



Improved method for the extraction of high-quality DNA from lignocellulosic compost samples for metagenomic studies

Ângela M. A. Costa¹ · Andréia O. Santos¹ · Joana Sousa¹ · Joana L. Rodrigues¹ · Eduardo J. Gudiña¹ · Sara C. Silvério¹ · Ligia R. Rodrigues¹

Received: 2 July 2021 / Revised: 7 October 2021 / Accepted: 13 October 2021 / Published online: 1 November 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The world economy is currently moving towards more sustainable approaches. Lignocellulosic biomass has been widely used as a substitute for fossil sources since it is considered a low-cost bio-renewable resource due to its abundance and continuous production. Compost habitats presenting high content of lignocellulosic biomass are considered a promising source of robust lignocellulose-degrading enzymes. Recently, several novel biocatalysts from different environments have been identified using metagenomic techniques. A key point of the metagenomics studies is the extraction and purification of nucleic acids. Nevertheless, the isolation of high molecular weight DNA from soil-like samples, such as compost, with the required quality for metagenomic approaches remains technically challenging, mainly due to the complex composition of the samples and the presence of contaminants like humic substances. In this work, a rapid and cost-effective protocol for metagenomic DNA extraction from compost samples composed of lignocellulosic residues and containing high content of humic substances was developed. The metagenomic DNA was considered as representative of the global environment and presented high quality (> 99% of humic acids effectively removed) and sufficient quantity (10.5–13.8 $\mu\text{g g}^{-1}$ of compost) for downstream applications, namely functional metagenomic studies. The protocol takes about 4 h of bench work, and it can be performed using standard molecular biology equipment and reagents available in the laboratory.

Key points/Highlights

- *Metagenomic DNA was successfully extracted from compost samples rich in humic acids*
- *The improved protocol was established by optimizing the cell lysis method and buffer*
- *Complete removal of humic acids was achieved through the use of activated charcoal*
- *The suitability of the DNA was proven by the construction of a metagenomic library*

Keywords Lignocellulosic · Compost · Metagenomic DNA · Humic acids · Activated charcoal

Introduction

The use of lignocellulosic biomass as a renewable feedstock holds great potential for biotechnology, biofuel and biomaterial industries (Guerriero et al. 2016). This kind of feedstock provides an interesting alternative to fossil sources since it is cheap, abundant, recyclable, renewable and widespread biomass on Earth, being its use compatible with food production (Tirado-Acevedo et al. 2010; Sheldon and Woodley 2018). Several innovative strategies have been developed for

the conversion of lignocellulose. In this context, more than 200 high value-added compounds have been reported (Isikgor and Becer 2015; Arevalo-Gallegos et al. 2017; Ali et al. 2020; Haldar and Purkait 2020; Ning et al. 2021; Patel and Shah 2021). Various fungi and bacteria able to enzymatically decompose lignocellulose were previously identified. However, the complete enzymatic hydrolysis of this feedstock is still a challenge due to the recalcitrant structure of these materials (Batista-García et al. 2016). Therefore, it is essential to continue the search for novel microbial enzymes with improved ability to fully convert lignocellulose (Montella et al. 2016; Wang et al. 2019).

Metagenomic approaches have opened a window to the unknown diversity of microorganisms present in the environment (Schmeisser et al. 2007). Functional

✉ Sara C. Silvério
sarasilverio@deb.uminho.pt

¹ CEB - Centre of Biological Engineering, Universidade Do Minho, 4710-057 Braga, Portugal

metagenomics is the state of art technique to identify new enzymes and biochemical mechanisms with desired biological activities (Mirete et al. 2016; Berini et al. 2017; Madhavan et al. 2017). An environment where lignocellulosic biomass is naturally processed represents a promising source of lignocellulose-degrading enzymes. Compost habitats are recognized as important bioreactors for the renewable bioenergy of the planet, sustaining an immense diversity of microorganisms specialized in the degradation of lignocellulosic residues (Wang et al. 2016). Therefore, this type of biomass represents a favourable source of biocatalysts with high stability and efficiency. Isolation of high molecular weight (HMW) metagenomic DNA is the critical step for metagenomic library construction for functional screening. Since the 1980s, several methods for DNA extraction have been published and continue to be revised (Lear et al. 2018). Numerous studies tried to establish the most suitable protocol for the extraction of total soil DNA (Philippot et al. 2010; Petric et al. 2011; Tanase et al. 2015). Nevertheless, a standard method that can be universally applied to all soil types to obtain DNA with the desired purity, quantity and HMW for metagenomic approaches is still lacking.

DNA extraction from compost samples represents a challenging procedure since these samples are rich in organic matter and humic material. Humic substances have a similar charge and size to DNA, generally resulting in co-precipitation during the extraction process, as may be evidenced by the frequent brown colour of some DNA extracts (Harry et al. 1999; Lakay et al. 2007). The presence of humic material in DNA extracts also inhibits the activity of some enzymes, thus limiting downstream applications (Harry et al. 1999; Dong et al. 2006). Some studies highlighted the difficulty to recover pure DNA and reported the development of several specific methodologies to increase purity (Harry et al. 1999; Rajendhran and Gunasekaran 2008; Young et al. 2014). Therefore, time-consuming methods like the use of caesium chloride density gradients, extensive and repetitive precipitation steps and also expensive chromatographic procedures have been described (Lakay et al. 2007; Rajendhran and Gunasekaran 2008; Sar et al. 2018). Nevertheless, these methodologies either are not cost-effective or require additional steps for the purification of DNA, namely the removal of humic acids and size selection.

In this work, a direct extraction method for metagenomic DNA isolation from compost or soil samples containing up to 50% of lignocellulosic material is described. This alternative protocol allows a rapid, efficient and cost-effective extraction of total DNA with the quantity, purity and HMW suitable for further metagenomic studies (provisional patent application PT116634). The methodology relies on the mechanical and chemical/enzymatic lysis of cellular material combined with the action of powdered activated charcoal

(PAC). Due to the pore volume and large surface area which results in an exceptional adsorption capacity, PAC is commonly used to eliminate several contaminants (Marsh and Rodríguez-Reinoso 2006).

Material and methods

Reagents

All chemicals and buffer components were of analytical grade and purchased from Sigma-Aldrich (St. Louis, USA) unless specified otherwise.

Compost samples

The compost samples were collected from two Portuguese composting units, namely Terra Fértil (Chamusca) and Lipor (Porto). The Terra Fértil unit generally handles agroforestry residues (around 50%) and municipal sludge (around 50%). The sample from Terra Fértil (approximately 1 kg) was obtained from a pile with approximately 4 weeks of composting, corresponding to the thermophilic phase of the process. The Lipor unit generally combines food wastes (50–70%) and agroforestry residues (30–50%). The sample from Lipor (approximately 1 kg) was obtained from a pile at the mesophilic phase of the process. Until being used for physicochemical characterization and DNA extraction, the compost samples were stored at 4 °C.

Compost sample characterization

Moisture content was determined gravimetrically after drying the samples in an oven at 105 °C until constant weight. The pH of the compost samples was determined from a mixture of water/compost in the ratio of 4:1 (w/w) using a pH meter with a calibrated glass electrode (Kalra 1995). Total organic carbon, nitrogen, hydrogen and sulphur content was determined through elemental analysis by automated dry combustion (Requimte/LAQV, Faculty of Sciences and Technology, Nova University of Lisbon, Portugal).

Metagenomic DNA isolation

The methodology was established based on previous studies from Zhou et al. (1996), Bergmann et al. (2014), Devi et al. (2015) and Verma et al. (2017) with key modifications and optimizations. Total DNA was extracted from compost samples using simultaneously mechanical and chemical/enzymatic methods. The compost sample (1 g) was mixed with 5 ml of a previously optimized lysis buffer (100 mM Tris-HCl, 100 mM Na EDTA, 1.5 M NaCl, 100 mM Na₂HPO₄, 100 mM CaCl₂·2H₂O, 1 mg/ml of proteinase K

(NZYtech, Lisbon, Portugal), 1 mg/ml of lysozyme, 0.2 mg/ml RNase A (NZYtech, Lisbon, Portugal) and 1% (w/v) PAC (4–8 mesh) and 1 g of glass beads (425–600 μm) in a 50-ml tube. Depending on the tested conditions, the mixture was further homogenized: (i) briefly in a vortex, (ii) at maximum speed for 2.5 min or (iii) at maximum speed for 5 min. The soil suspension was then incubated at 37 °C, 150 rpm, for 30 min. After that, 1 ml of 20% (w/v) sodium dodecyl sulphate (SDS) was added and the sample was incubated for 30 min at 65 °C (the tube was inverted to properly mix every 10 min). The obtained lysate was centrifuged at 3200 \times g, 4 °C, for 10 min, and the supernatant was transferred to a clean 15-ml tube and gently mixed with 1 volume of chloroform:isopropanol (C:I) (24:1 v/v). The mixture was then centrifuged at 7000 \times g, 4 °C, for 5 min. The aqueous phase (top phase) was transferred to fresh tubes followed by DNA precipitation through the addition of 1 volume of 3 M $\text{C}_2\text{H}_3\text{NaO}_2$ (pH 5.2) and 0.4 volumes of 30% (w/v) polyethylene glycol (PEG, MW-8000) and mixed by inversion. The tubes were kept at –20 °C for 20 min. The tubes were slowly thawed on ice, and then, the DNA pellet was precipitated at 12,000 \times g, 4 °C, for 10 min, and resuspended with 500 μl of Tris–EDTA buffer (TE) (10 mM Tris, 1 mM Na EDTA, pH 8.0). After that, 1 volume of C:I (24:1 v/v) was added. The tubes were gently inverted a few times and further centrifuged at 12,000 \times g, 4 °C, for 5 min. The aqueous phase was transferred to fresh tubes and 1 volume of cold isopropanol was added. After gently shaking by inverting the tubes, the mixture was allowed to precipitate for 5 min at 4 °C. The precipitated DNA was pelleted by centrifugation at 12,000 \times g, 4 °C, for 10 min, and washed twice with 500 μl of 70% (v/v) ethanol. After that, the DNA sample was centrifuged at 12,000 \times g, 4 °C, for 2 min. The supernatant was discarded and the pellet was air-dried for 10 min at room temperature (RT). Metagenomic DNA was dissolved in 100 μl of TE buffer and stored at 4 °C. Three replicates of each compost sample were used to evaluate the reproducibility of the methodology.

Additional experiments to evaluate the effectiveness of the method and validate its potential application under different conditions were performed. For the Terra Fértil sample, 10 mg of commercial humic acids was added to the sample. For the Lipor sample, the vortex time used for cell lysis was reduced to 2.5 min (Lipor 2.5) and a condition without mechanical lysis (Lipor 0) was also evaluated, as described above.

Yield and purity of the metagenomic DNA

DNA was quantified using the Nanodrop One (Thermo Scientific™, Waltham, USA), and the purity was determined by both absorbance ratios of $A_{260/280}$ and $A_{260/230}$. DNA recovery yield was estimated as an average value with standard

deviation (g DNA g^{-1} compost). To evaluate the quality of the extracted DNA, a 0.5% agarose (Grisp, Porto, Portugal) gel electrophoresis was run in 1 X Tris–acetate EDTA buffer (TAE), at 25 V, 4 °C, for 24 h. GeneRuler High Range DNA Ladder (ThermoFisher, Waltham, USA) was used as a molecular marker according to the manufacturer instructions. The gel was stained with 1 mg/ml Thiazol orange and visualized using the ChemiDoc™ XRS+ (Bio-Rad, Hercules, USA).

Humic acid determination

To estimate the concentration of the humic acids in both the original composting sample and the extracted DNA, a standard curve (0–500 $\text{ng } \mu\text{l}^{-1}$) was prepared using commercial humic acids (Alfa Aesar, Kandel, Germany) and the absorbance was measured at 340 nm (A_{340}) on a spectrophotometer reader (Bio-Tek from Izasa, Carnaxide, Portugal). Humic acids from both compost samples were extracted by acid precipitation as previously reported (Sar et al. 2018). Briefly, 1 g of compost sample was mixed with 9 ml of 0.1 M NaOH and stirred for 3 h at RT. Then, the mixture was centrifuged at 2500 \times g, RT, for 10 min. The supernatant was transferred to a fresh tube and acidified to pH 1.0 using concentrated HCl and incubated overnight at RT in the dark. The precipitated humic acids were pelleted by centrifugation at 2500 \times g, RT, for 10 min and air-dried for about 15 min. The pellet was resuspended in 1 ml of TE. Suitable dilutions were prepared for both the humic acids extracted from the original compost samples and the humic acids present in the purified DNA solutions. In all cases, the concentration of humic acids was similarly determined by measuring A_{340} and using the previously prepared standard curve. All the analyses were performed in triplicate.

PCR amplification

The purity of the DNA samples was evaluated by PCR to amplify a region of the prokaryotic 16S rRNA gene (Wu et al. 2009). For amplification of 1500 bp of the 16S rRNA gene, the universal primers pair 1492R (5'-TACCTTGTTACGACTT-3') and 27F (5'-AGAGTTTGATCCTGGCTCAG-3') were used. PCR reaction mixture (total volume of 25 μl) was prepared using 0.5 μl of each primer of a work solution prepared at 10 μM , 1 μl of template DNA (approximately 20 $\text{ng}/\mu\text{l}$), 12.5 μl of Speedy Supreme NZYtaq 2 \times Green Master Mix (NZYTech, Lisbon, Portugal) and 10.5 μl of water for each sample. Amplification was performed using the following cycles: 1 cycle at 95 °C for 2 min, followed by 35 cycles of 30 s at 95 °C, 40 s at 52 °C and 1 min at 72 °C followed by 5-min extension step at 72 °C. PCR was performed using Life ECO Thermo Cycler (BIOER, Hangzhou, China). The *Escherichia coli* strain

EPI300-T1^R (Epicentre Biotechnology, Madison, USA) and distilled water were used as positive and negative controls, respectively. The amplified products were separated by electrophoresis in a 0.8% agarose gel at 80 V for 45 min. NZYDNA Ladder III (NZYTech, Lisbon, Portugal, range size 200–10,000 bp) was used as a molecular weight standard. The gel was stained with 1 mg/ml Thiazole orange and visualized using the ChemiDoc™ XRS + (Bio-Rad, Hercules, USA).

Construction of a metagenomic library

To evaluate the suitability of the extracted HMW DNA for metagenomic approaches, a metagenomic library was constructed for both compost samples. An aliquot of the metagenomic DNA isolated from the compost sample was end-repaired to generate blunt-ended, 5'-phosphorylated DNA, according to the manufacturer's instructions of CopyControl™ fosmid library kit (Epicentre Biotechnology, Madison, USA). The reaction mixture was scaled up to a final volume of 20 µl. Since 90% of the end-repaired metagenomic DNA possessed a range of approximately 40 kb and a relatively tight band was observed (data not shown), it was directly used for the construction of the fosmid library. This library was constructed using the pCC1FOS™ vector with *E. coli* strain EPI300-T1^R as a host, according to the manufacturer's instructions of the CopyControl™ fosmid library kit. One fosmid clone was randomly picked, and the fosmid DNA was isolated using the FosmidMAX™ DNA Purification Kit (Epicentre Biotechnology, Madison, USA) according to the manufacturer's instructions. The fosmid DNA insert (around 10 µg) was verified by restriction digestion using the Fast Digest enzymes *NotI* (Thermo Fisher Scientific, Waltham, USA) and *HindIII* (Thermo Fisher Scientific, Waltham, USA), according to the manufacturer's instructions. The pattern of enzymatic restriction of the fosmid was evaluated by running a 0.5% agarose (Grisp, Porto, Portugal) gel electrophoresis in 1 X TAE, at 25 V, 4 °C for 24 h. λ DNA-Mono Cut Mix (NEB, Ipswich, USA) was used as a molecular marker. The gel was stained with 1 mg/ml Thiazole orange and visualized using the ChemiDoc™ XRS + (Bio-Rad, Hercules, USA).

DNA sequencing and taxonomic annotation

Metagenome sequencing, de novo assembly and open reading frames (ORFs) prediction and taxonomic annotation were performed by Novogene (Cambridge, UK). For the construction of the metagenomic library, the NEBNext™ ULTRA™ DNA LIBRARY PREP KIT was used following the manufacturer's recommendations (New England Biolabs, Ipswich, USA) and the shotgun sequencing was performed using the Illumina NovaSeq6000 Platform.

High-quality short reads of the DNA sample were initially assembled using an optimized SOAPdenovo protocol (Li et al. 2010) and the prediction of the ORF was performed using the MetaGeneMark (Gene Probe, Inc, San Diego, USA). The redundant ORFs were removed by CD-HIT (Li and Godzik 2006). The taxonomic annotation was performed by BLASTX using DIAMOND program (Buchfink et al. 2015) against the NCBI protein non-redundant (NR) database.

Results

Compost characterization

Two lignocellulose-rich compost samples with high humic acid content were used to optimize an efficient methodology for the isolation of HMW DNA suitable for metagenomic approaches. The results obtained for the physicochemical characterization of the compost samples are described in Table 1. The compost sample from Terra Fértil was collected at the thermophilic phase of the process at a temperature around 62.7 °C, presenting a pH of 6.96 and a moisture content of 69%. For each gramme of compost, approximately 10.6 mg of humic acids was extracted. The compost sample from Lipor was collected at the mesophilic phase of the process (20–40 °C), presenting a pH of 8.47 and a moisture content of 87.4%. For each gramme of compost, approximately 10.9 mg of humic acids was extracted.

Optimization of the DNA extraction method

The experimental procedure herein presented results from the combination of different strategies of DNA extraction previously reported (Zhou et al. 1996; Bergmann et al. 2014; Devi et al. 2015; Verma et al. 2017) with key modifications and improvements. Several experimental conditions (temperature, incubation and agitation time, buffer

Table 1 Physicochemical properties of the compost samples used in the DNA extraction procedure

Physicochemical properties	Compost sample	
	Terra Fértil	Lipor
Sampling temperature (°C)	62.7	20–40
Moisture content (%)	69.0 ± 2.4	87.4 ± 4.7
pH	6.96 ± 0.02	8.47 ± 0.01
Total carbon (%)	13.23 ± 0.41	17.91 ± 1.64
Total nitrogen (%)	1.15 ± 0.04	1.89 ± 0.17
Total hydrogen (%)	4.22 ± 0.21	3.69 ± 0.62
Total sulphur (%)	0.41 ± 0.01	0.17 ± 0.03
Humic acid (mg g ⁻¹ compost)	10.6 ± 0.38	10.9 ± 0.45

composition and addition of reagents) were previously tested (data not shown) to select the best combination towards a purer and HMW DNA while reducing the time and costs associated with the protocol. As an example, in Table 2, three preliminary tested protocols are presented comparing the main experimental conditions and results obtained. Additionally, the visual aspect of the DNA solutions after extraction with these preliminary protocols is shown in Fig. 1.

The optimized method successfully uses mechanical and chemical/enzymatic lysis to access the HMW DNA, and PAC to effectively remove contaminants, namely the humic acids (Table 2). A schematic representation of the improved methodology for metagenomic DNA isolation from compost samples is represented in Fig. 2. As illustrated in the figure, the optimized methodology comprises three main steps, namely cell lysis and humic acid removal, DNA recovery and DNA purification. The most relevant improvement noticed for this methodology resulted from the optimization of the cell lysis (procedure and buffer) and the addition of an effective agent to remove the undesired contaminants. Therefore, in the second step, after humic materials and other contaminants being successfully eliminated by centrifugation, the metagenomic

DNA was retrieved. In the last step, the crude DNA was ultimately purified. The visual aspect of the DNA solution obtained after extraction with the optimized method is presented in Fig. 1.

The efficiency of the isolation procedure and quality of the DNA

Total DNA was successfully extracted from both compost samples using the optimized method described above. With 5 min of mechanical lysis, an average yield of 10.5 ± 0.22 and 23.9 ± 0.82 μg of DNA per g of compost was obtained for Terra Fértil and Lipor, respectively. Furthermore, the 260/280 absorbance ratio was 1.80 ± 0.02 for Terra Fértil and 1.91 ± 0.01 for Lipor. Regarding the 260/230 absorbance ratio, 1.6 ± 0.03 and 2.0 ± 0.01 were obtained for Terra Fértil and Lipor, respectively. All these parameters indicate DNA with potential for further use in metagenomic studies (Table 3). However, the absorbance ratios were clearly affected by the addition of commercial humic acids (10 mg) to the Terra Fértil sample (Table 3).

The size and fragmentation of the isolated metagenomic DNA were determined by agarose gel electrophoresis (Fig. 3). A poor fragmentation of DNA was observed

Table 2 Comparison of the main experimental conditions and results obtained for some preliminary designed protocols for DNA extraction from compost samples and the optimized protocol

Protocol	Lysis buffer	Removal agent for humic substances	Mechanical lysis	Duration time	DNA
A	Tris-HCl; Na-EDTA; NaCl; SDS; lysozyme; proteinase K	-	-	24 h	No DNA
B	Tris-HCl; Na-EDTA; NaCl; SDS; lysozyme; proteinase K	PVPP+CTAB	-	24 h	Very low amount of genomic DNA
C	Tris-HCl; Na-EDTA; NaCl; Na ₂ HPO ₄ ; SDS; lysozyme; proteinase K	CTAB + CaCl ₂	-	3 days	Highly fragmented DNA with a discrete band ~ 40 kb
Optimized	Tris-HCl; Na-EDTA; NaCl; Na ₂ HPO ₄ ; SDS; lysozyme; proteinase K; Rnase A	PAC + CaCl ₂	Glass beads	4 h	> 40 kb

Fig. 1 Visual aspect of the DNA solutions after extraction with the protocols A, B, C and optimized method as described in Table 2

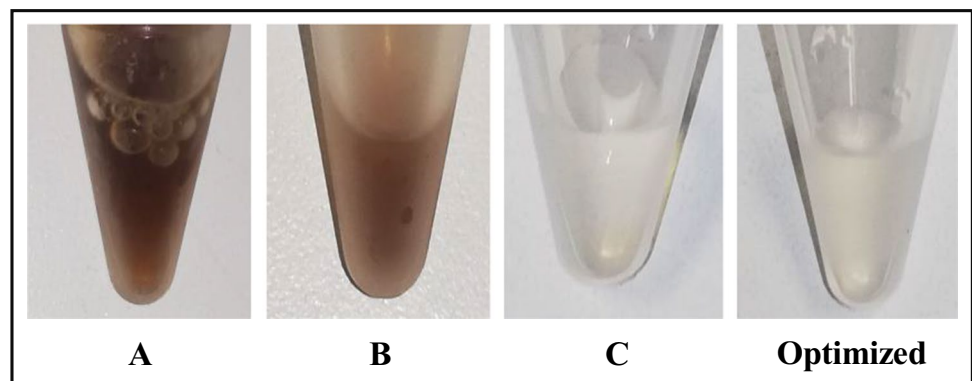


Fig. 2 Overview of the improved DNA extraction method comprising 3 main steps. PAC, powdered activated charcoal; SDS, sodium dodecyl sulphate; C:I, chloroform:isopropanol; PEG, polyethylene glycol (MW-8000); TE, Tris–EDTA buffer

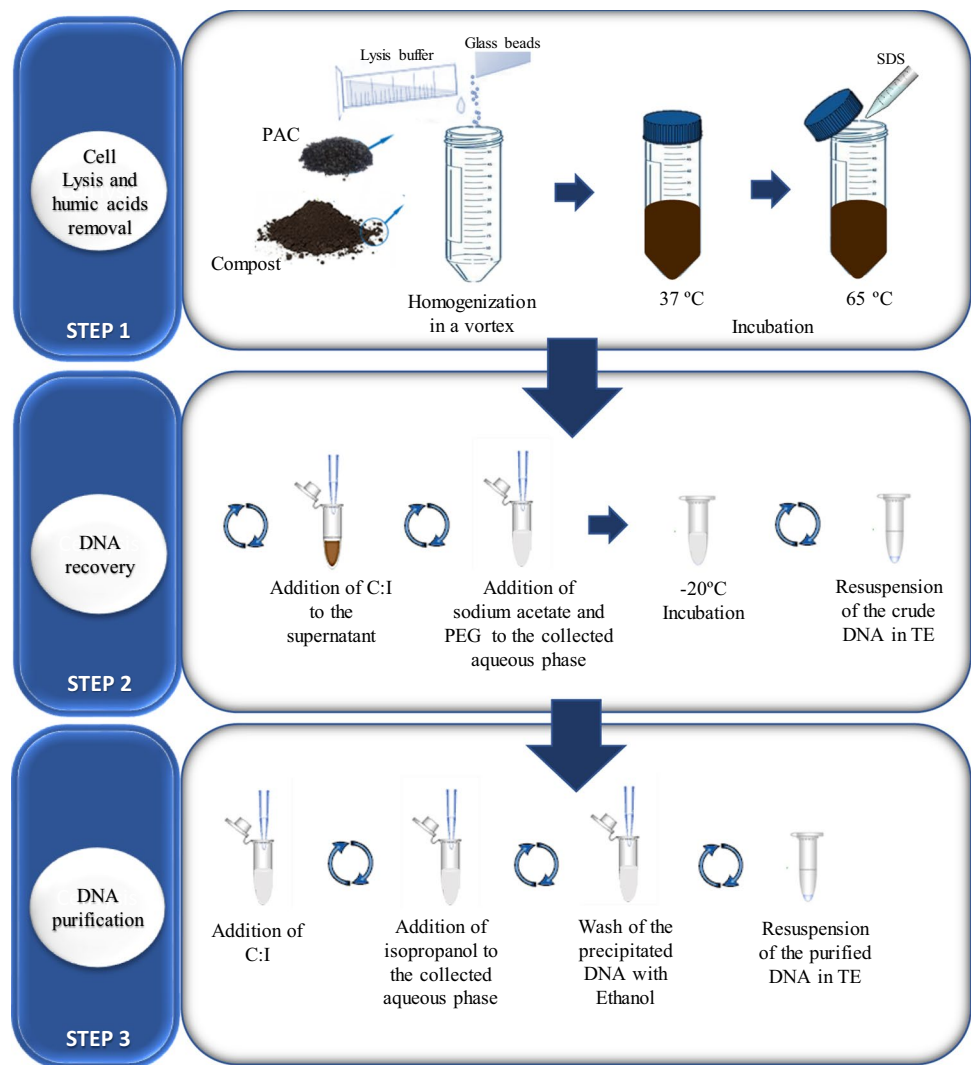


Table 3 Yield and purity of the isolated metagenomic DNA

Sample	Total DNA recovery ($\mu\text{g g}^{-1}$ of compost)	$A_{260/280}$	$A_{260/230}$	Total humic acid recovery ($\mu\text{g g}^{-1}$ of compost)	Humic acid removal (%)
Terra Fértil ^a	10.5 ± 0.22	1.80 ± 0.02	1.6 ± 0.03	3.01 ± 0.01	99.97 ± 0.020
Terra Fértil + 10 mg humic acids ^a	5.6 ± 0.18	1.64 ± 0.01	0.9 ± 0.01	3.73 ± 0.57	99.96 ± 0.005
Lipor 0 ^b	2.5 ± 0.03	1.84 ± 0.02	1.6 ± 0.02	1.54 ± 0.09	99.99 ± 0.001
Lipor 2.5 ^c	13.8 ± 0.04	1.85 ± 0.01	1.8 ± 0.01	2.38 ± 0.38	99.98 ± 0.004
Lipor ^a	23.9 ± 0.82	1.91 ± 0.01	2.0 ± 0.01	2.50 ± 0.45	99.98 ± 0.006

^aFive-minute agitation in the vortex for mechanical lysis; ^bbriefly agitation in the vortex (no mechanical lysis); ^c2.5-min agitation in the vortex for mechanical lysis

for both samples from Terra Fértil (lanes 1 and 2). However, the DNA extracted from the Lipor sample with 5 min of agitation in the vortex for mechanical lysis was highly fragmented (lane 5). When extracting DNA from

the Lipor sample without mechanical lysis (lane 3), only $2.5 \pm 0.03 \mu\text{g g}^{-1}$ of compost was recovered (Table 3). Nevertheless, when reducing the mechanical lysis time from 5 to 2.5 min, similar yields to the ones of Terra Fértil were

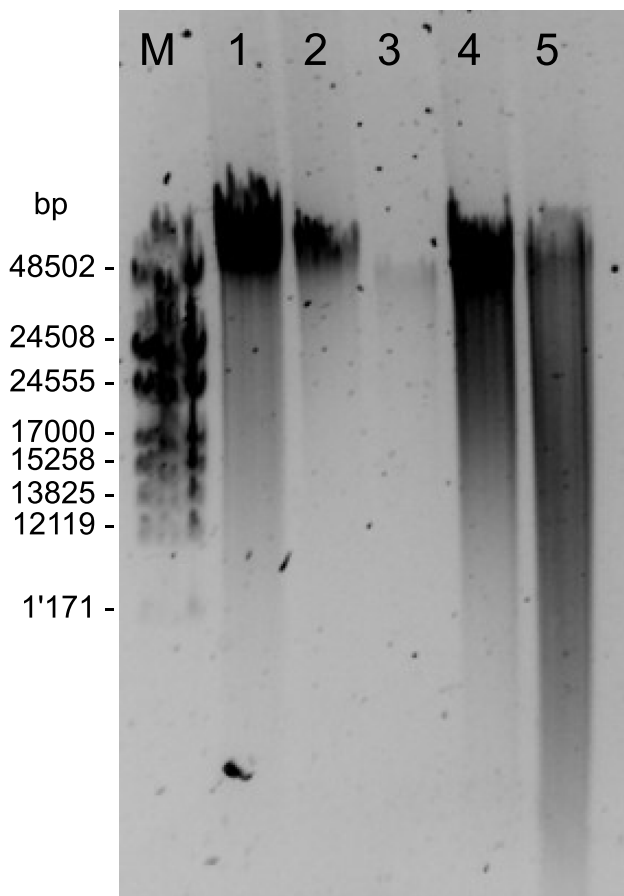


Fig. 3 Molecular weight characterization of the extracted DNA by gel electrophoresis. DNA was electrophoresed on 0.5% agarose gel, 4 °C for 24 h. Lane M—GeneRuler High Range DNA ladder; lane 1—Terra Fétil sample with 5-min agitation in the vortex for mechanical lysis; lane 2—Terra Fétil sample + 10 mg of humic acids with 5-min agitation in the vortex for mechanical lysis; lane 3—Lipor sample briefly mixed in the vortex (no mechanical lysis); lane 4—Lipor sample with 2.5-min agitation in the vortex for mechanical lysis; lane 5—Lipor sample with 5-min agitation in the vortex for mechanical lysis

obtained ($13.8 \pm 0.04 \mu\text{g g}^{-1}$ of compost) and the DNA fragmentation was visibly reduced (lane 4). With the addition of 10 mg of humic acids to the Terra Fétil sample, the amount of extracted DNA was reduced by half ($5.6 \pm 0.18 \mu\text{g g}^{-1}$ of compost). Additionally, for all the tested conditions, the humic acids were totally removed ($> 99.9\%$) using this optimized methodology (Table 3).

Validation of metagenomic DNA in downstream applications

The purity of the isolated metagenomic DNA and its suitability for downstream applications was validated by amplification of the conserved domain 16S rRNA and through the construction of a metagenomic library. For all the replicates studied, a product of 1500 bp was successfully amplified and no PCR product was observed in the negative control (Fig. 4).

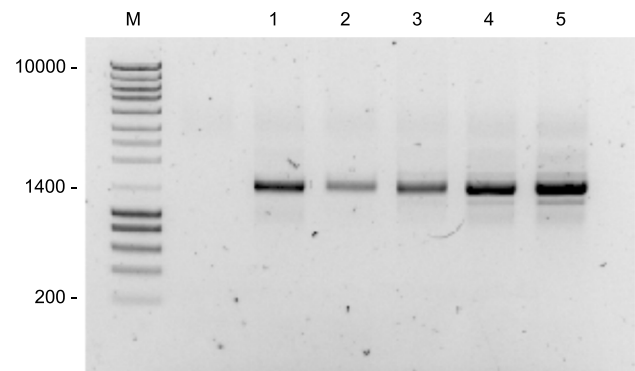


Fig. 4 PCR efficacy of the isolated metagenomic DNA by amplification of the 16S rRNA gene. Lane M—NZYDNA Ladder III. Lane 1—negative control (no template); lane 2—Terra Fétil sample with 5-min agitation in the vortex for mechanical lysis; lane 3—Terra Fétil sample + 10 mg of humic acids with 5-min agitation in the vortex for mechanical lysis; lane 4—Lipor sample mixed briefly in the vortex (no mechanical lysis); lane 5—Lipor sample with 2.5-min agitation in the vortex for mechanical lysis; lane 6—Lipor sample with 5-min agitation in the vortex for mechanical lysis

The metagenomic library was successfully constructed with approximately 6.5×10^{-3} colony-forming unit (cfu) ml^{-1} for both the Terra Fétil and Lipor samples. One randomly selected clone was purified and verified. The fosmid insert was submitted to restriction digestion using the enzyme *NotI*, which cut twice the pCC1FOSTM vector releasing the insert, and the enzyme *HindIII* that cuts once the pCC1FOSTM vector. The restriction pattern obtained is shown in Fig. S1 (supplementary material). The insert size presented about 40 kb, i.e. the ideal size for the construction of fosmid libraries.

Taxonomic annotation

The results of the taxonomic annotation at the kingdom level indicated the dominance of bacterial community for both samples. After analyzing the abundance at the phylum level within the bacterial community, it is possible to observe that the compost sample from Lipor presented the following composition: *Proteobacteria* > *Bacteroidetes* > *Actinobacteria* > *Firmicutes* (Fig. 5). The compost sample from Terra Fétil presented a similar composition, but an additional phylum was identified, namely the *Balneolaeota* (Fig. 5). For both samples, the genes annotated to other phyla with low abundance ($< 1\%$) and those that were not assigned in the NR database were included in the group Others.

Discussion

DNA isolation represents the basic and probably the most important step in molecular biology. The technology for DNA isolation from highly contaminated composting

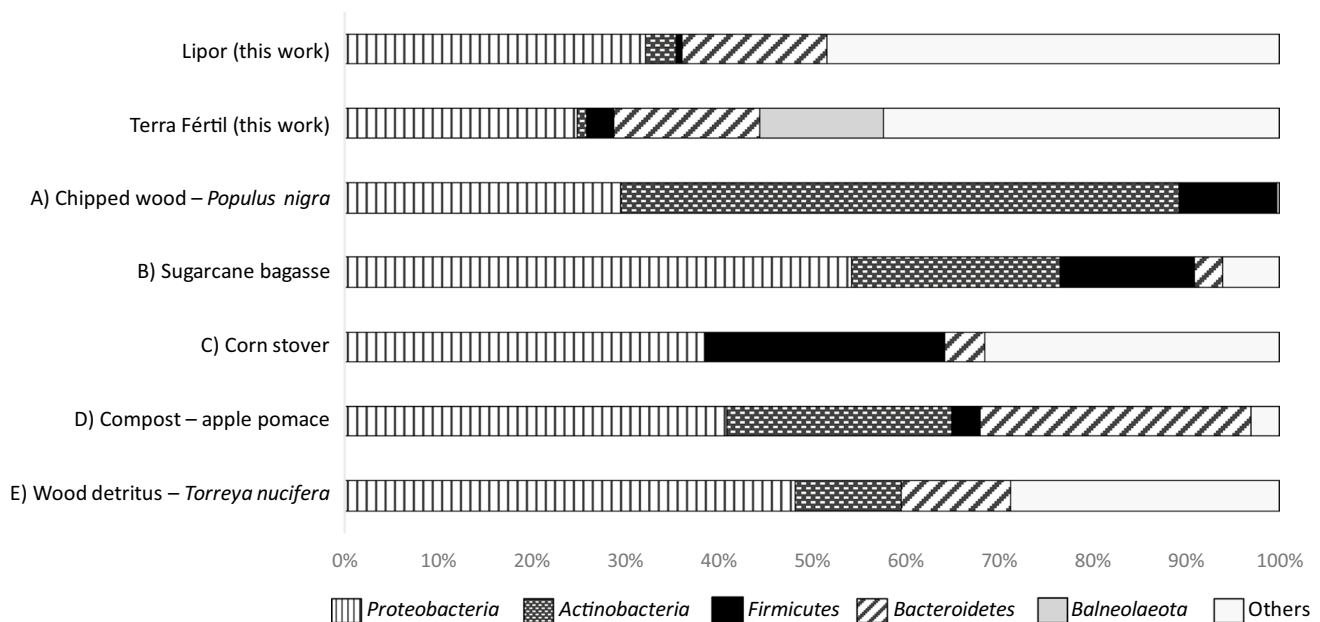


Fig. 5 Microbial composition (at the phylum level) obtained for the Terra Fétil and Lipor samples and also for similar samples containing (A) chipped wood (Montella et al. 2017), (B) sugarcane bagasse

(Soares et al. 2018), (C) corn stover (Zhu et al. 2016), (D) apple pomace (Zhou et al. 2017) or (E) wood detritus (Oh et al. 2017)

or soil-like samples is still currently a limiting factor in the meta-omic analysis. For the effective construction of metagenomic libraries, it is important to ensure that enough amount of highly pure and HMW DNA is isolated. Extracting DNA from soil-like samples requires the previous understanding of the remarkable complexity of the original samples and the identification of the multiple factors that may affect the performance of the extraction method (Miller et al. 1999). Commercially available DNA extraction kits have significant limitations in recovering high amounts of DNA with HMW since they have not been designed for this purpose. Additionally, the commercial kits are usually expensive and the UltraClean™ Soil DNA Isolation Kit, amply considered as one of the most effective for metagenomic DNA extraction from soil-like samples (Sharma et al. 2014), was discontinued. In general, the direct lysis method is more used than the indirect methods based on preliminary cell extraction. These indirect methods specifically target prokaryote DNA almost excluding eukaryotic DNA (Gabor et al. 2003); thus, the resulting DNA is not totally representative of the global environment. The major disadvantage of the direct lysis method is the potential co-extraction of humic acids, which can significantly compromise the success of soil metagenomic projects (Rajendhran and Gunasekaran 2008).

An environment where lignocellulosic biomass is naturally degraded, such as composting units, represents an abundant source of novel enzymes and an interesting sample for functional metagenomic studies (Gudiña et al. 2020).

In this work, two composting samples rich in lignocellulosic material and presenting a high content of humic acids were used to develop and optimize an isolation procedure for metagenomic DNA. The samples were obtained from two Portuguese compost units presenting around 10 mg g⁻¹ compost of humic acids, as illustrated in Table 1. Sar and co-workers (2018) studied the DNA extraction from other lignocellulosic samples, using indirect methods, and found a similar humic acid content in salt lake wood (9.7 mg g⁻¹). However, for other samples such as forest soil or rice straw compost, a lower humic acid content (5.4 and 6.5 mg g⁻¹, respectively) was reported (Sar et al. 2018). Additionally, higher levels of humic acids were found on forest arenosols (27.32 mg g⁻¹) and meadow histosol (34.66 mg g⁻¹) (Wnuk et al. 2020). Variations in the humic acid content are strongly dependent on the origin and composition of the sample. Since the compost samples analysed in this work are from different origins, they also present different elemental contents and physicochemical properties (Table 1). The Terra Fétil sample was collected at the thermophilic phase of the process (around 62.7 °C) and a C:N ratio of 11.5 was observed. The C:N ratio is considered typical for this phase of the decomposition due to some losses of carbon, mainly as carbon dioxide, and the increase of N content per unit of material (Goyal et al. 2005). The thermophilic phase was selected since it is the phase where it is expected to find more robust enzymes potentially involved in the degradation process of lignocellulose, such as cellulases, hemicellulases or xylanases (Berini et al. 2017; Wang et al. 2019).

The Lipor sample was collected at the mesophilic phase of the process and, as expected, a higher percentage of C was obtained since the composting process was initiating (Bernal et al. 1998). Nevertheless, the visible heterogeneity and variable particle size of the Lipor sample resulted in higher standard deviations associated with its elemental composition. These two compost samples were selected to cover the main steps (mesophilic and thermophilic phases) of the composting process.

Some conventional methodologies described in the literature (Zhou et al. 1996; Yeates et al. 1998; Pang et al. 2008) were initially followed to extract high-quality DNA from the compost samples. Nevertheless, using those protocols, the extracted DNA presented low quality, quantity and molecular weight (< 20 kb, highly fragmented, data not shown). Thus, some improvements were considered in the development of our methodology for DNA extraction from compost samples. Several modifications in crucial steps, such as the use or not of mechanical lysis, temperature and incubation time, buffer composition and addition of reagents and enzymes, were established and evaluated based on previous reports from Bergmann et al. (2014), Devi et al. (2015) and Verma et al. (2017). For the majority of the preliminary tested conditions, unsatisfactory results were obtained regarding reproducibility, DNA quality or molecular weight. The high content of organic carbon and humic acids in the compost samples may explain these unsatisfactory results. The three preliminary tested protocols presented in Table 2 allow comparing the main conditions and the obtained results. In protocol A, neither removal agent for humic acids nor mechanical lysis was used, and the lysis buffer was prepared according to Pang et al. (2008). Using these conditions, a brown solution with no genomic DNA was obtained (Fig. 1). The presence of contaminants, such as humic acids, can explain the visual appearance of the solution. To remove these contaminants, cetrimonium bromide (CTAB) and polyvinylpyrrolidone (PVPP), already described in the literature as effective removal agents (Zhou et al. 1996; Rajendhran and Gunasekaran 2008), were added to the lysis buffer used in protocol B, but again, a brown solution (Fig. 1) was obtained with a very low amount of genomic DNA. In protocol C, Na_2HPO_4 (used to keep DNA integrity) and CaCl_2 (flocculation agent) were added to the lysis buffer (Rajendhran and Gunasekaran 2008; Verma et al. 2017). With this protocol, great improvements were achieved and a highly clear solution was observed at the end of the DNA extraction procedure (Fig. 1). Nevertheless, the DNA was extremely fragmented. In addition, this procedure (protocol C) was extremely time consuming (approximately 3 days mainly due to the slow resuspension of DNA performed at 4 °C to avoid fragmentation) and exhibited poor reproducibility. All these previous attempts allowed us to acquire the necessary expertise to design the current

optimized methodology (Table 2) which resulted in an unusual but effective sequence of steps. In this case, a highly clear solution of metagenomic DNA (Fig. 1) with > 40 kb was obtained (Fig. 3).

The methodology herein proposed for DNA extraction is a three-step adjusted procedure (Fig. 2). The crucial approach of our methodology relied on the optimization of both the cell lysis method and buffer, and the addition of PAC, which effectively removed more than 99% of the humic acids (Table 3). In the first step, which corresponds to cell lysis, mechanical and chemical/enzymatic techniques were carefully combined using an optimized lysis buffer. This buffer resulted from the suitable combination of chemicals and enzymatic reagents, namely Na_2HPO_4 , CaCl_2 , proteinase K, lysozyme and RNase A, in addition to those compounds traditionally used in lysis buffers. The second step comprises the DNA recovery by precipitation after the elimination through centrifugation of the humic acids and other contaminants which can adsorb on PAC and/or flocculate with CaCl_2 . The use of PEG to precipitate the DNA also helps to increase its purity since it does not co-precipitate humic acids with the DNA as occurs with isopropanol or ethanol precipitation (LaMontagne et al. 2002). The third and final step corresponds to the purification of the crude DNA. This latter step quantitatively concentrates the DNA and also helps to eliminate additional contaminants that remain in the solution while DNA is selectively precipitated.

PAC is not the most frequently used removal agent for soluble humic components and other impurities in DNA extraction. However, it already proved its effectiveness in some previous works (Baker et al. 1992; Sharma et al. 2014; Devi et al. 2015), possibly due to its unique adsorption capacity. Other agents more often used to remove humic acids like CTAB and PVPP were also tested in this work. Although DNA with high purity was also extracted with these agents, much fewer amounts of DNA were obtained (not enough for further metagenomic studies) meaning that the DNA was probably lost during the extraction process.

The yield is an important factor in a DNA extraction process since DNA isolation from soil-like samples is often incomplete, suggesting that some DNA is inaccessible and therefore cannot be accounted for subsequent analysis. Recently, Sar et al. (2018) published an optimized protocol for soils contaminated with humic acids based on indirect methods of DNA extraction. However, using indirect methods, only bacterium was selected, thus excluding the other microbial diversity present in the soil. Therefore, the authors reported much lower DNA yields and purity ($0.33 \pm 0.04 \mu\text{g g}^{-1}$ with $A_{260/280}$ 1.45 ± 0.02 for salt lake wood; $0.55 \pm 0.05 \mu\text{g g}^{-1}$ with $A_{260/280}$ 1.46 ± 0.04 for forest soil; and $0.81 \pm 0.04 \mu\text{g g}^{-1}$ with $A_{260/280}$ 1.74 ± 0.02 for rice straw compost) when compared to those obtained with our optimized protocol (Table 3). Also, the fragmentation size

is a key aspect since small fragments (less than 30 kb) are not suitable either for the analysis of complete genomes/metagenomes or to study and explore the complete or virtually complete metabolic pathways. High DNA yields were generally obtained with the use of severe mechanical treatments, such as sonication and mechanical bead beating, to effectively lyse microbial cells. However, such treatments can significantly shear DNA to sizes of 5–10 kb or less, being frequently reported as not suitable to extract DNA for subsequent metagenomic approaches (Liesack et al. 1991). Nevertheless, some mechanical treatments have been used and as can be noted, especially in the Terra Fértil sample, a poor fragmentation of the DNA can be obtained (Fig. 3, lanes 1 and 2) when suitable conditions are applied. This result suggests that the mechanical lysis conditions used in our optimized extraction protocol (glass beads and 5 min vortex) did not cause severe DNA shearing in the Terra Fértil sample. However, depending on the origin and composition of the sample, some DNA fragmentation can occur, as observed for the Lipor sample (Fig. 3, lane 5). This compost sample exhibited high heterogeneity including the presence of small branches and pieces of wood which may act themselves as a kind of “beads” in the mechanical lysis, thus promoting a higher fragmentation of the DNA. Nevertheless, when the agitation time of the mechanical lysis was reduced to half (2.5 min), a significant improvement was obtained due to the decrease of DNA fragmentation (Fig. 3, lane 4). This evidence confirms the importance of a skilled evaluation of the natural characteristics of the compost samples (e.g. moisture, microbial composition, humic acid concentration or heterogeneity/homogeneity of the sample) to strategically modify and adapt crucial steps of the extraction methodology to each type of sample.

To evaluate the purity of the extracted DNA, the ratios of absorbance at 260/280 nm (Table 3) and also the absorbance at 340 nm were determined (to indirectly assess the humic acid content). The results indicated that the extracted DNA presented adequate purity for metagenomic studies. The presence of humic acids greatly impacts the success of the DNA extraction as observed when commercial humic acids were added to the Terra Fértil sample. This addition affected the absorbance ratios (Table 3), but not the DNA size and quality (Figs. 3 and 4). The final concentration of DNA was also affected when commercial humic acids were added, even if the protocol proved to be efficient on the removal of the humic acids (> 99%). The detrimental effect observed for a higher concentration of humic acids may have resulted from the ability of these compounds to establish interactions with both DNA and enzymes. Therefore, some DNA may have been lost after binding to the humic acids adsorbed on the PAC (Saeki et al. 2011; Wnuk et al. 2020) and/or inefficient enzymatic lysis could occur due to the formation of enzyme-humic

acid complexes (Li et al. 2013). However, despite the lower concentration of DNA, a strong and clean band with more than 48 kb was noted in the electrophoresis gel (Fig. 3, lane 2) and the DNA could be further used for downstream applications (Fig. 4). This means that this protocol has a wide range of applicability including using samples with different humic acid concentration.

The purity of the extracted DNA is an important aspect for further applications, since the presence of impurities, especially the co-extracted humic acids, strongly inhibit *Taq* polymerases (Tsai and Olson 1992). Therefore, the purity and quality of the metagenomic DNA from both compost samples were evaluated by PCR amplification of the bacterial 16S rRNA using universal primers. A fragment of 1400 bp of bacteria was successfully amplified for all the tested conditions (Fig. 4). Additionally, the purity and suitability of the extracted metagenomic DNA were investigated through the construction of fosmid libraries. The libraries were constructed for both Terra Fértil and Lipor (2.5 min of mechanical lysis) samples, and a fragment with approximately 40 kb was successfully inserted (Supplementary Fig. S1). The preparation of metagenomic libraries with large inserts will increase the chances of discovering novel biocatalysts and biosynthetic pathways, also leading to the discovery of new genes.

The microbial composition obtained for the isolated DNA is strongly dependent on both the sample origin/composition and the efficiency of the lysis/extraction methodologies. To evaluate the potential of the optimized protocol to access DNA considered as representative of the microbial communities commonly present in soil-like samples, a shotgun sequencing analysis was performed. After taxonomic annotation, it was possible to obtain the most dominant and assigned phyla. A comparable microbial composition was found for Terra Fértil and Lipor samples (Fig. 5). In both cases, the dominant phylum was *Proteobacteria* followed by *Bacteroidetes*. The phyla *Actinobacteria* and *Firmicutes* are also present but in a lower percentage. For the Terra Fértil sample, an additional phylum was detected, namely the recently proposed phylum *Balneolaeta* (Hahnke et al. 2016) identified in extreme environments (Sorokin et al. 2018). Furthermore, the microbial compositions obtained for Terra Fértil and Lipor samples are also comparable with those reported in the literature for similar samples containing chipped wood (Montella et al. 2017), sugarcane bagasse (Soares et al. 2018), corn stover (Zhu et al. 2016), apple pomace (Zhou et al. 2017) or wood detritus (Oh et al. 2017) (Fig. 5). Although the detected differences in terms of abundance, the dominant assigned phyla are the same which indicates that the optimized protocol, the classical methodologies (Zhou et al. 2017; Soares et al. 2018) and the commercial kits (Zhu et al. 2016; Montella et al. 2017;

Oh et al. 2017) have equivalent performance when accessing DNA from the microbial communities.

Another important feature to be highlighted in our optimized method of DNA extraction is the considerable time-consuming improvement. This method proved to be highly efficient and fast (only 4 h of bench work), not including additional and expensive steps of purification. All the procedure can be performed using commonly standard molecular biology equipment and reagents, thus reducing the associated costs. Furthermore, it was demonstrated that under suitable conditions of mechanical lysis, the DNA fragmentation can be minimized and the recovery yield increased.

In conclusion, the humic acids and phenolic compounds naturally present in soils are hard to remove, making DNA extraction and purification a critical step in the success of several metagenomic approaches. Our optimized methodology for DNA extraction was developed to provide a rapid, efficient and cost-effective procedure for total DNA extraction resulting in high purity, yield and molecular weight DNA. This novel protocol can be adapted and successfully applied to different compost samples rich in humic acids to obtain DNA representative of the microbial compositions and suitable for functional metagenomic studies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00253-021-11647-7>.

Acknowledgements AC and JS acknowledge her research grants UMINHO/BPD/37/2018 and UMINHOBIM/2020/28, respectively, under the scope of the project LIGNOZYMES (POCI-01-0145-FEDER-029773). The authors also acknowledge the Portuguese composting units Terra Fértil and Lipor for kindly supplying the compost samples.

Author contribution AC, AS and JS conducted the experiments. SS and EG conceived and designed research. JR, SS and LR analysed data. SS and LR obtained financial support and coordinated the research. AC wrote the manuscript. All authors read and approved the manuscript.

Funding This study was supported by the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UIDB/04469/2020 unit and the Project LIGNOZYMES—Metagenomics approach to unravel the potential of lignocellulosic residues towards the discovery of novel enzymes (POCI-01-0145-FEDER-029773).

Availability of data and material The authors confirm that the datasets supporting the findings and conclusions of this study are available within the article and its supplementary information file.

Code availability Not applicable.

Declarations

Ethics approval The article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

- Ali N, Zhang Q, Liu Z-Y, Li F-L, Lu M, Fang X-C (2020) Emerging technologies for the pretreatment of lignocellulosic materials for bio-based products. *Appl Microbiol Biotechnol* 104:455–473. <https://doi.org/10.1007/s00253-019-10158-w>
- Arevalo-Gallegos A, Ahmad Z, Asgher M, Parra-Saldivar R, Iqbal HMN (2017) Lignocellulose: a sustainable material to produce value-added products with a zero waste approach—a review. *Int J Biol Macromol* 99:308–318. <https://doi.org/10.1016/j.ijbiomac.2017.02.097>
- Baker FS, Miller CE, Repik AJ, Tolles ED (1992) Activated carbon. *Kirk-Othmer Encycl Chem Technol* 4:1015–1037. <https://doi.org/10.1002/0471238961.0103200902011105.a01>
- Batista-García RA, del Rayo Sánchez-Carbente M, Talia P, Jackson SA, O'Leary ND, Dobson ADW, Folch-Mallol JL (2016) From lignocellulosic metagenomes to lignocellulolytic genes: trends, challenges and future prospects. *Biofuels*, *Bioprod Biorefining* 10:864–882. <https://doi.org/10.1002/bbb.1709>
- Bergmann JC, Costa OYA, Gladden JM, Singer S, Heins R, D'haeseleer P, Simmons BA, Quirino BF (2014) Discovery of two novel β -glucosidases from an Amazon soil metagenomic library. *FEMS Microbiol Lett* 351:147–155. <https://doi.org/10.1111/1574-6968.12332>
- Berini F, Casciello C, Marcone GL, Marinelli F (2017) Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiol Lett* 364:fxn211. <https://doi.org/10.1093/femsle/fnx211>
- Bernal MP, Sánchez-Monedero MA, Paredes C, Roig A (1998) Carbon mineralization from organic wastes at different stages during their incubation with soil. *Agric Ecosyst Environ* 69:175–189. [https://doi.org/10.1016/S0167-8809\(98\)00106-6](https://doi.org/10.1016/S0167-8809(98)00106-6)
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>
- Devi SG, Fathima AA, Radha S, Arunraj R, Curtis WR, Ramya M (2015) A rapid and economical method for efficient DNA extraction from diverse soils suitable for metagenomic applications. *PLoS ONE* 10:e0132441. <https://doi.org/10.1371/journal.pone.0132441>
- Dong D, Yan A, Liu H, Zhang X, Xu Y (2006) Removal of humic substances from soil DNA using aluminium sulfate. *J Microbiol Methods* 66:217–222. <https://doi.org/10.1016/j.mimet.2005.11.010>
- Gabor EM, De Vries EJ, Janssen DB (2003) Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiol Ecol* 44:153–163. [https://doi.org/10.1016/S0168-6496\(02\)00462-2](https://doi.org/10.1016/S0168-6496(02)00462-2)
- Goyal S, Dhull SK, Kapoor KK (2005) Chemical and biological changes during composting of different organic wastes and assessment of compost maturity. *Bioresour Technol* 96:1584–1591. <https://doi.org/10.1016/j.biortech.2004.12.012>
- Gudiña EJ, Amorim C, Braga A, Costa Â, Rodrigues JL, Silvério S, Rodrigues LR (2020) Biotech green approaches to unravel the potential of residues into valuable products. In: Inamuddin M, Asiri AM (eds) Sustainable green chemical processes and their allied applications, 1st edn. Springer, Cham, pp 97–150
- Guerrero G, Hausman J-F, Strauss J, Ertan H, Siddiqui KS (2016) Lignocellulosic biomass: biosynthesis, degradation, and industrial

- utilization. *Eng Life Sci* 16:1–16. <https://doi.org/10.1002/elsc.201400196>
- Hahnke RL, Meier-Kolthoff JP, García-López M, Mukherjee S, Hunte-mann M, Ivanova NN, Woyke T, Kyrpidis NC, Klenk HP, Göker M (2016) Genome-based taxonomic classification of *Bacteroidetes*. *Front Microbiol* 7:2003. <https://doi.org/10.3389/fmicb.2016.02003>
- Haldar D, Purkait MK (2020) Lignocellulosic conversion into value-added products: a review. *Process Biochem* 89:110–133. <https://doi.org/10.1016/j.procbio.2019.10.001>
- Harry M, Gambier B, Bourezgui Y, Garnier-Sillam E (1999) Evaluation of purification procedures for DNA extracted from organic rich samples: interference with humic substances. *Analisis* 27:439–442. <https://doi.org/10.1051/analisis:1999270439>
- Isikgor FH, Becer CR (2015) Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polym Chem* 6:4497–4559. <https://doi.org/10.1039/c5py00263j>
- Kalra YP (1995) Determination of pH of soils by different methods: collaborative study. *J AOAC Int* 78:310–324. <https://doi.org/10.1093/jaoac/78.2.310>
- Lakay FM, Botha A, Prior BA (2007) Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J Appl Microbiol* 102:265–273. <https://doi.org/10.1111/j.1365-2672.2006.03052.x>
- LaMontagne MG, Michel FC Jr, Holden PA, Reddy CA (2002) Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. *J Microbiol Methods* 49:255–264. [https://doi.org/10.1016/S0167-7012\(01\)00377-3](https://doi.org/10.1016/S0167-7012(01)00377-3)
- Lear G, Dickie I, Banks J, Boyer S, Buckley HL, Buckley TR, Cruickshank R, Dopheide A, Handley KM, Hermans S, Kamke J, Lee CK, MacDiarmid R, Morales SE, Orlovich DA, Smissen R, Wood J, Holdaway R (2018) Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *N Z J Ecol* 42:10–50A. <https://doi.org/10.20417/nzjecol.42.9>
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20:265–272. <https://doi.org/10.1101/gr.097261.109>
- Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Li Y, Tan W, Koopal LK, Wang M, Liu F, Norde W (2013) Influence of soil humic and fulvic acid on the activity and stability of lysozyme and urease. *Environ Sci Technol* 47:5050–5056. <https://doi.org/10.1021/es3053027>
- Liesack W, Weyland H, Stackebrandt E (1991) Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. *Microb Ecol* 21:191–198. <https://doi.org/10.1007/BF02539153>
- Madhavan A, Sindhu R, Parameswaran B, Sukumaran RK, Pandey A (2017) Metagenome analysis : a powerful tool for enzyme bio-prospecting. *Appl Biochem Biotechnol* 183:636–651. <https://doi.org/10.1007/s12010-017-2568-3>
- Marsh H, Rodríguez-Reinoso F (2006) Activated carbons (origins). Activated carbon, 1st edn. Elsevier Science Ltd, Oxford, pp 13–86
- Miller DN, Bryant JE, Madsen EL, Ghiorse WC (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* 65:4715–4724. <https://doi.org/10.1128/aem.65.11.4715-4724.1999>
- Mirete S, Morgante V, González-Pastor JE (2016) Functional metagenomics of extreme environments. *Curr Opin Biotechnol* 38:143–149. <https://doi.org/10.1016/j.copbio.2016.01.017>
- Montella S, Amore A, Faraco V (2016) Metagenomics for the development of new biocatalysts to advance lignocellulose saccharification for bioeconomic development. *Crit Rev Biotechnol* 36:998–1009. <https://doi.org/10.3109/07388551.2015.1083939>
- Montella S, Ventrino V, Lombard V, Henrissat B, Pepe O, Faraco V (2017) Discovery of genes coding for carbohydrate-active enzyme by metagenomic analysis of lignocellulosic biomasses. *Sci Rep* 7:42623. <https://doi.org/10.1038/srep42623>
- Ning P, Yang G, Hu L, Sun J, Shi L, Zhou Y, Wang Z, Yang J (2021) Recent advances in the valorization of plant biomass. *Biotechnol Biofuels* 14:102. <https://doi.org/10.1186/s13068-021-01949-3>
- Oh HN, Lee TK, Park JW, No JH, Kim D, Sul WJ (2017) Metagenomic SMRT sequencing-based exploration of novel lignocellulose-degrading capability in wood detritus from *Torreya nucifera* in Bija Forest on Jeju Island. *J Microbiol Biotechnol* 27:1670–1680. <https://doi.org/10.4014/jmb.1705.05008>
- Pang M-F, Abdullah N, Lee C-W, Ng C-C (2008) Isolation of high molecular weight DNA from forest topsoil for metagenomic analysis. *Asia Pacific J Mol Biol Biotechnol* 16:35–41
- Patel A, Shah AR (2021) Integrated lignocellulosic biorefinery: gateway for production of second generation ethanol and value added products. *J Bioresour Bioprod* 6:108–128. <https://doi.org/10.1016/j.jobab.2021.02.001>
- Petric I, Philippot L, Abbate C, Bispo A, Chesnot T, Hallin S, Laval K, Lebeau T, Lemanceau P, Leyval C, Lindström K, Pandard P, Romero E, Sarr A, Schloter M, Simonet P, Smalla K, Wilke B-M, Martin-Laurent F (2011) Inter-laboratory evaluation of the ISO standard 11063 “soil quality - method to directly extract DNA from soil samples.” *J Microbiol Methods* 84:454–460. <https://doi.org/10.1016/j.mimet.2011.01.016>
- Philippot L, Abbate C, Bispo A, Chesnot T, Hallin S, Lemanceau P, Lindström K, Pandard P, Romero E, Schloter M, Simonet P, Smalla K, Wilke B-M, Petric I, Martin-Laurent F (2010) Soil microbial diversity: an ISO standard for soil DNA extraction. *J Soils Sediments* 10:1344–1345. <https://doi.org/10.1007/s11368-010-0265-8>
- Rajendhran J, Gunasekaran P (2008) Strategies for accessing soil metagenome for desired applications. *Biotechnol Adv* 26:576–590. <https://doi.org/10.1016/j.biotechadv.2008.08.002>
- Saeki K, Ihyo Y, Sakai M, Kunito T (2011) Strong adsorption of DNA molecules on humic acids. *Environ Chem Lett* 9:505–509. <https://doi.org/10.1007/s10311-011-0310-x>
- Sar A, Pal S, Dam B (2018) Isolation of high molecular weight and humic acid-free metagenomic DNA from lignocellulose-rich samples compatible for direct fosmid cloning. *Appl Microbiol Biotechnol* 102:6207–6219. <https://doi.org/10.1007/s00253-018-9102-6>
- Schmeisser C, Steele H, Streit WR (2007) Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* 75:955–962. <https://doi.org/10.1007/s00253-007-0945-5>
- Sharma S, Sharma KK, Kuhad RC (2014) An efficient and economical method for extraction of DNA amenable to biotechnological manipulations, from diverse soils and sediments. *J Appl Microbiol* 116:923–933. <https://doi.org/10.1111/jam.12420>
- Sheldon RA, Woodley JM (2018) Role of biocatalysis in sustainable chemistry. *Chem Rev* 118:801–838. <https://doi.org/10.1021/acs.chemrev.7b00203>
- Soares LA, Borges Silva Rabelo CA, Sakamoto IK, Delforno TP, Silva EL, Varesche MBA (2018) Metagenomic analysis and optimization of hydrogen production from sugarcane bagasse. *Biomass Bioenerg* 117:78–85. <https://doi.org/10.1016/j.biombioe.2018.07.018>
- Sorokin DY, Muntyan MS, Toshchakov SV, Korzhenkov A, Kublanov IV (2018) Phenotypic and genomic properties of a novel deep-lineage haloalkaliphilic member of the phylum Balneolaeota from soda lakes possessing Na⁺-translocating proteorhodopsin. *Front Microbiol* 9:2672. <https://doi.org/10.3389/fmicb.2018.02672>

- Tanase A-M, Mereuta I, Chiciudean I, Ionescu R, Milea L, Cornea CP, Vassu T, Stoica I (2015) Comparison of total DNA extraction methods for microbial community from polluted soil. *Agric Agric Sci Procedia* 6:616–622. <https://doi.org/10.1016/j.aaspro.2015.08.102>
- Tirado-Acevedo O, Chinn MS, Grunden AM (2010) Production of biofuels from synthesis gas using microbial catalysts. *Adv Appl Microbiol* 70:57–92. [https://doi.org/10.1016/S0065-2164\(10\)70002-2](https://doi.org/10.1016/S0065-2164(10)70002-2)
- Tsai Y, Olson BH (1992) Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Appl Environ Microbiol* 58:2292–2295. <https://doi.org/10.1128/aem.58.7.2292-2295.1992>
- Verma SK, Singh H, Sharma PC (2017) An improved method suitable for isolation of high-quality metagenomic DNA from diverse soils. *3 Biotech* 7:171. <https://doi.org/10.1007/s13205-017-0847-x>
- Wang C, Dong D, Wang H, Müller K, Qin Y, Wang H, Wu W (2016) Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of *Actinobacteria* in lignocellulose decomposition. *Biotechnol Biofuels* 9:22. <https://doi.org/10.1186/s13068-016-0440-2>
- Wang H, Hart DJ, An Y (2019) Functional metagenomic technologies for the discovery of novel enzymes for biomass degradation and biofuel production. *BioEnergy Res* 12:457–470. <https://doi.org/10.1007/s12155-019-10005-w>
- Wnuk E, Waśko A, Walkiewicz A, Bartmiński P, Bejger R, Mielnik L, Bieganowski A (2020) The effects of humic substances on DNA isolation from soils. *PeerJ* 8:e9378. <https://doi.org/10.7717/peerj.9378>
- Wu L, Li F, Deng C, Xu D, Jiang S, Xiong Y (2009) A method for obtaining DNA from compost. *Appl Microbiol Biotechnol* 84:389–395. <https://doi.org/10.1007/s00253-009-2103-8>
- Yeates C, Gillings MR, Davison AD, Altavilla N, Veal DA (1998) Methods for microbial DNA extraction from soil for PCR amplification. *Biol Proced Online* 1:40–47. <https://doi.org/10.1251/bpo6>
- Young JM, Rawlence NJ, Weyrich LS, Cooper A (2014) Limitations and recommendations for successful DNA extraction from forensic soil samples: a review. *Sci Justice* 54:238–244. <https://doi.org/10.1016/j.scijus.2014.02.006>
- Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62:316–322. <https://doi.org/10.1128/aem.62.2.316-322.1996>
- Zhou M, Guo P, Wang T, Gao L, Yin H, Cai C, Gu J, Lü X (2017) Metagenomic mining pectinolytic microbes and enzymes from an apple pomace-adapted compost microbial community. *Biotechnol Biofuels* 10:198. <https://doi.org/10.1186/s13068-017-0885-y>
- Zhu N, Yang J, Ji L, Liu J, Yang Y, Yuan H (2016) Metagenomic and metaproteomic analyses of a corn stover-adapted microbial consortium EMSD5 reveal its taxonomic and enzymatic basis for degrading lignocellulose. *Biotechnol Biofuels* 9:243. <https://doi.org/10.1186/s13068-016-0658-z>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.