

# The Bacterial Communities Associated with Retting: a Meta-Analysis

By: qsfl69

2025-09-18

## Contents

Background	1
Methods	3
Results and Discussion	12
Conclusion	14
References	16

## Background

The textile industry has changed greatly since the advent of polymer chemistry. Natural fibres like flax, kenaf, hemp, and jute, once some of the only material options, now constitute only 5% of global fibre production—with cotton comprising 20%.<sup>1</sup> This shift in fabric preference cannot be ignored as domestic laundering releases an annual 5 million tons of microplastic fibre into the environment.<sup>2</sup> Moreover, the production of synthetic fibres consumes ~98 million tons of oil annually, and releases far more greenhouse gas emissions when compared to the production of natural fibres.<sup>3</sup> Thus, there exists a need to reconsider the value of natural, sustainable fibres as a useful material.

The translation of flax's scientific name—"linen most useful"—or jute's colloquial title, the "golden fibre", emphasizes the popularity of natural fibres throughout history.<sup>4,5</sup> Crops like flax, jute, kenaf, and hemp have been used to produce clothing for millennia.<sup>4</sup> But, today, flax fibres account for only 0.4% of global clothing fibre production.<sup>6</sup> Pressure from increased amounts of inexpensive but resource intensive cotton and the emergence of synthetic materials have contributed to this dramatic reduction in flax, hemp, kenaf, and jute markets.<sup>5-9</sup> Producing strong, homogeneous fibres on an industrial scale is also challenging, which makes the use of synthetic fibres even more attractive.<sup>4</sup> The challenges associated with processing natural fibres in a consistent manner, then, must be overcome to make their adoption commercially viable.

A major problem with natural fibre processing is that of retting, in which a plant's bast fibres are separated from other, non-fibrous tissues.<sup>10</sup> Typically, retting uses microorganisms to degrade pectins, hemicelluloses, and other matrix polysaccharides which cement plant fibres to their stalk; some more recent retting tech-

niques involve the application of enzymes and other additives to enhance retting efficiency.<sup>4,11</sup> Retting is performed in one of two ways. The oldest practice—known as dew or field retting—gathers harvested stems into piles and allows the local bacteria and fungi to naturally decompose the non-fibrous tissues.<sup>4</sup> The second form—water-retting—submerges the harvested plant material in bodies of water for the duration of the retting period. These two forms of retting result in fibres of different quality, with water-retting producing longer, stronger fibres suitable for textile production in relatively little time.<sup>4</sup> The diminished and inconsistent quality of dew-retted fibres can be largely attributed to the fact that the process is dependent on variable weather conditions, farmer experience, and because the fibres are dirtied by soil and fungi.<sup>4,12</sup> Additionally, dew-retting is only feasible in locations with adequate humidities and temperatures, restricting the number of regions capable of dew-retting.<sup>4</sup> Despite these limitations, dew-retting has been the primary form of retting since the 1950's due to the expense, stench, and pollution involved with water-retting.<sup>13,14</sup> Thus, natural fibres have seen a decline in quality over the past 60 years.<sup>14</sup> In essence, the commercial adoption of natural fibres is limited by the poor efficacy of dew-retting which, in turn, limits their production to a few regions with amicable climates.<sup>4</sup>

Owing to a variety of factors—ranging from poor fibre quality, expensive consumption of land, and restrictive climates—it is unlikely that dew-retted fibres can meet global textile demand. However, natural fibre products can be used for purposes that do not necessitate the same quality demanded by clothing production. It has been shown that natural fibres and their waste products can serve as versatile biocomposites, sustainable feedstocks for aromatic compounds, and a source of activated carbon well suited for heavy metal absorption.<sup>15–18</sup> Even though dew-retting results in one-third of harvests not reaching textile grade quality, the process can still degum crops to an extent suitable for other commercially relevant, and sustainable, functions.<sup>4,18</sup> However, while the chemical and structural changes of fibrous corps during retting are well studied, the microbial communities responsible are still not fully understood.<sup>19–21</sup> Because microorganisms are responsible for the retting process, it is important to investigate the microbiota of retting crops. Specifically, understanding which microbe-derived Carbohydrate-Active enZYmes (CAZy) are produced at different stages of retting is valuable because of their principal role in retting. In addition to the functional profiles of retting communities, an examination of their phylogenetic profiles and primary members provides context for what factors contribute to successful retting conditions.<sup>22</sup>

Historically, microbiome research has depended on culture-based techniques that generally do not capture the dynamic nature of evolving microbial communities, being limited to a small percentage of culturable taxa.<sup>22</sup> Developments in hightthroughput sequencing (HTS) techniques, like *metabarcoding*, have allowed researchers to characterize microbial communities with high sensitivity and specificity.<sup>23</sup> In contrast with traditional DNA barcoding, which is designed to identify individual specimens within a community, metabarcoding interrogates the entirety of a community's diversity.<sup>24</sup> It does so by amplifying specific genetic elements that vary between species but are flanked by conserved regions that can be used as universal sites for PCR primers.<sup>25</sup> After amplifying the selected regions, the resulting amplicons can be used to calculate abundances of different taxa in a sample. The 16S and 18S/ITS regions of the rRNA genes are commonly used as markers for the identification of bacterial and fungal organisms, respectively.<sup>25</sup> Specifically, the V4-V6 sub-regions of the 16S rRNA gene provide ideal taxonomic resolution while both the ITS1 and ITS2 sub-regions have been shown to produce comparable results.<sup>26,27</sup> The method used to characterise a community's constituents using these barcoding regions is known as *targeted metagenomics*.<sup>22</sup> While potentially powerful, metabarcoding experiments are prone to bias and their results must be understood in the context of the provided—and unprovided—data.

The report presented here seeks to provide a meta-analysis of targeted metagenomic experiments that study the microbiota of retting flax, hemp, kenaf, and jute. The aims of this review are to compare the retting process between different species and treatments while surveying the amount of available data. This will both improve how retting is understood at a functional and community level, while assessing the field's current state. A synthetic aim of this study is to gauge how metabarcoding experiments and their data are reported to provide guidance for future research guidelines. The results presented here will hopefully accelerate the adoption of natural fibres for industrial and textile uses by uncovering the functions and taxonomies of retting communities and providing guidance for future studies. Ultimately, this review aims to improve the efficiency of retting by pooling the results of metabarcoding experiments that analysed the retting process with a variety of crops and treatments.

## Methods

### Systems:

All methods described below were performed using Durham University's Hamilton supercomputer utilizing Rocky Linux (v.8.10).<sup>28</sup> Sequence processing and classification was done using mothur.<sup>29</sup> Metagenomic predictions utilized PICRUSt2.<sup>30</sup> Output files from both softwares were prepared and plotted in RStudio (v.2025.05.1).<sup>31</sup> All scripts used for the data analysis pipeline are stored in a public GitHub repository.<sup>32</sup>

### Literature Review

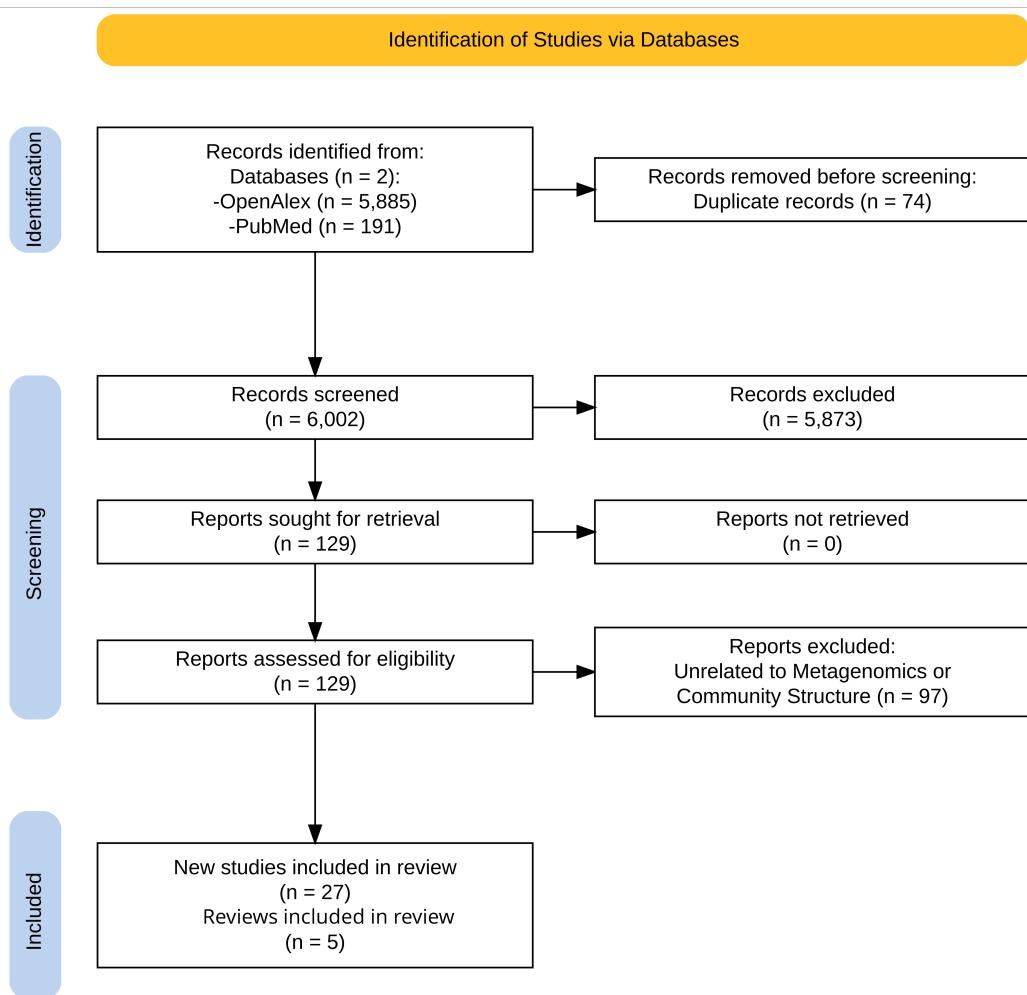
A systematic literature review was conducted to identify publications that examine the microbiology of retting. This approach was chosen in order to better understand the status of retting research and select sources of amplicon sequencing data.

The PubMed and OpenAlex databases were queried for journal articles, articles, and reviews containing specific key words. The queries were performed in RStudio using the openalexR (v.2.0.1) and pubmedR (v.0.0.3) packages.<sup>31,33,34</sup> The databases were searched for publications containing the term "retting" in their title or abstract on August 4, 2025. After removing duplicates (n=74) the resulting 6002 publications were filtered in RStudio. First, publications with notable authors such as *Law, Djemiel, Grec, Akin, Fuller, Norman, Ribeiro, Liu M., Orm, Brown, Sharma, Henriksson, and Fila* were extracted.<sup>19</sup> Publications with titles containing "MICROBIOLOGY OF RETTING" were also selected as papers of interest. The publications were then filtered on the basis of containing the words *hemp, flax, jute, kenaf, or bast* and *rDNA, 16S, 18S, fungi, fungi, fungal, metabar, meta-bar, metagen, microb, or community* in their title or abstract. Off-target results containing the words *Rett syndrome, autism, Sindrome de Rett, Vascular Responsiveness, banana, and pigs* were removed. The remaining 129 publications were then manually evaluated for their data availability and utility to the present study. Each step of the review was performed by a single reviewer using R software. A flow diagram detailing this literature review process was produced following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines using the R package PRISMA2020 and is presented in Figure 1.<sup>35</sup> A full list of the selected publications is also provided (Supplementary Table 1). The database queries were filtered in this way to ensure all papers concerning retting were selected. The additional search terms were chosen due to their associations with targeted metagenomics and microbial retting.

The final set of publications identified by the literature review contained a selection of reviews (n=5) and

studies (n=27) focusing on the microbial communities involved with retting. In addition to procuring a set of publications that largely describe the state of retting research, this literature review also sought to find viable sources of amplicon sequencing data. Of the 27 studies, 13 performed some type of amplicon sequencing targeting bacteria or fungi using kenaf (n=2), jute (n=2), hemp (n=6), and flax (n=3). A summary of these potential data sources is provided in Table 1.

Upon inspecting the data sources, both jute related studies and three hemp studies did not provide raw sequencing data: only publishing specific sequences as OTUs to NCBI's GenBank. Another hemp study was eliminated due to its use of single-end Nanopore reads which would bias and confuse results by producing noise due to poor sequence quality.<sup>36</sup> Two other hemp studies were eliminated because their data was not accessible. One reported that the data was available upon request, and the second reported the data was available under a BioProject accession, PRJNA1124099, that did not link to any publicly available data. Neither corresponding author responded to request. Additionally, a kenaf study was excluded because its data was lost on the MG-RAST server and was reportedly not able to be recovered locally.<sup>37</sup> Thus, the literature review process led to data being selected from five studies that performed targeted metagenomics on samples from flax dew-retting (n=2), flax water-retting (n=1), hemp dew-retting (n=1), and kenaf water-retting (n=1).



**Figure 1:** A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram detailing the steps involved with the literature review

#### Data Collection:

After performing the literature review, five papers were chosen that published raw sequencing data. The fasterq-dump tool from NCBI's SRA Toolkit was used to download FASTQ files from each selected accession from Table 2.<sup>38</sup> Because some experiments took soil and stem samples, focusing on both the ITS and 16S regions, it was decided for consistency's sake to limit the current study to sequences that target the 16S region from plant stem samples. Metadata describing each sample was manually curated from the published reports and are available in the provided GitHub repository. Seven metrics were used: stage of retting, retting period, whether the retting swath was turned, species type, cultivar (plant variety), retting method (water/dew), and treatment. The stage of retting was classified as either “early”, “middle”, or “late”. Samples’ stages were labelled according to the reported retting period and other observations from their respective studies. All treatment types were identified according to what was published and include: standard harvest, late harvest, low moisture, high moisture, high moisture with soil slurry (hereafter referred

Table 1: Studies that Performed Amplicon Sequencing

Crop	Source Country	Year	Retting Type	Raw Reads	Marker(s)	Data Availability and Accession	DOI	Selected
Kenaf:	China	2022	Water	Yes	16S	BioProject (PRJNA850647)	10.1186/s12870-022-03890-5	True
	USA	2013	Water	Yes	16S	Lost on MG-RAST (4513835.3-41.3), requested	10.1007/s10295-013-1242-1	False
Jute:	India	2008	Water	No	16S	GenBank (DQ252375-481, DQ238814-26, EF067808-39)	10.1007/s00248-007-9345-8	False
	Bangladesh	2020	Water	No	16S	GenBank (MH010052-195)	10.1038/s41598-020-61898-z	False
Hemp:	France	2017	Field	No	16S, ITS	GenBank (LT622055-69, LT622070-85)	10.1186/s13568-017-0355-8	False
	USA	2024	Water	Yes	16S	By request, requested	10.3390/pr12081725	False
	USA	2020	Field	Yes	16S	BioProject (PRJNA494847)	10.3390/agronomy10040492	True
	France	2023	Field	Yes	16S, 18S	BioProject (PRJNA911790)	10.1007/s00253-023-12582-5	False
	Italy	2024	Field	Yes	16S, ITS	Non-public BioProject (PRJNA1124099), requested	10.1007/s00253-024-13300-5	False
	France	2015	Field	No	16S, 18S	GenBank (KM507203-90)	10.1007/s00253-014-6356-5	False
Flax:	France	2020	Field	Yes	16S, ITS	BioProject (PRJEB27872)	10.1016/j.indcrop.2020.112255	True
	France	2017	Field	Yes	16S, ITS	BioProject (PRJEB20299)	10.3389/fmicb.2017.02052	True
	China	2016	Water	Yes	16S	BioProject (PRJNA309354)	10.1038/srep31812	True

to as “Soil Slurry”), and bacterial inoculation.

Each study was investigated to identify the primers and 16S regions used in their respective experiments; summarized in Table 2. Four experiments specified the 16S region targeted, three of which also specified the primers. One study did not report the primers or the targeted region used.

Table 2: Primer and 16S Region Summary

Study Accession	Forward Primer	Reverse Primer	Region
PRJEB20299	S-D-Bact-0341-a-S-17	S-D-Bact-0787-a-A-19	V3-V4
PRJEB27872	S-D-Bact-0341-a-S-17	S-D-Bact-0787-a-A-19	V3-V4
PRJNA309354	515F	907R	V4-V5
PRJNA494847	Unspecified	Unspecified	V4
PRJNA850647	Unspecified	Unspecified	Unspecified

### Bioinformatic Methods: Sequence Processing

Community analysis and taxonomic classification was performed using mothur’s MiSeq SOP (v.1.48.3), processing each experiment separately.<sup>29,39</sup> Paired-end FASTQ files from the selected studies were overlapped to form contiguous reads with a quality score threshold of 30, maximum homopolymer length of 8 bp, an

overlap of at least 30 bp, and an additional screen to ensure there were no differences between the known primers and their respective sequences. Specifying the primers used with the sequences from accession PRJNA309354 failed to produce viable contigs, but removing them from the pipeline's parameters fixed this behaviour. The following characteristics were also used to vet sequences: ambiguous bases, mismatches, a minimum/maximum length of 300/500 bp for the V3-V4 and V4-V5 regions, and a maximum length of 275 bp for the V4 region.<sup>39,40</sup> Sequences that targeted unknown regions were not screened for length. The sequences were then aligned to a SILVA (v.138.2) reference database customized for the appropriate region.<sup>41,42</sup> If no region was specified, then alignment proceeded with the original SILVA database. At this point, the accession PRJNA850647 was removed from this study due to its sequences having an average alignment length of 15 and an average start/end position of 23897/23975—likely targeting the V9 region. Moreover, the samples were sequenced with Illumina's NovaSeq platform—as opposed to the other experiments using Illumina MiSeq—and were too large for mothur to process with reasonable memory requirements (64 gigabytes of RAM with 128 processors).

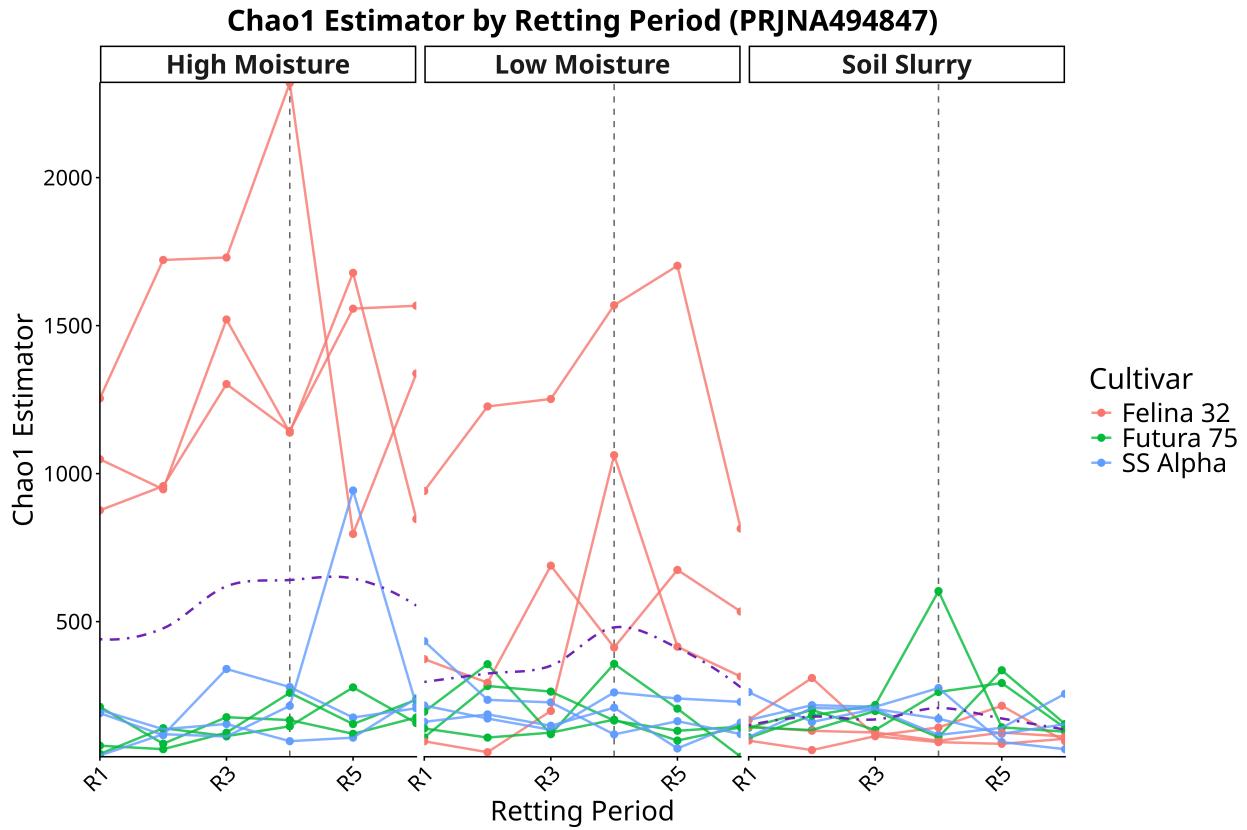
### Bioinformatic Methods: Community Analysis

When the sequences were sufficiently cleaned, they were ready to be classified into operational taxonomic units (OTUs) and subjected to community analysis. As recommended, a pre-clustering of the sequences was performed to reduce the number of erroneous reads. To account for sequencing errors and differences in read length, two mismatches were tolerated for experiments using the shorter V4 region and four differences were allowed for the longer V3-V4 and V4-V5 regions.<sup>41</sup> Next, chimeras that originated from amplification errors during PCR were detected by abundance and removed *de novo* using the UCHIME (v.4.2) algorithm.<sup>43,44</sup> Sequences were then classified with a Bayesian classifier using the Ribosomal Database Project's reference set.<sup>45</sup> Because each experiment used in this study targeted the bacterial rRNA gene, OTUs with lineages of Archaea, chloroplast, mitochondria, Eukaryota, or unknown, were removed.<sup>46</sup> Clustering of the aligned, non-chimeric sequences into OTUs was performed *de novo* using a column-formatted distance matrix with the OptiClust algorithm set at a 3% dissimilarity cut-off until complete convergence of the Matthews correlation coefficient.<sup>47</sup> The sequence with the smallest maximum distance between other sequences within an OTU was selected to be representative of that OTU.

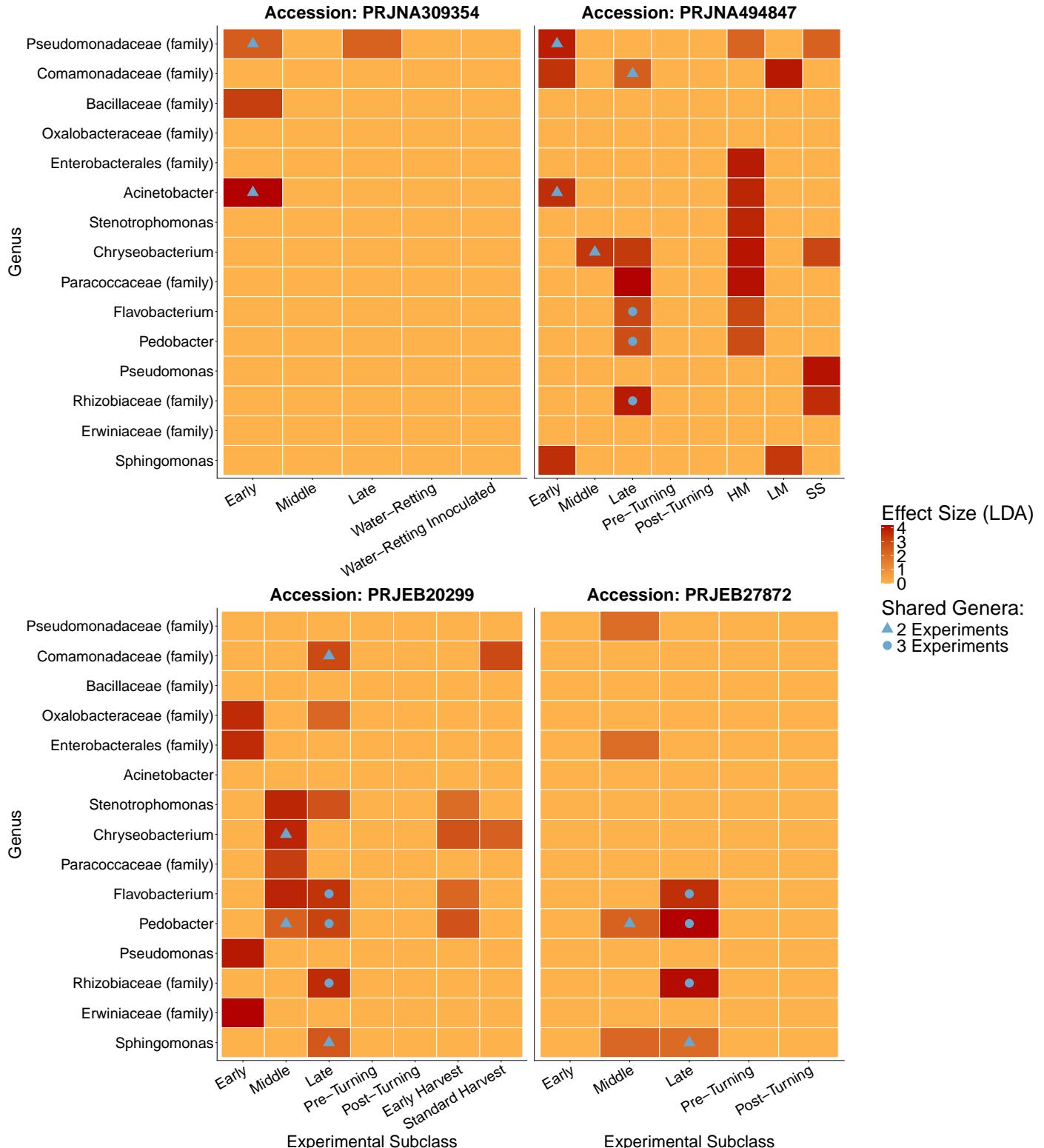
To normalize sequencing depth across samples, all datasets were rarefied to the size of the smallest library prior to diversity analyses. Alpha diversity, which describes the richness and evenness of microbial communities within individual samples, was assessed using the Chao1 estimator, Heip's evenness index, Shannon's diversity index, and the inverse Simpson's index. Community coverage—the extent to which sampling captured a sample's microbial diversity—was estimated using Good's coverage estimator and confirmed with rarefaction curves. Differences in alpha diversity between treatment groups were tested using the Mann-Whitney-Wilcoxon test. Treatment groups with fewer than seven samples were pooled to improve statistical power.

Beta diversity, which reflects differences in community composition between samples in the same experiment, was analyzed using non-parametric analysis of molecular variance (AMOVA) based on the Yue and Clayton theta similarity coefficient. To identify taxa that significantly distinguished experimental groups, Linear Discriminant Analysis Effect Size (LEfSe) was applied. To note, the LEfSe function in mothur (v.1.48.3) can not handle more than two experimental subgroups, so an older version (v.13.7.3) was used for the LEfSe step in the pipeline.

Graphs of the alpha-diversity metrics over time and a heatmap of the LEfSe results were generated in R (Figures 2, 3; Supplementary Figures 1-17).



**Figure 2:** An example plot of the Chao1 estimator for each week of retting coloured by cultivar (plant variety). The dashed vertical lines indicate when the swath was turned. The dotted line represents a smoothed line showing overall trends in richness over time.



**Figure 3:** A heatmap of the genera with the 15 largest effect scores, on average across all experiments. If the same genus was discriminant for a specific subgroup between two experiments it is marked with a circle. Those genera discriminant across three intra-experimental subgroups are marked with a triangle.”

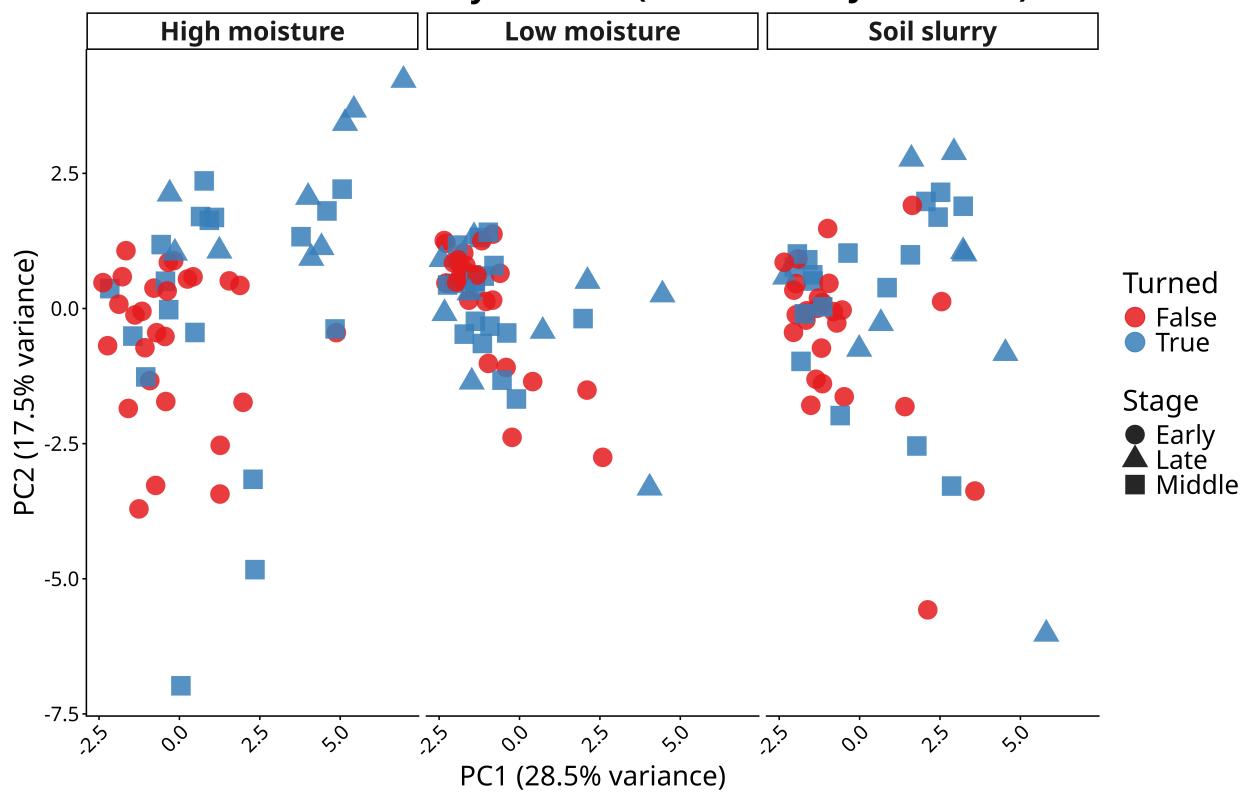
## Bioinformatic Methods: PICRUSt2

A pipeline was developed using PICRUSt2 (v.2.6.2) to predict the functional composition of each experiment's samples.<sup>30</sup> In brief, the pipeline involves three main steps: phylogenetic placement of reads, prediction of gene families per OTU, and pathway inference.<sup>48–53</sup> Essentially, the metabolic potential of a community can be predicted. The representative sequences for each sample's OTUs and their abundances were converted to BIOM formatted count tables and submitted to PICRUSt2. The count tables were denoised by retaining only OTUs that had relative abundances of at least 0.05% in a single experimental sample.<sup>54</sup> The accuracy of PICRUSt2's metagenome predictions was controlled by measuring the weighted Nearest Sequenced Taxon Index (NSTI) with a cutoff of 2. An expertly curated list of CAZy families associated with the degumming process was then used to focus on the most relevant parts of the metagenome for this study.<sup>54</sup> Mothur's taxonomic assignments for each sample's OTUs were also merged with the PICRUSt2 output for downstream analysis.

After determining the functional profiles of the experiments' samples, the functional abundances for each CAZy family in a sample were normalized to relative abundances to allow for meaningful comparisons between samples. Principal component analysis (PCA) was then conducted for each experiment to reduce the dimensionality of the data and to identify patterns of variation between treatment groups, turning statuses, and retting stages (Supplementary Figures 18-21). This unsupervised approach enabled the visualisation of how subgroups cluster across experiments (Figure 4). To parody this global view with statistical comparisons, differential abundance analysis (DAA) was applied to the raw, unnormalized abundance counts of CAZy families for each experiment. This analysis aimed to identify specific functions that were significantly enriched or depleted for any subgroup, including: treatment type, turned status, plant type, retting type, and retting stage. Four DAA methods—DESeq2 (v.1.48.2), edgeR (v.4.6.3), MaAsLin2 (v.1.22), and ALDEx2 (v.1.40.0)—were used to improve robustness and reduce method-specific bias. CAZy families that were identified as differentially abundant by at least two of these methods within any subgroup were retained for downstream analysis.

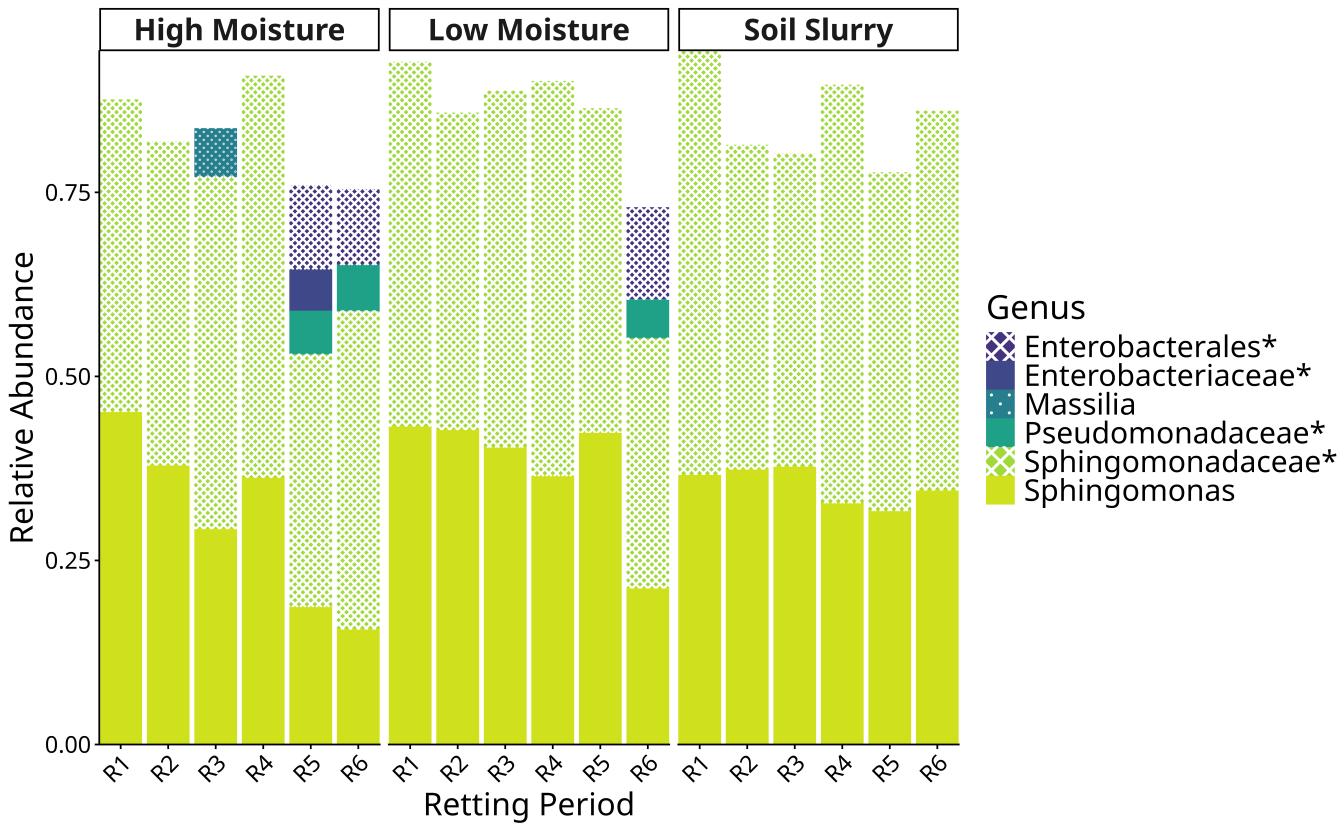
A final analysis was conducted that compared how different genera contribute to the metagenome. This began by determining the total functional abundances for cellulose, hemicelluloses, and pectin within a sample. If a sample had replicates they were combined. The abundances of each genus were then normalized by finding their relative abundance in respect to each substrate. Only the top 10 genera for each substrate were retained for visualisation. Bar graphs were plotted to illustrate genus/function trends over time and between treatments (Figures 5; Supplementary Figures 22-31).

### PCA of Selected CAZy Families (Accession PRJNA494847)



**Figure 4:** An example of a Principal Component Analysis (PCA) plot from the of predicted metagenomes of a retting experiment. Shapes denote the stage of retting and color describes the turning status. Because the PCA was performed globally for each experiment, each graph in the facet can be directly compared.

## Relative Functional Abundance of Pectinase Activity by Genus and Treatment



**Figure 5:** An example stacked bar chart describing how genera contribute to the functional potential of retting communities over time and under different conditions. The x-axis provides the retting period and the y-axis details the relative abundance of the respective genera. Genera marked with “\*” could only be classified by family using mothur.

## Results and Discussion

### Literature Review

The literature review accomplished two goals. It produced a selection of publications that describe the current state of retting research, specifically focused on the microbial communities involved in the retting process. Additionally, it provided a pool of sequencing data that can be used to compare the retting of various fibres under different conditions. It should be noted that every study identified in this review focused on the retting of a single fibrous crop which makes this report the first to compare bacterial communities between retting experiments. It is important to acknowledge that this review may not have surveyed all the relevant research pertaining to the microbiology of retting. While the systematic process presented here is transparent and repeatable, there is no guarantee that the databases queried, or the queries themselves, captured the totality of metabarcoding experiments involving retting. Additionally, this review provides an overview of retting research at a single point in time, and similar reviews in the future may have different

results.

### **Overall Data Quality:**

After vetting potential data sources, it was found that ~38% of the experiments did not provide raw sequence reads necessary for the comparative study presented in this report. An additional ~23% referenced data that is private or inaccessible. Of the available data, five experiments were chosen for analysis that published raw reads for: flax field retting ( $n=2$ ), hemp field retting ( $n=1$ ), flax water-retting ( $n=1$ ), and kenaf water-retting ( $n=1$ ). Among these experiments 60% specified the primers used, 80% reported the target region, and only 20% provided replicate samples (Table 2). No experiments included mock communities in their designs; which, combined with the general absence of replicates, severely undermines the usefulness of the data.<sup>55</sup> All figures exhibited in this report were generated from the experiment that used replicates and used three cultivar types for each treatment (BioProject accession: PRJNA494847). After aligning to the SILVA database, it was found that the sequences from the kenaf study were unusable and the experiment was dropped.

The sequence filtering steps in mothur were not analysed. However, it would be valuable to determine if different experiments or subgroups have higher rates of chimeras or mismatched sequences.

### **Alpha-Diversity:**

All experiments had an average Good's coverage value of 98% with the lowest score being 87%. In terms of differences in alpha-diversity, only samples from BioProject accession PRJNA494847 had any statistically significant results. In general, the inverse Simpson's index is lower during the early stages than when it is turned ( $p < 0.05$ ). The Shannon's index was also elevated for samples in the middle of retting compared to early or late samples ( $p < 0.05$ ). In regards to treatment groups, each treatment had statistically significant differences in their Heip's measure of evenness and Chao1 scores (Supplementary Table 1). Also, as seen in Figure 1, apparent outliers do exist within experiments. For this reason, it is important that studies provide replicates in order to understand whether differences in community structure are real or the result of sampling bias.

### **Beta-Diversity:**

The AMOVA results showed that the bacterial community structure of different intra-experimental subgroups is statistically different for three of the experiments (Supplementary Table 2). Expectantly, samples from different retting periods, stages, and turned statuses have distinct community structures ( $p < 0.05$ ). Additionally, the water-retting treatment group that was inoculated with a strain of pectinolytic bacteria proved to have a statistically different bacterial structure ( $p < 0.05$ ). Notably, the experiment with BioProject accession PRJEB27872 did not yield any significant results from AMOVA, likely limited by its small sample size ( $n=7$ ).

Interestingly, LEfSe showed that retting community subgroups do partly resemble each other. Figure 1 shows that nine out of the top 15 genera with the highest average effect size distinguish the same subgroup between different experiments. Specifically, the genera *Flavobacterium*, *Pedobacterium*, and the family *Rhizobiaceae* are shown to be distinguishing markers of dew-retted swaths in the late stage of retting. To note, Figure 1 does not show all the distinguishing taxa and the data could be used to find more similarities

between experimental subgroups; thereby acting as an indicator of retting the retting process.

#### Predictive Metagenomics:

It is important to acknowledge the limitations of using PICRUSt2 to predict the metagenomes of a community. Generally, this method is not advisable because the results can very easily be skewed by the abundances of different taxa.<sup>56</sup> However, some interesting results were found. First, one experiment reported finding 60 CAZy families within their sample, but the pipeline presented here only found 16.<sup>54</sup> This is most likely due to the fact that fungal communities were excluded from the present study. Additionally, the PCA plots—an example of which is provided in Figure 4—generally show that the metagenomes of samples at different stages of retting get increasingly more variable. Lastly, Figure 5 demonstrates a common trend within the functional abundances of bacterial communities in that ~70% of each samples' functional abundances for a substrate—pectin, cellulose, hemicellulose—can be attributed to ~10 genera or families. However, this finding could be skewed as the most functionally abundant genera were also the most abundant OTUs.

All 16 of the CAZy families identified from the selected list of families were found to be differentially expressed in at least one of the following subgroups: treatment, stage, plant, turned status, or retting type. The sign of these changes was not able to be interrogated.

## Conclusion

Bias and error are inherent to each step of metabarcoding experiments. Excluding the confounding effects that plague all bioinformatic analyses, there are many reasons why targeted metagenomic data can be misleading. Beginning with sample collection, samples may not be representative of heterogeneous environments, and variations in their biomass could inflate abundances of certain species while minimizing others.<sup>57,58</sup> Secondly, DNA extraction methods have been shown to have a differential impact on molecular analyses of microbial communities.<sup>20</sup> The amplification stage also produces variable results owing to: variable amplification rates, primer bias, the PCR protocol used, variations in thermocycler calibrations, reagent concentrations, batch effects, the presence of PCR inhibitors, and PCR errors such as polymorphisms, chimeras, and polymerase misincorporations.<sup>59–63</sup> Additionally, sequencing platforms result in different error profiles and bioinformatic pipelines yield different results.<sup>64,65</sup> The choice of barcoding region also effects error rates. Notably, the shorter V4 region produces significantly less spurious OTUs compared to the longer V3-V4 or V4-V5 regions because incomplete overlapping regions can result in spurious OTUs.<sup>66</sup> Mock communities, sequenced alongside experimental groups, can help qualify the quality of metabarcoding data, but this practice is not yet routine among microbiome researchers.<sup>67,68</sup> Despite the limitations of targeted metagenomics, the technique offers unparalleled insights into the structure and evolution of microbial communities. And, while the extent of metabarcoding's quantitative ability is debated, its results are generally proportionate to the underlying community members.<sup>58</sup> However, studies would benefit from the inclusion of mock communities and following more comprehensive reporting standards.<sup>55</sup>

Due to the small sample sizes of most experiments, the meta-analysis presented here can only speculate as to the nature of bacterial retting communities. Some interesting trends can be observed, such as all samples having >70% of their relevant metagenomic potentials being explained by only a few genera; and that the predicted metagenomes of retting communities generally change over time. An additional finding is that

LEfSe revealed that all three dew-retting experiments—for hemp and flax regardless of treatment group—had the same distinguishing genera/families in the late stage of retting. This could be a useful indicator for gauging the progress of retting.

This study is admittedly quite limited. In addition to being exploratory in nature, these results are based on data that are hard to compare directly. Moreover, it only focused on bacterial communities. It has been shown that fungal populations are crucial for successful dew-retting, and including ITS sequences in this pipeline would provide important context about the microbiology of retting.<sup>21</sup> In general, however, the integrity of future targeted metagenomic experiments would benefit from using replicates, mock communities, and carefully selected regions in their designs.

## References

1. Market Materials Report. (2025).
2. Wang, C. *et al.* Global microplastic fiber pollution from domestic laundry. *Journal of Hazardous Materials* **477**, 135290 (2024).
3. Chen, X., Memon, H. A., Wang, Y., Marriam, I. & Tebyetekerwa, M. Circular Economy and Sustainability of the Clothing and Textile Industry. *Mater Circ Econ* **3**, 12 (2021).
4. Akin, D. E. Linen Most Useful: Perspectives on Structure, Chemistry, and Enzymes for Retting Flax. *ISRN Biotechnology* **2013**, 1–23 (2013).
5. Roy, S. & Lutfar, L. B. Bast fibres: jute. in *Handbook of Natural Fibres* 39–59 (Elsevier, 2012). doi:10.1016/B978-0-12-818398-4.00003-7.
6. Kozłowski, R. M., Mackiewicz-Talarczyk, M., Wielgusz, K., Praczyk, M. & Allam, A. M. Bast fibres: flax. in *Handbook of Natural Fibres* 93–162 (Elsevier, 2020). doi:10.1016/B978-0-12-818398-4.00006-2.
7. Horne, M. R. L. Bast fibres: hemp cultivation and production. in *Handbook of Natural Fibres* 163–196 (Elsevier, 2020). doi:10.1016/B978-0-12-818398-4.00007-4.
8. Xu, J., Aifen Tao, Qi, J. & Wang, Y. Bast fibres: kenaf. in *Handbook of Natural Fibres* 71–92 (Elsevier, 2020). doi:10.1016/B978-0-12-818398-4.00005-0.
9. Jaczynska, K., Ruto, D., Orner, K. & Mehta, S. A comparative life cycle assessment of textile fiber production processes: Hemp versus cotton. *Cleaner Waste Systems* **11**, 100277 (2025).
10. Akin, D. E., Morrison, W. H., Rigsby, L. L. & Dodd, R. B. Plant Factors Influencing Enzyme Retting of Fiber and Seed Flax. *J. Agric. Food Chem.* **49**, 5778–5784 (2001).
11. Akin, D. E., Gamble, G. R., Morrison Iii, W. H., Rigsby, L. L. & Dodd, R. B. Chemical and Structural Analysis of Fibre and Core Tissues from Flax. *J. Sci. Food Agric.* **72**, 155–165 (1996).
12. Pisupati, A., Willaert, L., Goethals, F., Uyttendaele, W. & Park, C. H. Variety and growing condition effect on the yield and tensile strength of flax fibers. *Industrial Crops and Products* **170**, 113736 (2021).
13. *The Biology and Processing of Flax*. (M Publications, Belfast, 1992).
14. Sharma, H. S. S. & Faughey, G. J. Comparison of subjective and objective methods to assess flax straw cultivars and fibre quality after dew-retting. *Annals of Applied Biology* **135**, 495–501 (1999).
15. Marshall, W. E., Wartelle, L. H. & Akin, D. E. Flax Shive as a Source of Activated Carbon for Metals Remediation. *BioRes* **2**, 82–90 (2007).
16. Müssig, J., Amaducci, S., Bourmaud, A., Beaugrand, J. & Shah, D. U. Transdisciplinary top-down review of hemp fibre composites: From an advanced product design to crop variety selection. *Composites Part C: Open Access* **2**, 100010 (2020).
17. Elfaleh, I. *et al.* A comprehensive review of natural fibers and their composites: An eco-friendly alternative to conventional materials. *Results in Engineering* **19**, 101271 (2023).
18. Ross, K. & Mazza, G. Characteristics of Lignin from Flax Shives as Affected by Extraction Conditions. *IJMS* **11**, 4035–4050 (2010).
19. Bou Orm, E., Bergeret, A. & Malhautier, L. Microbial communities and their role in enhancing hemp fiber quality through field retting. *Appl Microbiol Biotechnol* **108**, 501 (2024).
20. Bou Orm, E. *et al.* Estimating the bias related to DNA recovery from hemp stems for retting microbial community investigation. *Appl Microbiol Biotechnol* **107**, 4665–4681 (2023).

21. Fernando, D., Thygesen, A., Meyer, A. S. & Daniel, G. Elucidating field retting mechanisms of hemp fibres for biocomposites: Effects of microbial actions and interactions on the cellular micro-morphology and ultrastructure of hemp stems and bast fibres. *BioRes* **14**, 4047–4084 (2019).
22. Djemiel, C. *et al.* Targeted Metagenomics of Retting in Flax: The Beginning of the Quest to Harness the Secret Powers of the Microbiota. *Front. Genet.* **11**, 581664 (2020).
23. Nizamani, M. M., Zhang, Q., Muhae-Ud-Din, G. & Wang, Y. High-throughput sequencing in plant disease management: a comprehensive review of benefits, challenges, and future perspectives. *Phytopathol Res* **5**, 44 (2023).
24. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**, 2045–2050 (2012).
25. Abdelfattah, A., Malacrino, A., Wisniewski, M., Cacciola, S. O. & Schena, L. Metabarcoding: A powerful tool to investigate microbial communities and shape future plant protection strategies. *Biological Control* **120**, 1–10 (2018).
26. Blaalid, R. *et al.* ITS 1 versus ITS 2 as DNA metabarcodes for fungi. *Molecular Ecology Resources* **13**, 218–224 (2013).
27. Yang, B., Wang, Y. & Qian, P.-Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* **17**, 135 (2016).
28. Rocky Linux 8.10.
29. Schloss, P. D. *et al.* Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**, 7537–7541 (2009).
30. Douglas, G. M. *et al.* PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* **38**, 685–688 (2020).
31. Posit team. RStudio: Integrated Development Environment for R. Posit Software, PBC (2025).
32. Vucelic-Frick, J. [git@github.com:brick1233/targetedMetagenomicsOfFibrousPlants.git](https://github.com/brick1233/targetedMetagenomicsOfFibrousPlants.git). (2025).
33. Aria, M. pubmedR: Gathering Metadata About Publications, Grants, Clinical Trials from ‘PubMed’ Database. 0.0.3 <https://doi.org/10.32614/CRAN.package.pubmedR> (2020).
34. Priem, J., Piwowar, H. & Orr, R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. Preprint at <https://doi.org/10.48550/arXiv.2205.01833> (2022).
35. Haddaway, N. R., Page, M. J., Pritchard, C. C. & McGuinness, L. A. PRISMA2020 : An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews* **18**, e1230 (2022).
36. Schloss, P. 16S microbial analysis of Nanopore data with Mothur. <https://forum.mothur.org/t/16s-microbial-analysis-of-nanopore-data-with-mothur/21512> (2022).
37. Allen, M. S. Kenaf Retting Bacterial Community - Data Inquiry.
38. SRA Toolkit Development Team. NCBI SRA Toolkit.
39. Schloss, P. MiSeq SOP. *MiSeq SOP* [https://mothur.org/wiki/miseq\\_sop](https://mothur.org/wiki/miseq_sop) (2019).
40. Djemiel, C., Grec, S. & Hawkins, S. Characterization of Bacterial and Fungal Community Dynamics by High-Throughput Sequencing (HTS) Metabarcoding during Flax Dew-Retting. *Front. Microbiol.* **8**, 2052 (2017).
41. SILVA Reference. [https://mothur.s3.us-east-2.amazonaws.com/wiki/silva.nr\\_v138\\_2.tgz](https://mothur.s3.us-east-2.amazonaws.com/wiki/silva.nr_v138_2.tgz).

42. Schloss, P. Customize your reference alignment for your favorite region. *Customize your reference alignment for your favorite region* <https://mothur.org/blog/2016/Customization-for-your-region> (2016).
43. Huse, S. M., Welch, D. M., Morrison, H. G. & Sogin, M. L. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology* **12**, 1889–1898 (2010).
44. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimaera detection. *Bioinformatics* **27**, 2194–2200 (2011).
45. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* **73**, 5261–5267 (2007).
46. Wang, Q. & Schloss, P. 16S rRNA reference (PDS).
47. Westcott, S. L. & Schloss, P. D. OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere* **2**, e00073–17 (2017).
48. Howard Hughes Medical Institute. HMMR.
49. Czech, L., Barbera, P. & Stamatakis, A. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* **36**, 3263–3265 (2020).
50. Barbera, P. *et al.* EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Systematic Biology* **68**, 365–369 (2019).
51. Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
52. Ye, Y. & Doak, T. G. A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. *PLoS Comput Biol* **5**, e1000465 (2009).
53. PICRUSt2-SC database.
54. Edgar, R. Singletons. <https://drive5.com/usearch/manual/singletons.html>.
55. Klymus, K. E. *et al.* The MIEM guidelines: Minimum information for reporting of environmental metabarcoding data. *MBMG* **8**, e128689 (2024).
56. Scloss, P. Compare Samples With Sequences of Different Lengths? <https://forum.mothur.org/t/compare-samples-with-sequences-of-different-lengths/22360> (2025).
57. Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J. & Burridge, C. P. A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology* **45**, 373–385 (2020).
58. Lamb, P. D. *et al.* How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* **28**, 420–430 (2019).
59. Shaffer, M. R. *et al.* Observation Bias in Metabarcoding. *Molecular Ecology Resources* e14119 (2025) doi:10.1111/1755-0998.14119.
60. Schloss, P. D., Gevers, D. & Westcott, S. L. Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. *PLoS ONE* **6**, e27310 (2011).
61. Yu, Y., Mai, Y., Zheng, Y. & Shi, L. Assessing and mitigating batch effects in large-scale omics studies. *Genome Biol* **25**, 254 (2024).
62. Potapov, V. & Ong, J. L. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLoS ONE* **12**, e0169774 (2017).
63. Huggett, J. F. *et al.* Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon. *BMC Res Notes* **1**, 70 (2008).

64. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics and Bioinformatics* **3**, lqab019 (2021).
65. Plummer, E. & Twin, J. A Comparison of Three Bioinformatics Pipelines for the Analysis of Preterm Gut Microbiota using 16S rRNA Gene Sequencing Data. *J Proteomics Bioinform* **8**, (2015).
66. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl Environ Microbiol* **79**, 5112–5120 (2013).
67. Marinchel, N. *et al.* Mock community experiments can inform on the reliability of eDNA metabarcoding data: a case study on marine phytoplankton. *Sci Rep* **13**, 20164 (2023).
68. Yeh, Y.-C., Needham, D. M., Sieradzki, E. T. & Fuhrman, J. A. Taxon Disappearance from Microbiome Analysis Reinforces the Value of Mock Communities as a Standard in Every Sequencing Run. *mSystems* **3**, e00023–18 (2018).