N., Hoashi, J., Morita, T., McNiven, S.J., Nakamura, H., Karube, I., 2001.  
671 Improvement of a mediator-type biochemical oxygen demand sensor for on-site  
672 measurement. *J. Biotechnol.* 88, 269–275. [https://doi.org/10.1016/S0168-](https://doi.org/10.1016/S0168-1656(01)00282-6)  
673 1656(01)00282-6
- 674 Yoshida, N., Nakamura, H., Karube, I., Yano, K., Morita, T., McNiven, S.J., 2000. A  
675 mediator-type biosensor as a new approach to biochemical oxygen demand estimation.  
676 *The Analyst* 125, 2280–2284. <https://doi.org/10.1039/b0059951>
- 677 Zhang, B., Xu, X., Zhu, L., 2018. Activated sludge bacterial communities of typical  
678 wastewater treatment plants: distinct genera identification and metabolic potential  
679 differential analysis. *AMB Express* 8, 184. <https://doi.org/10.1186/s13568-018-0714-0>  
680

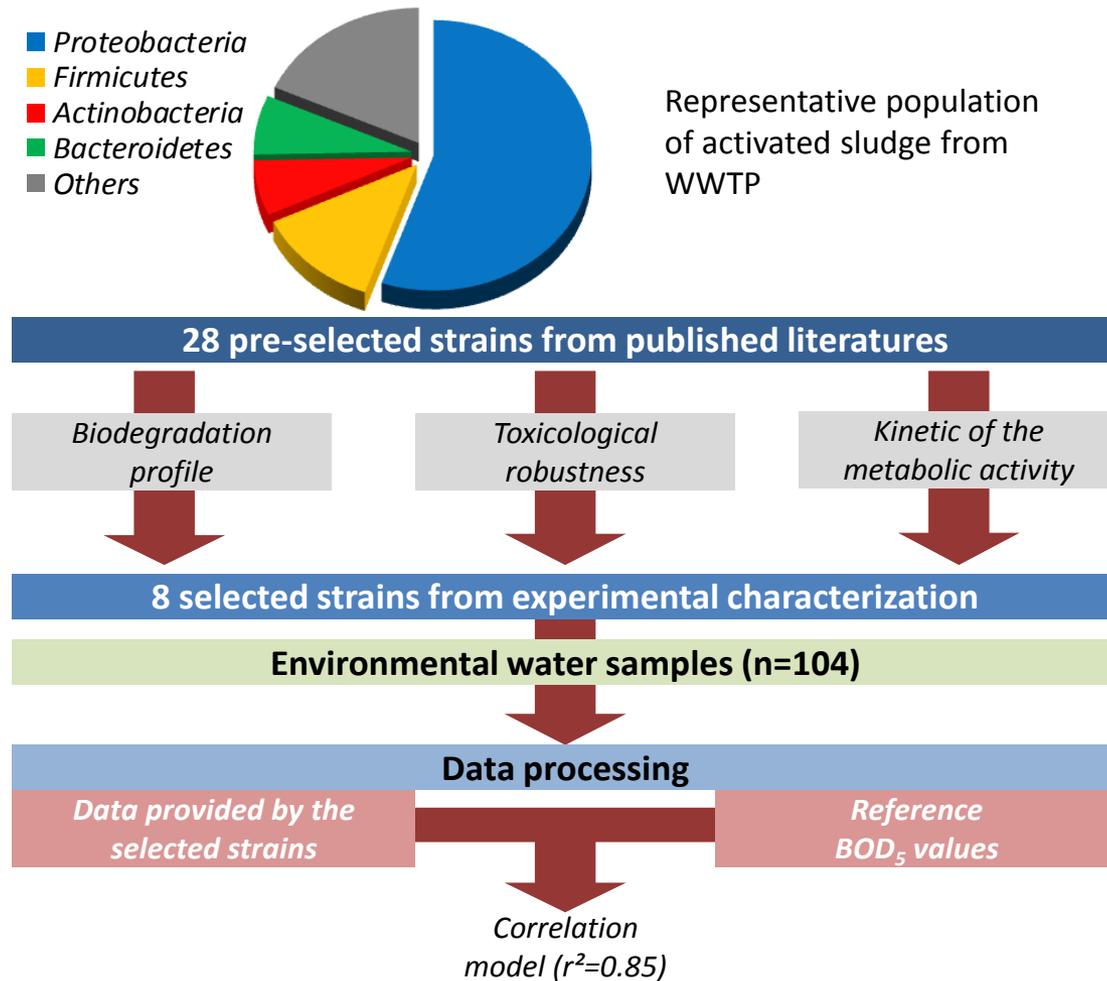


Fig.1. Strategy applied to the development of the BOD<sub>5</sub> assessment method (Isazadeh et al., 2016; Xia et al., 2010; Zhang et al., 2018).

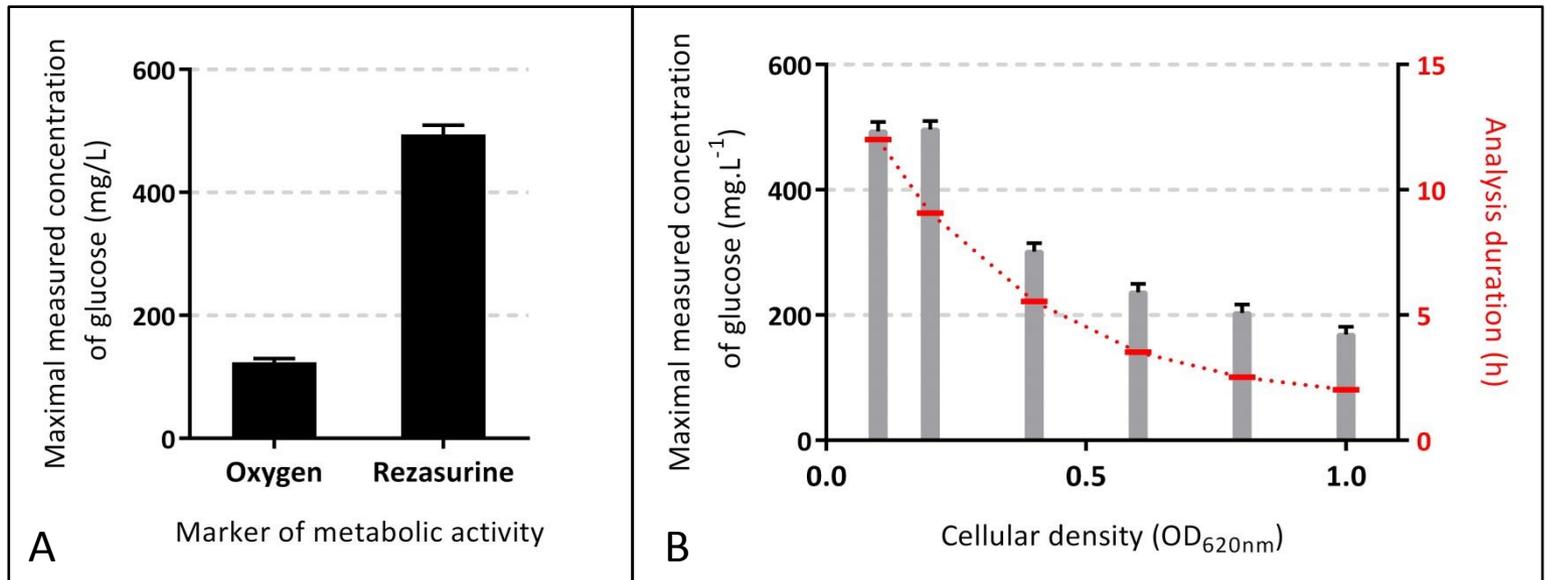


Fig 2. Optimization of the measurement conditions. (A) Comparison of two activity marker (oxygen or resazurin) on the maximal measured concentration of glucose (strain; *E.coli*; analysis duration: 12 hours; cellular density:  $OD_{620nm} = 0.1$ , standard deviation calculated from duplicates). (B) Effect of the cellular density ( $OD_{620nm} = 0.1/0.2/0.4/0.6/0.8/1$ ) on the analysis duration (red dotted line) and the measurement range with resazurin (grey bars) (standard deviation calculated from duplicates).

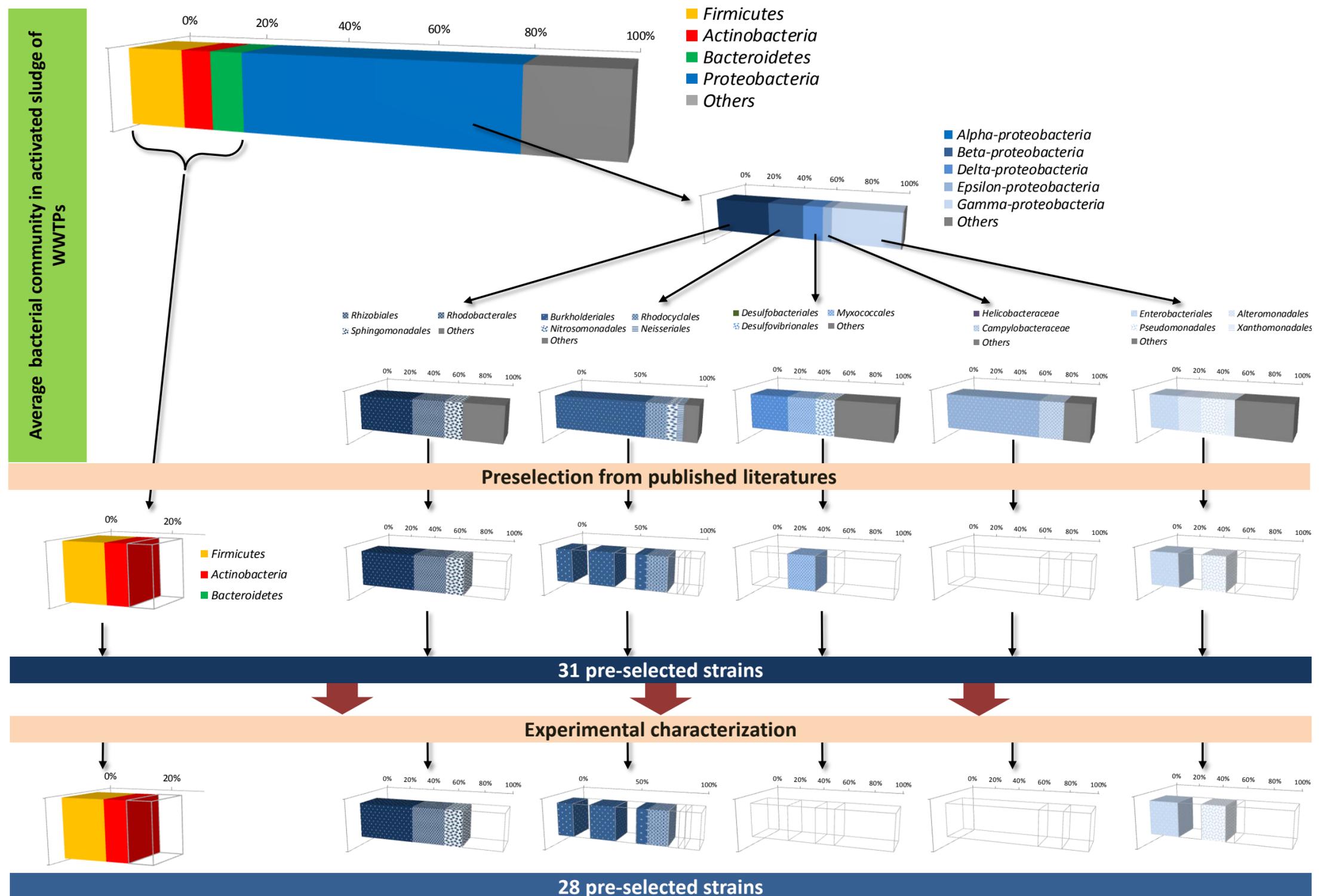


Fig 3. Overall preselection strategy of bacterial candidates.

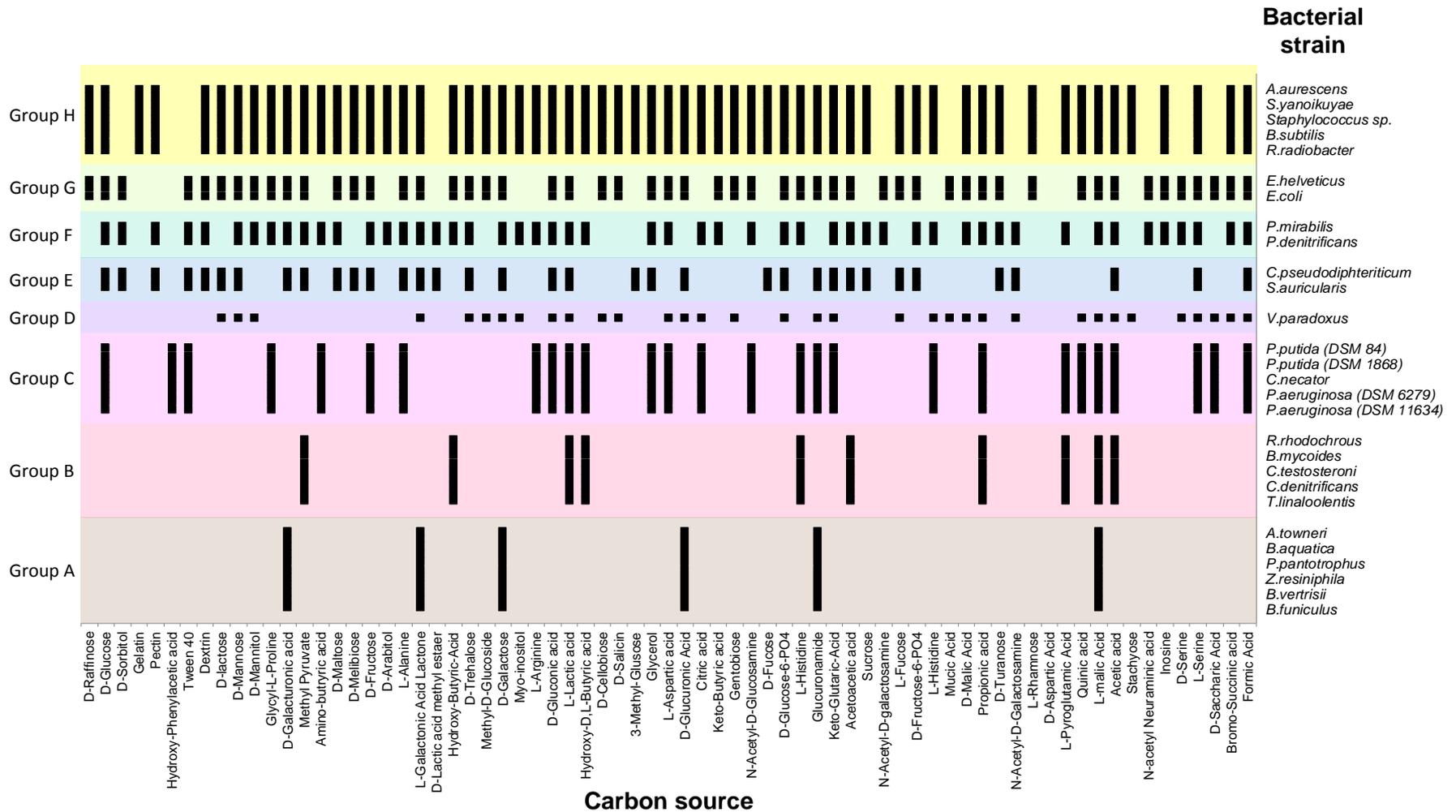


Fig 4. Average profiles obtained after clustering of the data collected from the 28 preselected bacterial strains (clustering carried out from 71 carbon sources). Black squares correspond to the carbon sources used as substrate by the bacteria strains.

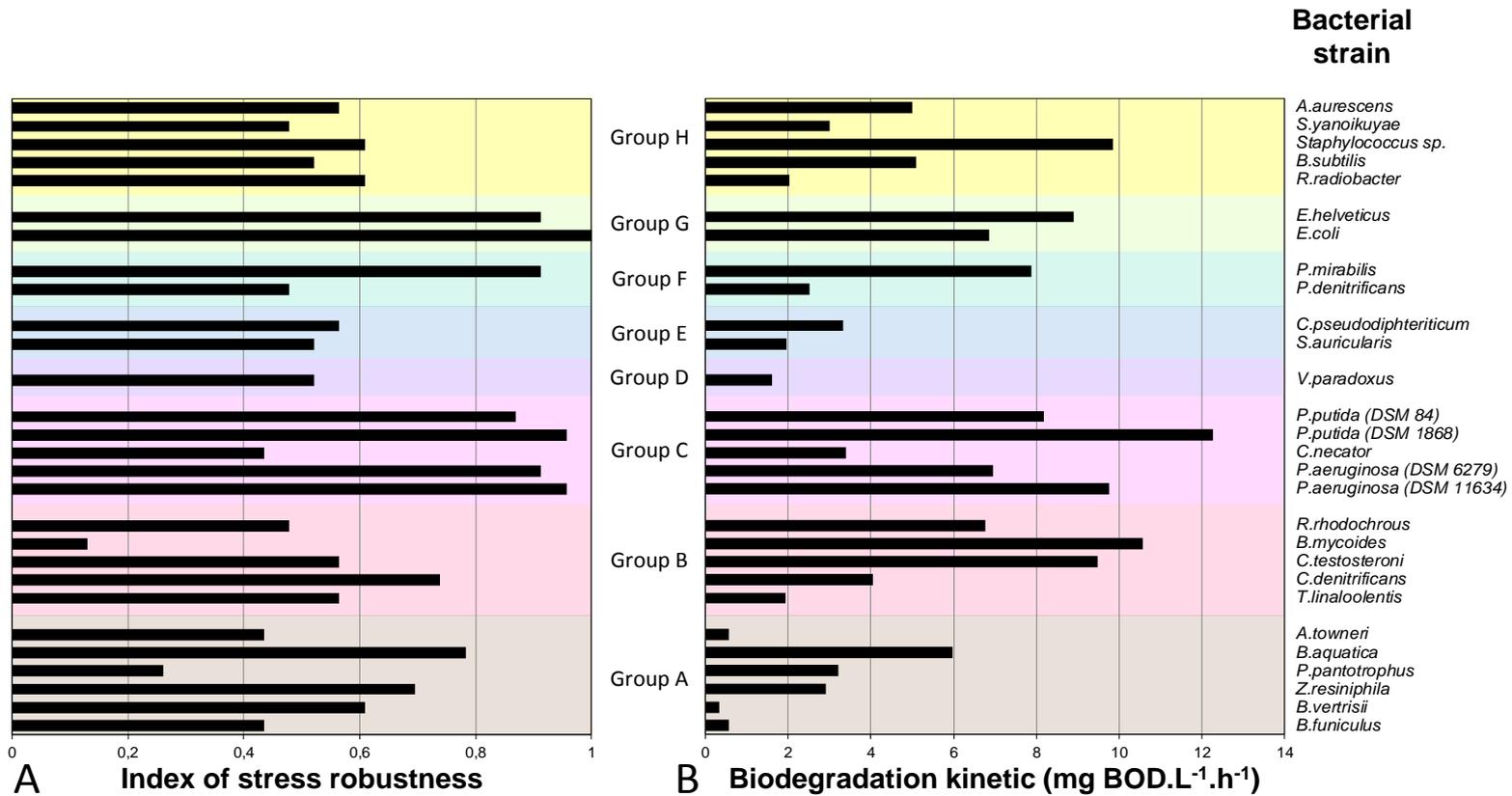


Fig 5. Biological characterization of pre-selected candidates. (A) Index of stress robustness of strains, (B) biodegradation kinetics obtained with a standardized wastewater. For the sake of clarity, the error bars are not represented but the standard deviations do not exceed 10% between the duplicates.

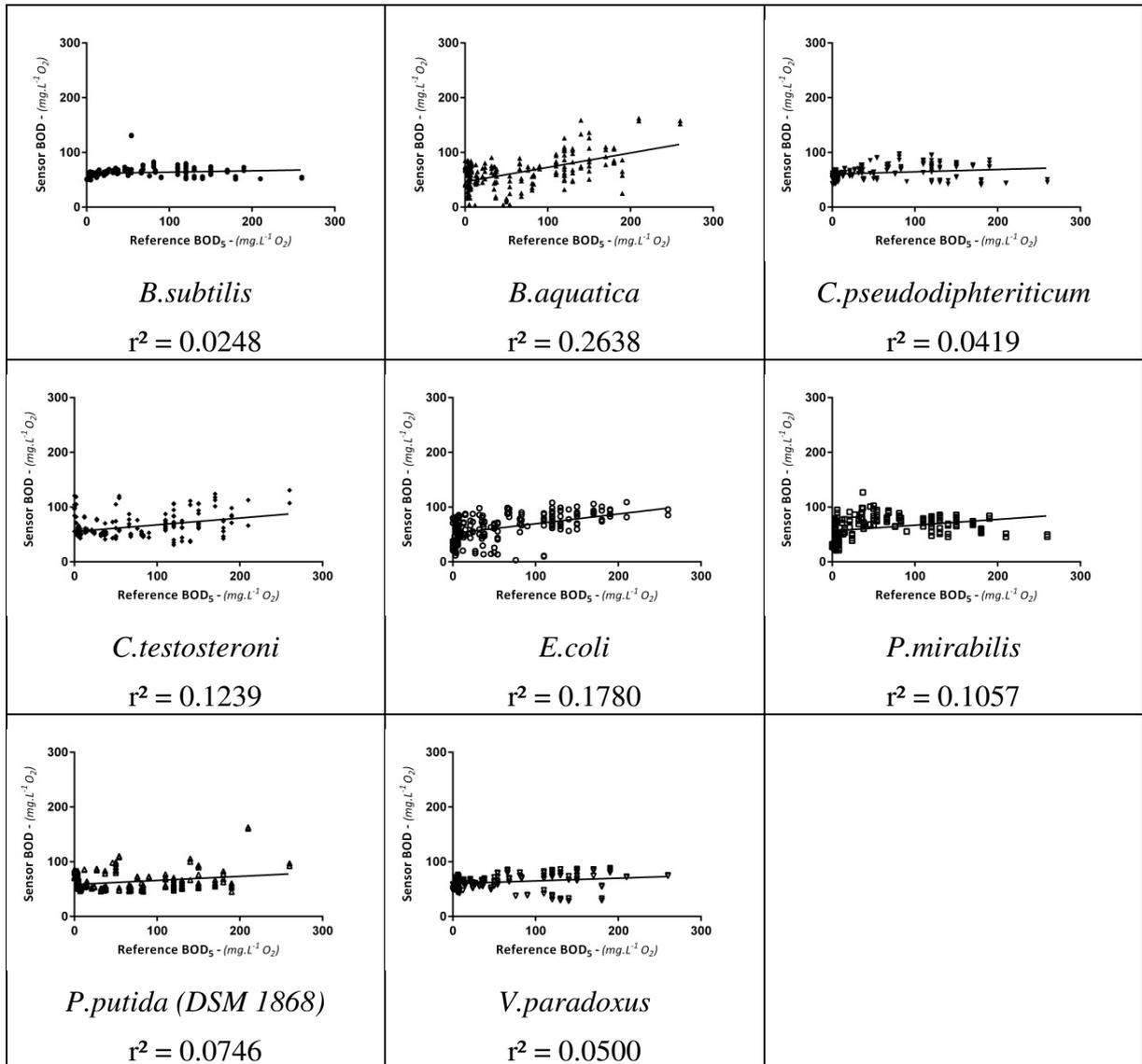


Fig. 6. Linear correlation between the fluorescence values and the corresponding reference values of BOD<sub>5</sub>; according to an individual approach. Sensor BOD : Fluorescence value provided by the strain convert in BOD<sub>5</sub> equivalent

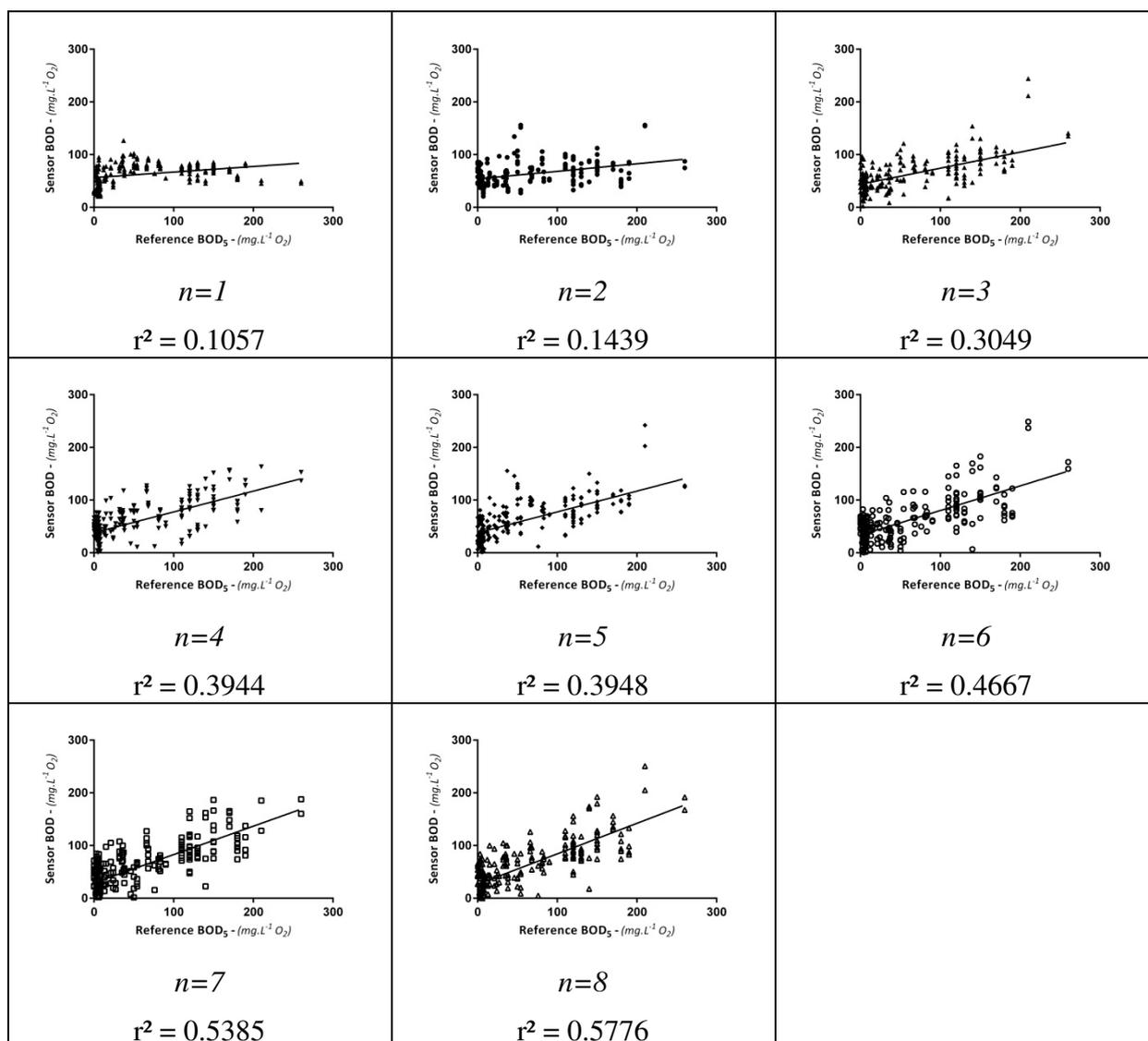


Fig. 7. Correlation rate (linear regression) obtained according to the number of used strains. Sensor BOD: BOD<sub>5</sub> equivalent calculated from fluorescence values of n strain(s). n = number of strains used in the linear models.

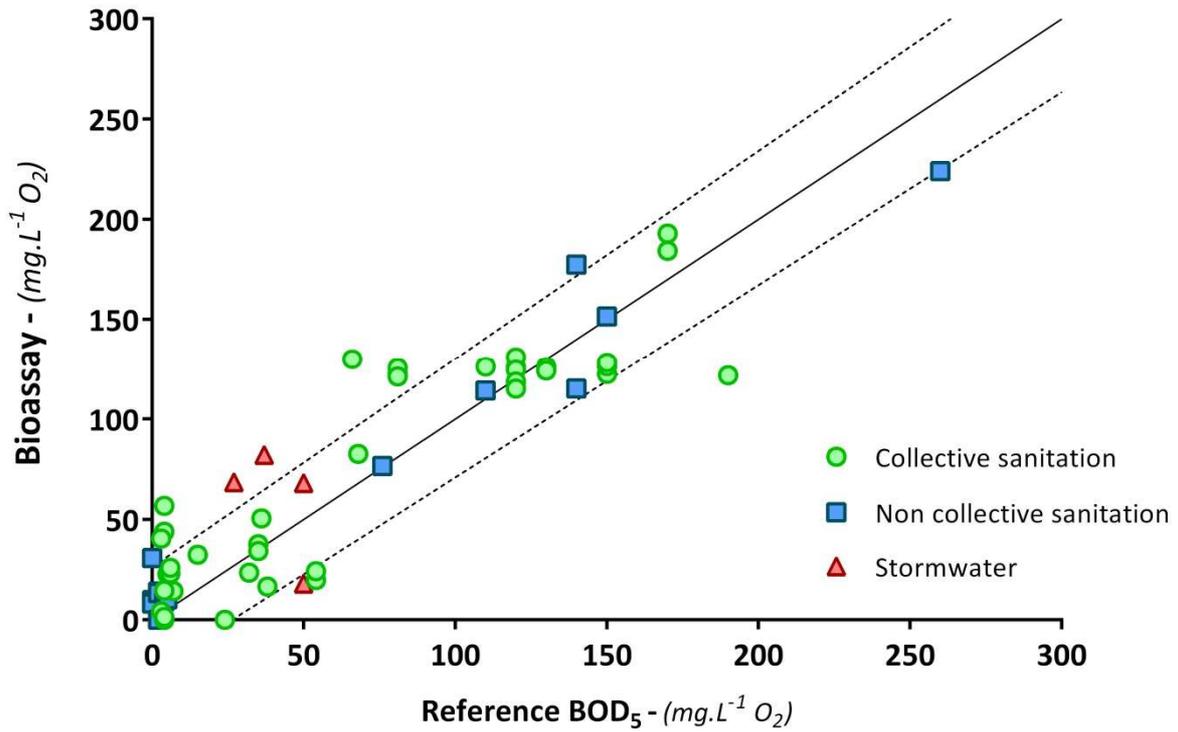


Fig. 8. Correlation between the reference values (5 days required) and the  $BOD_5$  estimation calculated from the bacterial set (3 hours required). Broken line: confidence interval of model. 1/1 correlation between calculated and reference values is shown with the center-line on the graph.

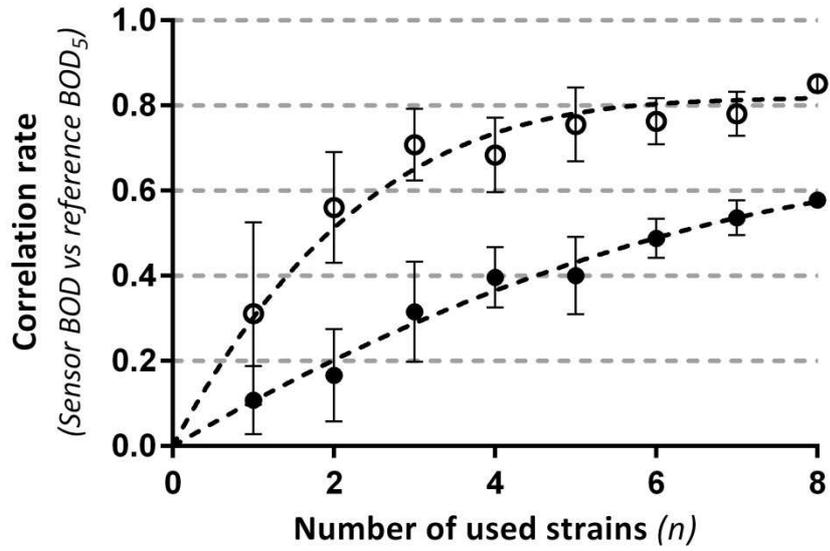


Fig. 9. Relationship between the correlation rate and the number of biological strains used to estimate the BOD<sub>5</sub>. Correlation rates and standard deviations were calculated from available models (a single model for the overall bacterial set). Black circle: Correlation rate obtained from linear regression. Empty circle: Correlation rate calculated from neural network models.