Lipids in benthic diatoms: A new suitable screening procedure

Eva Cointet, Gaëtane Wielgosz-Collin, Vona Méléder, Olivier Gonçalves

Lipids in benthic diatoms: a new suitable screening procedure

Short title: Screening and rapid selection of lipid rich benthic diatoms

Eva Cointet[1], Gaëtane Wielgosz-Collin[1], Vona Méléder[1], Olivier Gonçalves[2]

[1] Université de Nantes, Laboratoire Mer Molécules Santé, EA 21 60, BP 92208, 44322 Nantes, France

[2] Université de Nantes, GEPEA, UMR CNRS-6144, Bât.CRTT, 37 Boulevard de l'Université, BP406, F-44602 Saint-Nazaire Cedex, France

Abstract

The selection of suitable and indigenous microalgae species is a fundamental requirement in developing added-value bioactive compounds recoverable in the food, health, and cosmetics markets. In this work, an integrated screening approach was developed to characterize the lipid rate of 33 diatom species (including 15 species studied for the first time) belonging to 16 genera from the Nantes Culture Collection, with the main objective of discovering bioactive lipid producers. For that purpose, a simple reliable method for establishing growth kinetics of strains and semi-quantitative analysis of lipid rates was developed. Growth kinetics measurements were achieved by daily minimal measurement fluorescence (F0) whereas lipid rate analyses were performed by high-throughput Fourier Transform Infrared spectroscopy on entire cells and lipid extracts. Results indicated that the method could be used directly on entire cells in spite of the presence of silica for the FTIR approach (due to frustule). The total lipid rate was species-dependant and ranged from 3.7% to 30.5% DW. Six strains out of 33 were found to present a higher total lipid rate superior to 15% DW, and 11 showed medium lipid rates ranging from 10% to 15% DW. The results revealed that five diatom species i.e. *Amphora* sp. NCC169, *Nitzschia* sp. NCC109, *Nitzschia alexandrina* NCC33, *Opephora* sp NCC366 and *Staurosira* sp. NCC182 presented interesting growth capabilities and should be further investigated as potential sources for their original lipid rate.

Keywords: FTIR spectroscopy, growth kinetics, benthic Diatoms, lipid rate, bioactive fatty acids.

Highlights

- Minimal fluorescence measurement characterized growth capabilities
- FTIR spectroscopy on whole cells characterized lipid rates
- Lipid rates of 15 strains were highlighted for the first time
- Five strains of benthic diatoms were selected for furthers biotechnology applications

*Corresponding authors

Email: olivier.goncalves@univ-nantes.fr (Olivier Gonçalves), eva.cointet@univ-nantes.fr (Eva Cointet), vona.meleder@univ-nantes.fr (Vona Méléder), Gaetane.Wielgosz-Collin@univ-nantes.fr (Gaëtane Wielgosz-Collin).

Preprint submitted to Algal Research September 09, 2018

Conflict of interest

The authors declare that there is no conflict of interest. No conflicts, informed consent, human or animal rights are applicable to this work.

Author's contribution

Eva Cointet, Gaëtane Wielgosz-Collin, Vona Méléder and Olivier Gonçalves designed and supervised the research. Eva Cointet, Gaëtane Wielgosz-Collin, Vona Méléder and Olivier Gonçalves conducted experiments. Eva Cointet, Gaëtane Wielgosz-Collin, Vona Méléder and

Olivier Gonçalves analyzed and interpreted the data and drafted the manuscript. Eva Cointet, Gaëtane Wielgosz-Collin, Vona Méléder and Olivier Gonçalves critically reviewed the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

This work was supported by the regional Atlantic Microalgae research program (AMI), funded by the Pays de la Loire region. We also express our sincere thanks to GEPEA staff in particular Remy Coat and Delphine Kucma for support and advice on the FTIR spectrometer.

Abbreviation

ATR: Attenuated Total Reflectance

AW: Ash weight

DW: Dry weight

DHA: Docosahexaenoic acid

EPA: Eicosapentaenoic acid

FA: Fatty acid

F0: Minimum chlorophyll fluorescence

HTSXT-FTIR: Fourier-transform infrared spectroscopy high-throughput screening extension

LED: Late exponential day

LED biomass: Biomass corresponding to culture harvest days

MUFA: Monounsaturated fatty acid

NCC: Nantes Culture Collection

NDVI: Normalized Difference Vegetation Index

N: Nitrogen

PFD: Photon flux density

PAM: Pulse amplitude modulation

PAR: Photosynthetically active radiation

PUFA: Polyunsaturated fatty acid

ρ : Reflectance

SD : Standard deviation

Si : Silica

SFA: Saturated fatty acid

TAG: Triacylglycerol

μmax: Maximum growth rate

1 Introduction

In the last two decades, a large body of research has focused on finding new strains of microalgae capable of producing high lipid content for a wide range of applications including pharmaceutical, cosmetics and alternative biofuels [1–4]. In the kingdom of microalgae, diatoms are very accessible resources, since they are ubiquitously found in most aquatic environments (rivers, oceans, coastal areas). They constitute a unicellular eukaryotic group with a typical species-specific siliceous cell wall (also known as frustule). They present different life-forms that could be benthic (microphytobenthos) or planktonic (phytoplankton). Marine diatoms can grow quickly and store large amounts of lipids [5]. Their lipids are mainly composed of a neutral fraction with traces of sterols and polar lipids [6]. Neutral lipids constitute the reserve fraction, with triacylglycerol (TAG) accounting for more than 60% of the total lipids [7]. Their PUFAs are mainly composed of eicosapentaenoic acid (EPA, C20:5 n-3) [8] but some strains were also found to present docosahexaenoic acid (DHA, C22:6 n – 3) [9]. The biosynthesis of the lipids varies within the different diatom species, their growth stages, and environmental parameters [10,11]. Previous studies [5–9] have demonstrated their ability for lipid production, more specifically for the PUFA fraction (DHA and EPA), recognized for its broad spectrum bioactivities (anti-carcinogenic, immune modulation, anti-diabetic, anti-obesity and anti-thrombotic properties) [12]. Unfortunately, the Diatoms group is poorly studied and constitutes therefore an underexploited resource [13] even if the number of their genera and species is estimated to be between 250 and 100,000 [14]. Bioprospecting efforts should therefore be encouraged in order to assess this potential. Basically, bioprospecting would be achieved if the identified microalgae could be exploited at an industrial scale for their biomass or their high value lipid compounds [1]. Therefore, during the screening approach, specific focus should be performed on the efficient identification of the appropriate microalgae strains, i.e. those characterized by high productivity (biomass and lipids), high resistance to

contamination and high tolerance to a wide range of environmental parameters [15–18]. Native species adapt to local environmental changes and should be thus resilient and competitive enough regarding these criteria. However, systematic estimation of the growth rates requires a time-consuming series of measurements in order to estimate biomass evolution. The cell count approach is among the most widely used. Alternatively, other parameters could be measured as proxy for the cell numbers, if it could be shown to be linearly correlated. Typical proxy measurements are *in vivo* fluorescence [19], optical density and biomass direct estimation (such as dry weight, pigment content) [20] and *in vivo* or in solvent spectroscopy [21]. The concentrations of protein, carbohydrates and lipids in cultures could also be used as proxy measures depending of the robustness of their linear relationship with either cell numbers or biomass. Finally, the cell number is recognized as being a robust reference method if "counting methods" are easily available and could be applied to a given biological system. Only *in vivo* fluorescence and spectroscopy seem to be faster than cell counting and are easy to use. For these reasons, one of the objectives of this study is to test both techniques as an alternative to counting in the screening process. Conventional methods used for lipid determination, systematically require solvent extraction and gravimetric determination. These methods are time consuming, need extensive manipulations, use high amounts of biomass (10-15 mg [22]) and have a low throughput screening rates. Consequently, a faster measurement of the lipid content is needed [23]. As for counting methods, alternative approaches exist, that can be roughly classified into invasive and non-invasive techniques. Fluorescent based technologies are the most commonly used and require fluorescent dyes like Bodipy. However, it is an indirect measure that has several issues such as sample preparation before staining, careful choice of dye especially in microalgae due to the presence of chlorophylls within chloroplasts, leading to non-quantitative information [24]. Vibrational spectroscopy such as Raman [25] could be a good alternative for investigating lipid content since it can image and chemically identify the

lipids without labelling. However, autofluorescence signals from the chloroplast will overwhelm the Raman signal. Increasing the acquisition time or incident power can solve this issue, although irreversible photo damage may result, leading to loss of the semi-quantitative information, limiting therefore its use in the context of high-throughput screening. Infrared spectroscopy has advantages over the above techniques, since it limits photodamage, is not influenced by autofluorescence, and presents robust systems such as high-throughput screening for chemical spectra acquisition for large quantities of samples. Coat *et al.* demonstrated [26] that it was robust and sensitive enough to quantify the lipids and that directly on entire microalgae cells. The second objective of this study was to test this technique as an alternative to gravimetric determination of lipid content.

In the present work, we propose to explore the potential for original lipid sourcing of the benthic diatoms hosted in the Nantes Culture Collection (NCC) bank. First – to identify the NCC strains that could potentially produce interesting fatty acids – a bibliographic inventory of current and past knowledge of the principal genus and species hosted in the NCC bank was conducted. It focused on the families of molecules with high added value for pharmacology, health, nutrition, and cosmetics such as PUFAs, taking also into account the influence of cultural conditions [27,28]. The selected species were then, investigated to gain basic knowledge on their biological characteristics, to highlight their chemo diversity (in terms of protein, carbohydrates, silica and lipid rates), and finally on their potential for high-value fatty acid molecule biosynthesis. To select the most promising NCC strains, an original workflow was developed, integrating different steps including the analysis of the NCC strain growth and their rapid biochemical profiling through HTSXT-FTIR. The results obtained in the present study are discussed hereafter.

2 Material and methods

2.1 Strains cultivation

Diatom strains from the NCC were selected and cultivated in 250 mL Erlenmeyer flasks filled with 150 mL of F/2 culture medium, enriched in silica using 0.47 µm filtered natural sea water [29, 30]. Vitamins and carbonates were added before autoclave sterilization (Autoclave Vertical Lequeux AUV 100L), salinity was adjusted at 28 and pH was fixed at 7.8 to reduce nutrient precipitation. Inoculation of each strain was performed from the stock cultures with an initial concentration of 30 000 cells.mL^{-1}. The diatom strains were thus grown at 16°C with a photon flux density (PFD) of 127 µmol photons m^{-2} s^{-1} provided under continuous light to the bottom of flasks by flat led pannel (LP EPURE-Chateaugiron, France).

2.2 Growth parameter estimations

To retrieve the growth parameters on the screened strains, the growth rate (µmax in day^{-1}) and the end of the exponential phase were identified. The growth rate identifies the faster growing strains. The end of the exponential phase, known to be the period of lipid accumulation [31], identifies the harvest day for lipid analyses. These parameters were retrieved using daily biomass estimation. All erlen flasks and stir bars were autoclaved. Before sampling, flasks were agitated by magnetic stirring for 2 minutes allowing cell and nutrient homogenization and aggregate destruction.

Alternative techniques were tested to replace the cell counting by hemocytometer (here Neubauer; n≥300) which is a time-consuming technique.,

- Minimum chlorophyll fluorescence measurement (F0) by fluorometry PAM (Water-PAM, Waltz, Germany). This parameter, proportional to Chl *a* content, was used as a proxy of the vegetal biomass [32]. Measurements were made directly on the microalgal suspension.

- Reflectance (ρ) of the cells by spectroradiometry (Jaz, Ocean Optics, USA) in the PAR domain (400-700 nm). Reflectance values were used to calculate the Normalized Difference Vegetation Index (NDVI) following the equation. (1), known to be proportional to Chl *a* content and used as a biomass proxy [21]. Measurements were made on filtered microalgae using CML microfiber filters with a 25 mm diameter and a 0.7 µm pore diameter and a 25mm Whatman filter funnel.

$$NDVI = \frac{\lambda 750 - \lambda 675}{\lambda 750 + \lambda 675} \quad \text{eq. (1)}$$

With λ750: maximal reflectance wavelength, and λ675: Chl *a* absorption wavelength

Daily measurements were performed on each sample of culture to follow the growth kinetics (at the beginning of the growth phase and during three days, between 3 mL and 5 mL of cultures were sampled, and then 2 mL until the end of the growth). To extract and compare the growth parameters from the alternative techniques and the cell count approach, the Gompertz model [33] was used to fit the growth data (equation. 2) using MATLAB software. It consisted of a latency phase followed by an exponential phase and a stabilization of the curve at its maximum phase.

$$f(x) = A \times e^{-e(\mu max \times \frac{e^1}{A} \times (\lambda - x) + 1)} \quad \text{eq. (2)}$$

With *A*: maximum cell concentration in the natural logarithm of the biomass (number of cells. mL^{-1}, F0 or NDVI); µmax: Maximum growth rate (day^{-1}); λ : Latency (days).

The *A* parameter allows calculation of the biomass ($LED_{biomass}$) obtained at the late exponential day (LED) (equation. 3). This corresponded to the harvest day for the lipid analyses.

$$LED_{biomass} = exp(A + log(Bmin)) \quad \text{eq. (3)}$$

With *Bmin* = Initial biomass (number of cells.mL^{-1}, F0 or NDVI); $LED_{biomass}$ = biomass of the late exponential day (cell.mL^{-1}, F0 or NDVI)

206 The comparison of the three thechniques, and the selection of the fastest one to estimate the
207 growth parameters was done using a panel of 6 very different strains, chosen for their different
208 growth kinetics and their different morphology (aggregate of cells, cells in chains or solitary
209 cells) : *Amphora* sp. 1 NCC260, *Entomoneis* sp. 1 NCC350, *Entomoneis* sp. 6 NCC335,
210 *Entomoneis paludosa* NCC18.2, *Extubocellulus cribriger* NCC229 and *Navicula* sp. 2
211 NCC226.

212 ## 2.3 Diatom strain characterisation

213 At the end of the exponential phase (LED), when the cells were harvested for HTSXT-FTIR
214 analyses (see §2.4), several analysis were perforemd to collect more information about the
215 strains:
216 - the cell length and the width were estimated using the image of light microscopy (OLYMPUS
217 CH40, ×400; n=150).
218 - the dry weight (DW) of the biomass was estimated by filtering 150 mL of algal suspension
219 through a microfiber filter, Whatman 47 mm diameter, 0.7 μm pore. The filters containing the
220 cells were washed using 10 mL of ammonium formiate (3%) to the remove salt. The wet filters
221 were frozen at -80°C and freeze-dried under a vacuum. DW ($g.L^{-1}$) and µmax (day^{-1}) were then
222 used to estimate the strain productitivty (Px) in $g.L^{-1}.day^{-1}$ (equation. 4).

223 $$Px = \mu max \times DW \quad \text{eq. (4)}$$

224 - the strains total lipid rate was assessed by gravimetric assay, to compare it to the infrared semi-
225 quantitative measurments (HTSXT-FTIR, see § 2.4). Biomass filtered, washed and freeze-dried
226 for dry weight estimation were used for lipid content estimation. The filters were macerated in
227 flasks using 100 mL of solvent per gram of biomass (dichloromethane-methanol (1:1 V/V))
228 [34]. Maceration at ambient temperature was performed for 24H on a vibrating tray (Edmund
229 Bühler GmbH, SM-30). After maceration, the mixture was filtered on pleated filters, 190 mm
230 diameter, 10µm pore, to remove the filter debris and the silica fragments. The filtrates were

231 transferred into a separatory funnel with 20 mL of distilled water and shaken for 5 min. The
232 lipid fraction (organic phase) was then separated from the separatory funnel, dried using an
233 anhydrous sodium sulfate salt, filtered, evaporated and weighed to obtain the crude lipid extract
234 (CLE) value. Total lipid rate (TLR) was finally expressed in % of the DW (equation. 5).

235 $$TLR = \frac{CLE}{DW} \times 100 \qquad \text{eq. (5)}$$

236 - the silica content of the cells was also determined to ensure the normalization of the FTIR
237 semi-quantitative results (see § 2.4). Cultures were harvested by filtering new 150 mL of the
238 algal suspension using the same recovery cell procedure used for dry weight estimation. Filters
239 were then freeze-dried, weighed (DW) and heated at 400°C for four hours in a muffle oven and
240 weighed (AW) and the silica proportion evaluated in % of DW by equation. 6 :

241 $$Silica\ content = \frac{AW}{DW} \times 100 \qquad \text{eq. (6)}$$

242 ## 2.4 Molecular profiles measured by infrared spectroscopy

243 The FTIR spectra acquisition on entire cells was performed according to Coat et al. 2014
244 recommendations [26]. It consisted in concentrating the cells up to 10^6 - 10^8 cells.mL^{-1} by
245 centrifugation (10 000g for 5 min) using Sigma 3K30 centrifuge. The supernatant was removed,
246 and the pellet resuspended in an ammonium formiate isotonic solution (68 g.L^{-1}). This process
247 was repeated twice to wash out the cells from the growth medium in order to avoid medium
248 contribution on the FTIR spectra. The cells were thus resuspended in 1mL in an isotonic
249 solution. Ammonium formiate solution prevents cell lyses during washing [35]. For the FTIR
250 spectra acquisition, a Bruker tensor 27 FTIR spectrometer equipped with a HTSXT plate reader
251 module coupled to OpusLab v 7.0.122 software (Bruker Optics, Germany) was used. Rinsed
252 cell aliquots of 5 µL were deposited on a 384 well microplate, void-dried in a vacuum desiccator
253 for at least 24 hours. FTIR spectra were then recorded in transmission mode directly on the
254 microplate loaded with the dried samples. This method was chosen since it is fast and non-

255 invasive on intact diatom cells, their biochemical signatures expressed in term of total lipids,
256 total proteins and total carbohydrates superimposed partially with the silica signal of the diatom
257 frustul. To confirm the whole cells FTIR spectra results, estimation of the lipid rate was also
258 performed on crude lipid extract, using attenuated total reflectance sampling system (ATR),
259 that was more suitable forcrude organic extracts. For the acquisition of the ATR spectra, 10 µL
260 of crude lipid extract (see §2.3. [34]) were deposited directly on the Bruker tensor 27 FTIR
261 spectrometer lens. The absorbance spectra were all collected between 4 000 cm^{-1} and 700 cm^{-1}
262 with 30 scans and averaged. The spectra were analyzed by relatively straightforward methods
263 such as peak ratios or integral ratios [36]. The lipid signature was associated to the CH_2-CH_3
264 signal (~3000 – 2800 cm^{-1}) and the ester bond (Eb) signal (~1740 cm^{-1}). The carbohydrate
265 signature was associated to the C-O-C signal of the polysaccharides (~1200-900 cm^{-1}) [37]. The
266 protein signature was associated to the amide II band (~1540 cm^{-1}) of the N-H of the amides
267 associated to the proteins. The silica signature was associated to the Si-O signal of the silicate
268 frustule (~1068 cm^{-1}) [38]. To estimate the relative content of the lipids, carbohydrates and
269 proteins, their respective peak heights (i.e. ester bond + (CH_2 + CH_3), ~1159 cm^{-1} and amide II)
270 were standardized to the silica peak [39]. (equation 7). For the crude lipid extract spectra, the
271 ratio used for that purpose was the ester bond and the CH_2 + CH_3 signals standardized with the
272 total spectrum area (equation 8). FTIR and ATR ratio were expressed in arbitrary units
273 abundance (a.u).

274 $$\text{FTIR} = \frac{Peak\ height\ (S)}{Silica\ peak\ high} \qquad \text{eq. (7)}$$

275 S = lipids (eb+(CH_2+CH_3)) or carbohydrates (~1159 cm^{-1}) or amide II (~1540 cm^{-1})

276 $$\text{ATR} = \frac{Peak\ area\ (eb+CH_2+CH3)}{Total\ spectra\ area} \qquad \text{eq. (8)}$$

2.5 Data processing

The Pearson product-moment correlation was carried out to test the positive correlation between the growth curves obtained with the cell count (Cells.mL^{-1}), fluorescence (F0) and reflectance (NDVI).

Comparison of the growth rate and LED estimated from the three techniques for the selected species were performed using ANOVA when the data presented normal distribution or the Kruskal – Wallis test when the data distribution was not normal. It was systematically followed by the Tuckey post hoc test. The ANOVA was also performed on the growth rate (µmax), production (Px) and lipid rate (TLR) results to identify the strains with the highest performance.

A multivariate correspondence analysis was performed on the FTIR ratio normalized by silica for lipids, protein and carbohydrates to assess the dispersion of the biochemical information across the screened species and to identify if strains could be classified according to this information. This method was chosen since it analyzes binary, ordinal and nominal data without distributional assumptions (unlike traditional multivariate techniques) and also to preserve the categorical nature of the variables. The correspondence analysis provided a unique graphical display showing how the variable response categories were related [40].

The Pearson product-moment correlation was used to test the correlation between the calculation methods using the FTIR spectra semi-quantitative information with the total lipid quantification reference method (the gravimetric approach). The Pearson product-moment correlation and the comparison of the growth parameters were carried out using SigmaStat 3.1 software. The Past3 software was used for the correspondence analysis approach. All experiments were performed in triplicate.

Finally, all the information and tests were combined and used for the validation of the screening methodology to characterize the diatom strains of the NCC and the selection of candidates presenting the best potential in terms of growth capabilities and lipid rates.

3 Results

3.1 Determination of the growth parameters

The latency, exponential and stationnary phases were observable whatever the technique used to establish the growth curve for the 6 strains selected for their different kinetic behaviour (Fig. 1). Moreover, the cell count, NDVI and F0 data were all correlated (Tab.1) confirming the same growth pattern. Cell count and NDVI were positively correlated (p<0.001) as cell count and F0 (p<0.001). R values for the NDVI varied from 0.70 to 0.90 and for F0 measurements from 0.60 to 0.93.

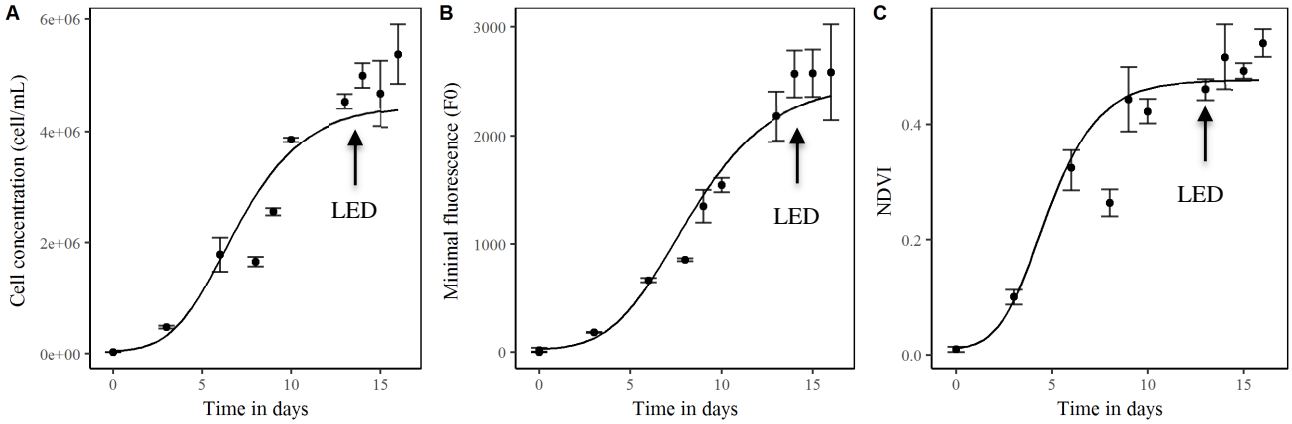


Figure1. Example of growth curves measured for *Amphora* sp 1 NCC260 by (A) cell count, (B) fluorometry PAM (F0) and (C) radiometry (NDVI). Points corresponding to cell concentration according to time. Line curves corresponding to the Gompertz model fitted to cell concentration as a function of time. n=3, vertical bar = SD. The arrows indicate the late exponential day estimated by the Gompertz model.

317 **Table 1.** Pearson product correlation between radiometry (NDVI) and cell count and
318 fluorometry PAM (F0) and cell count. Correlation was significant (p <0.001)

Species	NDVI	F0
Amphora sp. 1 NCC260	0.90	0.93
Entomoneis paludosa NCC18.2	0.86	0.82
Entomoneis sp. 1 NCC335	0.86	0.81
Entomoneis sp. 6 NCC350	0.80	0.60
Extubocellulus cf cribriger NCC229	0.70	0.90
Navicula sp. 2 NCC226	0.80	0.78

319

320 There were no significant differences ($p > 0.05$) between the technique used to estimate the
321 growth rate (µmax) and the late exponential day (LED) for the six tested species, except for the
322 µmax value estimated for *Entomoneis paludosa* (Tab. 2). For example, *Amphora* sp. 1 reached
323 the late exponential phase at the day c.a. 13 (LED, Fig. 1, Tab. 2) with a mean µmax of c.a.
324 0.81 (Tab. 2). At this time, the maximum biomass was reached by the cell count indicator with
325 $4.5 \pm 0.2 \times 10^6$ cells.mL^{-1} by PAM fluorometry with F0 values 2785 ± 609, and by radiometry
326 with NDVI reaching 0.47 ± 0.05. Because the fluorimetry is used extensively for the
327 measurement of extracted Chl *a*, for the estimation of the phytoplankton Chl *a in vivo* [19,41]
328 and do not need filtration of high amounts of culture (contrary to NDVI), it was selected as an
329 fast and reliable alternative approach to the cell count to determine the growth rates of all
330 selected diatom strains in this study. Regarding *Entomoneis paludosa*, a new experiment using
331 only PAM measurments, in the same culture conditions but without counting and NDVI
332 estimation were conducted. µmax values obtained were 0.52, 0.55, 0.55, leading to an average
333 value of 0.54 ± 0.016. A second statistical test performed on these new data concluded no
334 significant difference for µmax and LED ($p = 0.47$).

335 **Table 2**. Values of the growth parameters (maximum growth rate, µmax (in d^{-1}) and late
336 exponential day, LED (in days)) retreived from the Gompertz model using six species employed
337 for the comparison of the cell count approach with two alternative techniques: Fluorometry

PAM and radiometry. The calculated P value corresponded to the ANOVA test except for the (*) values that were obtained using the Kruskal-Wallis test. N=3, independent measurements, ± SD.

Species	Techniques	µmax (day^{-1})	LED (day)	P value (µmax)	P Value (LED)
Amphora sp. 1 NCC260	Cell count	0.76 ± 0.10	13 ± 1	P=0.8	P=0.16*
	Pam	0.78 ± 0.30	14 ± 1		
	Spectroradiometer	0.88 ± 0.25	12 ± 2		
Entomoneis paludosa NCC18.2	Cell count	0.50 ± 0.03	12 ± 1	P≤ 0.05[1]	P=0.07*
	Pam	0.95 ± 0.23[1]	10 ± 0		
	Spectroradiometer	0.47 ± 0.01	10 ± 0		
Entomoneis sp. 6 NCC335	Cell count	0.53 ± 0.10	14 ± 1	P=0.27	P=0.14*
	Pam	0.54 ± 0.12	12 ± 1		
	Spectroradiometer	0.32 ± 0.14	12 ± 2		
Entomoneis sp. 1 NCC350	Cell count	0.24 ± 0.01	16 ± 2	P=0.29	P=0.09
	Pam	0.37 ± 0.12	14 ± 1		
	Spectroradiometer	0.35 ± 0.06	14 ± 1		
Extubocellulus cf cribriger NCC229	Cell count	0.68 ± 0.04	14 ± 1	P=0.10	P=0.14
	Pam	0.81 ± 0.01	9 ± 0		
	Spectroradiometer	0.83 ± 0.07	10 ± 3		
Navicula sp. 2 NCC226	Cell count	0.91 ± 0.04	8 ± 0	P=0.42*	P=0.78*
	Pam	1.02 ± 0.23	8 ± 2		
	Spectroradiometer	0.87 ± 0.02	9 ± 2		

[1] Entomoneis paludosa µmax value for the second run was 0.54 and associated P value was 0.47

3.1.1 Diatom strain characteristics

Among the 68 screened strains, 33 were cultivated successfully and 36 did not grow (supplementary data 1). This information is reported in Tab. 3, (with the NCC reference, the sampling location, the cell size, the LED, the growth rate, the silica content, the DW and lipid rate). Globally the cell length varied from 58 ± 14 µm (*Craspedostauros* sp. 2 NCC218) to 4.8 ± 0.7 µm (*Extubocellulus cf cribriger* NCC229). The cell width varied from 22.6 ± 4.1 µm (*Lithodesmium sp* NCC353) to 2.9 ± 0.4 µm (*Fallacia* sp. 1 NCC303). The end of the exponential phase, corresponding to the harvest day (LED), varied from day 6 for *Surirella* sp. 1 NCC270 to day 16 for *Craspedostauros britannicus* NCC228 and *Navicula* sp. 1 NCC113 (ANOVA, $p < 0.001$). Three groups were therefore identified depending on their growth rate: group A, included *Nitzschia alexandrina* NCC33 and *Entomoneis* sp. 5 NCC302 with respective growth rates of 1.89 ± 0.10 and 1.50 ± 0.18 day^{-1}, group B with 13 species (i.e 40% of the total number of strain) presented a growth rate ranging from 1.19 ± 0.08 to 0.81 ± 0.09

355 day^{-1} and group C with 18 species (54%) had a growth rate below 0.8 day^{-1} with a minimum at
356 0.22 ± 0.05 day^{-1}.
357 To ensure the standardization of the FTIR results for inter species comparison purposes, the
358 total silica rate was evaluated for the 33 species. It showed a significant difference (ANOVA,
359 $p < 0.01$), where five species (i.e. 15% of the total number of strains) had a silica content of c.a.
360 40% of the dry weight with a maximum of 46.11 ± 4.58% for *Entomoneis* sp 1 NCC350. Three
361 species (i.e. 9% of the total number of strains) presented a silica content of c.a. 23% of the dry
362 weight with a minimum of 22.9 ± 2.6% for *Conticriba weissflogii* CCMP1336. In the other
363 species, representing more than 75% of the total number of strains, the differences were not
364 significant (ANOVA, $p = 0.055$), supporting a stable silica content situated at around 35% DW.
365 The dry weight (DW) biomass values of the assessed strains ranged from 0.502 ± 0.039 $g.L^{-1}$
366 (*Craspedostauros britannicus* NCC195) to 0.065 ± 0.001 $g.L^{-1}$ (*Entomoneis* sp. 3 NCC351).
367 For three strains (9% of the total number of strains) the DW was greater than 0.30 $g.L^{-1}$. For 27
368 strains (81%) the biomass was greater than 0.10 $g.L^{-1}$ but lower than 0.30 $g.L^{-1}$. For the three
369 remaining strains (9%) the biomass was lower than 0.10 $g.L^{-1}$.
370 The lipid rate estimated with the gravimetric method ranged from 30.5 ± 0.7% DW (*Nitzschia*
371 sp. 5 NCC109) to 3.7 ± 1.1% DW (*Brockmaniella brockmanii* NCC161). For three strains (9%)
372 the lipid rate was greater than 20% DW. For 14 strains (42%) the lipid rate was greater than
373 10% DW. In the remaining 16 strains (48%) the lipid rate was lower than 10% DW.
374 Biomass productivity varied substantially among the tested strains (Fig. 2A) and ranged from
375 0.36 ± 0.02 $g.L^{-1}.day^{-1}$ (*Nitzschia alexandrina* NCC33) to 1.4 ± 0.3 × 10^{-2} $g.L^{-1}.day^{-1}$
376 (*Entomoneis* sp. 3 NCC351). The results were found to be statistically different among those
377 strains (ANOVA, $p < 0.001$). Finally, the strains were clustered into three groups according to
378 the following parameters (Fig. 2): group 1 with productivity between 0.35 ± 0.02 and 0.19 ±
379 0.04 $g.L^{-1}.day^{-1}$; group 2 with productivity ranging from 0.17 ± 0.08 to 0.09 ± 0.02 $g.L^{-1}.day^{-1}$;

group 3 with lower productivity ranging from $7.5 \pm 0.4 \times 10^{-2}$ to $1.4 \pm 0.3 \times 10^{-2}$ ($g.L^{-1}.day^{-1}$). The most productive strains were those exhibiting the highest μmax associated to the highest DW. But they did not correspond to the richest in terms of total lipid rate (Fig. 2B).

Table 3. Characteristics and sampling locations of the investigated diatoms species. All data were obtained by experimental measurements. N=3, independent measurement, ± SD.

Species	NCC strain identification	Sampling location	Cell size (µm)		LED (day)	µmax		Silica content (%)	Biomass ($g.L^{-1}$)	TLR (% DW)	Ref
			Length	Width							
Nitzschia alexandrina	NCC33	*France, NW Atlantic coast*	11.0 ± 1.2	3.7 ± 0.4	11 ± 0	1.89 ± 0.10	A	27.2 ± 3.8	0.189 ± 0.004	12.7 ± 3.2	[15,42–49]
Entomoneis sp. 5	NCC302	*France, NW Atlantic coast*	22.4 ± 3.5	15.0 ± 2.2	9 ± 0	1.50 ± 0.18		38.9 ± 0.7	0.132 ± 0.025	6.9 ± 2.4	[50–54]
Entomoneis sp. 2	NCC20	*France, NW Atlantic coast*	30.5 ± 4.6	18.5 ± 3.2	8 ± 0	1.19 ± 0.08		34.3 ± 2.8	0.244 ± 0.042	10.3 ± 2.7	[50–54]
Staurosira sp.	NCC182	*France, NW Atlantic coast*	6.9 ± 1.4	5.5 ± 1.2	8 ± 2	1.17 ± 0.09		28.9 ± 5.1	0.262 ± 0.084	25.1 ± 1.2	[16]
Fallacia sp. 1	NCC303	*Greenland, east coast*	6.4 ± 0.7	2.9 ± 0.4	12 ± 2	1.12 ± 0.14		37.4 ± 4.5	0.154 ± 0.016	14.2 ± 1.9	-
Fallacia sp. 2	NCC304	*Greenland, east coast*	6.4 ± 0.8	3.1 ± 0.4	12 ± 0	1.07 ± 0.15		31.8 ± 4.9	0.265 ± 0.028	9.2 ± 3.3	-
Entomoneis sp. 7	NCC445	*France, NW Atlantic coast*	10.4 ± 1.5	4.1 ± 0.7	9 ± 2	1.02 ± 0.05		41.4 ± 2.9	0.122 ± 0.005	10.2 ± 2.8	[50–54]
Entomoneis paludosa	NCC18.1.1	*France, NW Atlantic coast*	18.4 ± 3.3	10.7 ± 1.9	11 ± 1	0.96 ± 0.03		37.8 ± 4.0	0.130 ± 0.016	5.3 ± 0.3	[50–54]
Entomoneis sp. 4	NCC301	*France, NW Atlantic coast*	23.9 ± 3.4	17.3 ± 2.2	7 ± 1	0.92 ± 0.14	B	26.7 ± 3.9	0.171 ± 0.012	6.7 ± 1.5	[50–54]
Pseudonitzschia americana	PNA06 KER	*France, NW Atlantic coast*	6.6 ± 0.8	4.5 ± 0.5	9 ± 2	0.89 ± 0.09		36.4 ± 0.6	0.099 ± 0.019	5.7 ± 1.6	-
Surirella sp. 1	NCC270	*France, NW Atlantic coast*	6.8 ± 1.3	5.0 ± 1.6	6 ± 1	0.85 ± 0.31		33.7 ± 3.5	0.173 ± 0.033	19.3 ± 2.8	-
Navicula sp. 2	NCC226	*France, NW Atlantic coast*	9.1 ± 1.3	5.6 ± 0.7	8 ± 2	0.82 ± 0.05		41.4 ± 1.5	0.131 ± 0.015	8.7 ± 2.5	[15,45,47,49,55,58]
Opephora sp. 1	NCC366	*France, NW Atlantic coast*	5.4 ± 0.8	3.7 ± 0.8	14 ± 1	0.82 ± 0.16		37.8 ± 8.4	0.233 ± 0.010	11.5 ± 3.1	-
Entomoneis paludosa	NCC18.2.1	*France, NW Atlantic coast*	18.8 ± 1.8	11.2 ± 2.5	12 ± 1	0.81 ± 0.05		36.0 ± 3.9	0.150 ± 0.003	7.3 ± 1.0	[50–54]
Extubocellulus cf cribriger	NCC229	*France, NW Atlantic coast*	4.8 ± 0.7	3.9 ± 0.5	9 ± 0	0.81 ± 0.09		33.7 ± 2.3	0.155 ± 0.003	14.2 ± 4.8	[52,59,60]
Amphora sp. 1	NCC260	*France, NW Atlantic coast*	9.8 ± 1.3	3.7 ± 0.5	13 ± 0	0.78 ± 0.30		36.2 ± 2.6	0.176 ± 0.041	18.5 ± 4.2	[49,55,58,61,62]
Conticriba weissflogii	CCMP1336	*USA, NE Atlantic coast*	15.6 ± 2.5	10.8 ± 1.4	12 ± 0	0.59 ± 0.15		22.9 ± 2.6	0.277 ± 0.035	7.2 ± 1.7	-
Brockmaniella brockmanii	NCC161	*France, NW Atlantic coast*	9.2 ± 2.4	4.3 ± 0.6	10 ± 0	0.58 ± 0.01		37.5 ± 0.2	0.185 ± 0.025	3.7 ± 1.1	-
Craspedostauros britannicus	NCC195	*France, NW Atlantic coast*	37.8 ± 8.8	11.0 ± 2.2	9 ± 0	0.53 ± 0.09		44.7 ± 0.1	0.502 ± 0.039	8.2 ± 1.8	-
Conticriba weissflogii	NCC133	*Morocco,Oum R'bia estuary*	14.3 ± 1.7	12.1 ± 1.7	7 ± 0	0.53 ± 0.08		34.6 ± 1.2	0.181 ± 0.015	6.4 ± 0.7	-
Licmophora sp. 1	NCC253	*France, NW Atlantic coast*	20.0 ± 3.9	12.4 ± 1.6	11 ± 0	0.51 ± 0.02		37.8 ± 5.7	0.161 ± 0.049	13.5 ± 1.1	-
Amphora sp. 2	NCC169	*France, NW Atlantic coast*	9.0 ± 1.2	5.8 ± 1.2	13 ± 3	0.48 ± 0.14		33.8 ± 4.8	0.157 ± 0.024	16.0 ± 2.6	[49,55,58,61,62]
Craspedostauros sp. 1	NCC228	*France, NW Atlantic coast*	55 ± 14	14.9 ± 4.2	16 ± 4	0.46 ± 0.22	C	25.2 ± 0.7	0.384 ± 0.042	5.6 ± 1.4	-
Craspedostauros sp. 2	NCC218	*France, NW Atlantic coast*	58 ± 14	13.6 ± 3.4	12 ± 0	0.45 ± 0.18		26.0 ± 4.1	0.178 ± 0.047	5.9 ± 1.4	-
Cymatosira belgica	NCC208	*France, NW Atlantic coast*	5.6 ± 0.6	4.7 ± 0.7	12 ± 0	0.44 ± 0.12		35.5 ± 3.4	0.118 ± 0.043	14.6 ± 3.0	-
Amphora acutiuscula	NCC216	*Viet Nam, South east coast*	10.0 ± 1.0	5.6 ± 0.8	11 ± 0	0.43 ± 0.07		32.4 ± 0.7	0.144 ± 0.041	13.7 ± 5.1	[49,55,58,61,62]
Lithodesmium sp	NCC353	*Australia, Moreton bay*	29.3 ± 5.4	22.6 ± 4.1	9 ± 1	0.43 ± 0.02		34.1 ± 6.1	0.129 ± 0.006	7.1 ± 2.9	-
Navicula sp. 1	NCC113	*Morocco,Oum R'bia estuary*	17.1 ± 2.9	5.1 ± 1.2	16 ± 0	0.42 ± 0.02		25.0 ± 1.0	0.168 ± 0.035	23.1 ± 6.4	[15,45,47,49,55,58]
Entomoneis sp. 6	NCC335	*France, Mediterranean Sea*	13.9 ± 1.9	7.7 ± 1.9	12 ± 1	0.38 ± 0.03		36.8 ± 4.7	0.226 ± 0.036	3.7 ± 1.1	[50–54]
Entomoneis sp. 1	NCC350	*France, Mediterranean Sea*	29.0 ± 4.0	17.5 ± 2.7	14 ± 0	0.36 ± 0.13		46.1 ± 4.6	0.098 ± 0.048	8.8 ± 0.1	[50–54]
Craspedostauros britannicus	NCC199	*France, NW Atlantic coast*	31.2 ± 3.2	8.8 ± 1.6	12 ± 2	0.35 ± 0.14		40.0 ± 2.9	0.421 ± 0.053	12.5 ± 1.2	-
Nitzschia sp. 5	NCC109	*France, NW Atlantic coast*	33.1 ± 4.1	10.1 ± 1.5	12 ± 1	0.31 ± 0.05		23.0 ± 1.5	0.239 ± 0.026	30.5 ± 0.7	[15,42–49]
Entomoneis sp. 3	NCC351	*France, Mediterranean Sea*	16.6 ± 3.7	12.3 ± 3.0	13 ± 0	0.22 ± 0.05		38.6 ± 5.8	0.065 ± 0.001	13.3 ± 4.6	[50–54]

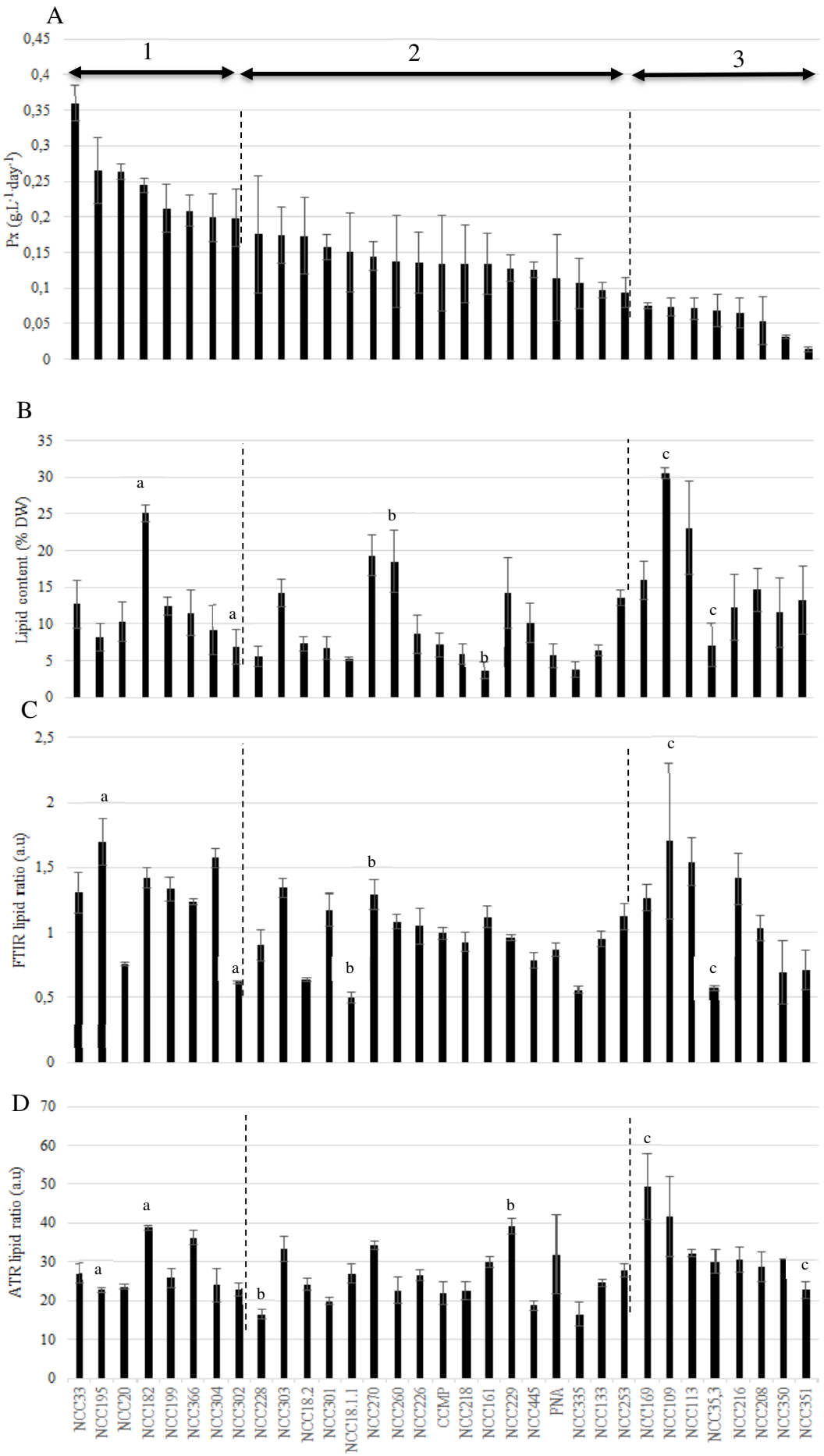

A
1
2
3
Px (g.L-1.day-1)
B
Lipid content (% DW)
C
FTIR lipid ratio (a.u)
D
ATR lipid ratio (a.u)
NCC33
NCC195
NCC20
NCC182
NCC199
NCC366
NCC304
NCC302
NCC228
NCC303
NCC18.2
NCC301
NCC18.1.1
NCC270
NCC260
NCC226
CCMP
NCC218
NCC161
NCC229
NCC445
PNA
NCC335
NCC133
NCC253
NCC169
NCC109
NCC113
NCC35,3
NCC216
NCC208
NCC350
NCC351

Figure 2. Values of the parameters measured for the screened strains includes strain productivity (A), lipid rates as measured with the gravimetric method (B), lipid ratio measured semi-quantitatively by the FTIR approaches, [(eb+CH3+CH3)/si] (C) and [area eb+CH2+CH3)/Total area] multiplied by 100 for scaling purposes (D). Notations a, b, and c correspond to the maximum and minimum values for groups 1, 2 and 3. N=3, independent measurements, ± SD.

3.2 FTIR analysis

3.2.1 FTIR spectrum interpretation

HTSXT-FTIR analysis was performed on the 33 assayed species (Tab. 3). The lipid rate was associated to three main signals on the recorded spectra (Fig. 3A), i.e. through the two vibrations of the fatty acid carbon chains (ν CH_2 and ν CH_3) (ν C-H ~ 2923 and 2852 cm^{-1}) [26] and of the ester bond function (Eb) (ν C=O ~ 1750 cm^{-1}) [63]. The other major bands corresponded to the principal cellular components such as the proteins (the amide I band ν C=O ~1650 cm^{-1}; the amide II band δ N=H ~1540 cm^{-1}), the nucleic acids (ν P=O ~1230 cm^{-1}) and the carbohydrates band superimposed on the silica band (~900 – 1200 cm^{-1}). For details see Wagner et al., 2010 [64]. Whereas the infrared signature obtained on the whole cells showed superimposed bands of silica and carbohydrates at 1078 cm^{-1} (Fig. 3A), the signature obtained on the crude lipid extract (Fig. 3B) did not exhibit this band, but a well defined ester bond (Eb) band at 1750 cm^{-1} and well-defined bands for the CH_2-CH_3 signature at 3000-2800 cm^{-1}.

The lipid ratio estimated from the FTIR data measured on the entire cells (Fig.2C) ranged from 1.70 ± 0.59 (*Nitzschia* sp. 5 NCC109) to 0.49 ± 0.04 (*Entomoneis paludosa* NCC18.1.1). It was thus possible to cluster the assessed strains into three groups. Group 1 showed a maximum ratio at 1.69 ± 0.18 for *Craspedostauros britannicus* NCC195 and a minimum at 0.61 ± 0.01 for *Entomoneis* sp 5 NCC302, group 2 showed a maximum ratio at 1.34 ± 0.07 for *Fallacia* sp 1 NCC303 and a minimum at 0.49 ± 0.04 for *Entomoneis paludosa*

NCC18.1.1 and group 3 showed a maximum ratio at 1.70 ± 0.59 for *Nitzschia* sp NCC109 and minimum at 0.56 ± 0.02 for *Lithodesmium sp* NCC35.3.

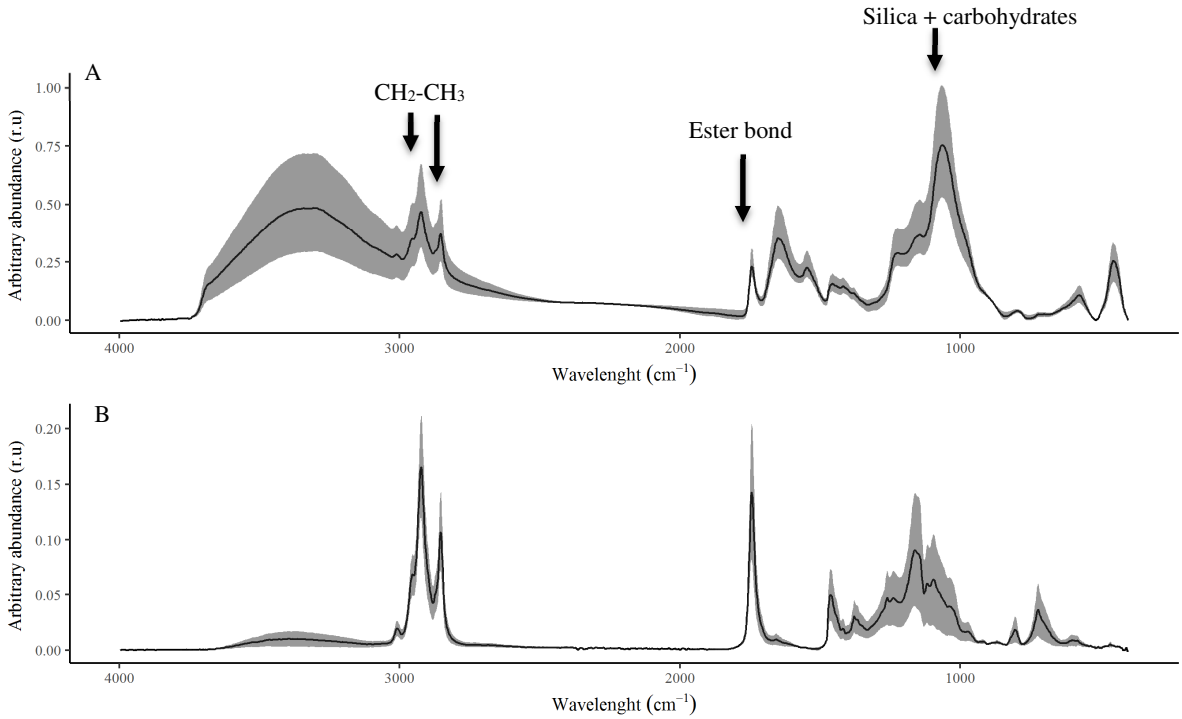


Figure 3. Example of averaged FTIR spectra recorded on entire cells or on the corresponding lipid extract. (A) *Staurosira* sp NCC182 FTIR signature recorded on the entire cells and (B) *Staurosira* sp NCC182 FTIR signature recorded on a crude lipid extract. The grey area corresponds to the variation of the FTIR signal associated to the standard deviation for n=3 independent measurements.

The lipid ratio estimated from the FTIR data recorded on the lipid extracts (Fig.2D) ranged from 49.3 ± 8.5 (*Amphora* sp. 2 NCC169) to 16.5 ± 3.1 (*Entomoneis* sp. 6 NCC335). Three groups were also proposed regarding this criterion, group 1 showed a maximum ratio at 38.84 ± 0.63 for *Staurosira* sp. NCC182 and a minimum ratio at 22.8 ± 1.8 for *Entomoneis* sp. 5 NCC302, group 2 a maximum ratio at 39 ± 2 for *Extubocellulus* sp. NCC229 and a minimum at 16.5 ± 1.2 for *Craspedostauros britannicus* NCC228 and group 3 with a maximum ratio at 49.3 ± 8.3 for *Amphora* sp. 2 NCC169 and a minimum at 23 ± 2 for *Entomoneis* sp. 3 NCC351.

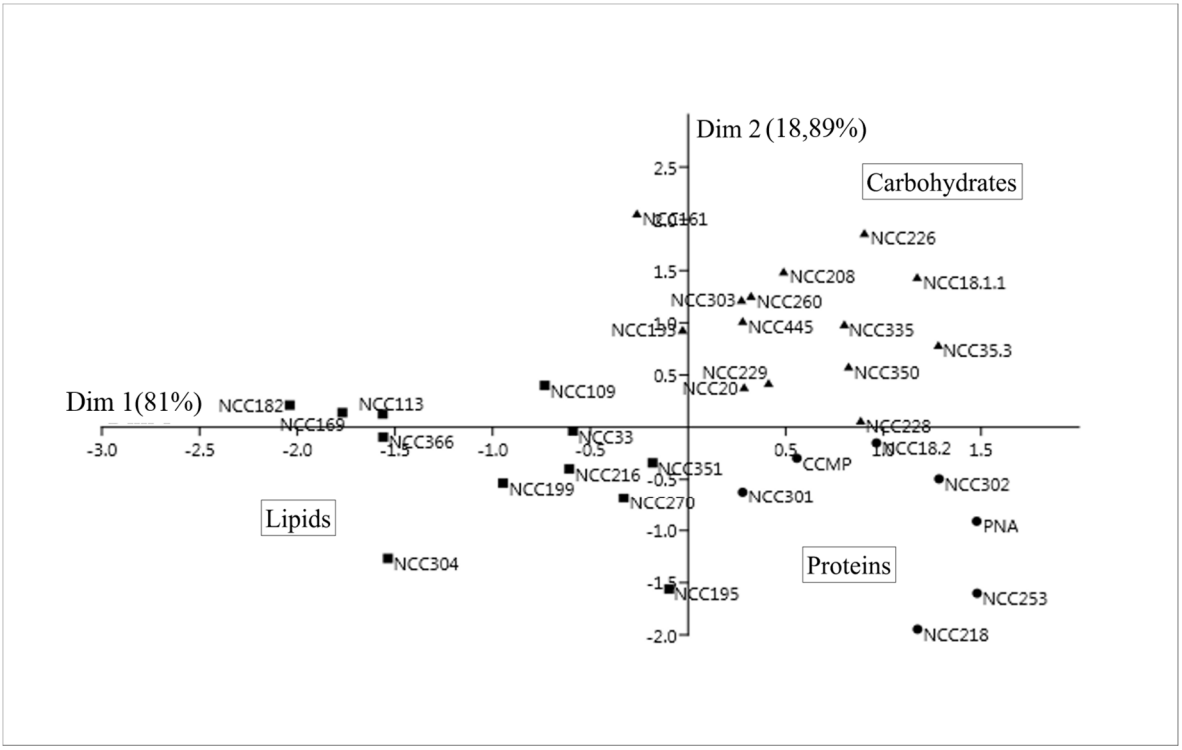
The silica amount did not appear to significantly impact on the lipid FTIR signature except in three species for which the FTIR lipid ratio fell between the two FTIR signature sampling methods (i.e. on entire cells or on lipid extract). These were Craspedostauros *brittanicus* NCC195, *Navicula* sp. 4 NCC113 and *Conticriba weissflogii* CCMP1336. For the other species with different silica content, the lipid signature remained stable in both FTIR sampling methods.

3.2.2 Multivariate analysis of the FTIR spectra recorded on entire cells

In order to assess the main differences in terms of biochemical composition among the 33 screened strains, a correspondence analysis approach using the lipid, protein and carbohydrate bands normalized to the silica amounts was performed. The resulting map (Fig. 4) is a classification of the data on two main dimensions. Dimension 1 represents 81% of the initial information and could be associated to the variation of the lipid composition of the assessed strains, ranging from the lowest to the highest amount of total lipids from right to left on the map. Two strains, *Staurosira* sp. NCC182 and *Amphora* sp. 2 NCC169 presented the highest amount in total lipids. Along dimension 2, representing 19% inertia, the strains associated to that dimension were mainly opposed on the basis of their protein and carbohydrate content. *Craspedostauros* sp. 2 NCC218 was rich in proteins, whereas in *Brockmaniella brockmanii* NCC161 the main fraction was associated to the carbohydrates.

These results indicate that the 2 most notable differences or largest deviations in the sample were observed first between *Staurosira* sp. NCC182, *Amphora* sp. 2 NCC169 and the other species for their lipid rates, and secondly between the strains *Craspedostauros* sp. 2 NCC218, *Brockmaniella brockmanii* NCC161 and the other species by their respective protein and carbohydrate composition. This analysis summarizes the main biochemical characteristics of the strains hosted in the NCC bank in a single step. Although the distance between the

459 macromolecular content and the species were not mathematically defined, their closeness on
460 the map could be used as a guideline to interpret their biochemical characteristics: the squares
461 correspond to the strains particularly rich in lipids, the triangles, the strains rich in
462 carbohydrates and the dots, the strains particularly rich in proteins.
463



464

465 **Figure 4.** Correspondence analysis map calculated on the basis of the macromolecular content
466 as evaluated by FTIR on all the assayed strains of the NCC. N=3 independent measurements.
467

468 ### 3.3 Comparison of the lipid amounts estimated by the gravimetric
469 method and the FTIR approaches.

470 A significant positive correlation was found for all the techniques (gravimetric and
471 FTIR) ($p < 1.10^{-7}$) with a Pearson correlation score R superior to 0.53. For the 33 analyzed
472 species, 14 presented a high lipid ratio (> 12% DW) using the gravimetric measurements:
473 *Nitzschia alexandrina* NCC33, *Staurosira* sp. NCC182, *Fallacia* sp. 1 NCC303, *Surirella* sp.

1 NCC270, *Extubocellulus cf cribriger* NCC229, *Amphora* sp. 1 NCC260, *Licmophora sp.1* NCC253, *Amphora sp. 2* NCC169, *Cymatosira belgica* NCC208, *Amphora acutiuscula* NCC216, *Navicula* sp. 1 NCC113, *Craspedostauros britannicus* NCC199*, Nitzschia* sp. 5 NCC109 and *Entomoneis sp.3* NCC351 (Fig. 2B, Fig. 5). The FTIR method applied on entire cells identified 12 species rich in lipids (FTIR lipid ratio > 1.20) with nine species identified in common with the gravimetric method: NCC33, NCC182, NCC303, NCC270, NCC169, NCC216, NCC113, NCC199 and NCC109 (Fig. 2C, Fig. 5). FTIR for the lipid extract analyses identified 12 species rich in lipids (ATR lipid ratio > 30) with eight species identified in common with the gravimetric approach: NCC182, NCC303, NCC270, NCC229, NCC169, NCC216, NCC113 and NCC109 (Fig 2D, Fig. 5).

The correspondence analyses (Fig.4) performed on the FTIR profiles obtained in the entire cells gave supplementary information and identified seven species particularly rich in lipids: NCC182, NCC270, NCC169, NCC113, NCC366, NCC109 and NCC33. These species were also identified by the gravimetric method with the exception of NCC366, only identified by the FTIR method on crude lipid extract.

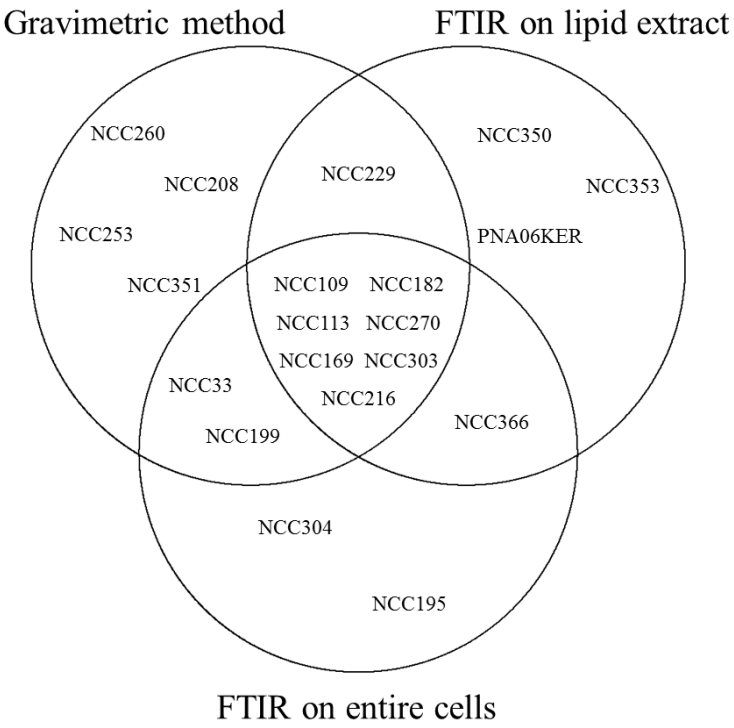


Figure 5 Venn diagram showing the degree of overlap among the different approaches used to identify the lipid rich diatoms. In the gravimetric method circle, 14 strains were identified as rich in lipids: 4 only with this method (NCC260, NCC208, NCC253, NCC351), one strains were also identified as rich and lipid by the FTIR on lipid extract (NCC229), and two strains were also identified as rich in lipids by the FTIR on entire cells (NCC33, NCC199). One strain was identified by FTIR on lipid extract and FTIR on whole cells (NCC366) and seven strains were identified as rich in lipids by all three methods (NCC109, NCC182, NCC113, NCC270, NCC169, NCC303, NCC216).

3.4 Strain selection

The selection of the strains exhibiting both high biomass productivity and high lipid rates, was performed using whole sample distribution based on the lipid rates as estimated by the FTIR approaches, gravimetry and strain biomass productivity as estimated by the fluorometry. On the boxplots summarizing this data (Fig. 6), the colored dots represent the species with the highest potential for biotechnology applications based on lipid molecules: *Nitzschia alexandrina* NCC33, *Staurosira* sp NCC182 and *Opephora* sp NCC366 presented a lipid ratio whatever the considered technique and productivity above the median. *Amphora* sp 2 NCC169 and *Nitzschia* sp 5 NCC109 were also identified with an above median lipid ratio, but with lower productivity. Both these species were finally selected for their high lipid rate, even though their productivity needs to be improved.

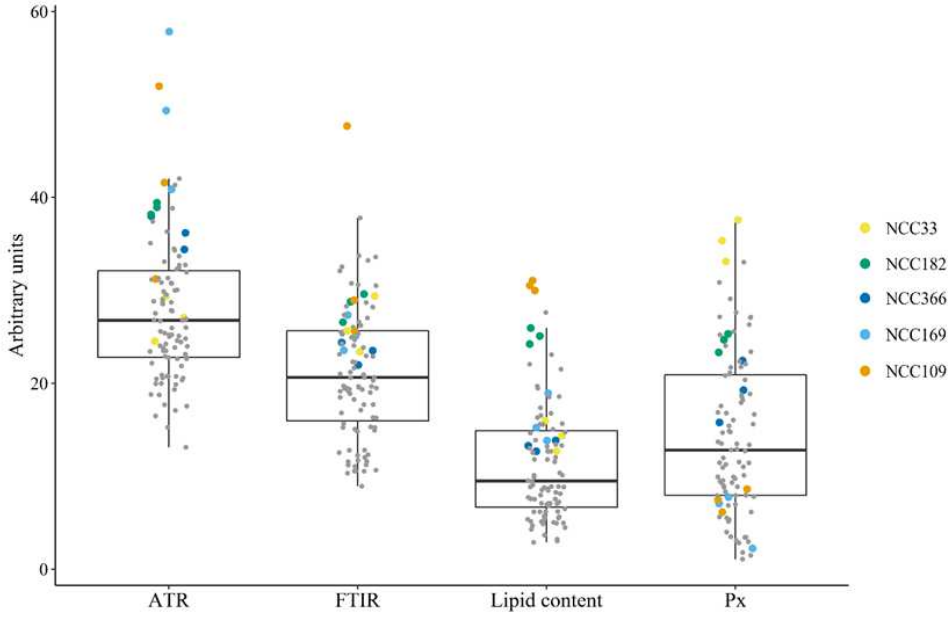


Figure 6. Boxplots summarizing the sample distribution criteria as measured with FTIR methods, gravimetry for the lipid rate and flurorimetry for productivity. FTIR data was expressed in arbitrary units. Lipid rate in %DW and Px in $g.L^{-1}.day^{-.1}$ FTIR results were

multiplied by 20 for scaling purposes. The Px was multiplied by 100 for scaling purposes. 33 strains were assayed in independent biological triplicates.

4 Discussion

4.1.1 Determination of growth by fluorimetry

In this study a rapid and precise method for the estimation of cell abundance was needed to efficiently establish the growth parameter of over 60 benthic diatom species. The use of microscope counting chambers is slow, tedious and imprecise. The use of particle counters is feasible only for those phytoplankton species within a certain size range that do not form chains and have long appendages [65]. *In vivo* fluorescence has been successfully used in the past to monitor the growth of phytoplankton culture [19, 20, 32, 41, 66]. *In vivo* fluorescence measurement is rapid, sensitive and can be used with all types of diatoms cell structures, i.e. on solitary cells such as *Entomoneis* and *Navicula* genera and on chain-forming diatoms, such as *Extubocellulus* genus or aggregates of cells such as *Amphora* genus. However, *in vivo* fluorescence is a measurement of the increase in chlorophyll and chlorophyll per cell varies greatly with light intensity and cellular nutritional status so it can only be used to measure growth when culture conditions are constant. Changes in light and other culture conditions can lead to a change in chlorophyll a content. In this case F0 will also change, introducing a bias in growth rate estimation. Reflectance can also be used to monitor the diatom growth rate, as demonstrated in this study; the NDVI was highly correlated with the cell count as *in vivo* fluorescence (R> 0.70). Méléder et al, 2003 [21] have already demonstrated this correlation on monospecific benthic diatom cultures with *Entomoneis paludosa* and *Navicula ramosissima*. Compared to fluorometry this technique required more culture sampling and a filtration step which took longer to set up. Fluorometry was more sensitive than spectroradiometry and required less material, which made it ideal for very low amounts of

biomass. *In vivo* fluorometry demonstrated several advantages: ease of use, real-time measurement, non-destructive sampling.

4.1.2 Determination of the lipid rate by FTIR

The total lipid evaluation by gravimetry has been used for more than 50 years [67], but this technique is clearly not compatible with screening efforts involving large numbers of samples because of processing time required particularly in the extraction phase. Furthermore, large volumes of samples are systematically needed for the measurements. It also requires sufficient amounts of dried biological sample, thus making it unsuitable for high frequency monitoring of small scale microalgal cultivation. Feng et al., (2013) [68] suggested that the presence of chlorophyll could also affect the accuracy of the method. In comparison to conventional chemical analysis, FTIR spectroscopy presented striking advantages due to its high reliability, sensitivity and speed of the measurement [64]. An IR spectrometer coupled to a microtiter plate reader open the possibility of high throughput analysis of a few nanograms of cell material [69]. Coat et al, (2014) [26] demonstrated that the repeatability of FTIR signal reached excellent values (10%), but only for a limited range of analyzed quantities of matter (10^6 to 10^8 cells. mL^{-1}).

In the present study we have demonstrated that FTIR was a suitable technique to evaluate the lipid ratio of diatoms. The use of the FTIR technique was even more rapid, due to the use of direct fresh biomass. The results obtained on entire cells and lipid crude extracts were similar, suggesting that measurement on entire cells did not improve the lipid quantification. Measurement on whole cells could be considered sufficient to get a first idea on their intracellular lipid rate. In addition, removing the silica did not seem necessary in view of the present results to improve the lipid semi-quantification.

However, there are some limitations with this non-invasive technique. Even though the FTIR results between entire cells and crude extracts were similar, we did not obtain a perfect linear

correlation with the traditional lipid extraction. Nevertheless, the FTIR method has the advantage of simultaneously detecting the relative amount of lipids, carbohydrates and proteins, even though those components overlapped, implying a certain degree of inaccuracy. The different steps required for the lipid extraction can also produce a bias in the evaluation of the lipid quantity, added to the inaccuracy of weighing of the lipid matter, making it less precise than the spectrometric approach. The heterogeneity observed in the present FTIR results could be also associated to the utilization of different strains (33 strains in the current study). These are not described in the previous studies evaluating lipid quantification with FTIR where only one species was used [26, 39, 70]. FTIR did not differentiate polar/apolar lipids, or the different types of fatty acids produced by the strains. Lipid analysis by chromatographic techniques coupled to mass spectrometry will be needed to further identify the presence of interesting molecules like EPA or DHA. Nevertheless, the present study opens the way to rapid and reliable semi-quantification of total amounts of intracellular lipids in diatoms using a fast, non-invasive approach.

4.1.3 Method assessments

This study highlighted the efficiency of PAM and FTIR measurements as fast techniques to characterize both the growth and lipid content of microalgae (Tab.4). The use of PAM fluorometry was three times faster than cell counting. Where 15 minutes were necessary to count 3 samples (8.25 hours for 99 samples) and 10 minutes to measure reflectance by spectroradiometry (5.5 hours for 99 samples), only 5 minutes were necessary to measure F0 by PAM fluorometry (2.75 hours for 99 samples).

Table 4 Evaluation of three methods for algal growth kinetic determination

Evaluation index	**Counting**	**NDVI**	**PAM**
Cultivation scale	Large	Large–small	Large–small
Monitoring frequency	High	Medium	High
Sample state	Liquid	Filtered	Liquid
Volume consumed	5 - 1 mL	5- 3 mL	>1 mL
Time consumed	8.25 h	5.5 h	2.75 h
Equipment	Counting chambers	Spectroradiometer	PAM fluorometer
Reliability	Low	High	High

For lipid extraction, dichloromethane methanol solvent is usually used [34, 71–73]. Even though extraction was very effective, this method is known to have environmental and health risks [71]. Moreover, it is expensive: for 99 extractions, 10 L of each solvent was necessary costing 234 € (9.68 €/L of methanol and 13.72 €/L of dichloromethane (Fischer Scientific). The time necessary for the lipid extraction using Bligh and Dyer method was time-consuming (Tab. 5) due to the time needed for maceration and lipid separation. 5.5 weeks were necessary to perform the extraction of 99 samples (ca. 7920 hours). Nile red, a lipid soluble fluorescent dye, is commonly used to evaluate the lipid content of animal cells, microorganisms, and especially microalgal strains. Required time to perform spectrophotometric measurements with Nile red and Bodipy was evaluated at 10 mins per sample (measurement before and after Nile red or Bodipy application) and 25 to 40 mins incubation time was necessary after

application of Nile red or Bodipy before reading spectrophotometric values [68,74]. In comparison FTIR is a fast and eco-friendly technique. Two plates of 384 well were necessary to evaluate the lipid content for the 33 strains (each strain needed 15 wells with 5 wells per replicate). It took 30 seconds to read each well to obtain the FTIR spectra leading to ca. and 4 hours to read an entire plate.

Table 5 Evaluation of three methods for algal lipid content determination

Evaluation index	Gravimetric	FT-IR	Nile Red / Bodipy 505/515
Cultivation scale	Large	Large–small	Large–small
Monitoring frequency	Low	Medium	High
Sample state	Dried	Liquid	Liquid
Biomass consumed	100 mg	5 μL of 1.0 mg mL^{-1}	3 mL of OD_{680} 0.4
Time consumed	>50 h	8 h	25 h
Equipment	Nitrogen evaporator	FT-IR	FS & FM [a]
Accuracy	Total lipid	Total lipid	Neutral lipid

[a] Fluorescence spectrophotometer and Fluorescence microscopy

Nile red and Bodipy 505/515 staining are powerful quantification tools in terms of time and cost of biomass [75, 76], high throughput quantification method of lipids with Nile red or Bodipy 505/515 fluorescence can hardly been seen as a method for screening different species of microalgae, as the staining protocol is species specific. The significant disadvantages of Nile red were its limited photostability, interference with chlorophyll [77, 78], and difficulty of permeation for some species. Bodipy 505/515 produced a better marker than Nile red for visualizing neutral lipid content in fluorescence microscopy [75, 79] but some authors have reported disadvantages with these techniques such as background fluorescence of the dye in the medium and failure to quantify neutral lipids between rich and low oil strains. When microalgae were cultured on a large scale with a low-frequency monitoring requirement, any of the three methods could be adopted, although gravimetric determination might be preferable as it was an absolute method for quantification of both crude and neutral lipids without the need of specialized equipment. For the general laboratory culture of microalgae,

634 the FT-IR method for simultaneous characterization of total lipid, carbohydrate and protein
635 content and the Nile Red method for both neutral lipid content and location can be used, both
636 of which are relative quantification methods, but require special equipment.
637 Although these analyses demonstrate that FTIR and Nile Red were equally effective at
638 measuring lipid accumulation, FTIR was likely to be a more efficient tool for this purpose
639 because of its much faster analysis time and high reproducibility of results [80]. Furthermore,
640 FTIR may also be more suitable than Nile Red for efficiently detecting large increases in lipid
641 concentration. Nile Red does not appear to be efficient at accurately quantifying lipid
642 concentration above 20 mg/ml [77] while FTIR can efficiently detect linear lipid
643 concentration changes up to at least 250 mg [81]. Measurements with FTIR were more precise
644 because of the technical quintuplicate performed for the acquisition of spectra.

645 ## 4.2 Screening for lipid rich benthic diatom strains

646 The aim of this study was to investigate the growth characteristics and the lipid rates of the
647 benthic marine diatom species hosted in the NCC bank in order to evaluate their potential for
648 original lipid bioproducts of potential economic interests. First, a selection based on the genera
649 identified in the literature was applied. Among the 134 strains hosted by the NCC,
650 corresponding to 40 genera and 101 species, 23 genera (corresponding to 77 species, 105
651 strains in the NCC) were largely studied [15, 16, 42, 43, 46,49, 52, 55, 58, 60, 82]. Among
652 these 23 genera, only 13 genera (42 species, 47 strains) with high productivity and/or
653 producing high lipid quantities were selected for the current study. Among these 47 strains,
654 only 18 strains, (6 genera, 17 species) grew successfully. Among the NCC's 40 genera, 17
655 genera (corresponding to 24 species, 29 strains) were not previously reported in the literature
656 in terms of productivity or ability to produce lipids or added value molecules. Thus, the 29
657 strains corresponding to these 17 genera were also selected and assayed for the current study.
658 Among these strains, 15 strains (13 species, 10 genera) did successfully grow. Finally, at the

659 end of this first screening step based on the growth rates, 33 strains (18 previously described
660 and 15 that have never been described before in the literature) were selected for the second
661 step of the screening process. As reported in the supplementary data, 43 strains failed to grow
662 which may be associated to shear stress; some species like *Rhizosolenia setigera* cells were
663 broken during agitation.
664 The second step consisted in the determination of the lipid rate on the selected strains. In
665 microalgae, it can typically vary from 1 to 85% of the dry weight under adverse conditions
666 [15, 83, 84]. Factors such as temperature, irradiance and most markedly nutrient availability
667 have been shown to affect both lipid composition and lipid rate [17, 57, 85]. In general, high
668 irradiance stimulates TAG accumulation [57], while under low irradiance, the polar lipids
669 (phospholipids and glycolipids), structurally and functionally associated to cell membrane,
670 are preferentially synthesized [17]. The lipid rate found in the current study for the *Amphora*
671 genus was similar to that estimated by Renaud et al., (1999) [46]. The authors proposed 19%
672 DW *vs.* 14 to 19% DW in the current study. This similarity could be explained by the similar
673 culture conditions for light and nutrient although the temperature was different (25°C *vs* 16°C
674 in the current study). According to Chtourou et al., 2015 [61], the temperature could be an
675 important factor for *Amphora* genus lipid rate which can achieve lipid rates up to 24% of DW
676 at 20°C. Media enrichment could also be an important factor: *Amphora* cells grown in media
677 enriched with macronutrients and trace metal can also achieve lipid rates up to 32% under low
678 light conditions (11.4 µmol photon m^{-2} s^{-1}) [62] or under nitrogen deficiency [55]. These
679 factors could be used as guideline to improve lipid production of the selected *Amphora* strains
680 hosted in the NCC. The biomass measurements for the *Amphora* genus in Zhao et al study
681 [49] were in accordance with our results. The authors found 0.13 $g.L^{-1}$ vs. 0.16 $g.L^{-1}$ in the
682 current study. According to their measured lipid productivity, the *Amphora* genus appeared
683 to be a good candidate for lipid based potential applications.

684 The lipid rate found in our study for the *Nitzschia* genus was similar to the results found by
685 Renaud et al., (1999) for one species [46] but very different for *Nitzschia sp 5* NCC109 in the
686 current study. They found a lipid rate of between 13 and 16% DW for the *Nitzschia* genus. In
687 the present study, *Nitzschia alexandrina* NCC33 showed a lipid rate of 13% DW but for
688 *Nitzschia sp 5* NCC109 it was estimated up to 30% DW. However, this rate could only be
689 found for *Nitzschia* genus under nitrogen or silica deficiency increasing up to 45-47% DW
690 [44,45,47]. Since the amount of nitrogen was not monitored in the culture, it is possible that
691 nitrogen depletion occurred, explaining the high lipid rate measured for *Nitzschia sp 5*
692 NCC109. Further analysis for this strain is necessary to establish whether this was due to the
693 absolute lipid richness of the strain, or to a bias in culture conditions.
694 The biomass found for the *Nitzschia* genus in Zhao et al (2016) [49] was lower than the
695 measurements in the present study. The biomass estimated in the current study ranged from
696 0.07 to 0.14 $g.L^{-1}$ vs. 0.19 to 0.23 $g.L^{-1}$. This difference could be due to the light/dark cycle
697 culture conditions, suggesting that under continuous light the biomass production would be
698 more significant for this specific genus [86].
699 It has been demonstrated that the species with a lipid rate of 30% DW and productivity under
700 non-optimized conditions of around 0.30 $g.L^{-1}$ could be potential strains for lipid production
701 [15, 87]. In the present study *Nitzschia sp 5* NCC109 had the highest lipid rate, 30.51% DW,
702 Nitzschia *alexandrina* NCC33 with the highest productivity 0.35 $g.L^{-1}$. Both strains were thus
703 selected as candidates for further analyses to assay their potential for lipid-based applications.
704 The *Navicula* genus also presented a lipid rate for the 2 tested species of between 9% and 23%
705 DW. Zhao et al, (2016) [49] found a lipid range from 5 to 30% for 3 *Navicula* species
706 (*Navicula ramosissima*, *Navicula molli* and *Navicula halophia*) and Scholz et al, (2013) [58]
707 between 18-25% for 6 *Navicula* species (*Navicula digito-radiata*, *Navicula forcipata*,
708 *Navicula gregaria*, *Navicula perminuta*, *Navicula phyllepta* and *Navicula salinicola*). The

709 data were estimated in spite of very different culture conditions (dark night cycle at 600 µmol
710 photon m^{-2} s^{-1}). This observation is important, since for that specific genus, the observed
711 variability in the range of lipid rate for the different tested species is the same regardless of
712 the culture conditions used. The biomass measurements found by Zhao et al., (2016) for the
713 *Navicula* genus are in accordance with the present study, ranging from 0.10 to 0.17 $g.L^{-1}$ vs.
714 0.13 to 0.17 $g.L^{-1}$ suggesting that this genus could produce the identical biomass quantity
715 under light/dark cycle or continuous light.
716 The lipid rate found for the *Extubocellulus* genus in Slocombe et al., (2015) [60] was 23%
717 DW and the biomass equalled 0.06 g. L^{-1}. The lipid rate for this genus, 15% DW in the actual
718 study was lower but the biomass obtained was higher 0.16 $g.L^{-1}$. The culture was grown under
719 light/dark cycle in the Slocombe et al. study suggesting that this genus could grow better under
720 continuous light but produce more lipids under light/dark cycle. Wahidin et al., (2013) [88]
721 found the same trend for the microalgae genus *Nannochloropsis*.
722 Knuckey et al, (2002) [52] found 27% DW of lipid rate for the *Entomoneis* genera. In the
723 present study 9 strains of *Entomoneis* presented lipid rates ranging from 3 to 13% DW. In the
724 Knuckey et al, (2002) study the pH and the nutrients were monitored. The culture conditions
725 were identical to the present study (continuous light, 120 µmol photon m^{-2} s^{-1}). It is therefore
726 possible, that for this genus a nutrient limitation occurred and instead of enhancing the lipid
727 production, it has diminished it. It has been reported that this species produces EPA which is
728 interesting for nutraceutical products [52].
729 In this study the species with the highest lipid rate (~20 %) and the greatest biomass
730 productivity (up to 0.24 $g.L^{-1}.day^{-1}$) was *Staurosira* sp NCC182. This species was similar in
731 productivity to the marine microalgae *Nannochloropsis* sp. with productivity higher than 0.21
732 $g.L^{-1}.day^{-1}$ and reached a total lipid rate of 30% DW, when cultivated in batch mode under
733 continuous light and reached 68% DW of lipid production under nitrogen deprivation [89].

734 *Nannochloropsis* sp. was investigated for algal biofuel production due to its ease of growth
735 and high oil rate. *Staurosira* sp was grown in raceway ponds by Huntley et al., (2015) [16]
736 and the lipid quantification demonstrated that these strains could reach a lipid rate of 45.5%
737 DW under low N content. In the present study, *Staurosira* sp. NCC182 presented all the
738 characteristics (good productivity and oil content) to be produced on a large scale and was
739 considered to be one of the most promising candidates for lipid-based applications.
740 Even if the microalgae oil yield is strain-dependent it is generally superior to other vegetable
741 crops [15, 87]. Oil content in Corn, Hemp, Soybean, Sunflower or Palm oil varied from 18
742 (Soybean) to 44% DW (Corn). Christie et al., (2007) [15] demonstrated that microalgae with
743 a lipid rate up to 30% of DW could produce 58 L/Ha of oil and microalgae up to 70% of DW
744 could produce 136 L/Ha of oil. Corn can produce 172 L/Ha, soybean 446 L/Ha and oil palm
745 5950 L/Ha but require a lot of land for production: 1540 M/Ha for Corn, 594 M/Ha for
746 Soybean and 45 M/Ha for oil palm. Microalgae production only requires 2 to 5 M/Ha. In the
747 NCC collection *Nitzschia sp 5* NCC109 and *Staurosira sp NCC182* had all the mandatory
748 features to be grown on a large scale: good productivity and high lipid rate. In the present
749 study, *Nitzschia alexandrina* NCC33 had the highest productivity in the NCC collection (0.36
750 $g.L^{-1}.day^{-1}$) and presented a lipid rate superior to 10% DW. Despite its low productivity of
751 0.08 $g.L^{-1}.day^{-1}$, *Amphora sp 2* NCC169 was chosen for its lipid rate, superior to 15% DW.
752 These 4 strains were selected for further analyses and to improve their productivity and lipid
753 rate with the objective of supplying new resources for lipid based applications. *Entomoneis*
754 *paludosa* was also selected despite its low lipid rate since it was already characterized for its
755 ability to produce EPA.
756 Among the strains not described in the literature, the measured lipid rate ranged from 3 to
757 20% DW and productivity from 0.05 to 0.27 $g.L^{-1}.day^{-1}$. Only one genus was selected for
758 further study: *Opephora sp 1* NCC366. This species showed a mandatory balance between

759 relatively high biomass productivity (0.23 $g.L^{-1}$) and a high intracellular lipid rate (above 10%
760 DW).

761 5 Conclusion

762 In this work, we focused on developing an easy to use screening method to explore the
763 NCC bank for diatom strains with the highest relative lipid content. The experimental results
764 showed that the combined use of water-PAM to estimate strain growth kinetics and FTIR on
765 whole cells to estimate the semi-quantitative strain macromolecular content and more
766 specifically lipids, could be rapid, reliable and accessible techniques. The developed
767 methodology opens the way to a systematic, fast, and convenient screening of microorganisms
768 (microalgae in this proof of concept). Moreover, the sensitivity and specificity of the method
769 makes it suitable for a reasonable amount of biomass. This method could also be used in
770 systematic studies for the optimization of culture conditions and to measure the influence of
771 the environment on the metabolic plasticity of the assessed organism. Using this screening
772 approach, 5 strains hosted in the NCC bank were selected for their high productivity and high
773 lipid rate: *Nitzschia alexandrina* NCC33, *Staurosira sp* NCC182, *Opephora sp 1* NCC366,
774 *Nitzschia sp 5* NCC109 and *Amphora sp 2* NCC169. The lipid rate achieved by theses strains
775 reached a maximum of 30% DW in the assayed cultivation conditions.
776
777 In order to improve the lipid quantity, the selected strains could be grown under different
778 culture conditions. The impact of light, temperature and nutrients, especially nitrogen, could
779 be assayed both in terms of lipid productivity and ecophysiology to ensure the highest growth
780 rate possible. Once optimal conditions are found for those strains, the production in
781 photobioreactors could be tested and productivity and lipid rates evaluated, in order to

782 estimate quantitatively if the selected strains can compete with the best ones found in the
783 literature [90–92]. Finally, depending on their oil quality and original lipid activities, those
784 strains may constitute new and original genetic resources that could have potential interesting
785 applications (biodiesel, pharmaceutical, etc.).
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801

802 # 6 References

803 [1] T.M. Mata, A.A. Martins, N.S. Caetano, Microalgae for biodiesel production and other
804 applications: A review, Renew. Sustain. Energy Rev. 14 (2010) 217–232.
805 doi:10.1016/j.rser.2009.07.020.
806 [2] P. Spolaore, C. Joannis-Cassan, E. Duran, A. Isambert, Commercial applications of
807 microalgae, J. Biosci. Bioeng. 101 (2006) 87–96. doi:10.1263/jbb.101.87.

[3] M.A. Borowitzka, High-value products from microalgae—their development and commercialisation, J. Appl. Phycol. 25 (2013) 743–756. doi:10.1007/s10811-013-9983-9.
[4] M.I. Khan, J.H. Shin, J.D. Kim, The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products, Microb. Cell Factories. 17 (2018). doi:10.1186/s12934-018-0879-x.
[5] Y.-F. Niu, M.-H. Zhang, D.-W. Li, W.-D. Yang, J.-S. Liu, W.-B. Bai, H.-Y. Li, Improvement of Neutral Lipid and Polyunsaturated Fatty Acid Biosynthesis by Overexpressing a Type 2 Diacylglycerol Acyltransferase in Marine Diatom Phaeodactylum tricornutum, Mar. Drugs. 11 (2013) 4558–4569. doi:10.3390/md11114558.
[6] Z. Yi, M. Xu, X. Di, S. Brynjolfsson, W. Fu, Exploring Valuable Lipids in Diatoms, Front. Mar. Sci. 4 (2017). doi:10.3389/fmars.2017.00017.
[7] E. Artamonova, J. Svenning, T. Vasskog, E. Hansen, H. Eilertsen, Analysis of phospholipids and neutral lipids in three common northern cold water diatoms: Coscinodiscus concinnus, Porosira glacialis, and Chaetoceros socialis, by ultra-high performance liquid chromatography-mass spectrometry, J. Appl. Phycol. 29 (2017) 1241–1249.
[8] K.W. Chew, J.Y. Yap, P.L. Show, N.H. Suan, J.C. Juan, T.C. Ling, D.-J. Lee, J.-S. Chang, Microalgae biorefinery: High value products perspectives, Bioresour. Technol. 229 (2017) 53–62. doi:10.1016/j.biortech.2017.01.006.
[9] G.A. Dunstan, J.K. Volkman, S.M. Barrett, J.-M. Leroi, S.W. Jeffrey, Essential polyunsaturated fatty acids from 14 species of diatom (Bacillariophyceae), Phytochemistry. 35 (1993) 155–161. doi:10.1016/S0031-9422(00)90525-9.
[10] Y.-C. Chen, The biomass and total lipid content and composition of twelve species of marine diatoms cultured under various environments, Food Chem. 131 (2012) 211–219. doi:10.1016/j.foodchem.2011.08.062.
[11] L. Chuecas, J.P. Riley, Component Fatty Acids of the Total Lipids of Some Marine Phytoplankton, J. Mar. Biol. Assoc. U. K. 49 (1969) 97. doi:10.1017/S0025315400046439.
[12] K. Nagao, T. Yanagita, Conjugated fatty acids in food and their health benefits, J. Biosci. Bioeng. 100 (2005) 152–157. doi:10.1263/jbb.100.152.
[13] M. Gross, The mysteries of the diatoms, Curr. Biol. 22 (2012) R581–R585. http://www.sciencedirect.com/science/article/pii/S0960982212008664 (accessed September 12, 2016).
[14] T. Lebeau, J.-M. Robert, Diatom cultivation and biotechnologically relevant products. Part II: Current and putative products, Appl. Microbiol. Biotechnol. 60 (2003) 624–632. doi:10.1007/s00253-002-1177-3.
[15] Y. Chisti, Biodiesel from microalgae, Biotechnol. Adv. 25 (2007) 294–306. doi:10.1016/j.biotechadv.2007.02.001.
[16] M.E. Huntley, Z.I. Johnson, S.L. Brown, D.L. Sills, L. Gerber, I. Archibald, S.C. Machesky, J. Granados, C. Beal, C.H. Greene, Demonstrated large-scale production of marine microalgae for fuels and feed, Algal Res. 10 (2015) 249–265. doi:10.1016/j.algal.2015.04.016.
[17] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, A. Darzins, Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, Plant J. 54 (2008) 621–639. doi:10.1111/j.1365-313X.2008.03492.x.
[18] T.A. Dempster, M.R. Sommerfeld, Effects of environmental conditions on growth and lipid accumulation in Nitzschia communis (Bacillariophyceae), J. Phycol. 34 (1998) 712–721.

[19] V. Vyhnalek, Z. Fišar, A. Fišarová, J. Komarkova, In vivo fluorescence of chlorophyll a: Estimation of phytoplankton biomass and activity in Římov Reservoir (Czech Republic), Water Sci. Technol. 28 (1993) 29–33. http://wst.iwaponline.com/content/28/6/29.abstract (accessed October 10, 2017).
[20] A.D. Steinman, G.A. Lamberti, P.R. Leavitt, D.G. Uzarski, Biomass and Pigments of Benthic Algae, in: Methods Stream Ecol. Vol. 1, Elsevier, 2017: pp. 223–241. doi:10.1016/B978-0-12-416558-8.00012-3.
[21] V. Méléder, L. Barillé, P. Launeau, V. Carrère, Y. Rincé, Spectrometric constraint in analysis of benthic diatom biomass using monospecific cultures, Remote Sens. Environ. 88 (2003) 386–400. doi:10.1016/j.rse.2003.08.009.
[22] L. Akoto, R. Pel, H. Irth, A.T. Udo, R.J. Vreuls, Automated GC–MS analysis of raw biological samples: Application to fatty acid profiling of aquatic micro-organisms, J. Anal. Appl. Pyrolysis. 73 (2005) 69–75.
[23] K.E. Cooksey, J.B. Guckert, S.A. Williams, P.R. Callis, Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red, J. Microbiol. Methods. 6 (1987) 333–345.
[24] J. Koreivienė, Microalgae Lipid Staining with Fluorescent BODIPY Dye, (2017).
[25] D. Jaeger, C. Pilger, H. Hachmeister, E. Oberländer, R. Wördenweber, J. Wichmann, J.H. Mussgnug, T. Huser, O. Kruse, Label-free in vivo analysis of intracellular lipid droplets in the oleaginous microalga Monoraphidium neglectum by coherent Raman scattering microscopy, Sci. Rep. 6 (2016) 35340.
[26] R. Coat, V. Montalescot, E.S. León, D. Kucma, C. Perrier, S. Jubeau, G. Thouand, J. Legrand, J. Pruvost, O. Gonçalves, Unravelling the matrix effect of fresh sampled cells for in vivo unbiased FTIR determination of the absolute concentration of total lipid content of microalgae, Bioprocess Biosyst. Eng. 37 (2014) 2175–2187.
[27] J. Fan, Y. Cui, M. Wan, W. Wang, Y. Li, Lipid accumulation and biosynthesis genes response of the oleaginous Chlorella pyrenoidosa under three nutrition stressors, Biotechnol. Biofuels. 7 (2014) 17. doi:10.1186/1754-6834-7-17.
[28] P. Metzger, C. Largeau, Botryococcus braunii: a rich source for hydrocarbons and related ether lipids, Appl. Microbiol. Biotechnol. 66 (2005) 486–496. doi:10.1007/s00253-004-1779-z.
[29] R.R. Guillard, Culture of phytoplankton for feeding marine invertebrates, in: Cult. Mar. Invertebr. Anim., Springer, 1975: pp. 29–60.
[30] R.R. Guillard, J.H. Ryther, Studies of marine planktonic diatoms: I. Cyclotella Nana Hustedt, and Detonula Confervacea (CLEVE) Gran., Can. J. Microbiol. 8 (1962) 229–239.
[31] Z.-K. Yang, Y.-F. Niu, Y.-H. Ma, J. Xue, M.-H. Zhang, W.-D. Yang, J.-S. Liu, S.-H. Lu, Y. Guan, H.-Y. Li, Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation, Biotechnol. Biofuels. 6 (2013) 67. https://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/1754-6834-6-67 (accessed October 6, 2017).
[32] C. Honeywill, D. Paterson, S. Hagerthey, Determination of microphytobenthic biomass using pulse-amplitude modulated minimum fluorescence, Eur. J. Phycol. 37 (2002) 485–492. doi:10.1017/S0967026202003888.
[33] B. Gompertz, XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c, Philos. Trans. R. Soc. Lond. 115 (1825) 513–583.
[34] E.G. Bligh, W.J. Dyer, A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION, Can. J. Biochem. Physiol. 37 (1959) 911–917. doi:10.1139/y59-099.

907 [35] G. Breuer, W.A. Evers, J.H. de Vree, D.M. Kleinegris, D.E. Martens, R.H. Wijffels, P.P.
908 Lamers, Analysis of fatty acid content and composition in microalgae, J. Vis. Exp. JoVE.
909 (2013).
910 [36] E.S. León, R. Coat, B. Moutel, J. Pruvost, J. Legrand, O. Gonçalves, Influence of
911 physical and chemical properties of HTSXT-FTIR samples on the quality of prediction
912 models developed to determine absolute concentrations of total proteins, carbohydrates
913 and triglycerides: a preliminary study on the determination of their absolute
914 concentrations in fresh microalgal biomass, Bioprocess Biosyst. Eng. 37 (2014) 2371–
915 2380.
916 [37] W. Zeroual, C. Choisy, S.M. Doglia, H. Bobichon, J.-F. Angiboust, M. Manfait,
917 Monitoring of bacterial growth and structural analysis as probed by FT-IR spectroscopy,
918 Biochim. Biophys. Acta BBA - Mol. Cell Res. 1222 (1994) 171–178. doi:10.1016/0167-
919 4889(94)90166-X.
920 [38] D.H. Williams, I. Fleming, Spectroscopic methods in organic chemistry, McGraw-Hill,
921 1980.
922 [39] I. Schaub, H. Wagner, M. Graeve, U. Karsten, Effects of prolonged darkness and
923 temperature on the lipid metabolism in the benthic diatom Navicula perminuta from the
924 Arctic Adventfjorden, Svalbard, Polar Biol. (2017) 1–15. doi:10.1007/s00300-016-
925 2067-y.
926 [40] N. Sourial, C. Wolfson, B. Zhu, J. Quail, J. Fletcher, S. Karunananthan, K. Bandeen-
927 Roche, F. Béland, H. Bergman, Correspondence analysis is a useful tool to uncover the
928 relationships among categorical variables, J. Clin. Epidemiol. 63 (2010) 638–646.
929 doi:10.1016/j.jclinepi.2009.08.008.
930 [41] C.J. Lorenzen, A method for the continuous measurement of in vivo chlorophyll
931 concentration, in: Deep Sea Res. Oceanogr. Abstr., Elsevier, 1966: pp. 223–227.
932 http://www.sciencedirect.com/science/article/pii/0011747166911028 (accessed October
933 10, 2017).
934 [42] G. Chen, Y. Jiang, F. Chen, Fatty acid and lipid class composition of the
935 eicosapentaenoic acid-producing microalga, Nitzschia laevis, Food Chem. 104 (2007)
936 1580–1585. doi:10.1016/j.foodchem.2007.03.008.
937 [43] M.C. Dalay, S.S. Gunes, others, Biodiesel from Microalgae: A Renewable Energy
938 Source, Middle East J. Sci. Res. 22 (2014) 350–355.
939 http://idosi.org/mejsr/mejsr21(12)14/15.pdf (accessed February 19, 2016).
940 [44] M.J. Griffiths, S.T.L. Harrison, Lipid productivity as a key characteristic for choosing
941 algal species for biodiesel production, J. Appl. Phycol. 21 (2009) 493–507.
942 doi:10.1007/s10811-008-9392-7.
943 [45] J. Johansen, P. Lemke, N. Nagle, P. Chelf, P. Roessler, R. Galloway, S. Toon, Addendum
944 to Microalgae Culture Collection 1986-1987, National Renewable Energy Lab.(NREL),
945 Golden, CO (United States), 1987.
946 [46] S.M. Renaud, L.-V. Thinh, D.L. Parry, The gross chemical composition and fatty acid
947 composition of 18 species of tropical Australian microalgae for possible use in
948 mariculture, Aquaculture. 170 (1999) 147–159.
949 [47] J. Sheehan, T. Dunahay, J. Benemann, P. Roessler, Look Back at the U.S. Department
950 of Energy's Aquatic Species Program: Biodiesel from Algae; Close-Out Report,
951 National Renewable Energy Lab., Golden, CO. (US), 1998. doi:10.2172/15003040.
952 [48] Z.-Y. Wen, F. Chen, Production potential of eicosapentaenoic acid by the diatom
953 Nitzschia laevis, Biotechnol. Lett. 22 (2000) 727–733.
954 http://link.springer.com/article/10.1023/A:1005666219163 (accessed February 18,
955 2016).

[49] F.-Y. Zhao, J.-R. Liang, Y.-H. Gao, Q.-Q. Luo, Y. Yu, C.-P. Chen, L. Sun, Variations in the Total Lipid Content and Biological Characteristics of Diatom Species for Potential Biodiesel Production, J. Fundam. Renew. Energy Appl. 2016 (2016). http://www.omicsonline.com/open-access/variations-in-the-total-lipid-content-and-biological-characteristics-of-diatom-species-for-potential-biodiesel-production-2090-4541-1000201.php?aid=66659 (accessed February 9, 2016).
[50] T. Jauffrais, S. Drouet, V. Turpin, V. Méléder, B. Jesus, B. Cognie, P. Raimbault, R.P. Cosson, P. Decottignies, V. Martin-Jézéquel, Growth and biochemical composition of a microphytobenthic diatom (Entomoneis paludosa) exposed to shorebird (Calidris alpina) droppings, J. Exp. Mar. Biol. Ecol. 469 (2015) 83–92.
[51] T. Jauffrais, B. Jesus, V. Méléder, V. Turpin, A.D.P.G. Russo, P. Raimbault, V.M. Jézéquel, Physiological and photophysiological responses of the benthic diatom Entomoneis paludosa (Bacillariophyceae) to dissolved inorganic and organic nitrogen in culture, Mar. Biol. 163 (2016). doi:10.1007/s00227-016-2888-9.
[52] R.M. Knuckey, M.R. Brown, S.M. Barrett, G.M. Hallegraeff, Isolation of new nanoplanktonic diatom strains and their evaluation as diets for juvenile Pacific oysters (Crassostrea gigas), Aquaculture. 211 (2002) 253–274. doi:http://dx.doi.org/10.1016/S0044-8486(02)00010-8.
[53] A.T. Soares, B.F. Silva, L.L. Fialho, M.A.G. Pequeno, A.A.H. Vieira, A.G. Souza, N.R. Antoniosi Filho, Chromatographic characterization of triacylglycerides and fatty acid methyl esters in microalgae oils for biodiesel production, J. Renew. Sustain. Energy. 5 (2013) 053111. doi:10.1063/1.4821289.
[54] T. Viriyayingsiri, P. Sittplangkoon, S. Powtongsook, K. Nootong, Continuous production of diatom Entomoneis sp. in mechanically stirred tank and flat-panel airlift photobioreactors, Prep. Biochem. Biotechnol. 46 (2016) 740–746. doi:10.1080/10826068.2015.1135460.
[55] F.J. Fields, J.P. Kociolek, An evolutionary perspective on selecting high-lipid-content diatoms (Bacillariophyta), J. Appl. Phycol. 27 (2015) 2209–2220. doi:10.1007/s10811-014-0505-1.
[56] G.E. Fogg, D.M. Collyer, The accumulation of fats as a characteristic of certain classes of algae, Proc 8 Int Bot Congr. 17 (1954) 28.
[57] P.G. Roessler, ENVIRONMENTAL CONTROL OF GLYCEROLIPID METABOLISM IN MICROALGAE: COMMERCIAL IMPLICATIONS AND FUTURE RESEARCH DIRECTIONS, J. Phycol. 26 (1990) 393–399. doi:10.1111/j.0022-3646.1990.00393.x.
[58] B. Scholz, G. Liebezeit, Biochemical characterisation and fatty acid profiles of 25 benthic marine diatoms isolated from the Solthörn tidal flat (southern North Sea), J. Appl. Phycol. 25 (2013) 453–465. doi:10.1007/s10811-012-9879-0.
[59] M. Islam, M. Magnusson, R. Brown, G. Ayoko, M. Nabi, K. Heimann, Microalgal Species Selection for Biodiesel Production Based on Fuel Properties Derived from Fatty Acid Profiles, Energies. 6 (2013) 5676–5702. doi:10.3390/en6115676.
[60] S.P. Slocombe, Q. Zhang, M. Ross, A. Anderson, N.J. Thomas, Á. Lapresa, C. Rad-Menéndez, C.N. Campbell, K.D. Black, M.S. Stanley, Unlocking nature's treasure-chest: screening for oleaginous algae, Sci. Rep. 5 (2015) 9844.
[61] H. Chtourou, I. Dahmen, A. Jebali, F. Karray, I. Hassairi, S. Abdelkafi, H. Ayadi, S. Sayadi, A. Dhouib, Characterization of Amphora sp., a newly isolated diatom wild strain, potentially usable for biodiesel production, Bioprocess Biosyst. Eng. 38 (2015) 1381–1392. doi:10.1007/s00449-015-1379-6.

[62] M. De la Pena, Cell growth and nutritive value of the tropical benthic diatom, Amphora sp., at varying levels of nutrients and light intensity, and different culture locations, J. Appl. Phycol. 19 (2007) 647–655.
[63] M. Giordano, M. Kansiz, P. Heraud, J. Beardall, B. Wood, D. McNaughton, Fourier transform infrared spectroscopy as a novel tool to investigate changes in intracellular macromolecular pools in the marine microalga Chaetoceros muellerii (Bacillariophyceae), J. Phycol. 37 (2001) 271–279.
[64] H. Wagner, S. Dunker, Z. Liu, C. Wilhelm, Subcommunity FTIR-spectroscopy to determine physiological cell states, Curr. Opin. Biotechnol. 24 (2013) 88–94. doi:https://doi.org/10.1016/j.copbio.2012.09.008.
[65] L.E. Brand, R.R. Guillard, L.S. Murphy, A method for the rapid and precise determination of acclimated phytoplankton reproduction rates, J. Plankton Res. 3 (1981) 193–201. http://plankt.oxfordjournals.org/content/3/2/193.short (accessed April 5, 2016).
[66] M. Gilstad, E. Sakshaug, Growth rates of ten diatom species from the Barents Sea at different irradiances and day lengths., Mar. Ecol. Prog. Ser. Oldendorf. 64 (1990) 169–173.
[67] C. Breil, M. Abert Vian, T. Zemb, W. Kunz, F. Chemat, "Bligh and Dyer" and Folch Methods for Solid–Liquid–Liquid Extraction of Lipids from Microorganisms. Comprehension of Solvatation Mechanisms and towards Substitution with Alternative Solvents, Int. J. Mol. Sci. 18 (2017). doi:10.3390/ijms18040708.
[68] G.-D. Feng, F. Zhang, L.-H. Cheng, X.-H. Xu, L. Zhang, H.-L. Chen, Evaluation of FT-IR and Nile Red methods for microalgal lipid characterization and biomass composition determination, Bioresour. Technol. 128 (2013) 107–112.
[69] K. Stehfest, J. Toepel, C. Wilhelm, The application of micro-FTIR spectroscopy to analyze nutrient stress-related changes in biomass composition of phytoplankton algae, Plant Physiol. Biochem. 43 (2005) 717–726. doi:10.1016/j.plaphy.2005.07.001.
[70] A.P. Dean, D.C. Sigee, B. Estrada, J.K. Pittman, Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae, Bioresour. Technol. 101 (2010) 4499–4507. doi:10.1016/j.biortech.2010.01.065.
[71] M. Mubarak, A. Shaija, T.V. Suchithra, A review on the extraction of lipid from microalgae for biodiesel production, Algal Res. 7 (2015) 117–123. doi:10.1016/j.algal.2014.10.008.
[72] Y. Li, F.G. Naghdi, S. Garg, T.C. Adarme-Vega, K.J. Thurecht, W.A. Ghafor, S. Tannock, P.M. Schenk, A comparative study: the impact of different lipid extraction methods on current microalgal lipid research, Microb. Cell Factories. 13 (2014) 14.
[73] S.J. Lee, B.-D. Yoon, H.-M. Oh, Rapid method for the determination of lipid from the green alga Botryococcus braunii, Biotechnol. Tech. 12 (1998) 553–556.
[74] D. Xu, Z. Gao, F. Li, X. Fan, X. Zhang, N. Ye, S. Mou, C. Liang, D. Li, Detection and quantitation of lipid in the microalga Tetraselmis subcordiformis (Wille) Butcher with BODIPY 505/515 staining, Bioresour. Technol. 127 (2013) 386–390. doi:10.1016/j.biortech.2012.09.068.
[75] H. De la Hoz Siegler, W. Ayidzoe, A. Ben-Zvi, R. Burrell, W. McCaffrey, Improving the reliability of fluorescence-based neutral lipid content measurements in microalgal cultures, Algal Res. 1 (2012) 176–184.
[76] T. Mutanda, D. Ramesh, S. Karthikeyan, S. Kumari, A. Anandraj, F. Bux, Bioprospecting for hyper-lipid producing microalgal strains for sustainable biofuel production, Bioresour. Technol. 102 (2011) 57–70. doi:10.1016/j.biortech.2010.06.077.

[77] W. Chen, C. Zhang, L. Song, M. Sommerfeld, Q. Hu, A high throughput Nile red method for quantitative measurement of neutral lipids in microalgae, J. Microbiol. Methods. 77 (2009) 41–47. doi:10.1016/j.mimet.2009.01.001.
[78] L.M. Laurens, E.J. Wolfrum, Rapid compositional analysis of microalgae by NIR spectroscopy, NIR News. 23 (2012) 9–11.
[79] M.S. Cooper, W.R. Hardin, T.W. Petersen, R.A. Cattolico, Visualizing "green oil" in live algal cells, J. Biosci. Bioeng. 109 (2010) 198–201. doi:10.1016/j.jbiosc.2009.08.004.
[80] J.N. Murdock, D.L. Wetzel, FT-IR Microspectroscopy Enhances Biological and Ecological Analysis of Algae, Appl. Spectrosc. Rev. 44 (2009) 335–361. doi:10.1080/05704920902907440.
[81] I. Dreissig, S. Machill, R. Salzer, C. Krafft, Quantification of brain lipids by FTIR spectroscopy and partial least squares regression, Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 71 (2009) 2069–2075. doi:10.1016/j.saa.2008.08.008.
[82] S.M. Renaud, H.C. Zhou, D.L. Parry, L.-V. Thinh, K.C. Woo, Effect of temperature on the growth, total lipid content and fatty acid composition of recently isolated tropical microalgae Isochrysis sp., Nitzschia closterium, Nitzschia paleacea, and commercial species Isochrysis sp.(clone T. ISO), J. Appl. Phycol. 7 (1995) 595–602. http://link.springer.com/article/10.1007/BF00003948 (accessed August 25, 2016).
[83] M.A. Borowitzka, L.J. Borowitzka, Micro-algal biotechnology, Cambridge University Press, 1988.
[84] H.A. Spoehr, H.W. Milner, THE CHEMICAL COMPOSITION OF CHLORELLA; EFFECT OF ENVIRONMENTAL CONDITIONS, Plant Physiol. 24 (1949) 120–149. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC437916/.
[85] I.A. Guschina, J.L. Harwood, Lipids and lipid metabolism in eukaryotic algae, Prog. Lipid Res. 45 (2006) 160–186. doi:10.1016/j.plipres.2006.01.001.
[86] L. Brand, R. Guillard, The effects of continuous light and light intensity on the reproduction rates of twenty-two species of marine phytoplankton, J. Exp. Mar. Biol. Ecol. 50 (1981) 119–132.
[87] P.J. le B. Williams, L.M. Laurens, Microalgae as biodiesel & biomass feedstocks: review & analysis of the biochemistry, energetics & economics, Energy Environ. Sci. 3 (2010) 554–590.
[88] S. Wahidin, A. Idris, S.R.M. Shaleh, The influence of light intensity and photoperiod on the growth and lipid content of microalgae Nannochloropsis sp., Bioresour. Technol. 129 (2013) 7–11. doi:10/f4mb36.
[89] L. Rodolfi, G. Chini Zittelli, N. Bassi, G. Padovani, N. Biondi, G. Bonini, M.R. Tredici, Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor, Biotechnol. Bioeng. 102 (2009) 100–112. doi:10.1002/bit.22033.
[90] E.S. Shuba, D. Kifle, Microalgae to biofuels: 'Promising' alternative and renewable energy, review, Renew. Sustain. Energy Rev. 81 (2018) 743–755. doi:10.1016/j.rser.2017.08.042.
[91] M.P. de Souza, M. Hoeltz, P.D. Gressler, L.B. Benitez, R.C.S. Schneider, Potential of Microalgal Bioproducts: General Perspectives and Main Challenges, Waste Biomass Valorization. (2018). doi:10.1007/s12649-018-0253-6.
[92] M.-H. Liang, J. Zhu, J.-G. Jiang, High-value bioproducts from microalgae: strategies and progress, Crit. Rev. Food Sci. Nutr. 0 (2018) 01–53. doi:10.1080/10408398.2018.1455030.

7 Figure legends

Figure. 1 Example of growth curves measured for *Amphora* sp 1 NCC260 by (A) cell count, (B) fluorometry PAM (F0) and (C) radiometry (NDVI). Points corresponding to cell concentration according to time. Line curves corresponding to the Gompertz model fitted to cell concentration as a function of time. n=3, vertical bar = SD. The arrows indicate the late exponential day estimated by the Gompertz model.

Figure. 2 Values of the parameters measured for the screened strains. This includes strain productivity (A), lipid content as measured from the gravimetric method (B), the lipid ratio as measured semi-quantitatively by the FTIR approaches, [(eb+CH3+CH3)/si] (C) or [area eb+CH2+CH3)/Total area] multiplied by 100 for scaling purposes (D). The a, b, c notation corresponded to the maximum and minimum values for groups 1, 2 and 3. N=3, independent measurements, ± SD.

Figure. 3 Example of averaged FTIR spectra recorded on entire cells or on the corresponding lipid extract. (A) *Staurosira* sp NCC182 FTIR signature recorded on the entire cells and (B) *Staurosira* sp NCC182 FTIR signature recorded on a crude lipid extract. The grey area corresponds to the variation of the FTIR signal associated to the standard deviation for n=3 independent measurements.

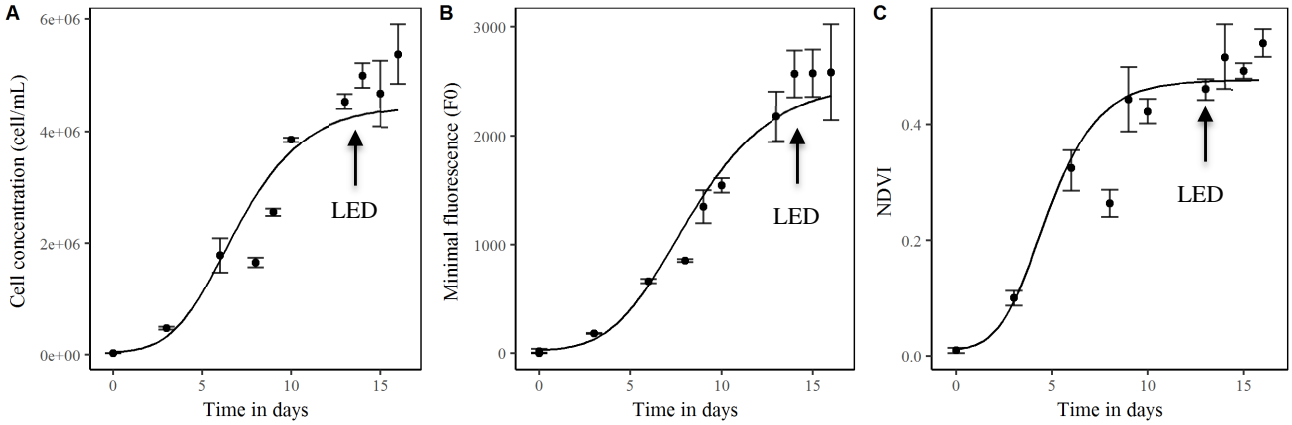
Figure. 4 Correspondence analysis map calculated on the basis of the macromolecular content as evaluated by FTIR on all the assayed strains of the NCC. N=3 independent measurements.

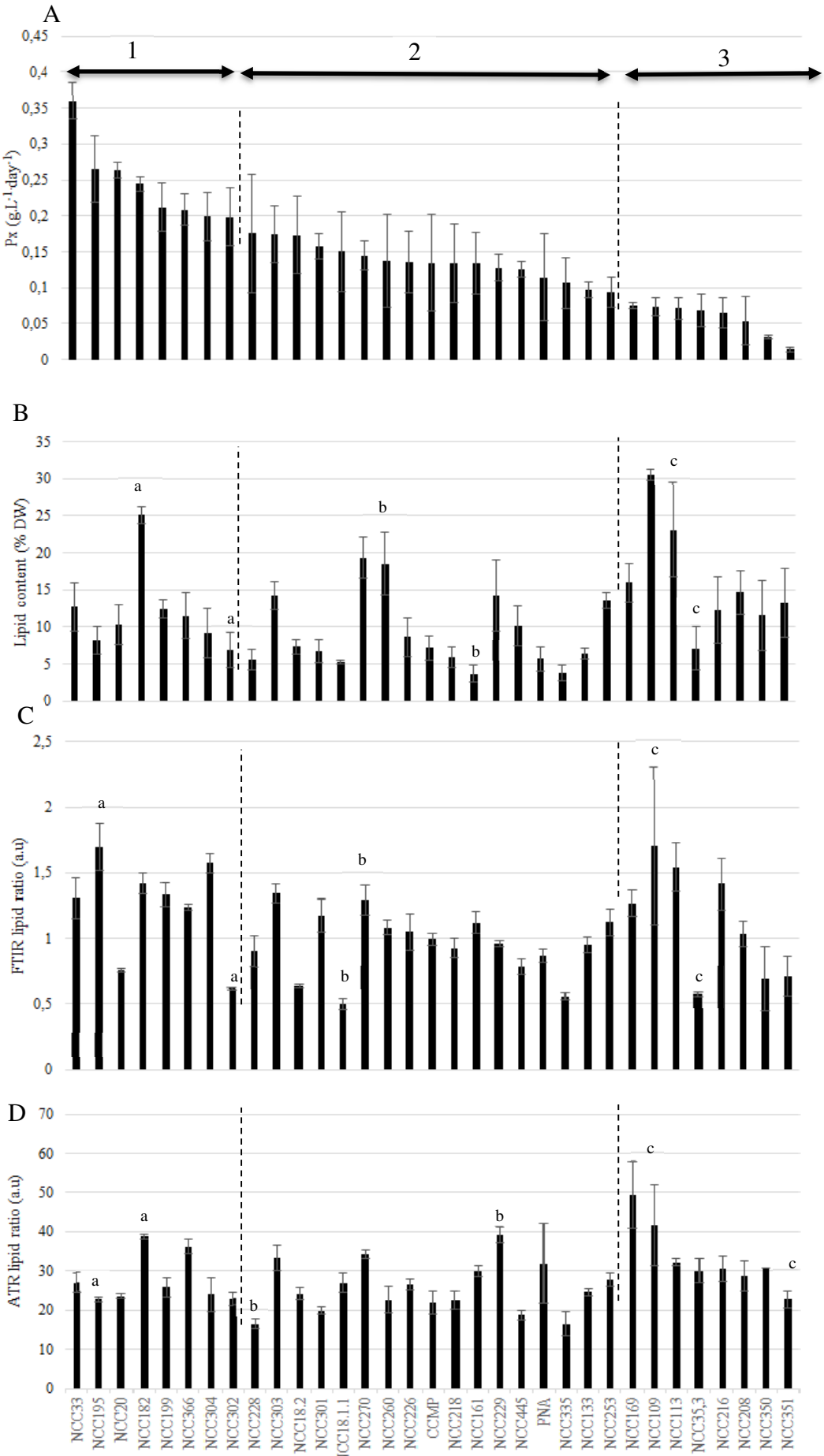
Figure 5 Venn diagram showing the degree of overlap among the different approaches used to identify the lipid rich diatoms. In the gravimetric method circle, 14 strains were identified as rich in lipids: 4 only with this method (NCC260, NCC208, NCC253, NCC351), one strains were also identified as rich and lipid by the FTIR on lipid extract (NCC229), and two strains were also identified as rich in lipids by the FTIR on entire cells (NCC33, NCC199). One strain was identified by FTIR on lipid extract and FTIR on whole cells (NCC366) and seven strains were identified as rich in lipids by all three methods (NCC109, NCC182, NCC113, NCC270, NCC169, NCC303, NCC216).

Figure 6. Boxplots summarizing the sample distribution criteria as measured with FTIR methods, gravimetry for the lipid rate and with flurorimetry for productivity. FTIR data was expressed in arbitrary units. Lipid rates in %DW and Px in $g.L^{-1}.day^{-}.^{1}$ FTIR results wasmultiplied by 20 for scaling purposes. The Px was multiplied by 100 for scaling purposes. 33 strains were assayed in independent biological triplicates.

Figures

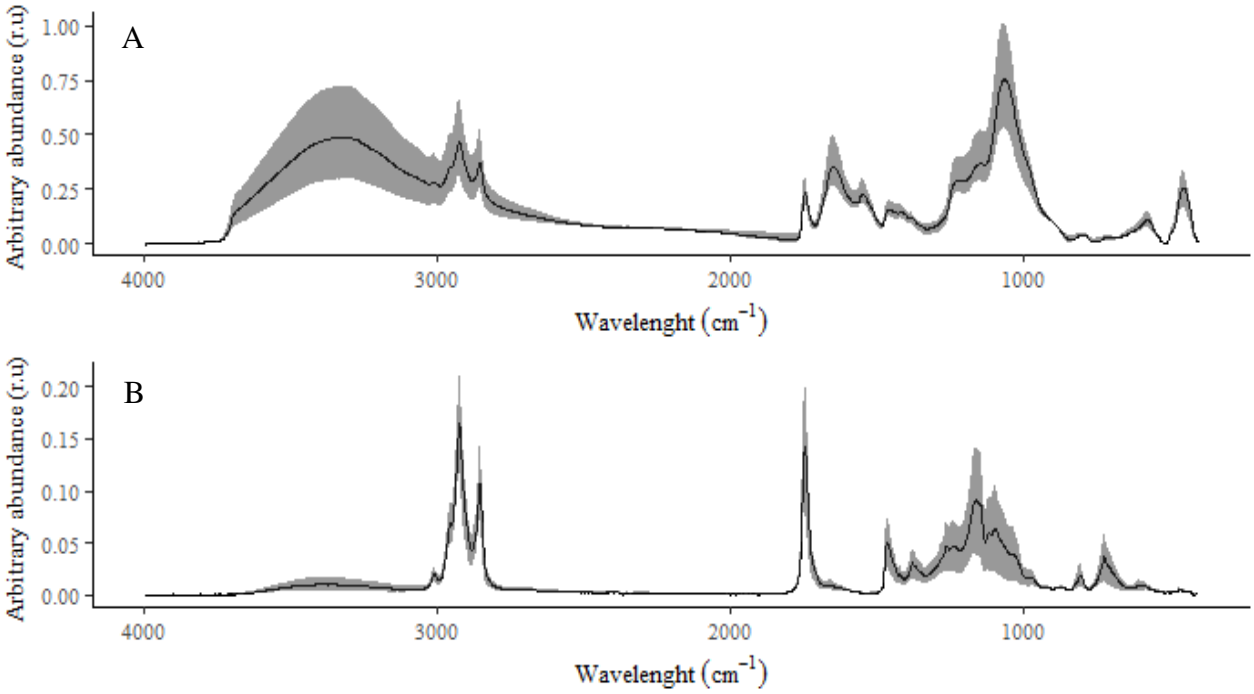
Figure 1.




A
1
2
3
Px (g.L-1.day-1)
B
Lipid content (% DW)
C
FTIR lipid ratio (a.u)
D
ATR lipid ratio (a.u)
NCC33
NCC195
NCC20
NCC182
NCC199
NCC366
NCC304
NCC302
NCC228
NCC303
NCC18.2
NCC301
NCC18.1.1
NCC270
NCC260
NCC226
CCMP
NCC218
NCC161
NCC229
NCC445
PNA
NCC335
NCC133
NCC253
NCC169
NCC109
NCC113
NCC35,3
NCC216
NCC208
NCC350
NCC351

1153

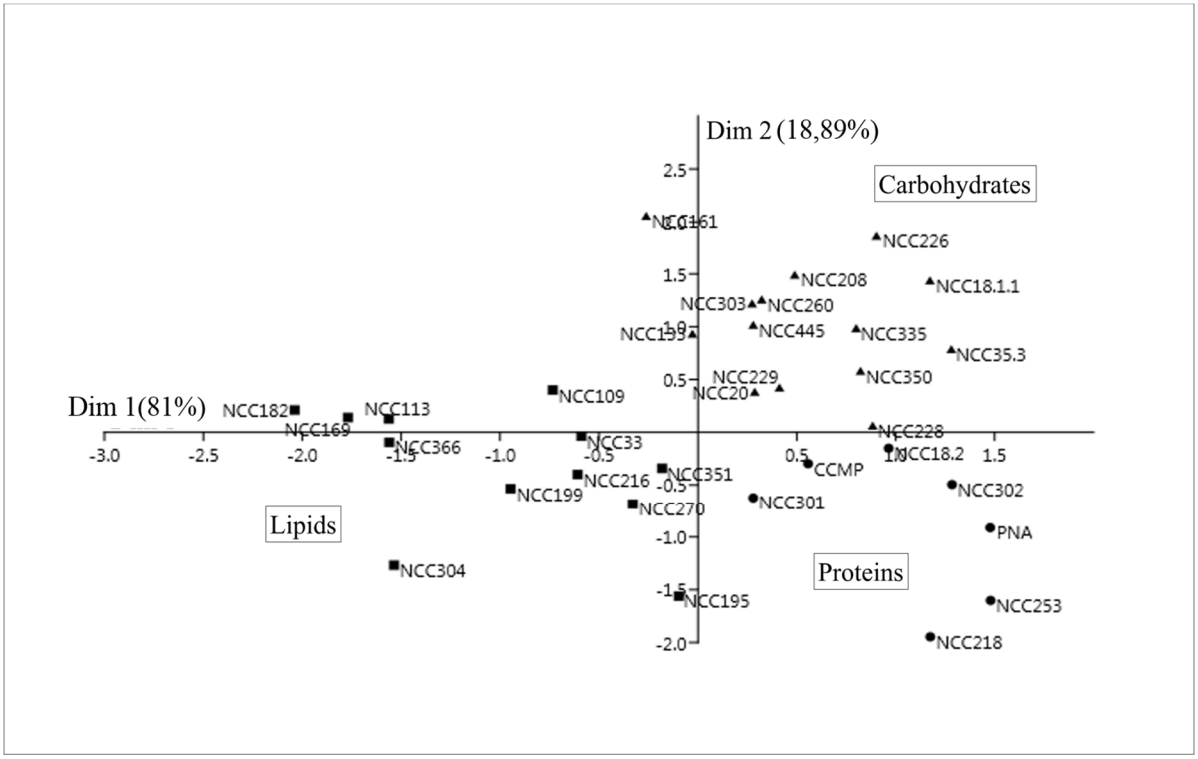
Fig. 3


A
Arbitrary abundance (r.u)
1.00
0.75
0.50
0.25
0.00
4000
3000
2000
1000
Wavelenght (cm^{-1})
B
Arbitrary abundance (r.u)
0.20
0.15
0.10
0.05
0.00
4000
3000
2000
1000
Wavelenght (cm^{-1})

1179 Fig. 4

1180

1181



1182

1183

1184

1185

1186

1187

1188

1189

1190

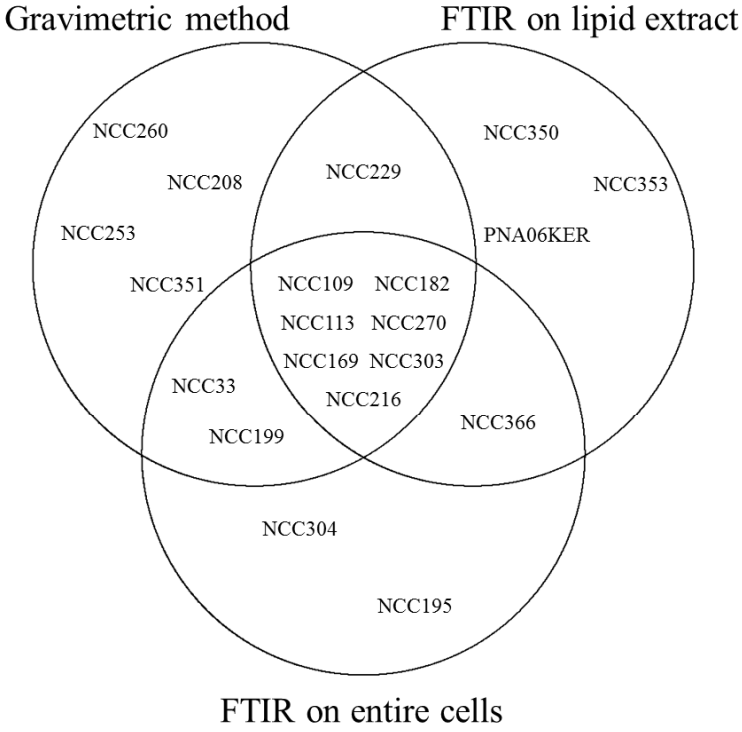
1191

1192

1193

1194

1195 Fig. 5



1196

1197

1198

1199

1200

1201

1202

1203

1204

1205

1206

1207

1208

Fig. 6

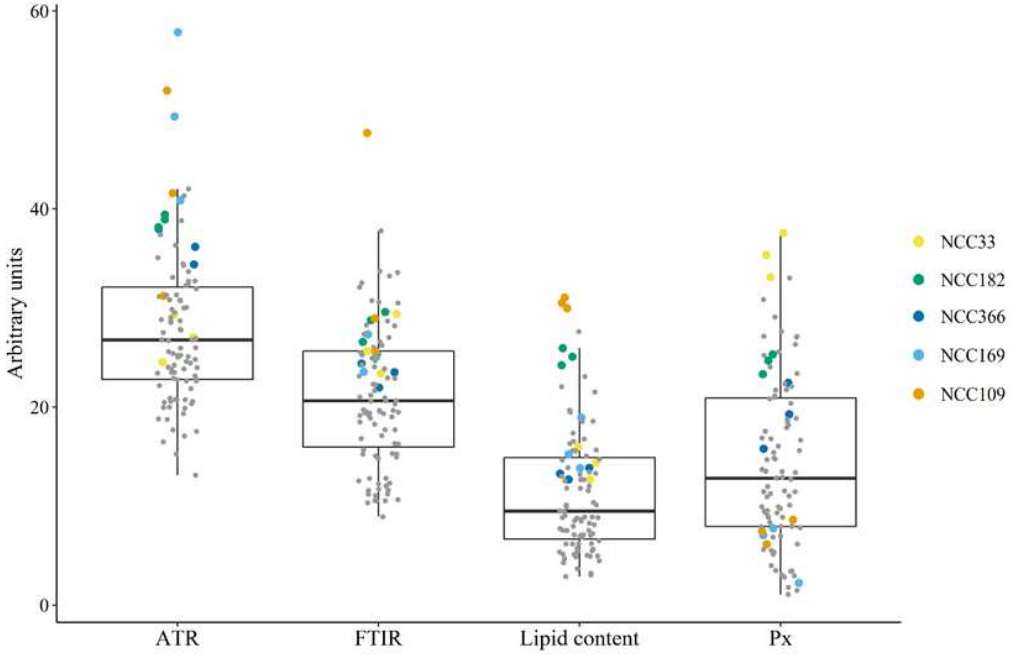

60
40
20
0
Arbitrary units
ATR
FTIR
Lipid content
Px
NCC33
NCC182
NCC366
NCC169
NCC109

Table S1. Supplementary data. Strains who did not grown.

Species	NCC strain identification	Sampling location
Amphora sp. B2	NCC261	*France,NW Atlantic coast*
Amphora sp. AC16	NCC410	*France,NW Atlantic coast*
Amphora sp. AE8	NCC413	*France,NW Atlantic coast*
Berkeleya rutilans	NCC210.2	*France,NW Atlantic coast*
Berkeleya rutilans	NCC309	*France, Mediterranean sea*
Brockmaniella brockmanii 2	NCC403	*France,NW Atlantic coast*
Caloneis sp. 1	NCC180	*France,NW Atlantic coast*
Catacombas sp. 1	NCC337	*France, Mediterranean sea*
Cocconeis scutellum 1	NCC209.1	*France, NW Atlantic coast*
Cocconeis scutellum 2	NCC209.2	*France, NW Atlantic coast*
Cocconeis scutellum 2	NCC209.3	*France, NW Atlantic coast*
Craspedostauros sp. 3	NCC57	*France, NW Atlantic coast*
Craspedostauros sp. 4	NCC58	*France, NW Atlantic coast*
Craspedostauros sp. 5	NCC204	*France, NW Atlantic coast*
Entomoneis alata 1	NCC16	*France, NW Atlantic coast*
Entomoneis alata 2	NCC448	*Portugal, NW Atlantic coast*
Entomoneis sp. BAB2	NCC415	*France, NW Atlantic coast*
Gyrosigma sp. 1	NCC411	*France, NW Atlantic coast*
Gyrosigma sp. 2	NCC412	*France, NW Atlantic coast*
Gyrosigma tenuissimum	NCC258	*France, NW Atlantic coast*
Halamphora coffeaformis	UTCC58	*Canada, NW Atlantic coast*
Helicotheca tamesis 1	NCC59	*, France, Mediterranean sea*
Helicotheca tamesis 2	NCC60	*France,NW Atlantic coast*
Lampriscus sp.	NCC347	*France, Mediterranean sea*
Leptocylindrus danicus 1	NCC205	*France,NW Atlantic coast*
Leptocylindrus danicus 2	NCC206	*France,NW Atlantic coast*
Melosira nummuloîdes 1	NCC25	*France,NW Atlantic coast*
Melosira nummuloîdes 2	NCC25.1	*France,NW Atlantic coast*
Navicula sp. Z4	NCC224	*France,NW Atlantic coast*
Navicula sp. e1	NCC269	*France, NW Atlantic coast*
Navicula cf ramosissima	NCC449	*France, NW Atlantic coast*
Nitzschia laevis	NCC39	*France, NW Atlantic coast*
Nitzschia salinicola	NCC41	*France, NW Atlantic coast*
Nitzschia sp. B4	NCC114	*France, NW Atlantic coast*
Opehora sp. 2	NCC365	*France, NW Atlantic coast*
Paralia sulcata	NCC177	*France, NW Atlantic coast*
Pleurosigma sp. K	NCC339	*Ukraine, Black sea*
Pleurosigma sp. LM	NCC404	*France, NW Atlantic coast*
Pleurosigma sp. BC1	NCC423	*France, NW Atlantic coast*
Pleurosigma sp. BC7	NCC425	*France, NW Atlantic coast*
Pleurosigma sp. BC15	NCC428	*France, NW Atlantic coast*
Rhizosolenia setigera	NCC127	*France, NW Atlantic coast*
Tabularia tabulata	NCC338	*France, NW Atlantic coast*

Table S2: Lipid quantification evaluated by FTIR and ATR method for all the species.

Species	code	FTIR					ATR	
		Carb/si	Prot/si	Eb/si	CH2/si	CH3/si	eb	CH2 + CH3
Amphora acutiuscula	NCC216	53.6 ± 8.2	45.8 ± 5.7	22.1 ± 4.9	65.7 ± 8.0	53.5 ± 7.8	6.2 ± 0.9	24.3 ± 2.4
Amphora sp. 1	NCC260	63.6 ± 2.9	40.0 ± 2.0	12.2 ± 2.3	57.3 ± 3.3	39.2 ± 0.6	4.5 ± 0.7	18.1 ± 3.1
Amphora sp. 2	NCC169	58.1 ± 4.3	42.4 ± 1.1	17.2 ± 3.4	59.7 ± 3.4	49.7 ± 3.0	7.1 ± 0.7	42.1 ± 11.0
Brockmaniella brockmanii	NCC161	51.2 ± 5.5	25.2 ± 6.4	12.8 ± 4.0	57.7 ± 5.5	41.8 ± 5.5	4.1 ± 1.6	21.1 ± 5.6
Conticriba weissflogii	NCC133	47.4 ± 0.5	40.2 ± 0.8	5.03 ± 1.5	48.4 ± 2.6	41.7 ± 1.9	4.7 ± 0.3	20.0 ± 0.7
Conticriba weissflogii	CCMP1336	48.3 ± 2.8	43.5 ± 3.3	9.8 ± 1.8	48.7 ± 3.0	40.9 ± 2.7	5.0 ± 0.7	16.9 ± 2.2
Craspedostauros britannicus	NCC195	60.7 ± 2.4	69.8 ± 11	23.2 ± 3.3	80.7 ± 8.0	66.0 ± 6.9	4.8 ± 0.4	19.2 ± 1.8
Craspedostauros britannicus	NCC199	55.1 ± 3.4	47.5 ± 5.7	26.9 ± 3.4	62.2 ± 2.7	44.2 ± 3.3	5.4 ± 0.6	20.5 ± 1.9
Craspedostauros sp. 1	NCC228	44.3 ± 0.1	37.8 ± 4.3	6.2 ± 0.4	46.3 ± 6.4	37.8 ± 6.0	3.9 ± 0.8	15.2 ± 3.4
Craspedostauros sp. 2	NCC218	54.6 ± 2.6	49.4 ± 3.5	7.5 ± 0.3	46.0 ± 3.8	39.0 ± 4.0	4.7 ± 0.3	17.9 ± 2.5
Cymatosira belgica	NCC208	41.6 ± 4.8	25.4 ± 3.3	6.7 ± 1.1	55.4 ± 5.1	41.5 ± 3.7	5.9 ± 0.7	22.7 ± 3.2
Entomoneis paludosa	NCC18.1.1	39.6 ± 0.3	26.0 ± 2.7	2.9 ± 0.1	27.3 ± 2.4	19.4 ± 1.8	3.9 ± 0.7	23.1 ± 1.8
Entomoneis paludosa	NCC18.2.1	45.7 ± 0.6	40.9 ± 1.4	6.1 ± 0.1	34.5 ± 0.9	22.5 ± 0.6	3.1 ± 0.4	21.6 ± 1.1
Entomoneis sp. 1	NCC350	39.9 ± 0.1	30.4 ± 0.8	5.5 ± 1.5	28.6 ± 0.9	21.2 ± 1.7	2.7 ±0.1	30.6 ± 4.3
Entomoneis sp. 2	NCC20	42.5 ± 0.5	32.5 ± 0.3	9.5 ±0 .3	40.0 ± 0.4	26.0 ± 0.6	4.7 ± 0.3	18..6 ± 0.4
Entomoneis sp. 3	NCC351	45.5 ± 2.9	41.4 ± 4.2	14.4 ± 3.4	37.2 ± 0.8	28.3 ± 0.6	3.9 ± 1.7	15.8 ± 5.6
Entomoneis sp. 4	NCC301	62.5 ± 4.8	56.7 ± 6.4	12.5 ± 2.5	64.0 ± 7.9	46.7 ± 5.6	2.9 ± 0.1	17.0 ± 0.8
Entomoneis sp. 5	NCC302	37.6 ± 0.7	36.8 ± 0.8	3.43 ± 0.1	33.9 ± 0.7	23.9 ± 0.5	3.3 ± 0.3	19.6 ± 1.5
Entomoneis sp. 6	NCC335	50.5 ± 0.6	35.3 ± 2.7	6.6 ± 0.7	29.9 ± 1.9	19.0 ± 1.1	3.1 ± 0.2	13.4 ± 3.0
Entomoneis sp. 7	NCC445	40.6 ± 0.9	27.0 ± 2.7	8.5 ± 0.2	37.7 ± 0.4	29.2 ± 0.3	4.1 ± 0.9	17.5 ± 4.3
Extubocellulus cf cribriger	NCC229	49.8 ± 3.7	38.1 ± 2.6	10.1 ± 1.2	48.0 ± 1.9	38.0 ± 1.1	4.9 ± 0.3	34.2 ± 1.7
Fallacia sp. 1	NCC303	64.7 ± 1.8	74.4 ± 7.0	14.2 ± 2.5	64.5 ± 4.1	55.5 ± 5.4	6.9 ± 0.3	26.4 ± 2.9
Fallacia sp. 2	NCC304	46.0 ± 6.2	45.7 ± 3.2	31.3 ± 4.0	72.8 ± 2.7	53.3 ± 2.1	5.1 ± 1.0	18.9 ± 3.4
Licmophora sp. 1	NCC253	53.1 ± 2.7	65 ± 10	4.4 ± 1.1	57.9 ± 5.4	50.3 ± 5.2	3.6 ± 0.3	22.6 ± 0.6
Lithodesmium sp	NCC353	34.2 ± 2.0	25.8 ± 0.8	2.4 ± 0.3	31.1 ± 1.1	23.1 ± 0.4	2.1 ± 0.3	28.0 ± 3.2
Navicula sp. 1	NCC113	63.3 ± 4.7	43.8 ± 8.9	33.4 ± 4.7	68.2 ± 9.6	52.4 ± 5.6	3.4 ± 0.4	28.8 ± 1.2
Navicula sp. 2	NCC226	54.2 ± 4.3	27.0 ± 3.0	5.2 ± 0.9	53.1 ± 3.8	39.5 ± 2.8	4.2 ± 0.7	22.4 ± 0.7
Nitzschia sp. 5	NCC109	62.0 ± 2.8	42.4 ± 6.7	22.0 ± 2.3	72.7 ± 4.6	61.3 ± 4.8	6.9 ± 0.4	45.1 ± 28.9
Nitzschia alexandrina	NCC33	58.2 ± 7.2	44.0 ± 4.6	20.7 ± 3.3	56.1 ± 4.0	45.8 ± 0.7	5.7 ± 0.3	21.3 ± 2.3
Opephora sp. 1	NCC366	58 ± 12	40.1 ± 8.8	31.6 ± 6.1	50.6 ± 4.5	42.0 ± 2.5	6.9 ± 1.1	26.3 ± 4.4
Pseudonitzschia americana	PNA06 KER	56.6 ± 1.4	60.6 ± 3.9	3.9 ± 1.0	44.0 ± 2.8	39.1 ± 3.2	4.0 ± 0.3	38 ± 28
Staurosira sp.	NCC182	47.7 ± 6.6	30.7 ± 4.1	30.9 ± 5.4	61.6 ± 4.0	49.0 ± 2.3	7.3 ± 0.1	31.6 ± 0.6
Surirella sp. 1	NCC270	58.5 ± 2.4	54.4 ± 8.4	22.0 ± 4.6	56.1 ± 5.3	51.3 ± 5.4	7.1 ± 0.4	27.1 ± 0.6