

# SCIENTIFIC REPORTS

**OPEN**

## Microbial community structure and dynamics in thermophilic composting viewed through metagenomics and metatranscriptomics

Received: 03 August 2016  
Accepted: 14 November 2016  
Published: 12 December 2016

Luciana Principal Antunes<sup>1</sup>, Layla Farage Martins<sup>1</sup>, Roberta Verciano Pereira<sup>1</sup>, Andrew Maltez Thomas<sup>1,2</sup>, Deibs Barbosa<sup>1,2</sup>, Leandro Nascimento Lemos<sup>1,2</sup>, Gianluca Major Machado Silva<sup>1,2</sup>, Livia Maria Silva Moura<sup>1,2</sup>, George Willian Condomitti Epamino<sup>1,2</sup>, Luciano Antonio Digiampietri<sup>3</sup>, Karen Cristina Lombardi<sup>1</sup>, Patricia Locosque Ramos<sup>4</sup>, Ronaldo Bento Quaggio<sup>1</sup>, Julio Cezar Franco de Oliveira<sup>5</sup>, Renata Castiglioni Pascon<sup>5</sup>, João Batista da Cruz<sup>4</sup>, Aline Maria da Silva<sup>1,2,\*</sup> & João Carlos Setubal<sup>1,2,6,\*</sup>

Composting is a promising source of new organisms and thermostable enzymes that may be helpful in environmental management and industrial processes. Here we present results of metagenomic- and metatranscriptomic-based analyses of a large composting operation in the São Paulo Zoo Park. This composting exhibits a sustained thermophilic profile (50 °C to 75 °C), which seems to preclude fungal activity. The main novelty of our study is the combination of time-series sampling with shotgun DNA, 16S rRNA gene amplicon, and metatranscriptome high-throughput sequencing, enabling an unprecedented detailed view of microbial community structure, dynamics, and function in this ecosystem. The time-series data showed that the turning procedure has a strong impact on the compost microbiota, restoring to a certain extent the population profile seen at the beginning of the process; and that lignocellulosic biomass deconstruction occurs synergistically and sequentially, with hemicellulose being degraded preferentially to cellulose and lignin. Moreover, our sequencing data allowed near-complete genome reconstruction of five bacterial species previously found in biomass-degrading environments and of a novel biodegrading bacterial species, likely a new genus in the order Bacillales. The data and analyses provided are a rich source for additional investigations of thermophilic composting microbiology.

Typical aerobic composting is a self-heating process in which microbial metabolism drives the temperature above 50 °C, followed by sustained high temperatures between 60–80 °C, and then followed by gradual cooling of the compost pile<sup>1,2</sup>. The biological decomposition of organic matter is performed by mesophilic and thermophilic microbial consortia with distinct physiological requirements and tolerances, consistent with the continuously changing environment throughout composting<sup>3–7</sup>. Bacterial phyla including Proteobacteria, Firmicutes, Bacteroidetes and Actinobacteria are routinely found in composting, being more or less abundant depending on the starting materials and the composting procedure<sup>1,4,7–12</sup>. Generally, fungi are not detected in composting piles

<sup>1</sup>Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil. <sup>2</sup>Programa de Pós-Graduação Interunidades em Bioinformática, Universidade de São Paulo, São Paulo, Brazil. <sup>3</sup>Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, Brazil. <sup>4</sup>Fundação Parque Zoológico de São Paulo, São Paulo, Brazil. <sup>5</sup>Departamento de Ciências Biológicas, Universidade Federal de São Paulo, São Paulo, Brazil. <sup>6</sup>Biocomplexity Institute of Virginia Tech, Blacksburg, VA, USA. \*These authors jointly supervised this work. Correspondence and requests for materials should be addressed to A.M.d.S. (email: [almsilva@iq.usp.br](mailto:almsilva@iq.usp.br)) or J.C.S. (email: [setubal@iq.usp.br](mailto:setubal@iq.usp.br))

Day of collection	Sample ID Collection Date	Average Temperature <sup>a</sup> (°C)	$\Delta$ Temperature <sup>b</sup> (°C)	pH	Moisture (%)
1	D01 08/06/2013	66.2 ± 7.5	18	6.0	71.0
3	D03 08/08/2013	68.4 ± 2.6	6	5.5	67.0
7	D07 08/12/2013	70.8 ± 3.5	9	6.4	70.3
15	D15 08/20/2013	68.2 ± 5.8	14	6.9	63.8
30	D30 09/04/2013	69.5 ± 2.3	5.6	6.9	61.8
64 <sup>*</sup>	D64 10/08/2013	57.9 ± 8.9	21.2	6.8	60.7
67	D67 10/11/2013	70.5 ± 4.8	10.7	7.4	n.d.
78	D78 10/22/2013	61.2 ± 3.2	8.0	7.8	58.6
99	D99 11/12/2013	47.8 ± 3.3	8.3	7.9	58.7

**Table 1. Description of samples collected from ZC4 composting cell.** <sup>a</sup>Mean plus standard deviation of temperature measurements taken at the four edges and in the center of the composting pile. <sup>b</sup>Variation between lower and higher temperature measurements. n.d., not determined. \*Turning was performed on day 63.

above 65 °C, suggesting that their degradative activities during the thermophilic stages of composting are minor compared to that of bacteria<sup>13,14</sup>. Therefore, the current understanding is that bacteria are the dominant degraders in thermophilic composting processes<sup>13,15,16</sup>, while fungi come into play during the cooling and curing phases<sup>17</sup>.

Composting is widely considered as a promising source of new thermophilic bacteria<sup>5,7,18</sup> and novel thermostable enzymes, particularly those related to biomass degradation, which have many advantages for industrial applications<sup>15,19–23</sup>. Despite intensive studies to investigate thermophilic composting ecosystems, using culture-dependent<sup>16,24</sup> and culture-independent approaches<sup>4,10,11,13,15,25</sup>, information regarding functional aspects of associated microbiota is still limited. In this regard, metagenomics and metatranscriptomics are valuable approaches to expand the repertoire of known biodegrading microorganisms and their active functional metabolic potential during thermophilic composting.

The goal of this study was to perform a comprehensive investigation of the structure, dynamics, and metabolic functions of the microbiota in a thermophilic composting operation at the São Paulo Zoo Park, which is located within the urban area of the São Paulo city (Brazil) and includes a remnant Atlantic rain forest patch. This composting facility was designed to compost all organic waste produced in the park. In a previous study<sup>13</sup>, we have shown that the São Paulo Zoo composting harbors considerable microbial diversity. Here we extend the previous work in three important ways: first, in addition to shotgun sequencing data, we obtained 16S rRNA amplicon data; second, we also obtained metatranscriptomics (RNA-seq) data; and third, these data were generated based on several time-series samples of the 99-day-long process. To our knowledge, this is the first study combining all three data types generated from time-series samples from a full-scale composting operation.

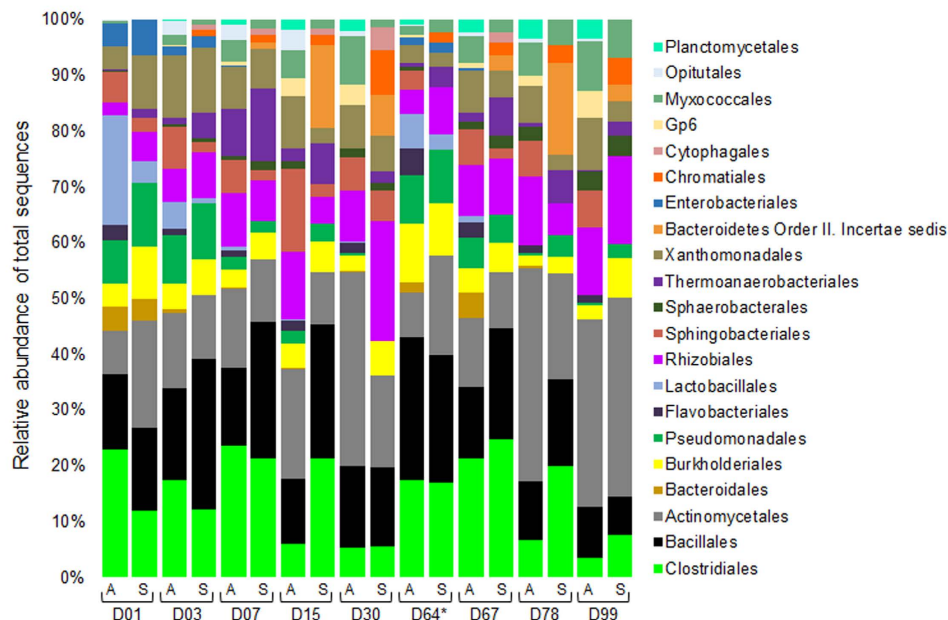
## Results and Discussion

**Composting time-series sampling and high-throughput sequencing.** We used shotgun DNA, 16S rRNA gene amplicon, and metatranscriptome high-throughput sequencing of time-series samples collected at the composting facility of São Paulo Zoo Park, São Paulo, Brazil. Supplementary Fig. S1 online shows the workflow used to generate the results we discuss here. We have collected nine time-series samples from composting cell ZC4 (days 1, 3, 7, 15, 30, 64, 67, 78 and 99) and five from composting cell ZC3 (days 1, 30, 64, 78 and 99). Sequence dataset accession numbers are listed in Supplementary Table S1. The main characteristics of samples collected from the two composting cells are summarized in Table 1 (ZC4) and Supplementary Table S2 (ZC3). Because ZC4 is the more extensive dataset, we base most of our discussion on it, referencing ZC3 when necessary.

ZC4 and ZC3 exhibited a sustained thermophilic profile (Supplementary Fig. S2), with average temperatures ranging from 50 °C to 75 °C. The amplitude of temperature variation between collection points in ZC4 and ZC3 cells, which in some days reached ~20 °C (Table 1 and Supplementary Table S2), suggests a heterogeneity of the semi-static composting process, reflecting differences in compaction and aeration of the pile as well as in the active microorganism populations. When the temperature dropped below 55 °C, turning of the pile was performed to restore aeration and thermophilic conditions (Supplementary Fig. S2). ZC4 and ZC3 presented C/N ratio in the range of 15:1 (Supplementary Tables S3 and S4), which is half of the recommended value (30:1)<sup>1</sup> estimated when the composting cells were built. Nevertheless, the decrease in C/N ratio was within the recommended values (<0.75), reflecting an acceptable loss of carbon as CO<sub>2</sub> as well as the recycling of nitrogen, instead of loss as ammonia gas or through leaching<sup>1,2,26</sup>.

Metagenomic DNA extracted from the time-series samples listed on Table 1 and Supplementary Table S2 online was submitted to shotgun and to 16S rRNA gene amplicon high-throughput sequencing. Eight of nine samples from ZC4 were also used for metatranscriptomic profiling aiming to reveal active metabolic pathways during thermophilic composting. Detailed metrics of sequencing and assembly data for ZC4 and ZC3 samples are presented in Supplementary Tables S5 and S6, respectively.

**Variation of microbial community composition and diversity during composting.** All compost samples we have collected are dominated by bacterial species (Supplementary Tables S7 and S8), regardless of composting age. The bacteria domain represented nearly 100% of classified reads, and at least 84% of all reads, in all samples.



**Figure 1. Relative abundance of bacteria in ZC4 samples at the order level.** Taxonomic assignments were obtained from 16S rRNA amplicon (A) and shotgun (S) sequencing datasets using RDP classifier and MG-RAST (M5NR) database, respectively. Only orders with relative abundance  $\geq 2\%$  are shown. Unassigned 16S (10–30%) and DNA shotgun reads (10.1–14.6%) were excluded. Samples are referred to by the letter D followed by the collection day. Asterisk indicates one day after the turning procedure.

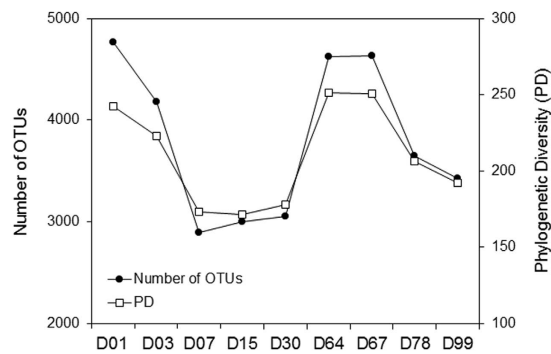
Analysis of the bacterial community structure at the levels of phylum and order by shotgun DNA as well as by 16S rRNA gene amplicon sequencing yields two main results: 1) shotgun DNA and 16S amplicon results by and large agree with each other; 2) the phyla and orders that are most abundant agree with those found in previous studies<sup>1,4,7–12</sup>. The four most abundant phyla throughout the composting process are Firmicutes, Proteobacteria, Bacteroidetes and Actinobacteria (Supplementary Fig. S3). These four phyla account for at least 85% of all classified reads in all samples.

At the order level there is far more diversity within samples as well as significant variation among them (Fig. 1). Noteworthy is the relative high abundance of orders Clostridiales, Bacillales, and Actinomycetales in most samples, regardless of the sequencing dataset analyzed (16S amplicon or shotgun). Among the most abundant genera within these orders (Supplementary Table S9), we highlight *Clostridium*, *Symbiobacterium*, and *Thermaerobacter* for Clostridiales, and *Geobacillus*, *Bacillus*, and *Ureibacillus* for Bacillales. We also feature *Thermopolyspora*, *Thermobispora*, and *Thermomonospora* among the abundant genera within Actinomycetales. These genera comprise bacterial species that can survive the thermophilic stages of composting and are hypothesized to play an important role in biomass degradation during the composting process<sup>4,12,13,16,25,27–30</sup>.

Similar profiles for phylum (Supplementary Fig. S4) and order (Supplementary Fig. S5) were observed for ZC3. Indeed, ZC4 and ZC3 appear to share much the same groups of microorganisms according to distance comparisons of corresponding time-series samples (Supplementary Fig. S6), despite some differences in the relative abundance over time for a few groups, as exemplified by Clostridiales (more abundant in D01 in ZC3 but abundant throughout in ZC4; compare Fig. 1 and Supplementary Fig. S5). On the other hand, the distance analyses showed a significant variation for the more abundant groups among the time-series samples of the same cell, indicating the highly dynamic nature of the composting process.

The variation in number of Operational Taxonomic Units (OTUs) and phylogenetic diversity index over time based on 16S data (Fig. 2) shows that the highest diversity was found in sample D01 and in samples D64 and D67 (right after the turning procedure). These results suggest the existence of two stages in the composting process we analyzed: the first from the beginning (D01) to D30, and the second from D64 to the end (D99). The second stage seems to recapitulate the first in terms of these measures. Moreover, pairwise distances calculated using two distance metrics (Bray-Curtis and Unifrac) and clustered using average linkage revealed three groups: D01, D03 and D64; D07, D15 and D67; D30, D78 and D99 (Supplementary Fig. S7).

This separation in two stages is most likely caused by the turning procedure performed just before sample D64 was collected. We hypothesize that turning restores to a certain extent conditions prevalent at the beginning of composting. This is in line with the observation that aeration of compost piles impacts microbial activity and temperature<sup>16,17</sup>. At the beginning there is an abundance of easily degradable organic nutrients and oxygen<sup>31</sup>. These conditions favor organisms that are aerobic and can most efficiently use the nutrients available, such as members from the order Lactobacillales found in the samples D01 and D64<sup>3,4,32</sup>. As the degradation progresses, both oxygen and easily degradable nutrients would become progressively scarcer and/or more difficult to access, especially in deeper regions of the chamber. This environment probably constrains the microbial community,



**Figure 2. ZC4 phylogenetic diversity variation.** Profile of the number of OTUs estimated by 16S rRNA amplicon sequencing and phylogenetic diversity index (alpha rarefaction) during ZC4 composting. Intervals on the horizontal axis do not represent chronological time. Samples are referred to by the letter D followed by the collection day.

and likely increases selection of more specialized microbial populations for degradation of remaining lignocellulosic materials. When turning is performed, oxygen and easily degradable nutrients that were in outer layers of the composting pile become accessible, probably easing the selection constraint. Based on these results we hypothesize that nutrient access and oxygen availability are the major determinants of changes in the composting microbiota composition, which agrees with the concept that accessibility by decomposer organisms is a constraint on biomass decomposition<sup>33</sup>.

**Abundant bacterial species in composting based on shotgun data.** At the level of species, using classifications provided by the program MyTaxa<sup>34</sup> on shotgun DNA reads, the most abundant organisms identified in ZC4 were *Rhodothermus marinus* (2.5% of reads), *Thermobispora bispora* (2.1%), *Symbiobacterium thermophilum* (1.5%), *Sphaerobacter thermophilus* (1.1%) and *Thermobifida fusca* (0.7%) (Supplementary Table S10). We reconstructed genomes for each of these species and obtained values for marker gene similarity between 86 and 100% (Supplementary Table S11).

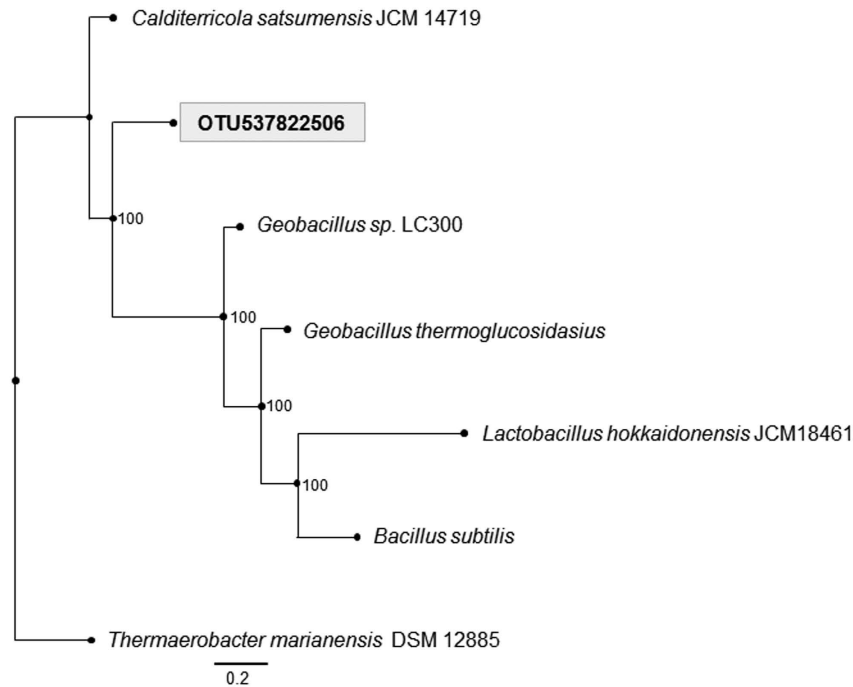
All species listed in Supplementary Table S10 online have been found in biomass-degrading environments: *S. thermophilus*<sup>20,35,36</sup>, *T. bispora*<sup>29,37</sup>, *T. fusca*<sup>38,39</sup>, *S. thermophilum*<sup>40–42</sup>, and *R. marinus*. For the latter, IMG strains to which our sequences are more similar are described as “compost feedstock-adapted isolates” (Supplementary Table S12)<sup>15,20</sup>.

The variation in abundance over time of the five species described above shows interesting patterns (Supplementary Fig. S8). The species with the highest relative abundance is *R. marinus*, and its relative abundance variation agrees with the pattern of phylogenetic diversity variation shown in Fig. 2, albeit with a small lag, both at the beginning and right after the turning procedure (Supplementary Fig. S8a). The other four species present relative abundance variation curves with less obvious patterns (Supplementary Fig. S8b). *T. bispora* and *S. thermophilum* show a pronounced peak on D03 followed by a sharp decrease on the following sampling days, suggesting that conditions at the beginning of the process are favorable to these species. The one species that most deviates from the two-stage pattern is *T. fusca*, since it starts with a low relative abundance, and slowly increases until D64. This suggests that *T. fusca* is able to proliferate only after other species have consumed initially available nutrients. This species is widely identified in lignocellulosic compostings and is well-known as a cellulose and lignin degrader<sup>43</sup>.

Several of the most abundant species identified in ZC4 shotgun DNA reads were also identified as most abundant in metatranscriptome data using classifications provided by the program MyTaxa. Specifically, the top six most abundant species in metatranscriptome data (Supplementary Table S13) are also among the top ten most abundant according to shotgun metagenomics (data not shown). In addition, in the case of *R. marinus*, the variation in abundance over time is very similar in both DNA and mRNA data (Supplementary Fig. S9). These observations give us confidence that our results based on DNA shotgun data reflect actual microbial activity in the composting process.

**Partial genome recovery of a novel bacterial genus guided by 16S data.** An analysis of 16S data at the genus level unveiled a microbial composition structure in stark contrast to that of the shotgun DNA just presented (Supplementary Table S14); none of the five most abundant OTUs seem to correspond to the species described in the previous section. We believe that the explanation for this discrepancy is that our analysis of shotgun DNA was entirely based on reference complete genomes. This means that organisms present in the composting process for which there are no complete genomes available (i.e. that are part of the so-called “microbial dark matter”) would not have been identified in the analyses of our shotgun DNA metagenomics dataset. However, the possibility still exists that we can uncover the DNA shotgun fragments of these abundant OTUs by some indirect method. This is what we attempted to do for the most abundant OTU, as described next.

A BLAST analysis of the 16S fragment that represents OTU537822506 (the most abundant) indicated 100% identity with *Calditerricola yamamurae*, which is a Firmicutes in the family Bacillaceae. There is no complete or even draft genome available for this species. On the other hand, our analysis of shotgun data indicated that



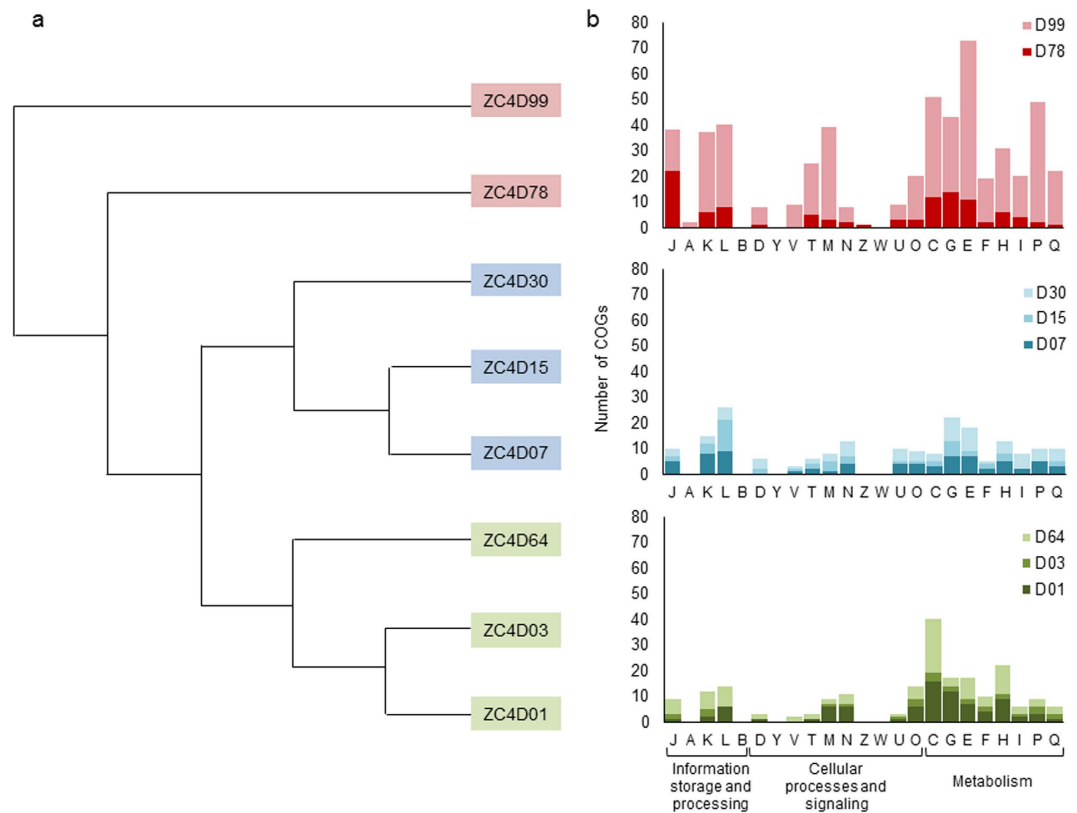
**Figure 3. Maximum likelihood phylogenetic tree based on nucleotide sequences of 113 ortholog genes.** Node numbers represent bootstrap support. Scale is number of substitutions per site. *Thermaerobacter marianensis* (clostridiales) was chosen as outgroup; all other species are members of the Bacillaceae family.

in all of our samples there was a fairly large fraction of reads (varying from 8% to 19%) identified as “unsigned Firmicutes”. Moreover, the variation in relative abundance over time for these reads closely followed that of OTU537822506 (data not shown). Based on this evidence, we applied a special assembly strategy (see Supplementary Methods) on these reads and succeeded in recovering what we estimate to be a near-complete (92%) genome of OTU537822506 (2,367,546 bp); average reads coverage for the assembled genome was 51.2 fold (Supplementary Table S15). A phylogenetic analysis based on 113 orthologous genes (Fig. 3) suggests that OTU537822506 is a new genus in the family Bacillaceae.

**Metabolic potential based on metagenomic and metatranscriptomic data.** To obtain an overall profile of the gene functions throughout the composting process, coding sequences (CDSs) from ZC4 and ZC3 assembled shotgun reads were classified using COGs (Cluster of Orthologous Groups<sup>44</sup>). Among the 15 functions most abundant in the ZC3 and ZC4 time-series metagenomes (Supplementary Table S16) we found COGs associated with cell maintenance and proliferation (e.g. carbohydrate transport and metabolism; amino acid transport and metabolism; DNA replication and repair; cell wall biogenesis), signal transduction (e.g. response regulator; histidine kinase) and defense mechanisms (e.g. efflux pump). These functions are present in both ZC3 and ZC4, despite some differences in their relative abundances in the time-series metagenomes.

Next, we analyzed the dynamics of COG functions during the composting process based on ZC4 metatranscriptomic data, to ensure that our results reflect actual functional activity. As a sanity check, we first mapped metatranscriptome reads onto shotgun DNA contigs; the vast majority of reads could be mapped (Supplementary Fig. S10), confirming that by and large the metatranscriptome reads are a subset of the shotgun DNA data. Using COGs, we classified CDSs found in contigs obtained from assembled metatranscriptome reads. Hierarchical clustering of ZC4 samples based on this COG classification resulted in three groups: D01, D03 and D64 (group 1), D07, D15 and D30 (group 2), and D78 and D99 (group 3) (Fig. 4a). These groups correspond to the groups obtained from pairwise distances mentioned previously (Supplementary Fig. S7). These results show that the effects of the turning procedure can be observed both in the microbial composition structure and in the functional profiles.

We interpret the three groups and their COG category differential abundance levels (Fig. 4b) as follows: group 1 represents the beginning (or restart) of the composting process and is characterized by energy production/conversion (COG category C), transport/coenzyme metabolism (H), carbohydrate transport/metabolism (G) and amino acid transport/metabolism (E). These categories include functions related to the microbial metabolism dedicated to degrading and utilizing easily degradable organic nutrients in an environment still rich and available in these compounds. Group 2, representing the middle of the process, is characterized by replication/recombination/repair (L), carbohydrate transport/metabolism (G) and amino acid metabolism (E) categories. At this stage, an intense depolymerizing activity has been occurring as well as the fast utilization of solubilized carbon compounds. When most of the easily degradable nutrients have been consumed, microorganisms able to degrade polymeric carbon become predominant. The higher expression levels of genes in category L could be



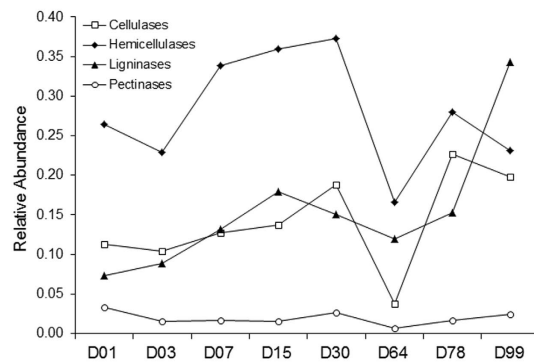
**Figure 4. Functional profile of the metatranscriptomes based on COG categories.** (a) Hierarchical clustering of metatranscriptome CDSs. The tree was generated using an IMG/M tool, selecting the COG pathways method. Samples are referred to by the letter D followed by the collection day. (b) Number of COGs (y axis) of each COG functional category (x axis) that had differential representation, based on relative abundance of CDSs. A COG functional category was considered differentially represented if the relative abundance value for a given sample was at least 1.5 times the interquartile range for the third quartile (i.e., only overrepresented outliers were noted) considering relative abundance values for that category in all samples. Each histogram corresponds to the same-color group of samples to its left.

explained by shifts in microbial composition between the beginning/restart and middle groups, also observed in the taxonomic analysis presented below. Group 3, representing the end of the composting process, is enriched in COGs related to amino acid metabolism (E), energy production/conversion (C), inorganic ion transport/metabolism (P), carbohydrate transport/metabolism (G), and categories dedicated to information, storage, and signaling processes (J, K, L and M). We hypothesize that the oxygen and nutrient limitations favor bacteria specialized in lignocellulosic recalcitrant material (such as lignin) degradation. It is interesting to note that, among abundant CDSs classified in category P, there are sequences encoding superoxide dismutases (SOD). Besides their important role in cellular antioxidant defense, some of these enzymes act as microbial lignin-oxidizers, as shown by manganese-SOD from *Sphingobacterium sp.* T2<sup>45</sup>.

### Transcriptional profile of genes predicted to be involved in lignocellulose degradation.

Deconstruction of lignocellulosic biomass can be achieved by diverse mechanisms, and the repertoire of known enzymes and microbial species capable of doing so is rapidly increasing<sup>46,47</sup>. We investigated the presence and abundance variation of key lignocellulose-degrading enzymes as given by COG classifications of the ZC4 metatranscriptome dataset. Hemicellulases, cellulases, pectinases, and ligninases were detected (Supplementary Table S17), and variation over time in their respective relative abundance levels indicates that hemicellulose, cellulose, pectin, and lignin are degraded throughout, with turning causing a temporary slowdown of the degradation process (Fig. 5). This is in agreement with the two-stage process mentioned previously. Note that ligninase abundance reaches a peak only at the end (D99). These results indicate that lignocellulosic biomass deconstruction occurs synergistically and sequentially, with hemicellulose being degraded preferentially to cellulose and lignin, as also seen in previous studies<sup>16,23,46</sup>.

To explore the repertoire of the lignocellulose-degrading enzymes in more detail, we compared all CDSs of ZC4 metatranscriptome against sequences in the CAZy database<sup>48</sup>. We found hits to six carbohydrate-active enzyme classes and associated modules: glycoside hydrolases (GHs) (29.5–34.2%), carbohydrate-binding modules (CBM) (21.5–39.1%), glycosyl transferases (GTs) (15.6–27.1%), carbohydrate esterases (CEs) (10.3–15.3%), polysaccharide lyases (PLs) (0.8–2.4%) and auxiliary activities (AAs) (1.6–8.4%) (Supplementary Fig. S11). Among the most abundant enzymes and associated domains, we highlight the surface-layer homology (SLH)



**Figure 5. Analysis of CDSs involved in lignocellulosic degradation during ZC4 composting process.** Relative abundance (%) of the CDSs annotated as hemicellulases, cellulases, ligninases and pectinases in the metatranscriptome contigs throughout the composting process, according to the IMG-M pipeline. Samples are referred to by the letter D followed by the collection day. Intervals on the horizontal axis do not represent chronological time.

domain and CBM50. The SLH domain is present in cellulosome scaffolding proteins anchoring the degradation complex to the bacterial cell surface<sup>49</sup> and in some secreted GHs (such as xyn-b39 and endo1,4 $\beta$  glucanase CelD)<sup>50,51</sup>. The CBM50 domain, which binds N-acetylglucosamine residues in bacterial peptidoglycans and in chitin, is found attached to enzymes from GH18, GH19, GH23, GH24, GH25, and GH73 families<sup>48</sup>. Along with these six GH families, endoglucanases (GH5 and GH9) and exo-cellobiohydrolases (GH6, GH48 and GH7), necessary to the complete hydrolysis of cellulose, are also represented in our metatranscriptome dataset (Supplementary Tables S17 and S18). Putative cellulase-encoding enzymes, mostly belonging to families GH5 and GH9, were also observed in bagasse, switchgrass-adapted compost, and in rice straw-adapted microbial consortia metagenomes<sup>19,23,30</sup>.

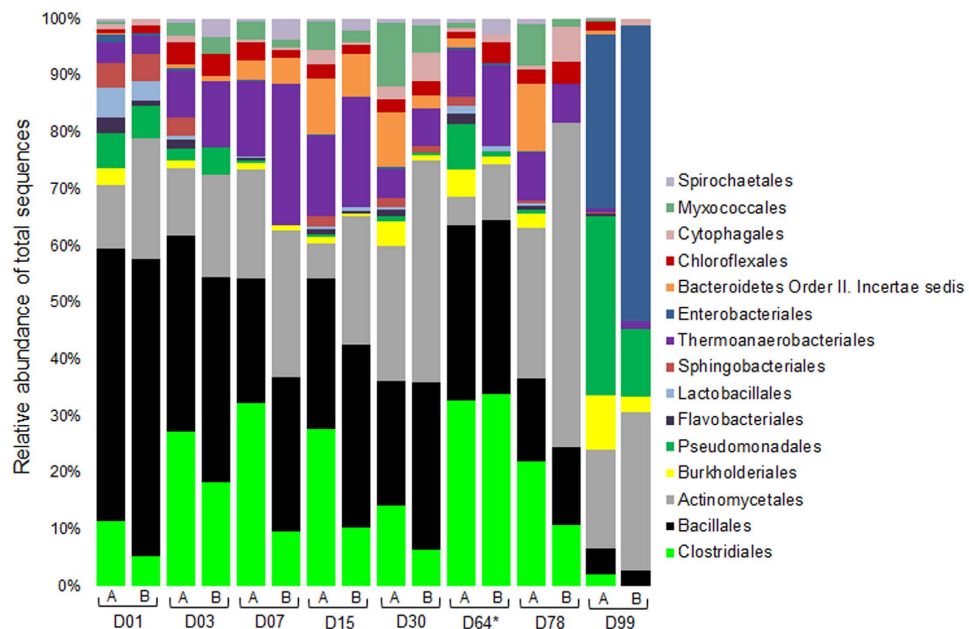
The ZC4 metatranscriptome also encodes ligninolytic enzymes and lytic polysaccharide mono-oxygenases (AA enzymes). The majority of these enzymes, found in all samples, belongs to classes AA2 (7–28% of all AAs) and AA6 (32–66%), followed by moderate amounts of AA3, 4, 7, 9 and 10 (Table S18). Sample D99 contains the highest number of CDSs assigned to this class.

Notably, we did not find any CDSs that could be annotated as fungal laccases, manganese peroxidases, versatile peroxidases, and lignin peroxidases. Wang *et al.*<sup>30</sup>, studying biomass-degrading consortia isolated from composting, have not found them either. This is additional evidence that lignocellulose degradation in a thermophilic composting process is primarily or even exclusively the result of bacterial enzymatic activity. Thus, lignocellulolytic enzymes of bacterial origin may be of special interest in industrial applications in which fungal enzymes cannot be used<sup>30,52,53</sup>.

**Taxonomic assignment of CDSs related to lignocellulosic biomass degradation.** Here we revisit the question of which microorganisms are mostly responsible for lignocellulosic biomass degradation based on metatranscriptome data. Figure 6 shows the most abundant orders. Except for D99, members from the orders Bacillales, Clostridiales, Actinomycetales, and Thermoanaerobacterales abound throughout the composting process, and most CDSs related to plant biomass degradation were classified as belonging to members of these orders. Actinomycetales are commonly found in compost, particularly in the thermophilic and mature stages<sup>11,16,25,54</sup>, and also in thermophilic microbial consortia enriched from compost<sup>30,55</sup>. Indeed, in our data the relative abundance of CDSs assigned to this order involved with lignocellulosic biomass degradation is largest at D30 and D78 (34.4% and 49%, respectively). Members of orders Clostridiales and Bacillales are known to have genes encoding enzymes involved in cellulose and hemicellulose degradation and were implicated as the major plant biomass degrading microbes in peat swamp forests<sup>56,57</sup>.

We can also observe orders that seem to play roles related to plant biomass deconstruction only during certain stages of the process. Members of Bacteroidetes Order II, incertae sedis are clearly present in the first stage (from D03 to D30), but fade thereafter. Members of Lactobacillales are more active at the beginning (D01) and after turning (D64). At the end of the process (D99), besides Actinomycetales, members of Enterobacteriales and Pseudomonadales are most abundant and seem to contribute to the continued degradation of remnant material. Enterobacter spp. are commonly described at the early stages of composting and are associated with lower temperatures (< 60 °C)<sup>32,58</sup>. Indeed, high levels of CDSs assigned to this bacterial order become more prominent when temperature decreases to less than 50 °C (D99). Most (69.2%) of the D99 Enterobacteriales CDSs were assigned to a single species, *Klebsiella pneumoniae*, which is known to perform cellulose and hemicellulose degradation, nitrogen fixation, and has been associated with wood, termite gut, and composting ecosystems<sup>59,60</sup>. Other studies also recorded the marked presence of members of the orders Enterobacteriales, especially *Klebsiella* species, in wheat straw degrading microbial consortia<sup>61,62</sup>.

A COG-based analysis of the reconstructed genome of OTU537822506 revealed two CDSs related to lignin breakdown (COG1496 and COG2132-multicopper oxidases) and one to pectin hydrolysis (COG3866-pectate lyase). CAZy analysis revealed 63 genes encoding 12 CBMs, 11 GHs, 17 GTs, 10 CEs, 1 PL and 4 AAs (Supplementary Table S19). These results indicate that the genome of this organism encodes enzymes related to bioconversion of all



**Figure 6. Taxonomic profile of the ZC4 metatranscriptome.** Columns labeled by A show taxonomic profiles considering all CDSs annotated with COGs. Columns labeled by B show taxonomic profiles considering CDSs annotated with COGs related to plant biomass (lignocellulose and pectin) degradation. The CDSs were classified with program myTaxa. Samples are referred to by the letter D followed by the collection day. The asterisk indicates one day after the turning procedure.

components in plant biomass, with emphasis on hemicellulose degradation (represented by the following CAZy families: CBM13, CBM37, GH11, GH16, CE4 and CE1). Variation over time of the shotgun reads used to reconstruct the genome as well as metatranscriptome reads mapped to the OTU537822506 genome show high levels at the beginning of composting (Supplementary Fig. S12), suggesting that this bacterial species is more active during the initial stages.

## Conclusions

We have presented a detailed study of a thermophilic composting process, from a microbial molecular standpoint. Thanks to time-series sampling and high-throughput sequencing, we were able to observe the variation of the microbiota that orchestrate the composting process as well as the variation of their enzymatic activities.

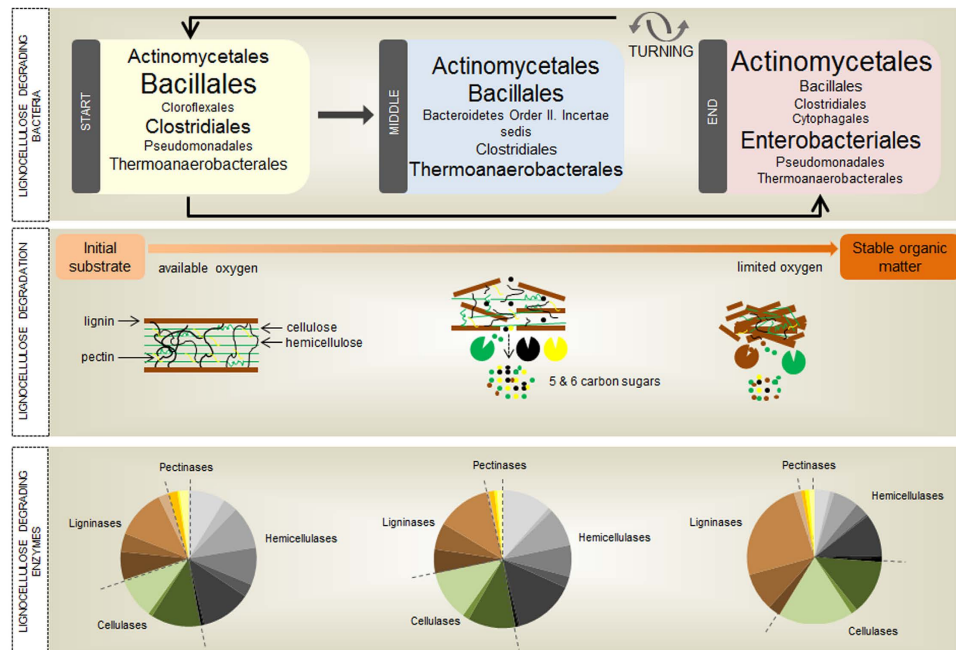
The results described here, together with other studies on composting microbiota<sup>16,23,57</sup>, allow us to propose a core group of bacterial microorganisms and metabolic functions primarily responsible for lignocellulosic biomass degradation, as sketched in Fig. 7.

The impressive variety of bacterial microorganisms and metabolic functions so far found to be active in thermophilic composting warrant this engineered ecosystem as valuable source for continued survey of new bacteria and metabolic functions adapted to an extreme and complex environment. We have identified and obtained a near-complete genome of an organism that plays a role in biomass degradation and likely represent a new genus. This is probably just the tip of an iceberg, and additional microbial molecular studies of composting systems should uncover a whole treasure trove of new microorganisms and their enzymes useful for environment management and biofuel production.

## Methods

**Composting.** The composting process was carried out at the composting facility of São Paulo Zoo Park, São Paulo, Brazil (23°38'56.9"S 46°37'18.7"W) with minor modifications from a previously described procedure<sup>63</sup>. The facility has several 8 m<sup>3</sup> open concrete cells (1.6 × 2.0 × 2.5; height × width × depth) and is designed to compost four tons/day of all organic waste produced in the park comprising mainly shredded tree branches, leaves and grass from the maintenance of park green areas, plus manure, beddings and food residues from about 400 species of zoo animals (mammals, avian and reptiles). The facility area is protected from sunshine and rain. To build the composting pile the substrates were layered within the concrete cell to roughly reach a Carbon:Nitrogen (C/N) ratio of 30:1. Adequate aeration of the pile was maintained by aeration pipes placed at the bottom of concrete cell and by arranging the substrates to limit excessive compaction. Daily reported temperatures of the composting piles were averages of measurements taken at the four edges and in the center. Proper humidity was maintained by watering the composting pile. When temperature dropped below 55 °C, usually after ~60–65 days of composting, the pile was turned using a BobCat skid-steer loader (Bobcat Company, USA) to restore aeration and thermophilic conditions. After ~100 days, the composting process within the concrete cell was considered finished, the compost material was removed and aged for an additional 10 days in windrows.





**Figure 7. Molecular overview of a thermophilic composting operation.** The top panel shows the three stages identified in this work: start (D01, D03 and D64), middle (D07, D15 and D30), and end (D78 and D99). The box for each stage shows the microbiota composition at the order level. The relative abundance of microorganisms is roughly proportional to the font size, and only orders with relative abundance  $\geq 2.5\%$  are shown. The middle panel depicts the degradation process, with vertical correspondence to the stages shown in the top panel. The bottom panel shows relative abundance of COG functions associated to lignocellulolytic enzymes, also with vertical correspondence to the stages shown in the top panel. The results in this figure were based on ZC4 metatranscriptomes.

**Sample collection.** Samples were collected throughout the composting process from two cells named ZC3 and ZC4, which were built, respectively, on 06/27/2011 and 08/05/2013. After completion of the composting pile, samples were collected on days 1, 30, 64, 78 and 99 for ZC3, and on days 1, 3, 7, 15, 30, 64, 67, 78 and 99 for ZC4. Samples are referred to by the letter D followed by the collection day (e.g. D01 is day one). The turning procedure was performed on day 65 for ZC3 and on day 63 for ZC4. To obtain a representative sample of the pile at each sampling day, sub-samples were collected from five different points (at all depths in the four edges and center)<sup>13</sup>. The five sub-samples (~100 g each) were pooled, thoroughly mixed, distributed into sterile 50 mL Falcon tubes and immediately frozen in dry ice before being transported to the laboratory, where they were stored at  $-80^{\circ}\text{C}$  until DNA extraction or chemical analysis. Samples for moisture analysis and pH were freshly processed. For total RNA extraction, ~1 g of each composite sample was immediately transferred to 15 mL Falcon tubes containing the twice the volume of LifeGuard Soil Preservation Solution (MoBio Laboratories, USA) and the suspensions were kept at room temperature until RNA extraction no longer than 1 week. Tubes and spatulas were previously treated with RNase Away (Life Technologies, USA).

**Physicochemical analyses.** Temperature was measured at a depth of 60 cm from the top of pile using digital thermometer with long-handled stainless-steel probe in the four edges and center of the pile. Moisture content determination was performed by microwave oven drying as previously described<sup>63</sup>. The pH was determined from suspensions of fresh compost samples in 0.9% sodium chloride using a pH electrode. Total Carbon, Hydrogen and Nitrogen content were determined using PerkinElmer 2400 series II CHNS/O analyzer (Perkin-Elmer, USA). Elemental analysis (Al, Fe, Mg, P, As, Cd, Cr, Cu, K, Ni, Pb, Se and Zn) was performed by inductively-coupled plasma optical emission spectrometry using Spectro Arcos ICP-OES analyzer (Spectro, Germany).

**DNA extraction and shotgun metagenomic library preparation.** Metagenomic DNA extraction was performed with MoBio DNA Power Soil kit. Prior to DNA extraction,  $-80^{\circ}\text{C}$  stored samples were subjected to lyophilization, and ~2 g lyophilized material were thoroughly grounded in a sterile mortar. Purified metagenomic DNA was subjected to a final clean-up step using QIAamp mini spin columns (Qiagen, USA) and stored  $-80^{\circ}\text{C}$ . DNA purity and concentration were evaluated on a ND-1000 spectrophotometer (Nano Drop Technologies, USA) at 260 nm, 280 nm and 230 nm. Further quantification was performed with Quant-iT Picogreen dsDNA assay kit (Life Technologies, USA). DNA integrity was examined with DNA 7500 chip using 2100 Bioanalyzer. DNA samples were mostly enriched in fragments higher than 10 kbp.

Shotgun metagenomic libraries were prepared using an Illumina Nextera DNA library preparation kit (Illumina, Inc., USA) with total DNA input of 20–35 ng. The resulting DNA fragment libraries were cleaned

up with Agencourt AMPure XP beads (Beckman Coulter, Inc., USA) and fragment size within the range of 400–700 bp was verified by running in the 2100 Bioanalyzer using Agilent High Sensitivity DNA chip.

**16S Metagenomic Sequencing Library Preparation.** PCR reactions were performed using a primer pair based on the sequences of primers S-D-Bact-0341-b-S-17 and S-D-Bact-0785-a-A-21<sup>64</sup> for amplification of the variable regions V3 and V4 of 16S rRNA gene and including adapters as suggested on the Illumina workflow for 16S Metagenomic Sequencing Library Preparation. Forward primer 341F, 5'-TCGTCGGCAGCGTCAGATGTGTAT AAGAGACAGCCTACGGGNGGCWGCAG and reverse primer 785R, 5'-GTCTCGTGGGCTCGGAGATG TGTATAAGAGACAGGACTACHVGGGTATCTAATCC were used in PCR reactions (25 µL) performed with 12.5 ng of metagenomic DNA samples, KAPA HiFi HotStart ReadyMix (Kapa Biosystems, USA), 200 nM of each primer at 95 °C for 3 min followed by 25 cycles at 95 °C for 30 sec, 55 °C for 30 sec and 72 °C for 30 sec plus 72 °C for 5 min. Amplicon expected size of ~550 bp was verified with Bioanalyzer DNA 1000 chip. PCR product cleanup was performed with AMPure XP beads (Beckman Coulter, Inc., USA). Dual indexes were attached using the Nextera XT Index Kit and after a second round of PCR cleanup with AMPure XP beads, V3–V4 16S indexed amplicons libraries were validated by running on a Bioanalyzer High Sensitivity DNA chip to verify the expected size of ~630 bp.

**RNA extraction, rRNA depletion and RNA-seq library preparation.** Total RNA was extracted from compost samples stored in LifeGuard Soil Preservation Solution (MoBio Laboratories, USA) using RNA MoBio PowerSoil Total RNA isolation kit. Extracted total RNA was stored at –80 °C. RNA purity and concentration were evaluated on a ND-1000 spectrophotometer (Nano Drop Technologies, USA) at 260 nm, 280 nm and 230 nm. Further quantification was performed with Quant-iT Ribogreen RNA Assay kit (Life Technologies, USA). RNA integrity number (RIN) was determined using RNA 6000 Nano kit in the 2100 Bioanalyzer (Agilent Technologies, USA). Only samples with RIN scores >7.5 were further processed. For removal of contaminant genomic DNA, total RNA samples were DNase I-treated with Illustra RNASpin mini kit (GE Healthcare Life Sciences, USA) and re-evaluated regarding their RIN scores. Complete removal of genomic DNA was confirmed by PCR with primers for variable region V3 and V4 of 16S rRNA gene. Depletion of rRNA was performed using a Ribo-Zero rRNA Removal kit/Bacteria (Epicentre, USA) and 5 µg of DNase-treated RNA. The final purification of the rRNA-depleted RNA (90 µL) was performed by adding 90 µL of RNase-free water, 18 µL of 3M-sodium acetate, 2 µL of glycogen (10 mg/mL) and 600 µL of cold 100% ethanol. After 2 h incubation at –20 °C, rRNA-depleted RNA was collected by centrifugation at 16,000 ×g for 30 min at 4 °C. After two washes with cold 70% ethanol, the rRNA-depleted RNA pellet was dried at room temperature for 5 min and immediately used for RNA-seq library preparation with ImProm-II Reverse Transcription System (Promega, USA) and TruSeq RNA Library Preparation Kit v2 (Illumina, Inc., USA). Complete depletion of rRNA was verified using Bioanalyzer RNA 6000 Pico chip. Average fragment size of ~300 bp of the RNA-seq libraries were verified using a Bioanalyzer High Sensitivity DNA chip.

**Sequencing.** Quantification of Illumina sequencing libraries with KAPA Library Quantification Kit, normalization, and pooling were performed following standard protocols for sequencing in the Illumina MiSeq platform. Pooled libraries were subjected to 2–3 runs using the MiSeq Reagent kit v2 (500-cycle format, paired-end (PE) reads) and resulting sequences for each library were combined. RNA-seq libraries were also run on a single-sequencing lane on Illumina HiSeq2500 (200-cycle format) using TruSeq SBS v3 kits, and resulting sequences were combined with the MiSeq PE sequences. On average, Illumina PE read1 and read2 presented, respectively, >80% and >75% of bases with quality score at least 30 (Q30).

DNA samples purified from ZC3 chamber were also submitted to pyrosequencing following standard Roche 454 GS FLX Titanium protocols (Roche Applied Science) as previously described<sup>13</sup>.

**Bioinformatics analyses.** We processed raw PE sequencing reads and 454-reads in various ways. Unassembled reads were submitted for automatic processing to the MG-RAST metagenomics analysis server<sup>65,66</sup> using their default quality control pipeline. We also submitted data to the IMG/M annotation pipeline<sup>67</sup> using the following steps. Raw PE sequencing reads data were quality-filtered to remove reads shorter than 50 bp or reads with average quality score lower than phred 20 using SICKLE<sup>68</sup>, FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) and fastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Assembly of quality-filtered PE reads and 454-reads for each shotgun metagenomic or RNA-seq library was performed with SOAPdenovo2<sup>69</sup> followed by assembly quality checking using QUAST<sup>70</sup>. The resulting sets of contigs were submitted to IMG/M<sup>67</sup>.

Methods employed in the analyses of metagenomic and metatranscriptomic sequences, taxonomic classification based on shotgun reads<sup>34,65,66</sup> and 16S rRNA amplicon sequencing<sup>71</sup>, microbial community and diversity analyses, reconstruction of bacterial genomes, genome annotation and phylogenetic tree construction are presented in Supplementary Methods.

## References

1. Ryckeboer, J. *et al.* A survey of bacteria and fungi occurring during composting and self-heating processes. *Annals of Microbiology* **53**, 349–410 (2003).
2. Kumar, S. Composting of municipal solid waste. *Crit Rev Biotechnol* **31**, 112–136 (2011).
3. Alfreider, A., Peters, S., Tebbe, C. C., Rangger, A. & Insam, H. Microbial community dynamics during composting of organic matter as determined by 16S ribosomal DNA analysis. *Compost Sci Util* **10**, 303–312 (2002).
4. Partanen, P., Hultman, J., Paulin, L., Auvinen, P. & Romantschuk, M. Bacterial diversity at different stages of the composting process. *BMC Microbiology* **10** (2010).
5. Federici, E. *et al.* Two-phase olive mill waste composting: Community dynamics and functional role of the resident microbiota. *Bioresour Technol* **102**, 10965–10972 (2011).

6. Kuok, F., Mimoto, H. & Nakasaki, K. Effects of turning on the microbial consortia and the *in situ* temperature preferences of microorganisms in a laboratory-scale swine manure composting. *Bioresource Technology* **116**, 421–427 (2012).
7. Jurado, M. *et al.* Exploiting composting biodiversity: Study of the persistent and biotechnologically relevant microorganisms from lignocellulose-based composting. *Bioresource Technology* **162**, 283–293 (2014).
8. Zainudin, M. H. M. *et al.* Bacterial Community Structure and Biochemical Changes Associated With Composting of Lignocellulosic Oil Palm Empty Fruit Bunch. *Bioresources* **9**, 316–335 (2014).
9. Storey, S., Ni Chualain, D., Doyle, O., Clipson, N. & Doyle, E. Comparison of bacterial succession in green waste composts amended with inorganic fertiliser and wastewater treatment plant sludge. *Bioresource Technology* **179**, 71–77 (2015).
10. Neher, D. A., Weicht, T. R., Bates, S. T., Leff, J. W. & Fierer, N. Changes in Bacterial and Fungal Communities across Compost Recipes, Preparation Methods, and Composting Times. *PLoS One* **8** (2013).
11. De Gannes, V., Eudoxie, G. & Hickey, W. J. Prokaryotic successions and diversity in composts as revealed by 454-pyrosequencing. *Bioresource Technology* **133**, 573–580 (2013).
12. Zhang, L. L., Zhang, H. Q., Wang, Z. H., Chen, G. J. & Wang, L. S. Dynamic changes of the dominant functioning microbial community in the compost of a 90-m<sup>3</sup> aerobic solid state fermentor revealed by integrated meta-omics. *Bioresource Technology* **203**, 1–10 (2016).
13. Martins, L. F. *et al.* Metagenomic analysis of a tropical composting operation at the Sao Paulo Zoo park reveals diversity of biomass degradation functions and organisms. *PLoS One* **8**, e61928 (2013).
14. Langarica-Fuentes, A. *et al.* Fungal succession in an in-vessel composting system characterized using 454 pyrosequencing. *Fems Microbiology Ecology* **88**, 296–308 (2014).
15. D'Haeseleer, P. *et al.* Proteogenomic Analysis of a Thermophilic Bacterial Consortium Adapted to Deconstruct Switchgrass. *PLoS One* **8** (2013).
16. Lopez-Gonzalez, J. A. *et al.* Dynamics of bacterial microbiota during lignocellulosic waste composting: Studies upon its structure, functionality and biodiversity. *Bioresource Technology* **175**, 406–416 (2015).
17. Neher, D. A., Weicht, T. R. & Dunseith, P. Compost for Management of Weed Seeds, Pathogen, and Early Blight on Brassicas in Organic Farmer Fields. *Agroecology and Sustainable Food Systems* **39**, 3–18 (2015).
18. Yang, G. Q., Chen, M., Yu, Z., Lu, Q. & Zhou, S. G. *Bacillus composti* sp nov and *Bacillus thermophilus* sp nov., two thermophilic, Fe(III)-reducing bacteria isolated from compost. *Int J Syst Evol Micr* **63**, 3030–3036 (2013).
19. Allgaier, M. *et al.* Targeted Discovery of Glycoside Hydrolases from a Switchgrass-Adapted Compost Community. *PLoS One* **5** (2010).
20. Gladden, J. M. *et al.* Glycoside hydrolase activities of thermophilic bacterial consortia adapted to switchgrass. *Appl Environ Microbiol* **77**, 5804–5812 (2011).
21. Dougherty, M. J. *et al.* Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. *BMC Biotechnol* **12**, 38 (2012).
22. Habbeche, A. *et al.* Purification and biochemical characterization of a detergent-stable keratinase from a newly thermophilic actinomycete *Actinomadura keratinilytica* strain Cpt29 isolated from poultry compost. *Journal of Bioscience and Bioengineering* **117**, 413–421 (2014).
23. Mhuantong, W., Charoensawan, V., Kanokratana, P., Tangphatsornruang, S. & Champreda, V. Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. *Biotechnol Biofuels* **8** (2015).
24. Dees, P. M. & Ghiorse, W. C. Microbial diversity in hot synthetic compost as revealed by PCR-amplified rRNA sequences from cultivated isolates and extracted DNA. *FEMS Microbiology Ecology* **35**, 207–216 (2001).
25. Tkachuk, V. L. *et al.* Targeted 16S rRNA high-throughput sequencing to characterize microbial communities during composting of livestock mortalities. *Journal of Applied Microbiology* **116**, 1181–1194 (2014).
26. Lopez-Gonzalez, J. A. *et al.* Tracking organic matter and microbiota dynamics during the stages of lignocellulosic waste composting. *Bioresource Technology* **146**, 574–584 (2013).
27. Yabe, S., Kato, A., Hazaka, M. & Yokota, A. *Thermaerobacter composti* sp nov., a novel extremely thermophilic bacterium isolated from compost. *J Gen Appl Microbiol* **55**, 323–328 (2009).
28. Zhang, L. L. *et al.* *Thermomyces lanuginosus* is the dominant fungus in maize straw composts. *Bioresource Technology* **197**, 266–275 (2015).
29. Steger, K., Jarvis, A., Vasara, T., Romantschuk, M. & Sundh, I. Effects of differing temperature management on development of Actinobacteria populations during composting. *Res Microbiol* **158**, 617–624 (2007).
30. Wang, C. *et al.* Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol Biofuels* **9** (2016).
31. Said-Pullicino, D., Erriquens, F. G. & Gigliotti, G. Changes in the chemical characteristics of water-extractable organic matter during composting and their influence on compost stability and maturity. *Bioresource Technology* **98**, 1822–1831 (2007).
32. Chandna, P., Nain, L., Singh, S. & Kuhad, R. C. Assessment of bacterial diversity during composting of agricultural byproducts. *BMC Microbiol* **13**, 99 (2013).
33. Dungait, J. A. J., Hopkins, D. W., Gregory, A. S. & Whitmore, A. P. Soil organic matter turnover is governed by accessibility not recalcitrance. *Global Change Biol* **18**, 1781–1796 (2012).
34. Luo, C., Rodriguez, R. L. & Konstantinidis, K. T. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res* **42**, e73 (2014).
35. Demharter, W., Hensel, R., Smida, J. & Stackebrandt, E. *Sphaerobacter thermophilus* Gen-Nov, Sp-Nov - a Deeply Rooting Member of the Actinomycetes Subdivision Isolated from Thermophilically Treated Sewage-Sludge. *Syst Appl Microbiol* **11**, 261–266 (1989).
36. Pati, A. *et al.* Complete genome sequence of *Sphaerobacter thermophilus* type strain (S 6022). *Stand Genomic Sci* **2**, 49–56 (2010).
37. Liolios, K. *et al.* Complete genome sequence of *Thermobispora bispora* type strain (R51). *Stand Genomic Sci* **2**, 318–326 (2010).
38. Lykidis, A. *et al.* Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *J Bacteriol* **189**, 2477–2486 (2007).
39. Toth, A. *et al.* Draft Genome Sequence of the Lignocellulose Decomposer *Thermobifida fusca* Strain TM51. *Genome announcements* **1** (2013).
40. Ueda, K. *et al.* Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. *Nucleic Acids Res* **32**, 4937–4944 (2004).
41. Ohno, M. *et al.* *Symbiobacterium thermophilum* gen. nov., sp nov., a symbiotic thermophile that depends on co-culture with a *Bacillus* strain for growth. *Int J Syst Evol Micr* **50**, 1829–1832 (2000).
42. Yabe, S. *et al.* *Sphingobacterium thermophilum* sp nov., of the phylum Bacteroidetes, isolated from compost. *Int J Syst Evol Micr* **63**, 1584–1588 (2013).
43. Rahmanpour, R., Rea, D., Jamshidi, S., Fulop, V. & Bugg, T. D. Structure of *Thermobifida fusca* DyP-type peroxidase and activity towards Kraft lignin and lignin model compounds. *Arch Biochem Biophys* **594**, 54–60 (2016).
44. Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* **43**, D261–D269 (2015).
45. Rashid, G. M. M. *et al.* Identification of Manganese Superoxide Dismutase from *Sphingobacterium* sp. T2 as a Novel Bacterial Enzyme for Lignin Oxidation. *Acs Chem Biol* **10**, 2286–2294 (2015).

46. Cragg, S. M. *et al.* Lignocellulose degradation mechanisms across the Tree of Life. *Curr Opin Chem Biol* **29**, 108–119 (2015).
47. Hensworth, G. R., Johnston, E. M., Davies, G. J. & Walton, P. H. Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol* **33**, 747–761 (2015).
48. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233–238 (2009).
49. Hyeon, J. E., Jeon, S. D. & Han, S. O. Cellulosome-based, Clostridium-derived multi-functional enzyme complexes for advanced biotechnology tool development: advances and applications. *Biotechnology advances* **31**, 936–944 (2013).
50. Zhao, Y. Y. *et al.* Molecular and biochemical characterization of a new alkaline active multidomain xylanase from alkaline wastewater sludge. *World J Microb Biot* **29**, 327–334 (2013).
51. Velikodvorskaya, G. A. *et al.* Family 28 carbohydrate-binding module of the thermostable endo-1,4-beta-glucanase CelD from *Caldicellulosiruptor bescii* maximizes enzyme activity and irreversibly binds to amorphous cellulose. *Mol Biol+* **47**, 581–586 (2013).
52. Bugg, T. D. H., Ahmad, M., Hardiman, E. M. & Rahmanpour, R. Pathways for degradation of lignin in bacteria and fungi. *Nat Prod Rep* **28**, 1883–1896 (2011).
53. Chandra, R., Abhishek, A. & Sankhar, M. Bacterial decolorization and detoxification of black liquor from rayon grade pulp manufacturing paper industry and detection of their metabolic products. *Bioresource Technology* **102**, 6429–6436 (2011).
54. Zhang, X. *et al.* Diversity and dynamics of the microbial community on decomposing wheat straw during mushroom compost production. *Bioresource Technology* **170**, 183–195 (2014).
55. Simmons, C. W. *et al.* Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnol Biofuels* **7**, 495 (2014).
56. Kanokratana, P. *et al.* Insights into the Phylogeny and Metabolic Potential of a Primary Tropical Peat Swamp Forest Microbial Community by Metagenomic Analysis. *Microbial Ecology* **61**, 518–528 (2011).
57. Ventorino, V. *et al.* Exploring the microbiota dynamics related to vegetable biomasses degradation and study of lignocellulose-degrading bacteria for industrial biotechnological application. *Scientific Reports* **5** (2015).
58. Gbolagade, J. S. Bacteria associated with compost used for cultivation of Nigerian edible mushrooms *Pleurotus tuber-regium* (Fr.) Singer, and *Lentinus squarrosulus* (Berk.). *Afr J Biotechnol* **5**, 338–342 (2006).
59. Doolittle, M., Raina, A., Lax, A. & Boopathy, R. Presence of nitrogen fixing *Klebsiella pneumoniae* in the gut of the *Formosan subterranean* termite (*Coptotermes formosanus*). *Bioresource Technology* **99**, 3297–3300 (2008).
60. Droffner, M. L., Brinton, W. F. & Evans, E. Evidence for the Prominence of Well Characterized Mesophilic Bacteria in Thermophilic (50–70-Degrees-C) Composting Environments. *Biomass Bioenergy* **8**, 191–195 (1995).
61. Jimenez, D. J., Dini-Andreote, F. & van Elsas, J. D. Metatranscriptomic profiling and prediction of functional behaviour of wheat straw degrading microbial consortia. *Biotechnol Biofuels* **7** (2014).
62. Jimenez, D. J., Chaves-Moreno, D. & van Elsas, J. D. Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. *Scientific Reports* **5** (2015).
63. Rynk, R. *et al.* On-Farm Composting Handbook, Northeast Regional Agricultural Engineering Service – Cooperative Extension, Ed., NRAES-54, 185 pag., (Northeast Regional Agricultural Engineering Service, 1992).
64. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**, e1 (2013).
65. Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D. & Meyer, F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* **2010**, pdb prot5368 (2010).
66. Meyer, F. *et al.* The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386 (2008).
67. Markowitz, V. M. *et al.* IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* **42**, D568–D573 (2014).
68. Sickel: A sliding-window, adaptive, quality-based trimming tool for FastQ files v. Version 1.33 (Available at <https://github.com/najoshi/sickle>) (2011).
69. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
70. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
71. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336 (2010).

## Acknowledgements

Funding for this research was provided by grant 2011/50870-6 from the São Paulo Research Foundation (FAPESP) and by the Provost's Office for Research of the University of São Paulo. AMT, GMMS, GWCE, LLN, LPA, and RVP were supported by fellowships from FAPESP. AMT was supported by a fellowship from the Coordination for the Improvement of Higher Education Personnel (CAPES). AMDS, DB, JCS, KCL, LAD, LJ and RCP received Research Fellowship Awards from National Council for Scientific and Technological Development (CNPq). The funders had no role in study design, data collection, analysis, decision to publish or preparation of the manuscript. We thank the staff from São Paulo Zoo Compost Facility for their help with sampling. We also thank Felipe Prata Lima and Carlos Moraes for help in computational analyses. We are indebted to Prof. Luiz Juliano Neto, Prof. Sergio Verjovski-Almeida, and Dr. Paulo Bressan for constructive criticisms and continued support for this project.

## Author Contributions

A.M.D.S., J.C.S., and L.P.A. conceived the study and designed the experiments; K.C.L. and L.P.A. collected the samples; L.F.M. and L.P.A. processed the samples and performed DNA sequencing; G.M.M.S., G.W.C.E., L.A.D., L.J., P.L.R., and J.B.S. contributed reagents, materials and/or analysis tools; A.M.D.S., A.M.T., D.B., J.C.F.O., L.F.M., L.L.N., L.M.S.M., L.P.A., R.B.Q., R.V.P., and J.C.S. analyzed data; A.M.D.S. and J.C.S. wrote the manuscript with input from L.P.A., L.F.M. and R.V.P.

## Additional Information

**Accession codes:** All sequencing data described in this work comprising unassembled sequence reads, resulting sets of contigs (including singlets) and 16S rRNA V3–V4 amplicon sequences are available at MG-RAST, IMG/M or SRA databases under accession numbers listed in Supplementary Table S1. The Whole Genome Shotgun project related to the genome of OTU537822506 has been deposited at DDBJ/ENA/GenBank under the accession LWLU00000000. The version described in this paper is version LWLU01000000.

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Antunes, L. P. *et al.* Microbial community structure and dynamics in thermophilic composting viewed through metagenomics and metatranscriptomics. *Sci. Rep.* **6**, 38915; doi: 10.1038/srep38915 (2016).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016