Università degli Studi del Piemonte Orientale
"Amedeo Avogadro"
Dipartimento di Scienze e Innovazione Tecnologica

Corso di Dottorato di Ricerca in
Chemistry & Biology - Energy, environmental and food sciences
XXX^ ciclo

TESI

Lignocellulose degradation: a proteomic and metagenomic study

SSD: BIO/10

Dottoranda: Dr.sa Sara Icardi

Coordinatore: Prof. Luigi Panza

Tutor: Prof.sa Maria Cavaletto

Index

Introduction

1.1 Circular economy and biorefineries

Several environmental concerns, such as fossil fuel depletion and climate changes, have triggered development of the bio-based economy, determining growing interest on biorefineries and bioproducts (Clauser *et al*., 2016). Currently, most of the utilized fuels and chemicals are predominantly derived from non-renewable fossil resources, that are gradually being exhausting, despite the world increasing demand for energy and chemicals. Moreover, their consumption causes greenhouse gas emission with a negative impact over the global climate change. In this context, the innovative valorization of renewable resources represents a corner stone in the pursuit of a sustainable future and bio-based economy, also permitting to reduce solid and liquid wastes. Various government initiatives have been launched at international level to support the biofuel, bioenergy and other biochemical production. On December 2015, the European Commission put forward a package to support the transition to a circular economy and several nations, even outside of Europe, such as USA, Japan and China, shared and adopted the circular economy approach (Pratt *et al*., 2016). The hub of such economy is the maintenance of the value of products and materials for as long as possible. As a consequence, waste production, emissions, and resource use are minimized and when a product reaches the end of its life, it is used again to create further value, creating an energy loop, in contrast with the approach "take, make, dispose" of the linear economy (Fig. 1). The circular economy seems to be more sustainable than the linear economic system mainly adopted in the past, since it could reduce resource depletion and environmental pollution (https://www.ellenmacarthurfoundation.org/Circular Economy). On March 2017 in Brussels, the Commission and the European Economic and Social Committee organized the "Circular Economy Stakeholder Conference" and the constitution of the "European Circular Economy Stakeholder Platform" was announced. The Circular Economy Package, adopted by the European

Commission, includes revised legislative proposals on waste to stimulate Europe's transition towards a circular economy. It represents an Action Plan for the Circular Economy that provides a concrete and ambitious program, with measures covering the whole cycle: from production and consumption to waste management and secondary raw materials marketing. The proposed actions will contribute to increase recycling and re-use, benefiting to environment and economy.

Clear targets for reduction, management and recycling of waste were defined:

- recycle respectively 65% and 75% of municipal and packaging waste by 2030;
- reduce landfill to maximum of 10% of municipal waste by 2030;
- promotion of economic instruments to discourage landfilling, banning landfilling of separately collected waste;
- introduction of concrete measures to promote re-use and stimulate industrial symbiosis, converting industrial by-products into raw materials;
- boost the commercialization of greener products and support recovery and recycling schemes, by mean of economic incentives (http://ec.europa.eu/environment/circular-economy/index_en.htm).

Just like petroleum-based refineries, biorefineries can produce several products, like fuels, chemicals and materials but starting from biomass instead of fossil sources. Oil replacement with biomass as raw material for fuels and chemical production represents the driving force for the development of biorefineries. Biorefineries have great potential to utilize several feedstocks, including pulp and paper, food, agricultural, agro-industrial and municipal wastes, algae and dedicated cultures known as energy crops and convert them in value-added products, avoiding the land use competition with food crops (Gavrilescu, 2014). Regarding organic carbon resources, biomass represents

the most abundant renewable resource in nature and its features, like recyclability, ease of accessibility and cheapness make it an eco-attractive and petro-alternative candidate (Zhang *et al*., 2017; Arevalo-Gallegos *et al*., 2017). In the modern "biorefinery" concept, biomasses are not only devoted to energy production but represent important sources for chemicals and biomaterial production with a low environmental impact, improving agro-industrial field activities value. Biorefineries are regarded as the cornerstone of a bioeconomy (Farzad *et al*., 2017) and several economic assessments have been published on biofuel or value-added chemicals and polymer production from lignocellulosic waste biomass (Gallezot, 2012; Sheldon, 2014; Isikgor & Becer, 2015). Environmental aspects are not less important so research and development efforts attempt to maximize profits, protecting product quality, operational safety and respect for the environment.

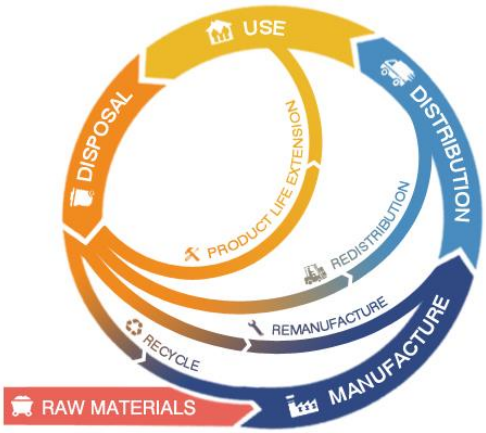


Figure 1: schematic overview of the circular economy.

1.2 Lignocellulose

Lignocellulose is a complex polymeric matrix and represents the most abundantly available biopolymer in nature. Depending on the origin, lignocellulosic biomass (LCB) can be classified in: woods (softwoods and hardwoods) and shrubs, non-food agricultural crops and residues (wheat straw, rise husk, sugar cane bagasse, winemaking residues, etc.) and municipal solid wastes (derived for example from thinning, gardening and road maintenance) (Balat *et al*., 2008; Sanchez & Cardona, 2008). Plants presents at least 35 different cell types, that vary in composition, structure and ultrastructure (Cosgrove, 2005). A common feature however is the presence of a cell wall (0.1-10 μm thick) which gives rigidity to the cell and protection from pathogens attack. Cell walls are made up of three different layers: the middle lamella, the primary cell wall and the secondary cell wall. Primary cell walls are ubiquitous in plant cells, while secondary cell wall matures generally after the end of the growth process. The composition of LCB varies according to the plant species and, even within a single plant, it varies among different plant tissues, stages of growth and other conditions. However, it consists primarily of cellulose (35-50%), hemicelluloses (20-35%), lignin (10-25%) and pectins. Proteins, lipids, soluble sugars and minerals are minor components (Sandgren *et al*., 2005; Pauly & Keegstra, 2008). Lignocellulose is considered the primary building block of plant cell walls, characterized by a complex hierarchical structure (Lee *et al*., 2014) (Fig. 2).

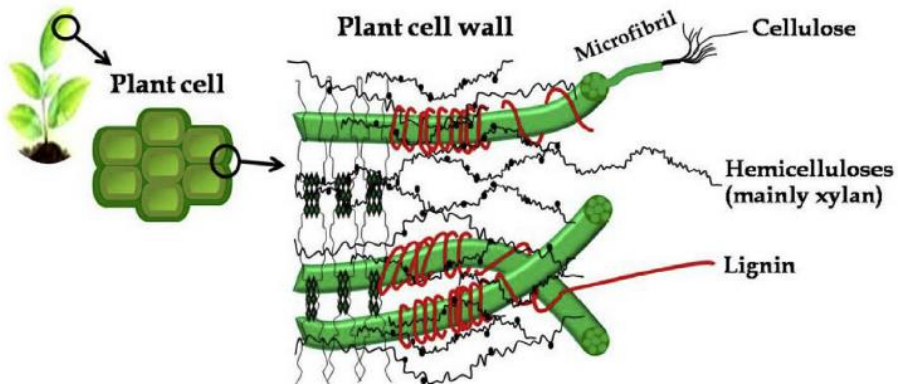


Figure 2: schematic representation of LCB complex structure (adapted from Ratanakhanokchai *et al*., 2013).

1.2.1 Cellulose

Cellulose represents the most abundant organic polymer on the Earth, being the major component of plant biomass. It confers structural support to plant cell walls and it's also present in bacteria, fungi and algae. It can be defined as a linear homopolysaccharide made up of anhydrous β-D-glucose units linked by β-1,4-glycosid bonds (Fernandes *et al*., 2012). The smallest repetitive unit of cellulose is cellobiose, that can be converted into glucose residues (Kumar *et al*., 2008) (Fig. 3a). The degree of cellulose polymerization depends on its source and can vary between 100 and 10.000, being higher in secondary than in primary cell walls (O'Sullivan, 1997; Somerville *et al*., 2004). Cellulose chains are chiral, presenting a reducing and a non-reducing end. The reducing end presents an anomeric carbon not linked to another glucose residue, while in the non-reducing end the anomeric carbon is involved in a glycosidic bond with a glucose residue. Multiple hydroxyl groups on the glucose residues from one chain can form intra- or inter-molecular hydrogen bonds with oxygen atoms on the same or another cellulose chain, respectively, strictly holding chains side-by-side and forming microfibrils responsible for the high tensile strength of cell walls, where cellulose microfibrils

cooperate to the creation of a polysaccharide matrix (Fig. 3b). Immediately after synthesis, in fact, the cellulose chains (20-300) are grouped to form microfibrils, bundled together to form cellulose fibers. The long-chain cellulose polymers are linked by hydrogen, hydrophobic interactions and van der Waals bonds, which cause the cellulose to be packed into microfibrils. This highly organized microfibrillar packing makes it much more resistant to hydrolysis than starch (α-1,4-linked glucan polymer). In plants, the micro-fibril units are about 3 nm wide and contain around 36 cellulose chains, but they are often tightly packed in larger, 20-100 nm microfibril bundles in the secondary cell wall (Persson *et al*., 2004; Himmel *et al*., 2007). In biomass, cellulose is present in both crystalline and amorphous form. Crystalline cellulose represents the major proportion and the most recalcitrant fraction of cellulose. Crystalline regions are separated by less ordered regions, known as ”amorphous cellulose”, that can be broken with strong acid treatments and are more susceptible to enzymatic degradation. The crystallinity of cellulose varies from 50% to 90%, also depending on its source (Hon, 1994). In most conditions, cellulose is wrapped by hemicellulose and lignin (Chen, 2014; Bajpai, 2016).

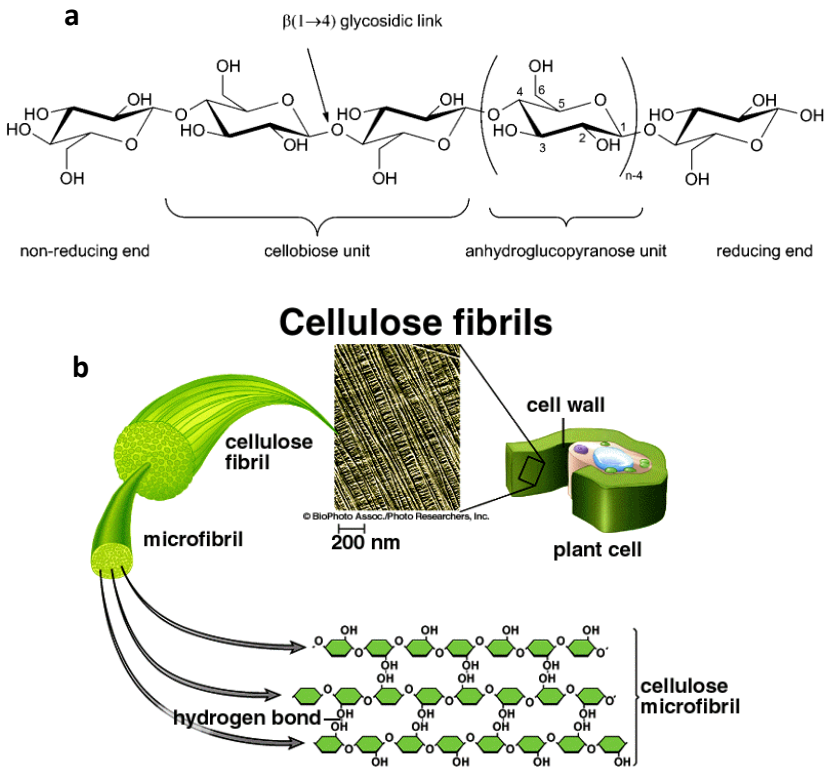


Figure 3: Cellulose chemical structure (a) and cellulose chains interactions for microfibril formation (b) (adapted from "Polysaccharides", General Biology Hub: Learning Resource 3, 2010).

1.2.2 Hemicellulose

Hemicellulose is the second most abundant renewable component of LCB (Kumar *et al*., 2008) and is defined as the cell wall fraction extractible by alkaline solutions (Mohr & Schopfer, 1995). It is not chemically homogeneous like cellulose, representing a heterogeneous mixture of different polymers of pentoses (xylose, arabinose), hexoses (mannose, glucose, galactose) and sugar acids (Saha, 2003; Shallom & Shoham, 2003; Kumar *et al*., 2008). Unlike cellulose, hemicellulose structure and composition change among different cell tissues and plant species due to possible variations of glycosidic bond type, side chain composition and degree of polymerization (Fengel & Wegener, 1989; Jeffries, 1994). The backbone of hemicellulose is either a homopolymer or a heteropolymer with short branches linked by β-1,4 (and occasionally β-1,3)-glycosid bonds. Moreover, it can show even some degree of acetylation, such as in heteroxylan. Hemicelluloses have lower molecular weight compared to cellulose and branches with short lateral chains that are easily hydrolysed (Saha, 2003; Fengel & Wegener 1984). The individual chains in hemicellulose in fact are shorter than in cellulose, with a degree of polymerization of 100-200, usually (Timell, 1967). Unlike cellulose, hemicelluloses change even in composition. For example, in agricultural biomass like straw and grasses, they are mainly composed of xylan while glucomannan predominates in softwood hemicelluloses. Hemicelluloses play an important structural role in cell wall regulating the spatial distribution of principal macromolecular components (cellulose and lignin) and providing their compatibility. In particular, they "coat" cellulose fibrils and it has been proposed that at least 50% of hemicellulose should be removed to significantly increase cellulose digestibility. Byproducts from hemicelluloses degradation, such as furfurals and hydroxymethyl furfurals, can inhibit fermentation processes (Palmqvist & Hahn-Hägerdal, 2000a; Palmqvist & Hahn-Hägerdal, 2000b). The most abundant hemicelluloses are xylans (Fig. 4) followed

by mannans and galactans. Xylans are heteropolysaccharides with homopolymeric backbone chains of 1,4-linked β-D-xylopyranose units. Besides xylose, xylans may contain arabinose, glucuronic acid or its 4-O-methyl ether, and acetic, ferulic, and *p*-coumaric acids. The frequency and composition of branches are dependent on its source (Saha, 2003).

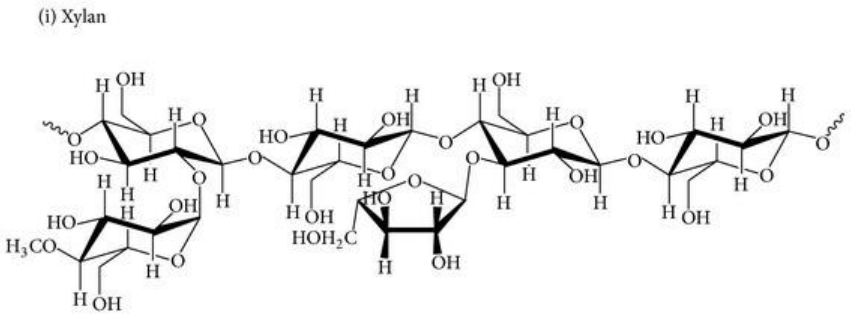


Figure 4: chemical structure of xylan (adapted from Lee *et al*., 2014).

1.2.3 Lignin

Lignin is the most abundant non-carbohydrate component of LCB. It is deposited in the cell wall carbohydrate network during secondary cell wall growth. It is a complex, heterogeneous aromatic polymer, derived from the radical condensation of the aromatic monolignol precursors coumaryl, coniferyl and sinapyl alcohol (Fig. 5). The building blocks of lignin are apparently random linked, creating a very complicated structure. Lignin is also covalently bound to hemicellulose, primarily with ester bonds to xylan, forming a complex matrix that surrounds the cellulose microfibrils. The lignin matrix gives strength to the plant cell wall, and also protects the cell wall from the oxidative stress and the attack by cellulolytic microorganisms (Mohr & Schopfer, 1995). Generally, herbaceous plants such as grasses have the lowest content of lignin, whereas softwoods have the highest

lignin contents (Hendricks & Zeeman 2009). Lignin may be considered as the glue that binds the different components of lignocellulosic biomass together, thus making it insoluble in water. Because of its association with cellulose microfibrils, lignin represents an important deterrent to enzymatic and microbial hydrolysis of LCB (Avgerinos & Wang, 1983) and its removal is essential to increase biomass digestibility (Chang & Holtzapple, 2000). Lignin presence in LCB involves negative effects on biomass susceptibility to enzyme degradation, not just because its role as physical barrier, but even because of non-specific adsorption of hydrolytic enzymes, giving non-productive binding of cellulolytic enzymes, and toxicity of its derivatives to microorganisms. Biomass pretreatments aim to remove lignin, like hemicellulose, to improve biomass digestibility.

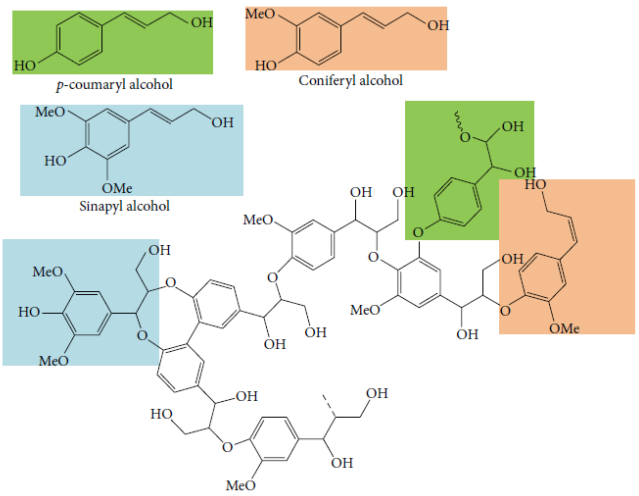


Figure 5: chemical structures of lignin and its precursors (*p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol) (adapted from Lee *et al.*, 2014).

1.2.4 Pectins

Pectins are the third main structural polysaccharide group of higher plants cell walls and represent a major component of the middle lamella. They can be found in abundance even in the primary cell walls of dicotyledonous plants and they play a fundamental role in cell growth (Palin & Geitmann, 2012), mechanical strength (Wolf *et al*., 2009) and mechanisms of defense (Lionetti *et al*., 2012). Moreover, pectin influences various cell wall properties such as porosity, surface charge, pH, and ion balance and therefore is very important for the ion transport (McNeil *et al*., 1984). They are abundant in sugar beet pulp and fruit, like citrus an apple fruit, where they can form up to half of the polymeric content of the cell wall. In contrast with hemicelluloses, pectins can be extracted with hot water and chelating agents (Schulze, 1891). Their use as gelling, thickening and emulsifying agents is widespread in several applications, from food to pharmaceutical products. Pectin is a polymer made up of at least 17 different monosaccharides interconnected through more than 20 different linkages (Ridley *et al*., 2001) so it may have a very complex structure (Fig. 6). The pectin polymer comprises different structural domains covalently linked (Harholt *et al*., 2010). Pectin is predominantly constituted by homogalacturonans, that show a backbone of homo-galacturonic acid regions with neutral sugar side chains made from L-rhamnose, arabinose, galactose and xylose (Kumar *et al*., 2008). Whereas the network of pectins provides the cell wall with the ability to withstand compression, the cross-linking hemicellulose increases the tensile strength of the cellulose and it is organized into a network with the cellulose microfibrils.

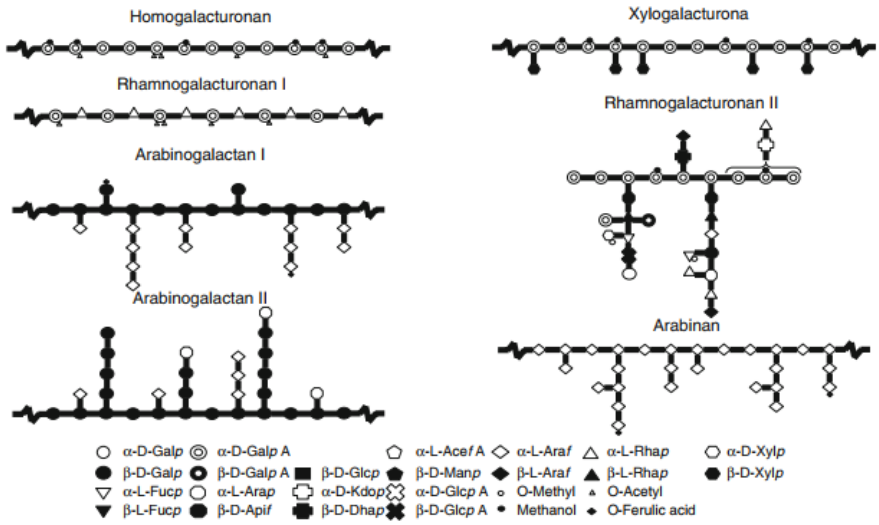


Figure 6: schematic representation of pectin structural elements (adapted from Voragen *et al*., 2009).

1.3 Lignocellulose degrading microorganisms

In nature, several microorganisms including fungi and bacteria show the ability to degrade lignocellulose. In a typical cellulose-degrading ecosystem, several cellulolytic bacteria and fungi cooperate to convert insoluble cellulosic substrates into soluble sugars, successively assimilated by the cells (Wang *et al*., 2016). For this reason, they produce a variety of enzymes, collectively known as cellulases (Bayer *et al*., 1998).

The microorganisms involved in plant cell wall breakdown are both aerobic and anaerobic and two main different strategies of cellulose degradation are known:

- complexed cellulase systems: typical of anaerobic microorganisms, that elaborate high molecular weight enzymatic complexes known as cellulosomes;
- non-complexed cellulase systems: prevalently adopted by aerobic microorganisms, that secrete "free" extracellular enzymes (Himmel *et al*., 2010).

Among anaerobic bacteria, the most studied are *Clostridium thermocellum* (that provided the first evidence of cellulosome), *Clostridium cellulovorans*, *Clostridium cellulolyticum*, *Ruminococcus albus* and *Ruminococcus flavefaciens*. They are commonly found in soil, decaying plant material, herbivorous rumen, gut of some insects, like termites, compost piles, paper mills or sewage sludges, where their presence is enriched by human activities. Anaerobic fungi indeed are found in gastrointestinal tract of ruminants and belong to the phylum *Chytridomycetes*. They use multi-enzymatic complexes, similar to bacterial cellulosomes. In these systems, different cellulose degrading enzymes are assembled on the non-catalytic scaffoldin subunits by means of strong non-covalent protein interactions between dockerins and cohesins domains. Scaffoldins usually contain multiple cohesin modules enabling the assembly of numerous different enzymes and anchor the entire complex to the cell surface through a carbohydrate-binding module (Himmel *et al*., 2010; Hasunuma *et al*., 2012). The cellulosomes allow concerted enzyme activity and enable an optimal

synergism between cellulases; moreover, they minimize the distance over which hydrolytic products could diffuse, enhancing the efficacy of their uptake by the microbial cell (Schwarz, 2001).

Aerobic cellulosolytic bacteria live in soil, humus, water, animal feces or leaf litter and among them are counted *Cellulomonas fimi*, *Cellulomonas flavigena*, *Streptomyces lividans*, *Acidothermus cellulolyticus*, *Thermobifida fusca* (Anderson *et al.*, 2012). Various bacterial strains produce cellulosomes both in aerobic and anaerobic conditions, for example *Rhodospirillum rubrum*, *Clostridium stercorarium*, *Bacillus polymyxa*, *Pyrococcus furiosus*, *Acidothermus cellulolyticus* and *Saccharophagus degradans* (Weber *et al*., 2001; Taylor *et al.*, 2006; Das *et al*., 2007).

Aerobial fungi play an important role in the degradation of lignocellulosic biomasses, can secrete large amounts of cellulases and are found both in soil and in decaying woods. They are the subject of great industrial interest (Lu *et al*., 2013) and they predominantly belong to the phylum *Ascomycota* (genera *Trichoderma, Aspergillus, Penicillium* and *Chaetomium).* Fungi in the genera *Mucor* are the only known *Zygomycota* that can secrete cellulases, like those in the genera *Phanerochaete*, belonging to *Basidiomycota*. Enzymes secreted by the fungi *Trichoderma reesei* and *Aspergillus niger* are the most widely used by the industry (Borin *et al*., 2015).

Regarding lignin, despite its strong resistance to degradation, due to its aromatic polymeric structure and high molecular weight, a complete degradation is mainly operated by microorganisms, especially white rot fungi, like *Phanerochaete chrysosporium* and *Trametes versicolor* and brown-rot fungi, like *Fomitopsis palustris* (Bugg *et al*., 2011). Many bacterial stains are capable to solubilize and metabolize lignin, mainly under aerobic conditions, but bacterial systems are less oxidatively powerful in respect to lignolytic fungi (Brown & Chang, 2014). Currently, the few bacterial species, known to be lignin degraders, belong to the genera *Pseudomonas, Cellulomonas, Streptomyces* and other in the order

Actinomycetales and they secrete extracellular laccases and peroxidases (Lynd *et al*., 2002; Pérez *et al*., 2002; Woo *et al*., 2014; Lotfi, 2014).

1.3.1 Cellulomonas fimi

Bacteria in the genus *Cellulomonas* are the only known for their ability to degrade plant cell wall biomass under both aerobic and anaerobic conditions. *Cellulomonas fimi* is a Gram positive, facultative anaerobe, mesophilic soil bacterium and its genome has recently been sequenced (Christopherson *et al*., 2013). Genome sequences analysis of representative species belonging to *Cellulomonas* genus suggested that Cellulomonads exploit a "secreted enzymes" approach under both aerobic and anaerobic conditions, representing an exception. In fact, no evidence of traditional cellulosome components were found in any studied genome. *C. fimi* genome analysis highlights its capacity to degrade both cellulose and hemicellulose, by meaning of secreted and intracellular cellulases and hemicellulases, acting in synergy (Stackebrandt *et al*., 2006). Several studies confirmed its lignocellulose degrading capabilities. Cellulases encoded by *C. fimi* include:

- two cellobiohydrolases, CbhA (Meinke *et al*., 1994) and CbhB (Shen *et al*., 1995), that remove cellobiosyl units from the non-reducing and the reducing ends of cellulose chains, respectively (Mansfield & Meder, 2003);
- three endoglucanases, CenA (Tomme *et al*., 1996), CenB (Tomme *et al*., 1996) and CenD (Meinke *et al*., 1993);
- an endoglucanase/exoglucanase, CenC (Tomme *et al*., 1996);
- an exoglucanase/xylanase, Cex (Gilkes *et al*., 1991; Hekmat *et al*., 2007);
- at least two intracellular β-glucosidase (Kim & Pack, 1989; Wakarchuk *et al*., 1984).

Cellulomonads species probably transport cellobiose and cleave it intracellularly by means of cellobiose phosphorylase. *Cellulomonas fimi* can even secrete a lytic polysaccharide monooxygenase, responsible for cellulose degradation with an oxidative mechanism.

The hemicellulolytic system of *C. fimi* includes several both secreted and intracellular enzymes, including mannanase (Man26A) (Stoll *et al*., 1999), β-mannosidase (Man2A) (Stoll *et al*., 1999) and xylanase (XynD) (Laurie *et al*., 1997).

Many of the cellulases and hemicellulases codified are multi-domain proteins, showing cellulose-binding capabilities: CenA, CenD, CbhA, CbhB and Cex have a cellulose binding domain family 2 (CBD2) that binds to crystalline and amorphous cellulose. CenB presents an internal family 3 and a C-terminal family 2 CBD. CenC has two CBD4 at the N-terminus and appears to be progressive endoglucanase, hydrolyzing internal glucosidic bonds, then removing cellobiosyl units from the non-reducing end at the initial site of hydrolysis (Tomme *et al*., 1996). XynD has two CBD2, one that binds to xylan but not cellulose, and the other one that binds to crystalline but not to amorphous cellulose (Black *et al*., 1995). *C. fimi* genome also codifies glycosyl transferases, pectate lyases, carbohydrate esterases and enzymes responsible for hexoses and pentoses fermentation. However, the lack of several pentose isomerases allows the only utilization of D-xylose, but not ribitol, arabitol or arabinose (Christopherson *et al*., 2013). Moreover, genes codifying for flagellar protein components were found, in agreement with *C. fimi* reported mobility, considered an important factor in the Cellulomonads cellulolytic strategy (Kenyon *et al*., 2005).

1.4 Biorefinery processes for lignocellulose conversion

The complex LCB structure is responsible for its recalcitrance because strong lignin layers, cross-linking hemicelluloses and high cellulose crystallinity reduce cellulose accessibility to chemicals and enzymes, inhibiting the digestibility of the biomass for sugar monomers production as well as for cellulose solid fraction extraction. Cellulose nanofibers, called nanocellulose, have in fact recently drawn attention, revolutionizing bio-based materials production. Biorefineries technologies thus have yet to overcome recalcitrance issue; therefore, multistep processes are necessary to deconstruct non-cellulosic components of biomasses. The presence of lignin in lignocellulosic biomass is the prior obstacle of biomass recalcitrance and its removal is necessary to enhance biomass digestibility. Hemicelluloses affects cellulose accessibility too, as they coat cellulose fibrils but they are more sensitive to thermochemical treatments. Finally, biomass pretreatments aim to deconstruct LCB matrix, even converting crystalline cellulose in the amorphous phase (Fig. 7).

1.4.1 Biomass pretreatment

Generally, a biomass pretreatment step is necessary to ensure the separation of cellulose component from the other tight bound polymeric constituents of lignocellulosic biomass (Lee *et al*., 2014). The pretreatment step is referred to as the technological bottleneck for anaerobic digestion bioprocesses starting from cost effective lignocellulosic feedstocks. At least 20% of the total production cost is represented by the pretreatment phase, considered the most expensive process step (Yang & Wyman, 2008).

Generally, pretreatment process aims to: (i) promote the immediate formation of sugars, (ii) avoid sugar loss or degradation, (iii) minimize the production of inhibitors for the following processes, (iv) reduce energy consumption and (v) minimize costs (Sarkar *et al*., 2011).

Four different processes can be applied: mechanical/physical, physicochemical, biological, and chemical. The most efficient and cost-effective method for biomass deconstruction is represented by chemical pretreatment. Mechanical pretreatment (chipping, grinding and milling) and thermal methods are less efficient and more energy consuming than chemical ones, while biological pretreatments are proved to be expensive and time consuming (Pedersen & Meyer, 2010).

Regarding physicochemical treatments, the most applied is steam explosion, based on the use of high temperatures and pressures to disrupt biomass internal structures. The chipped biomass is treated with high-pressure saturated steam and then the pressure is swiftly reduced, thus making the materials undergo an explosive decompression. Steam explosion is typically initiated at a temperature of 160-260 °C (corresponding pressure 0.69 - 4.83 MPa) for several seconds to a few minutes before the material is exposed to atmospheric pressure. The process causes hemicellulose hydrolysis, releasing free sugars and lignin transformation. An advantage of this pretreatment is the relatively low dilution of the sugars. Limitations, on the contrary, include destruction of a portion of the xylan fraction, incomplete disruption of the lignin-carbohydrate matrix, and generation of compounds that may be inhibitory to microorganisms used in downstream processes (Mosier *et al*., 2005).

Biological processes by microorganisms and their enzymes selected for delignification are conducted at lower temperatures, since fermentations and anaerobic processes do not require high temperatures. The most used microorganisms are aerobic fungi white rot, brown rot (phylum *Basidiomycota*) and soft rot (phylum *Ascomycota*). Brown rot fungi, like *Coniphora puteana* and *Trichoderma viride*, mainly decompose cellulose and hemicellulose components, having a weak effect on lignin. They slightly modify the lignin molecules, by demethylation or oxidation and hydroxylation but they cannot efficiently decompose lignin aromatic ring (Gao & Xu, 2004; Chen, 2014). Both white rot

and soft rot fungi, including *Poria subacida*, *Polyporus versicolor*, *Pleurotus ostreatus* and *Cheatomium globosum*, act on lignin. Soft-rot fungi can degrade even cellulose but have a slow degradation rate (Zhang *et al*., 2006; Chen, 2014). Mechanical processes do not change biomass composition or state, but reduce its dimensions or segregate its constituents. These processes usually anticipate the proper biorefining, for which a reduction of starting material dimensions is mandatory.

Chemical treatments indeed change the starting material composition, by meaning of chemical reagents. They are represented by: alkaline hydrolysis, oxidation agent, organosolv and ionic liquids. Different types of chemical pretreatment selectively solubilize hemicellulose or lignin components. However, all these chemical treatments effectively remove and recover most of the hemicellulose portions as soluble sugars in aqueous solution.

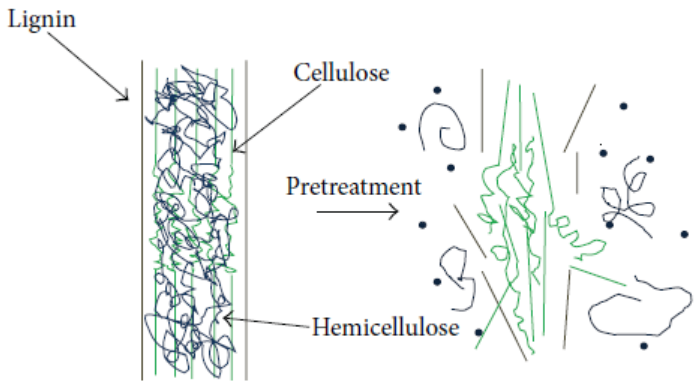


Figure 7: deconstruction of lignocellulose operated by biomass pretreatment.

1.4.2 Biomass hydrolysis

Pretreated LCB is subjected to the hydrolysis process in order to obtain cellulose-derived building blocks, convertible in value-added products (i.e. sugar monomers and nanocellulose fractions). Solubilized hemicellulose is converted into soluble monomeric sugars (hexoses and pentoses), just like the solid fraction of cellulose, from which besides glucose is also obtained nanocellulose, a nano-structured cellulose with extraordinary properties, classified according to physical features, in nanocrystalline cellulose (NCC) and nanofibrillated cellulose (NFC). Nanocellulose has a rigid rod-shaped structure, with a diameter between 1 and 100 nm and a length between tens and hundred nanometers (Sirò & Plackett, 2010; Lavoine *et al*., 2012; Lee *et al*., 2014).

Two different types of hydrolysis are nowadays available: acid hydrolysis and enzymatic hydrolysis. Acid hydrolysis technologies consist of the use of concentrated or diluted acids (i.e. H_2SO_4, HCl) in single-phase or multiphase reactions, respectively. These methods have many problems, such as low chemical recovery, reactor and equipment corrosion, construction costs and extra cost for waste effluents treatment, then are not economically and environmentally sustainable. Moreover, a main disadvantage of acid hydrolysis is represented by the generation of several by-products (i.e. formic acid, acetic acid, levulinic acid, furfural, 5-hydroxymethyl furfural and benzene compounds) with inhibiting effects on subsequent fermentation. Therefore, this process had been gradually replaced by the enzymatic technologies and became the main technology for chemical pretreatments (Lloyd & Wyman, 2005). In contrast to acid technologies, enzymatic hydrolysis exploits cellulosolytic enzymes action, with lower energy consumption and mild operative conditions. It brings several advantages, like high specificity, minor by-products production, high sugar yield, simple equipment and no need for corrosion-resistant materials and acid recovery devices, no polluting substances generation and low waste disposal costs. It’s

suitable for miniature local production where raw material is produced. Therefore, the study of the enzymatic hydrolysis process has recently aroused growing attention (Sun & Cheng, 2002; Taherzadeh & Karimi, 2007; Ferreira *et al.*, 2009; Chen, 2014). Microbial ability to decompose cellulose is known since 1850 and its study has attracted growing attention when it was found in 1906 that cellulase in the snail's digestive juice could break down cellulose. As previously introduced (see paragraph 1.3), the term "cellulases" refers to a group of enzymes that degrade lignocellulose to generate sugar monomers, representing a highly specific biocatalyst for LCB hydrolysis. Generally, fungi, especially belonging to the genera *Trichoderma* and *Aspergillus*, are the major producers of cellulases, that can be also produced by bacteria. Bioconversion of renewable natural lignocellulose represents one of the most advanced technologies able to solve such current world problems as food shortages, energy crises and environmental pollution in view of a future exhaustion of coal, oil, and other mineral raw materials. Anyhow, some issues are still unsolved and search efforts attempt to improve the enzymatic hydrolysis rate and reduce costs, especially due to high enzyme prices (Al-Zuhair *et al.*, 2011). Lignocellulosic waste conversion operated by biological biomass treatments does not require additional energy or chemicals so minimizes the negative impact on the environment and meets the need of eco-friendly refineries (Liguori & Faraco, 2016).

1.4.2.1 Enzymes for plant cell wall polysaccharides degradation

Because of the complicated structure of LCB, it is difficult for each single enzyme to efficiently hydrolyze lignocellulose, thereafter the saccharification process needs enzymatic cocktails, constituted by a large variety of enzymes. Table 1 represents the principal enzymes required to degrade lignocellulose to monomers.

Table 1: Principal enzymes for lignocellulose degradation

Lignin	Laccase, Manganese peroxidase, Lignin peroxidase
Pectin	Pectin methyl esterase, pectate lyase, polygalacturonase, rhamnogalacturonan lyase
Hemicellulose	Endo-xylanase, acetyl xylan esterase, β-xylosidase, endomannanase, β-mannosidase, α-L-arabinofuranosidase, α-glucuronidase, ferulic acid esterase, α-galactosidase, p-coumaric acid esterase
Cellulose	Cellobiohydrolase, endoglucanase, β-glucosidase, lytic polysaccharide monooxigenase

1.4.2.1.1 Cellulases

The enzymatic degradation of cellulose to glucose is generally operated by four distinct classes of enzymes, acting in synergy:

- Exo-1,4-β-glucanases, also known as cellobiohydrolase (CBH), that attack the ends of cellulose chains and hydrolyze the glycosidic bonds in a processive manner, releasing cellobiose or glucose monomers. They have a strong substrate specificity and may show a preference for acting on the reducing (Cbh I) or non-reducing ends (Cbh II) of cellulose chains. Their degradation capacity decreases with the shortening of the oligosaccharide chain (Van Dyk & Pletschke, 2012; Bok *et al*., 1998).
- Endo-1,4-β-glucanases, or endoglucanases (EG), that cleave cellulose chains in the middle and reduce the degree of polymerization. They cut at random at internal amorphous sites in the cellulose polysaccharide chain, generating oligosaccharides of various lengths and consequently new chain ends. The specificity of EG is not strong, so they could hydrolyze water-soluble substituted cellulose derivatives, like carboxymethil-cellulose (CMC). For this reason, they are also known as CMCases (Chen, 2014).
- β-glucosidases (BG), or cellobiases, only active on cello-oligosaccharides and cellobiose, releasing glucose monomers from the cellobiose (Kumar *et al*., 2008). BGs can be extracellular, cell wall-associated and intracellular. Their hydrolysis rate increases with the decrease of the size of the substrate and cellobiose has the highest hydrolysis rate, so BGs can reduce the feedback inhibition of cellulases caused by cellobiose (Banerjee *et al*., 2010; Chen & Li, 2002).
- Lytic polysaccharide monooxygenases (LPMOs), that oxidatively cleave glycosidic linkages, rendering cellulose more susceptible to the hydrolysis operated by the aforementioned conventional cellulases. Their recent

discovery is considered an important breakthrough in the enzymatic degradation of cellulose (Hemsworth *et al*., 2015; Villares *et al*., 2017).

Most of the cellulases and other biomass degrading enzymes show a bimodular organization, having small independently folded carbohydrate-binding modules (CBMs) linked to the proper catalytic domain. Several types of CBMs, with specific binding properties, are known. Generally, cellulose binding domains (CBDs) interact with the surface of crystalline cellulose, leading the catalytic domain on the substrate and enhancing its activity, since the interaction with the substrate is very stable (Bayer *et al*., 1998; Guillen *et al*., 2010).

1.4.2.1.2 Hemicellulases

Hemicellulose is more variable than cellulose in term of composition and therefore various enzymes are required for its complete degradation. In general, hemicellulose degrading enzymes are either glycoside hydrolases (GHs) or carbohydrate esterases (CEs) (Shallom & Shoham, 2003) and can be divided in two different categories:

- Depolymerizing enzymes, that cleave the polysaccharidic backbone and can be divided in endo-acting and exo-acting enzymes, according to their mode of action.
- Substituents removing enzymes, that cleave linkages between substituent groups and the main chain.

Xylan is the most abundant hemicellulose and one of the most abundant resources in nature besides cellulose. In nature, the most abundant forms of xylan are acetyl xylan in hardwood and arabinoxylan in softwood. Their complete hydrolysis is operated by several synergistic enzymes. Among them, the most important are xylanases, including β-1,4-(eso)xylanase, β-1,4-endoxylanase and xylosidase, that degrade xylan to oligosaccharides and xylose.

The degradation of mannan is indeed operated by endo-mannanases and β-mannosidase. Both xylanases and mannosidases are inhibited by the presence of the side-chain substituents, generally linked to xylans and mannans backbone, so different glycosidases, like α-L-arabinosidase, α-galactosidase, α-L-arabinofuranosidases, α-D-glucuronidase, acetyl xylan esterase and feruroyl esterase, are needed to hydrolyze the glycosidic linkage between the main chain and side-chain-substituted groups (Jørgensen *et al*., 2003; Van Dyk *et al*., 2012; Chen, 2014).

1.4.2.1.3 Pectinases

Several enzymes are involved in pectin degradation and belong to hydrolases and lyases class, acting by hydrolysis or trans-elimination, respectively (Turner *et al.*, 2007; Kumar *et al*., 2008). According to their mode of action, pectinases are classified in:

- pectin esterases, that catalyze the de-esterification of pectin methoxyl group, generating pectic acid;
- hydrolases, such as polygalacturonases and polymethylgalacturonases, that respectively catalyze the hydrolytic cleavage of α-1,4-glycosidic bonds in pectic acid and pectin;
- lyases, like polygalacturonate lyase and polymethylgalacturonate lyase, that respectively catalyze the cleavage of α-1,4-glycosidic bonds in pectic acid and pectin, with a trans-elimination reaction, forming unsaturated galacturonates and methyl galacturonates, respectively (Garg *et al*., 2016).

1.4.2.2 Carbohydrate-Active enZymes database: CAZy database

Enzymes that build and breakdown complex carbohydrates and glycoconjugates are collectively designated as "Carbohydrate-Active enZymes", or CAZymes, and annotated in the online resource CAZy database (http://www.cazy.org), since 1998 (Cantarel *et al*., 2009). This database is the only complete source where CAZymes are classified according to their aminoacidic sequence similarity, protein fold and catalytic mechanism.

Currently (September, 2017), CAZy database counts 386 sequence-based protein families, grouped in 5 enzyme classes, according to the related enzyme activities:

- Glycoside hydrolases (GHs), including glycosidases and transglycosidases (Henrissat, 1991; Henrissat & Bairoch, 1993). These enzymes are classified into 145 families that are responsible for hydrolysis and/or transglycosylation of glycosidic bonds and because of their widespread importance for biotechnological and biomedical applications, represent the best biochemically characterized set of enzymes in the database (Cantarel *et al*., 2009).
- Glycosyltransferases (GTs), responsible for the biosynthesis of glycoside bonds from phospho-activated sugar donors (Yip & Withers, 2006) and represented by 104 sequence-based families in CAZy database.
- Polysaccharide lyases (PLs), that cleave the glycosidic bonds of uronic acid-containing polysaccharides by a β-elimination mechanism (Yip & Withers, 2006) and are ascribed to 27 families in CAZy.
- Carbohydrate esterases (CEs), that remove ester-based modifications present in mono-, oligo- and polysaccharides facilitating the action of GHs on complex polysaccharides. They are associated to 16 families in the database.

- Auxiliary activities (AAs), including oxidative enzymes acting in synergy with other CAZymes. This class has been recently added to the database (2013), after the discovery that members of CBM33 and GH61 families were Cu-dependent lytic polysaccharide mono-oxigenases (LPMOs) (Forsberg *et al.*, 2011; Quinlan *et al.*, 2011). LPMOs catalyze the oxidative degradation of cellulose by means of low molecular weight reducing agents, such as ascorbate or reduced glutathione (Westereng *et al*., 2011); moreover, they deconstruct cellulose crystalline regions creating new more accessible sites for the following action of hydrolytic cellulases (Levasseur *et al*., 2013). They act on recalcitrant polysaccharides by a combination of hydrolytic and oxidative mechanism, giving rise to oxidized and non-oxidized chain ends (Dimarogona *et al*., 2013). AA class groups 13 enzymatic families and includes ligninolytic enzymes (families AA1-AA8) and LPMOs (families AA9-AA13). In fact, even if lignin is not a polysaccharidic polymer, it is strictly associated with plant cell wall polysaccharides and ligninolytic enzymes cooperate with classical polysaccharide depolymerases (Levasseur *et al*., 2013).

Finally, the last category found in CAZy database is referred to the associated modules "Carbohydrate-binding modules" (CBMs). CBMs are autonomously folding and functioning protein fragments that have no enzymatic activity but are known to potentiate the activity of many enzyme activities described above by targeting to and promoting a prolonged interaction with the substrate. CBMs are most often associated to the other carbohydrate-active enzyme catalytic modules in the same polypeptide and can target different substrate forms depending on different structural characteristics. This feature set them apart from other sugar binding proteins, like lectins or transporters. Moreover, they can be present in isolated or tandem forms not coupled with any enzyme. 81 families are classified as CBMs.

1.4.3 Fermentation/chemical conversion and value-added products production

Cellulose and hemicelluloses can be hydrolyzed to glucose, xylose and other sugars, that can be converted into value-added products (such as ethanol, acetone, butanol, acetic acid, butanediol, and other liquid fuels and chemical raw materials), exploiting microbial fermentations, both in aerobic and anaerobic conditions. In particular, they represent important starting raw materials for bioenergy and chemical industry fields. The different LCB components in fact can generate different functional products and a complete exploitation of each component is an important challenge for biomass bioconversion (Chen, 2014). Although the biological production of chemicals is fundamental for the sustained development of biorefining technologies, it is not a new technology. In the first half of the 20th century, several commodity products were produced by fermentation, i.e. acetic acid, citric acid, lactic acid, and itaconic acid (Lipinsky, 1981; Dodds & Gross, 2007; FitzPatrick *et al*., 2010).

1.4.3.1 Sugar monomers and bioenergy applications

Lignocellulosic biomass-derived sugar monomers can be fermented into conventional liquid and gaseous fuels: bioethanol, biohydrogen and biogas. Biofuels are defined as "liquid or gaseous transport fuels, such as biodiesel and bioethanol, which are made from biomass" and represent a promising renewable alternative to fossil fuels in the transport sector, contributing to reduce greenhouse gas emissions. By 2020, the European Union aims to obtain the 10% of the transport fuel used in each Union member country from renewable sources, such as biofuels (https://ec.europa.eu/energy/en/topics/renewable-energy/biofuels).

According to the source of sugar monomers used for their production, biofuels are classified in:

- First generation biofuels (1G): derived from edible resources, like crop plants providing energy-containing molecules like sugars, oils and cellulose. Their biofuel yields are limited and they show a negative impact on food security.
- Second generation biofuels (2G): derived from lignocellulosic, non-food materials, such as straw, bagasse, forest residues and energy crops, growing on marginal lands. The replacement of edible sugars with lignocellulosic biomass as feedstocks can reduce the cost of biofuels, in particular of bioethanol, overcome ethical issues and decrease emissions of greenhouse gasses.
- Third generation biofuels (3G): derived from algal biomass and not competing with agricultural food and feed production (Behera *et al*., 2015).
- Fourth generation biofuels (4G): based on the use of genetically modified microorganisms. 4G are just emerging and are still at a basic research level (Aro *et al*., 2016).

Nowadays, 1G and 2G account for 99% of the global biofuels production (Chen, 2014).

1.4.3.1.1 Bioethanol

Bioethanol is one of the most attractive biofuels, since it can be easily produced in large amounts and blended with gasoline or used pure as a "green" fuel. Its higher oxygen content determines a better oxidation of the gasoline and reduces particulate and CO emissions. Moreover, it is a building block to produce several other chemicals, usually petrochemical-based, like acetaldehyde, ethane, ethylene, propylene, butadiene, carbon monoxide or hydrogen (Idriss & Seebauer,

2000; Yu *et al*., 2009; Lippits & Nieuwenhuys, 2010; Oakley & Hoadley, 2010; Song *et al*., 2010).

Bioethanol production consists of four major steps: biomass pretreatment, hydrolysis, fermentation and product recovery through separation or distillation (Zhang *et al*., 2016a; Zhang *et al*., 2016b). In human society, alcoholic fermentation is commonly used since long time in alcohol, traditional brewing fermentation industry and modern industry and the most widely exploited microorganism is the yeast *Saccharomyces cerevisiae*. Since the 91% of energy contained in glucose is converted into ethanol by alcoholic fermentation, the use of lignocellulose to produce ethanol represents an excellent energy conversion technology (Society JE, 2002; Chen, 2014). Xylose metabolism is more complex than glucose metabolism and currently pentose fermentation attracts growing attention, since several microorganisms are not able to ferment xylose, that is often converted into other by-products in the metabolic process. Mono- or co-cultures of several microorganisms are used for bioethanol production, such as *Saccharomyces cerevisiae*, *Escherichia coli*, *Pichia stipitis*, *Zymomonas mobilis*, *Candida shehatae* and genetically-modified microorganisms, able to ferment both pentoses and hexoses, have been developed to improve yields (Bauban *et al*., 2010). Four different types of fermentation process are known:

- separate hydrolysis and fermentation (SHF): hydrolysis and fermentation are sequentially conducted, so each process may occur at its optimal conditions. However, ethanol yield is affected by cellobiose and glucose accumulation, that inhibits cellulose hydrolytic enzymes (Saha *et al*., 2005; Wang *et al*., 2010).
- simultaneous saccharification and fermentation (SSF): cellulose saccharification and fermentation occur together in the same reactor. The feedback inhibition of hydrolytic products is avoided but the process must be conducted at a temperature that is different from both the optimal

conditions of hydrolytic and fermenting enzymes. Then, much search efforts are addressed to identify adapted microorganisms or enzymes, like heat-resistant yeast and bacteria. Generally, in fact the optimal temperature of cellulases is higher than the *optimum* for fermentation.

- simultaneous saccharification and cofermentation (SSCF): enzymatic hydrolysis and xylose and glucose fermentation occur in a single bioreactor, by means of mixed microorganisms (Sanchez & Cardona, 2008) or microorganisms with engineered xylose metabolism, like the xylose-utilizing yeast *Saccharomyces cerevisiae* RWB222 (Wang *et al*., 2010; Zhang *et al*., 2009).
- direct microbial conversion (DMC) or consolidated bioprocessing (CBP): combination in a single step of cellulases production, hydrolysis and fermentation of hexoses and pentoses. Despite the cost-effectiveness of this process, the final ethanol yields obtained are still low because of low strains ethanol tolerance and several by-products production. Among the microorganisms, the most studied for DMC are *Clostridium thermocellum*, *Clostridium thermohydrosulphaircum*, *Thermoanaerobacter ethanolicus* and *Zymomonas mobylis* recombinants (Bjerre *et al*., 1996; Chen, 2007; Chen, 2014; Yang *et al*., 2016).

1.4.3.1.2 Biohydrogen

Hydrogen represents one of the most abundant elements in the universe and is regarded as one of the most promising alternative energy carriers as well as a viable energy option without CO_2 emissions (Chen *et al*., 2013). It can be produced by chemical methods, water electrolysis and biological method (biohydrogen). It's a clean, non-toxic fuel with good combustion properties and the best thermal conductivity of all gasses. Unlike fossil fuels, its combustion releases only water and small amounts of nitride hydrogen. The demand for hydrogen is rapidly expanding and its production must be environmental friendly

and independent from fossil fuels. Hence, the use of lignocellulose as source for hydrogen production represents an important perspective, particularly for agricultural countries. The major biological processes for hydrogen gas production are bio-photolysis of water by algae and dark and photo-fermentation of organic materials, usually carbohydrates, by bacteria (Kapdan & Kargi, 2006). So, biohydrogen can be obtained with different microorganisms: photosynthetic bacteria, cyanobacteria and green algae, non-photosynthetic organisms, as obligate or facultative anaerobic bacteria and archaeas. All these organisms bring a special hydrogen metabolism system. *Chlamydomonas reinhardtii* is one of the well-known hydrogen producing algae (Ghirardi *et al*., 2000; Melis, 2002). The algal hydrogen production is affected by strong inhibiting effect of generated oxygen on hydrogenase enzymes. Moreover, other disadvantages are represented by the low hydrogen production potential and no waste utilization. Therefore, dark and photo-fermentations are considered more advantageous due to simultaneous waste treatment and hydrogen gas production (Kapdan & Kargi, 2006). Non-photosynthetic organisms can produce hydrogen by dark fermentation (i.e. *Clostridium butyricum*, *Clostridium acetobutylicm*, *Escherichia coli*, *Enterobacter aerogenes*, *Azotobacter chroococcum*, *Ruminococcus albus*). Species belonging to the genera *Enterobacter* and *Clostridium* are the most studied fermentative hydrogen-producing strains. They can utilize several substrates to produce hydrogen with reactions catalyzed by nitrogenase or hydrogenase. These substrates include formic acid, lactic acid, pyruvic acid, short-chain fatty acids, glucose, starch, xylose, cellobiose, sulfide and others. Consolidated bioprocessing for hydrogen production permits to reduce costs and improve process yields. Currently, many studies on CBP-based hydrogen production mainly focus on using co-cultures of the thermophilic cellulolytic bacteria *Clostridium thermocellum* and *Caldicellulosiruptor saccharolyticus* (Liu *et al*., 2008; Ivanova *et al*., 2009). However, recent studies have highlighted that

strains in the genus *Thermoanaerobacterium* could also be used for cellulose degradation and biohydrogen production (Cao *et al*., 2014).

1.4.3.1.3 Biogas

The term "biogas" typically refers to a gaseous mixture obtained from the biological breakdown of organic matter in anaerobic conditions. Primarily biogas is constituted by methane and carbon dioxide and may contain small amounts of hydrogen sulfide (H_2S), moisture and CO; however, its chemical composition varies depending on the materials and conditions applied for its production. Biogas is a renewable energy source that can be produced by anaerobic digestion or fermentation of biodegradable materials. Biogas fermentation is a complex process. It can be artificially divided into three basic stages. Initially, the hydrolytic and fermenting bacterial groups, such as *Bacteroides*, *Clostridium* and *Acetivibrio*, convert organic materials to smaller compounds, such as monosaccharides, amino acids, glycerin and fat. In the second phase, the products from hydrolysis are converted mainly into various organic acids (i.e. acetic, propionic, butyric, succinic and lactic acid), alcohols, carbon dioxide and hydrogen through various fermentation reactions. The last step is represented by the methanogenesis, a process involving the fermentation of organic compounds (i.e. organic acids, CO_2 and H_2) producing mainly methane, along with other end products like carbon dioxide, hydrogen and traces of other gases (Goswami *et al*., 2016).

1.4.3.2 Sugar monomers and chemical industry

Cellulose and hemicellulose are polysaccharidic polymers that can be hydrolyzed to glucose, xylose and other sugars, convertible into acetone, butanol, acetic acid, butanediol and other chemical raw materials by microbial fermentation. Even lignin phenylpropanoid derivatives can be further transformed into other chemical products and used as basic organic chemical raw materials for the production of advanced materials, polymers and aromatic aldehydes (Sanchez & Cardona, 2008; Zhang *et al*., 2008; Sannigrahi *et al*., 2010; Santos *et al*., 2011). In Figure 8 is reported a schematic overview of chemical raw materials that could be obtained from natural cellulosic feedstocks.

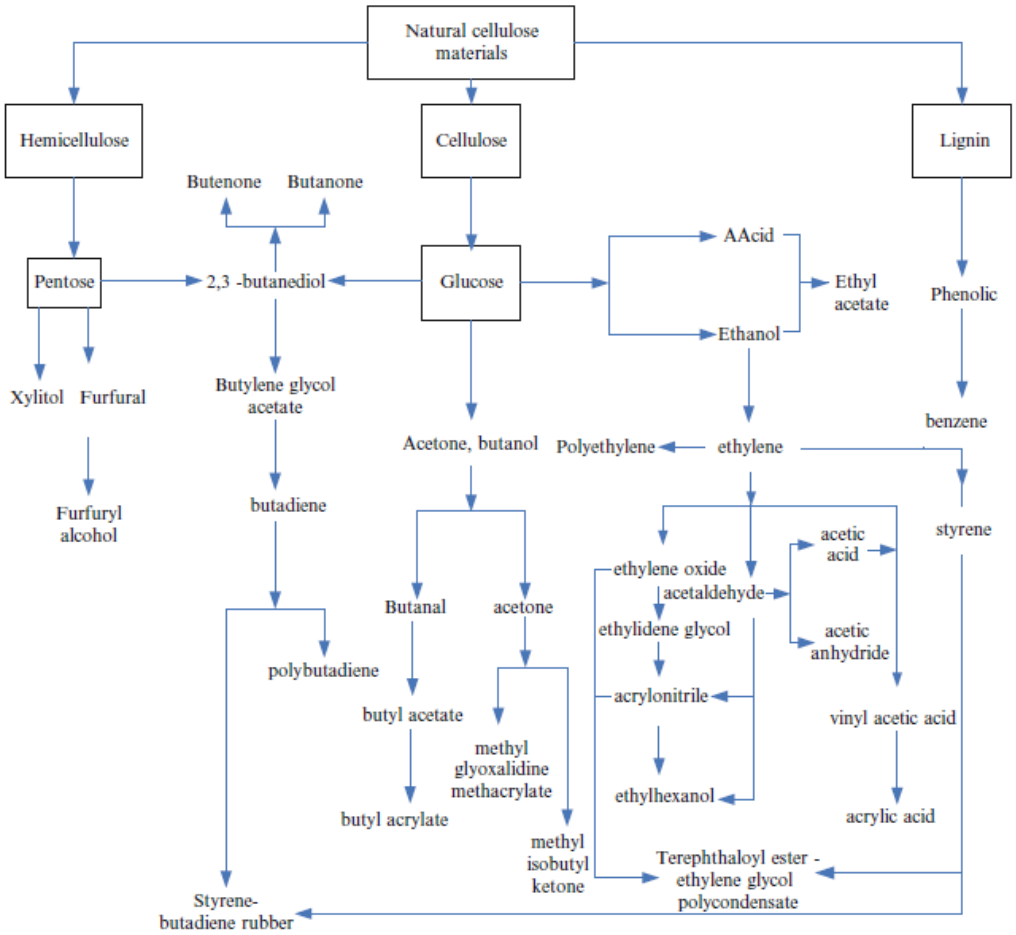


Figure 8: chemical raw materials obtainable from the main lignocellulosic components (adapted from Danner & Brown, 1999).

1.4.3.2.1 Acetone and butanol

Acetone and butanol are important chemicals and organic solvents with many applications, i.e. spray paint, explosives, plastics, pharmaceuticals, plant extract, plexiglass and synthetic rubber. Butanol represents a promising alternative to ethanol, due to its higher calorific value, low vapor pressure and water miscibility. Moreover, it shows almost the same energy value of gasoline and can be used as fuel in existing automobile engines without any modification (Chen, 2014). Butanol may be produced by microbial fermentation by anaerobic organisms such as *Clostridium butyricum* and *Clostridium acetobutylicum*, which provides mixtures of acetone, 1-butanol and ethanol (3:6:1) by ABE (Acetone-Butanol-Ethanol) fermentation; other species indeed produce 1-butanol exclusively (Wu *et al*., 2016). Acetic acid, butyric acid and ethanol are also produced, with the release of CO_2 and H_2. Unlike ethanol fermentation, in butanol fermentation xylose derived from hemicellulose degradation is also able to be used, thus representing an effective way to exploit the hemicellulosic fraction.

1.4.3.2.2 Organic acids

Organic acids are widely distributed in nature and they can be found in animals, plants, and microorganisms *in vivo*. Organic acids like lactic acid, citric acid, gluconic acid, malic acid, succinic acid and acetic acid are widely used in industry and most of them are produced by fermentation from glucose and sucrose.

Diacids like succinic acid are useful in the polymer industry, and fermentation processes using engineered *Escherichia coli* or other organisms are being developed to produce C4 and C6 diacids. Acetic acid can be produced both by chemical synthesis and by fermentation processes. In bio-catalytic processes, it is produced either by oxidative fermentation of ethanol using *Acetobacter* or via the direct fermentation of sugar to acetic acid (Wu *et al*., 2016).

1.4.3.2.2.1 Lactic acid

Lactic acid (LA) is an organic acid and represents an important material for chemical industries. It's recoverable from lactate, obtained from renewable substrates such as glucose, sucrose, lactose, or other sugars after fermentation processes by *Lactobacillus* and *Lactococcus sp.* strains (*L. lactis, L. casei, L. plantarum, L. buchneri*) (Datta & Henry, 2006; Ni *et al*., 2015). Lactic acid can be chemically converted to several important chemicals including methyl lactate, lactide, and polylactic acid (PLA) (FitzPatrick *et al*., 2010). PLA is a biodegradable polymer, non-existing in nature, which is the representative bio-based plastic used in packaging, stationery and containers. This material shows the same barrier property as polyester and the same gloss, transparency and processing ability as polystyrene, with good biodegradability and biocompatibility. The *Lactobacillus* strains can convert cellobiose into lactic acid and this capability represent an important advantage in removing bottlenecks like feedback inhibition by glucose and cellobiose during enzymatic hydrolysis of biomasses (Adsul *et al*., 2007). However, these strains are not able to ferment pentoses and this fact represents a major obstacle for development of PLA industry. The main breakthrough is the recent isolation of the natural *Enterobacter mundtii* QU 25 strain which efficiently metabolized the xylose into lactic acid with a good productivity (Abdel-Rahman *et al*., 2016). Moreover, metabolic engineering could help to solve some problems, like broadening the range of carbon sources used by a microorganism or improving its product yield and productivity (Mazzoli *et al*., 2014).

1.4.3.2.2.2 Levulinic acid

Levulinic acid is an important chemical product, obtained from the decarboxylation of 5-hydroxymethylfurfural (5-HMF), derived from cellulosic and hemicellulosic sugars (Fig. 9). The soluble components obtained after the hydrolysis of cellulose and hemicellulose are mainly saccharides (xylose, glucose, cellobiose and other pentoses from hemicelluloses), sugar aldehydes (i.e. furfural, 5-HMF) and organic acids (such as levulinic acid, formic acid, acetic acid). 5-HMF is obtained both from hexoses and pentoses, derived from cellulose and hemicellulose hydrolysis, respectively. Hexoses are dehydrated to form 5-HMF and then decarboxylated into levulinic acid. Pentoses are subjected to intermolecular dehydration and cyclization, generating furfural after the removal of three molecules of water. Furfural is then oxidized to produce furfuryl alcohol, generating a small amount of levulinic acid in a further degradation process. Levulinic acid can be used as a platform chemical to produce a wide range of value-added products, like methyltetrahydrofuran (MTHF), a solvent and fuel extender (Bozell *et al*., 2000) and polyhydroxyalkanoate (PHA), polymers used in bioplastic production (Wang *et al*., 2013).

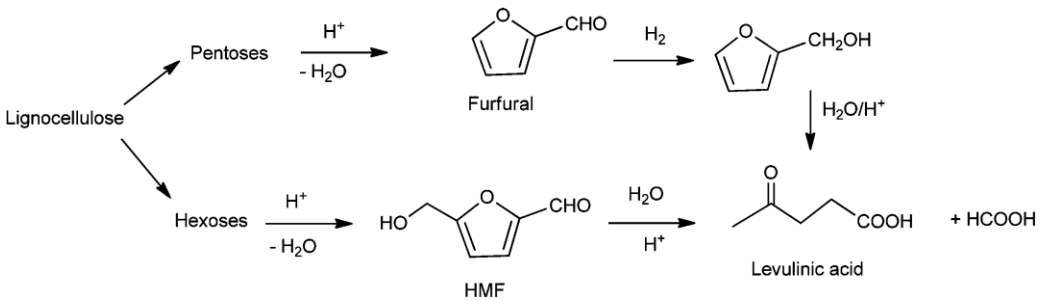


Figure 9: production of levulinic acid from lignocellulose (adapted from Climent *et al*., 2014).

1.4.3.2.2.3 Citric acid

Citric acid is one of the major metabolites in almost all the industrially used microorganisms and is widely used in food, drug and chemical fields. Many microbial species can produce citric acid, such as *Mucor piriformis*, *Paecilomyces divaricatum*, *Penicillium lufeum* and *P. citrinum*, *Trichoderma virid*, *Ustilago vulgaris* and several species belonging to the genus *Aspergillus* (Xu, 1991). However, three main species have particularly industrial value for their higher rates of acid production and their ability to use different sugars as carbon source: *Aspergillus niger*, *Aspergillus awamori* and *Saito aspergillus* (Zhang *et al*., 2001; Chen, 2014).

1.4.3.2.3 2,3-butanediol

2,3-Butanediol is a valuable liquid fuel. It is a chiral compound, colorless and odorless and exists in three stereoisomers: dextro-, levo-, and meso-isomers. Due to its peculiar structure, its chemical synthesis is so expensive and, although the 2,3-butanediol fermentation process has internationally reached the level of the alcohol industry, industrial production has not been realized yet because of the high costs. 2,3-butanediol is also an important chemical building block for the synthesis of several compounds, for example organic solvents, like methyl ethyl ketone, fuel and food additives, rubber monomers, polymer compounds, antifreeze agents, styrene, octane, polybutylene terephthalate resin, γ-butyrolactone and spandex fibers (Chen, 2014). Esterified forms of 2,3-butanediol are precursors in the synthesis of polyimide, applied to drugs, cosmetics and lotions. Because of its wide range of uses, the demand for 2,3-butanediol in the international market is constantly growing. Currently 2,3-butanediol is mainly produced by bacteria (Ji, 2006), including *Klebsiella pneumoniae* and *K. oxytoca* (Garg, 1995), *Bacillus subtilis*, *B. polymyxa* (de Mas *et al*., 1988) and *B. licheniformis*, *Serratia marcescens*, *Aeromonas hydrophila* and *Pseudomonas*

spp. These strains ferment glucose and generate, along with 2,3-butanediol, other important and useful by-products, like ethanol, acetic acid, lactic acid, and succinic acid (Fig. 10) (Chen, 2014).

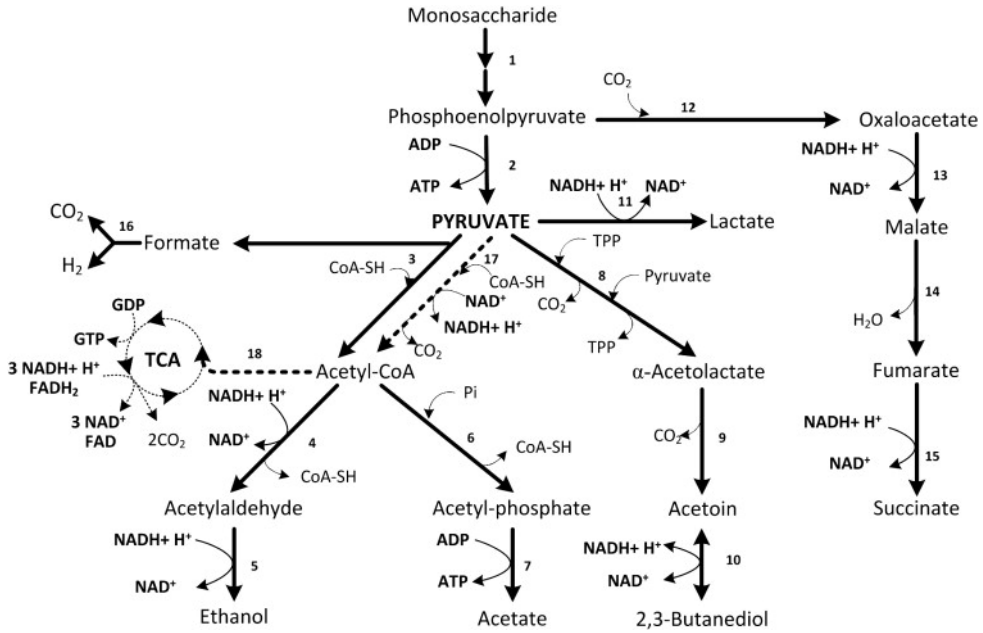


Figure 10: microbial 2,3-butanediol production (adapted from Ji *et al*., 2011).

1.4.3.2.4 Xylitol

Xylitol represents one of the top 12 platform chemicals from agricultural sources identified by the US Department of Energy (Su *et al*., 2015). It's a five-carbon sugar alcohol with broad application prospects in pharmaceutical, alimentary and chemical industries. Its sweetness equals that of sucrose but it is not metabolically involved with insulin, so it can adjust abnormal glucose metabolism *in vivo*, finding use as a nutritional and therapeutic agent for diabetic patients. Moreover, xylitol has strong anti-ketone ability, good thermal stability and anti-caries properties. It's industrially produced by xylose hydrogenation and represents an important derivative of hemicellulosic hydrolysate, just like arabinitol, obtained

from arabinose (Saha, 2003; Nair & Zhao, 2010). Xylose metabolism is well known in microbial cells and starts with an isomerization. First, xylose is converted to xylulose by xylose isomerase and to phosphate xylulose by xylulose kinase. Then, it enters the pentose phosphate cycle (Fig. 11). A series of biochemical reactions convert xylose into ethanol and other metabolites. Yeasts are the best xylitol producers in nature, as only a few bacteria can generate xylitol and filamentous fungi have a low efficiency. In xylose isomerization, metabolic pathways of bacteria, filamentous fungi and yeasts are different. In most bacteria and actinomycetes, it consists in a single reaction catalyzed by xylose isomerase. In yeasts and filamentous fungi, indeed, xylose is initially reduced to xylitol by a NADPH-dependent xylose reductase, and then xylitol is oxidized to xylulose by NAD-dependent xylitol dehydrogenase. The most performant yeast strains for xylitol production mainly belong to the genus *Candida* (*C. guilliermondii*, *C. tropicalis*, *C. mogii*, *C. parasilosis*) (Chen, 2014).

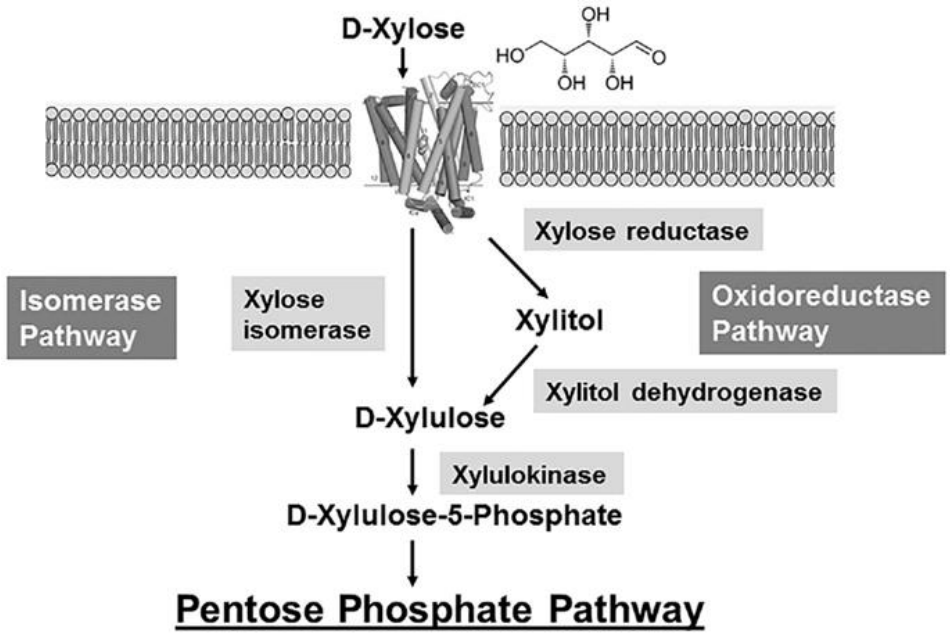


Figure 11: metabolic pathways of D-xylose metabolism (adapted from Nieves *et al.*, 2015).

1.4.3.2.5 Furfural

Furfural is a heterocyclic organic compound, which is a kind of liquid with colorless to yellow color that can be slightly dissolved in water but is soluble in hot water, ethanol, ethyl ether, and benzene (Ren, 2001; Zeitsch, 2000). Its molecular structure consists of a furan ring and an aldehydic group, that give it specific properties, such as hydrogenation, oxidation, chlorination, nitration and condensation. Due to these properties, furfural is used as a raw material for several number of derivative products. Furfural derives from pentoses dehydration and cyclization (Fig. 12) and is mainly used for the synthesis of fibers, synthetic resins, perfumes, paint, and drugs (Wang & Shi, 2008). Two methods are industrially used for furfural synthesis: the one-step method consists in the co-occurring hydrolysis and pentose dehydration cyclization in the same reactor, while the two-steps method is more complex and expensive than the former and consists in the separation between the two processes.

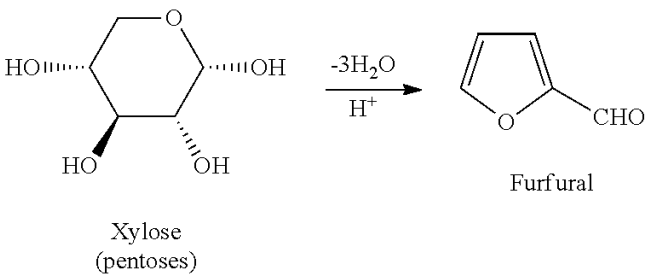


Figure 12: Acid catalyzed dehydration of xylose to furfural.

1.4.3.2.6 Xanthan gum

Xanthan gum is a type of microbial extracellular heteropolysaccharide constituted by glucose, mannose and glucuronic acid in a molar ratio of 2,8:2:2. The main chain of xanthan gum is a repeating unit of β-1,4-glycosidic bonds formed by the two D-glucose molecules, similarly to cellulose. Xanthan gum is a green, non-toxic, safe substance, widely used in the food industry and one of the most superior biological glues. Its most significant properties include high viscosity at low concentrations and stability for a wide range of temperatures and pH, even in the presence of salts (de Mello Luvielmo *et* al., 2016). It can be produced in batch, semibatch or continuous fermentation using several species in the genus *Xanthomonas*, like *Xanthomonas campestris, X*. *phaseoli*, *X*. *malvacearum* and *X*. *carotae* (Krishna Leela & Sharma, 2000).

1.5 Lignocellulosic biomasses utilized in this study

1.5.1 Wheat straw

Wheat straw represents an agricultural by-product, derived from wheat plant, a grass widely cultivated for its seeds, belonging to the genera *Triticum*. The most widely cultivated variety is *Triticum aestivum* (Fig. 13). Straw, in particular, is constituted by the dry stalks of cereal plants after the threshing and the removal of grains and generally accounts for a half of aerial biomass weight. It can be used by breeders as integrative element of forage or as stables litter. Sometimes it's buried and used as fertilizer, contributing to increase organic matter content of lands. Wheat straw is produced globally in large quantities (Saini *et* al., 2015) and in the last year was the second most-produced cereal after maize, due its world production of 749 million tonnes (World food situation: FAO cereal supply and demand brief, 2016). Although its chemical composition can vary (according to plant cultivar or harvesting method, for example), wheat straw is composed by 38-40% cellulose, 21-26% hemicellulose and 11-23% lignin (Khan & Mubeen, 2012). Hence, it represents an attractive substrate for second generation bioethanol production not competing with food production. Much research is devoted to develope optimized biorefining technologies for its pretreatment, enzymatical saccharification and fermentation (Collins *et al*., 2014).

Figure 13: *Triticum aestivum*

1.5.2 Sugar cane bagasse

Sugar cane bagasse is the fiber residue obtained from sugar cane after the milling process for sugar production (Loh *et al*., 2013). Sugar cane is a large tropical perennial grass belonging to the genera *Saccharum*. It grows in clumps constituted by many strong unbranched stems. Several varieties are widespread cultivated but the most representative one is *Saccharum officinarum* (Fig. 14), that can accumulate large amounts of sucrose in its stems. The first producer of sugar cane all over the world is Brazil, where this plant is used both in alimentary and energetic sector for the extraction of sucrose and for first and second-generation biofuels production. In Brazil, sugar cane has been defined as the feedstock showing the best performance. The reasons rely in its efficient biochemical photosynthesis pathway, the C4 route (Pessarakli, 1997), its growth in a tropical environment, where sun and water are abundant throughout the year and its ready to use soluble carbohydrate (Miranda, 2008; Villela Fillho *et al*., 2011). Sugar cane bagasse is generally constituted by 50% cellulose, 25% hemicellulose and 25% lignin (Pandey *et al.*, 2000; Modani & Vyawahare, 2013) and represents a raw material for second generation biofuels and biomaterials production (Premjai *et al*., 2014) and paper and building industry (Gartside *et al.*, 1981; Fairbairn *et al*., 2010; Poopak & Reza, 2012; Madurwar *et al.*, 2014); moreover, it can be used as fuel in biorefineries when burned (Modani & Vyawahare, 2013).

Figure 14: *Saccharum officinarum*

1.5.3 *Arundo donax*

Arundo donax, commonly known as giant cane, is one of the largest herbaceous grasses, belonging to the *Poaceae* family (Fig. 15). It's an invasive perennial plant, widely found in subtropical and warm temperate regions all over the world. It is adapted to a wide variety of ecological conditions and even if it's generally associated with riparian and wetland systems, *A. donax* is almost ubiquitous and shows a strong metal tolerance. Given its capacity to grow vigorously even in marginal lands and the high biomass production, *A. donax* is since long considered a promising energy crop (Mariani *et al*., 2010; Lemons & Silva, 2015). Carbohydrates account for about 83% of its dry weight. In particular, *Arundo donax* contains an average of 22% lignin (Shatalov *et al*., 2001), 42% cellulose and 27% hemicelluloses (even if these values could vary between different parts of the plant). Besides its use as energy crop, *A. donax* is also considered a promising medicinal plant (Al-Snafi, 2015).

Figure 15: *Arundo donax*

References

Abdel-Rahman MA, Tashiro Y, Zendo T, Sakai K, Sonomoto K. Highly efficient L-lactic acid production from xylose in cell recycle continuous fermentation using Enterococcus mundtii QU 25. *RSC Advances*. 2016; 6: 17659-17668.

Adsul M, Khire J, Bastawde K, Gokhale D. Production of Lactic Acid from Cellobiose and Cellotriose by Lactobacillus delbrueckii Mutant Uc-3. *Appl Environ Microb*. 2007; 73(15): 5055-5057. DOI:10.1128/AEM.00774-07.

Al-Snafi AE. The constituents and biological effects of *Arundo donax* - a review. *International Journal of Phytopharmacy Research*. 2015; 6(1): 34-40.

Al-Zuhair S, Ramachandran KB, Farid M, Aroua MK, Vadlani P, Ramakrishnan S, Gardossi L. Enzymes in Biofuels Production. 2011; *Enzyme Research*. DOI:10.4061/2011/658263.

Anderson I, Abt B, Lykidis A, Klenk HP, Kyrpides N, Ivanova N. Genomics of Aerobic Cellulose Utilization Systems in Actinobacteria. *PLoS One*. 2012; 7(6): e39331. DOI:10.1371/journal.pone.0039331.

Arevalo-Gallegos A, Ahmad Z, Asgher M, Parra-Saldivar R, Iqbal HMN. Lignocellulose: A sustainable material to produce value-added products with a zero waste approach- A review. *Int J Biol Macromol*. 2017; 99: 308-318.

Aro EM. From first generation biofuels to advanced solar biofuels. *Ambio*. 2016; 45: S24–S31.

Avgerinos GC, Wang DIC. Selective delignification for fermentation of enhancement. *Biotechnol Bioeng*. 1983; 25: 67-83.

Balat M, Balat H, Oz C. Progress in bioethanol processing. *Prog Energ Combust*. 2008; 34: 551-573.

Banerjee S, Mudliar S, Sen R, Giri B, Satpute D, Chakrabarti T, Pandey RA. Commercializing lignocellulosic bioethanol: technology bottlenecks and possible remedies. *Biofuel Bioprod Bior*. 2010; 4: 77-93.

Bauban B, Inoue H, Yano S, Tanapongpipat S, Ruanglek V, Champreda V, Pichyangkura R, Rengpipat S, Eurwilaichitr L. Bioethanol production from ball milled bagasse using an on-site produced fungal enzyme cocktail and xylose-fermenting Pichia stipitis. *J Biosci Bioeng*. 2010; 110: 18-25.

Bayer EA, Chanzyt H, Lamed¢ R, Shoham Y. Cellulose, cellulases and cellulosomes. *Curr Opin Struc Biol.* 1998; 8: 548-557.

Behera S, Singh R, Arora R, Sharma NK, Shukla M, Kumar S. Scope of algae as third generation biofuels. *Front Bioeng Biotechnol*. 2015; 2(90): 1-13. DOI: 10.3389/fbioe.2014.00090.

Bjerre AB, Olesen AB, Fernqvist T. Pretreatment of wheat straw using combined wet oxidation and alkaline hydrolysis resulting in convertible cellulose and hemicellulose. *Biotechnol Bioeng*. 1996; 49: 568-577.

Black GW, Hazelwood GP, Millward-Sadler SJ, Laurie JI, Gilbert HJ. A modular xylanase containing a novel noncatalytic xylan-specific binding domain. *Biochem J*. 1995; 307: 191–95.

Bok JD, Yernool DA, Eveleigh DE. Purification, characterization, and molecular analysis of thermostable cellulases CelA and CelB from Thermotoga neapolitana. *Appl Environ Microbiol*. 1998; 64: 4774-81.

Borin GP, Sanchez CC, de Souza AP, de Santana ES, de Souza AT, Paes Leme AF, Squina FM, Buckeridge M, Goldman GH, Oliveira JV. Comparative secretome analysis of Trichoderma reesei and Aspergillus niger during growth on sugarcane biomass. *PLoS One*. 2015; 10(6): e0129275. DOI: 10.1371/journal.pone.0129275.

Bozell JJ, Moens L, Elliott DC, Wang Y, Neuenscwander GG, Fitzpatrick SW, Bilski RJ, Jarnefeld JL. Production of levulinic acid and use as a platform chemical for derived products. *Resour Conserv Recy*. 2000; 28(3-4): 227-239. DOI: 10.1016/S0921-3449(99)00047-6.

Brown ME, Chang MCY. Exploring bacterial lignin degradation. *Curr Opin Chem Biol*. 2014; 19: 1-7.

Bugg TDH, Ahmad M, Hardiman EM, Singh R. The emerging role for bacteria in lignin degradation and bio-product formation. *Curr Opin Biotech*. 2011; 22: 394-400.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: The Carbohydrate-Active EnZyme database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*. 2009; 37: 233-238.

Cao G, Zhao L, Wang A, Wang Z, Ren N. Single-step bioconversion of lignocellulose to hydrogen using novel moderately thermophilic bacteria. *Biotechnol Biofuels*. 2014; 7: 82.

Chang VS, Holtzapple MT. Fundamental factors affecting biomass enzymatic reactivity. *Appl Biochem Biotechnol*. 2000; 84: 5–37.

Chen H. Biotechnology of Lignocellulose - Theory and Practice. 2014. Springer Dordrecht Heidelberg New York London (Ed.), ISBN 978-94-007-6897-0. DOI: 10.1007/978-94-007-6898-7.

Chen HZ, Li ZH. Lignocellulosical microorganisms and biomass total utilization. *Biotechnol Inf*. 2002; 2: 25-9.

Chen M. Study on key technologies in ethanol production from corn stover [dissertation]. Zhejiang University; 2007.

Chen R, Wang Y, Liao Q, Zhu X, Xu T. Hydrolysates of lignocellulosic materials for biohydrogen production. *BMB Reports*. 2013; 46(5): 244-251. DOI: 10.5483/BMBRep.2013.46.5.038.

Christopherson MR, Suen G, Bramhacharya S, Jewell KA, Aylward FO, Mead D, Brumm PJ. The genome sequences on Cellulomonas fimi and "Cellvibrio gilvus" reveal the cellulolytic strategies of two facultative anaerobes, transfer of "Cellvibrio gilvus" to the genus Cellulomonas, and proposal of Cellulomonas gilvus sp. nov. *PLoS One*. 2013; 8: e53954. DOI: 10.1371/journal.pone.0053954.

Clauser NM, Gutiérrez S, Area MC, Felissia FE, Vallejos ME. Small-sized biorefineries as strategy to add value to sugarcane bagasse. *Chem Eng Res Des*. 2016; 107: 137-46.

Climent MJ, Corma A, Iborraa S. Conversion of biomass platform molecules into fuel additives and liquid hydrocarbon fuels. *Green Chem*. 2014; 16: 516-547.

Collins SRA, Wellner N, Martinez Bordonado I, Harper AL, Miller CN, Bancroft I, Waldron KW. Variation in the chemical composition of wheat straw: the role of tissue ratio and composition. *Biotechnol Biofuels*. 2014; 7: 121.

Cosgrove DJ. Growth of the plant cell wall. *Nat Rev Mol Cell Biol*. 2005; 6: 850-861.

Danner H, Braun R. Biotechnology for the production of commodity chemicals from biomass. *Chem Soc Rev*. 1999; 28: 395–405.

Das M, Royer TV, LeV LG. Diversity of fungi, bacteria, and actinomycetes on leaves decomposing in a stream. *Appl Environ Microbiol*. 2007; 73: 756-767.

Datta R, Henry M. Lactic acid: recent advances in products, processes and technologies - a review. *J Chem Technol Biotechnol*. 2006; 81: 1119-1129.

de Mas C, Jansen NB, Tsao GT. Production of optically active 2,3-butanediol by Bacillus polymyxa. *Biotechnol Bioeng*. 1988; 31: 366-77.

de Mello Luvielmo M, Dellinghausen Borges C, de Oliveira Toyamac D, C Tondo Vendruscolo, Scamparini ARP. Structure of xanthan gum and cell ultrastructure at different times of alkali stress. *Braz J Microbiol*. 2016; 47: 102–109. DOI: 10.1016/j.bjm.2015.11.006.

Dimarogona M, Topakas E, Christakopoulos P. Recalcitrant polysaccharide degradation by novel oxidative biocatalysts. *Appl Microbiol Biot*. 2013; 97: 8455-8465. DOI: 10.1007/s00253-013-5197-y.

Dodds DR, Gross RA. CHEMISTRY: chemicals from biomass. *Science*. 2007; 318: 1250–1251.

Fairbairn EM, Americano BB, Cordeiro GC, Paula TP, Toledo Filho RD, Silvoso MM. Cement replacement by sugar cane bagasse ash: CO2 emissions reduction and potential for carbon credits. *J Environ Manage*. 2010; 91(9):1864-71. DOI: 10.1016/j.jenvman.2010.04.008.

Farzad S, Mandegari MA, Guo M, Haigh KF, Shah N, Görgens JF. Multi‐product biorefineries from lignocelluloses: a pathway to revitalization of the sugar industry?. *Biotechnol Biofuels*. 2017; 10: 87.

Fengel D, Wegener G. Wood chemistry, ultrastructure, Reactions. Walter de Gruyter, Berlin. 1984.

Fengel D, Wegener G. Wood: Chemistry, Ultrastructure, Reactions. Berlin: Walter de Gruyter. 1989; 613.

Fernandes DLA, Pereira SR, Serafim LS, Evtunguin DV, Xavier AMRB. 2012. Second Generation Bioethanol from Lignocellulosics: Processing of Hardwood Sulphite Spent Liquor, Bioethanol, Prof. Marco Aurelio Pinheiro Lima (Ed.), ISBN: 978-953-51-0008-9.

Ferreira S, Durate AP, Ribeiro MHL, Queiroz JA, Domingues FC. Responce surface optimization of enzymatic hydrolysis of Cistus ladanifer and Cytisus striatus for bioethanol production. *Biochem Eng J*. 2009; 45: 192-200.

FitzPatrick M, Champagne P, Cunningham MF, Whitney RA. A biorefinery processing perspective: Treatment of lignocellulosic materials for the production of value-added products. *Bioresource Technol*. 2010; 101: 8915-8922. DOI: 10.1016/j.biortech.2010.06.125.

Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenstrøm Y, MacKenzie A, Sørlie M, Horn SJ, Eijsink VG. Cleavage of cellulose by a CBM33 protein. *Protein Sci*. 2011; 20: 1479-1483.

Gallezot P. Conversion of biomass to selected chemical products. *Chem Soc Rev*. 2012; 41:1538-58.

Gao PJ, Xu P. Resources and environmental microbial technology. *Beijing: Chemical Industry. Press*; 2004.

Garg G, Singh A, Kaur A, Singh R, Kaur J, Mahajan R. Microbial pectinases: an ecofriendly tool of nature for industries. *3 Biotech*. 2016; 6:47. DOI 10.1007/s13205-016-0371-4.

Garg SK, Jain A. Fermentative production of 2,3-butanediol: a review. *Bioresource Technol*. 1995; 51: 103-9.

Gartside G, Langford-Csiro NG. Developing the potential of Australian bagasse as a paper resource. *Proceedings of Australian society of sugar cane technologists*. 1981; 232-238.

Gavrilescu M. Biorefinery systems: an overview. In: Gupta VG, Tuohy M, Kubicek CP, Saddler J, Xu F, editors. Bioenergy Research: Advances and Applications. 2014; 219-241.

Ghirardi ML, Zhang L, Lee JW, Flynn T, Seibert M, Greenbaum E, Melis A. Microalgae: a green source of renewable H2. *Tibtech*. 2000; 18: 506–11.

Gilkes NR, Henrissat B, Kilburn DG, Miller RC Jr, Warren RA. Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiology Review*. 1991; 55: 303-315.

Goswami R, Chattopadhyay P, Shome A, Banerjee SN, Chakraborty AK, Mathew AK, Chaudhury S. An overview of physico-chemical mechanisms of biogas production by microbial communities: a step towards sustainable waste management. *3 Biotech*. 2016; 6: 72.

Guillen D, Sanchez S, Rodigruez-Sanoja R. Carbohydrte-binding domains: multiplicity of biological roles. *Appl Microbiol Biot*. 2010; 85: 1241-1249.

Harholt J, Suttangkakul A, Scheller HV. Biosynthessis of pectin. *Plant Physiol*. 2010; 153: 384-395.

Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresource Technol*. 2013; 135: 513-522. DOI: 10.1016/j.biortech.2012.10.047.

Hekmat O, Florizone C, Kim YW, Eltis LD, Warren RA, Withers SG. Specificity fingerprinting of retaining beta-1,4-glycanases in the Cellulomonas fimi secretome using two fluorescent mechanism-based probes. *Chembiochem: a European journal of chemical biology*. 2007; 8(17): 2125–32. DOI: 10.1002/cbic.200700481.

Hemsworth GR, Johnston EM, Davies GJ, Walton PH. Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol*. 2015; 33: 747–761. DOI: 10.1016/j.tibtech.2015.09.006.

Hendricks AT, Zeeman G. Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technol*. 2009; 100:10-108.

Henrissat B, Bairoch A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*. 1993; 293: 781-788.

Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*. 1991; 280: 309-316.

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science*. 2007; 315:804-807.

Himmel ME, Xu Q, Luo Y, Ding S, Lamed R, Bayer EA. Microbial enzyme systems for biomass conversion: emerging paradigms. *Biofuels*. 2010; 1(2): 323-341.

Hon D. Cellulose: a random walk along its historical path. *Cellulose*. 1994; 1:1-25.

Idriss H, Seebauer EG. Reactions of ethanol over metal oxides. *J Mol Catal A Chem*. 2000; 152: 201-212.

Isikgor FH, Becer CR. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polym Chem*. 2015; 6: 4497-559.

Ivanova G, Rákhely G, Kovács KL. Thermophilic biohydrogen production from energy plants by Caldicellulosiruptor saccharolyticus and comparison with related studies. *Int J Hydrogen Energ*. 2009; 34: 3659-3670.

Jeffries TW. Biodegradation of lignin and hemicelluloses. In: Biochemistry of Microbial Degradation, ed. C Ratledge, Dordrecht, Netherlands: Kluwer Academic Publishers. 1994; 233-77.

Ji XJ, Huang H, Ouyang P. Microbial 2,3-butanediol production: A state of the art review. *Biotechnol Adv*. 2011; 29: 351-364.

Ji XJ, Zhu JG, Gao Z. Microbial fermentation production of 2,3-butanediol research progress. *Mod Chem Ind*. 2006; 26: 23-7.

Jørgensen H, Kutter JP, Olsson L. Separation and quantification of cellulases and hemicellulases by capillary electrophoresis. *Anal Biochem*. 2003; 317: 85–93.

Kapdan IK, Kargi F. Bio-hydrogen production from waste materials. *Enzyme Microb Tech*. 2006; 38(5): 569-582. DOI: 10.1016/j.enzmictec.2005.09.015.

Kenyon WJ, Esch SW, Buller CS. The curdlan-type exopolysaccharide produced by Cellulomonas flavigena KU forms part of an extracellular glycocalyx involved in cellulose degradation. *Antonie Van Leeuwenhoek*. 2005; 87: 143–148.

Khan TS, Mubeen U. Wheat straw: a pragmatic overview. *Current Research Journal of Biological Science*s. 2012; 4(6): 673-675.

Kim HK, Pack MY. Cloning and expression of Cellulomonas fimi β-glucosidase genes in Escherichia coli. *Enzyme Microb Tech*. 1989; 11: 313-316.

Krishna Leela J, Sharma G. Studies on xanthan production from Xanthomonas campestris. *Bioprocess Biosyst Eng*. 2000; 23: 687-9.

Kumar R, Singh S, Singh OV. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J Ind Microbiol Biot*. 2008; 35: 377-391.

Laurie JI, Clarke JH, Ciruela A, Faulds CB, Williamson G, Gilbert HJ, Rixon JE, Millward-Sadler J, Hazlewood GP. The NodB domain of a multidomain xylanase from Cellulomonas fimi deacetylates acetylxylan. *FEMS Microbiol Lett.* 1997; 148: 261-264.

Lavoine N, Desloges I, Dufresne A, Bras J. Microfibrillated cellulose—its barrier properties and applications in cellulosic materials: a review. *Carbohyd Polym*. 2012; 90(2): 735–764.

Lee HV, Hamid SBA, Zain SK. Conversion of Lignocellulosic Biomass to Nanocellulose: Structure and Chemical Process. *The Scientific World Journal*. 2014; DOI: 10.1155/2014/631013.

Lee HV, Hamid SBA, Zain SK. Conversion of Lignocellulosic Biomass to Nanocellulose: Structure and Chemical Process. *Sci World J*. 2014; DOI: /10.1155/2014/631013.

Lemons e Silva CF, Schirmer MA, Maeda RN, Barcelos CA, Pereira Jr N. Potential of giant reed (Arundo donax L.) for second generation ethanol production. *Electron J Biotechno*. 2015; 18: 10-15.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels*. 2013; 6: 41.

Liguori R, Faraco V. Biological processes for advancing lignocellulosic waste biorefinery by advocating circular economy. *Bioresource Technol*. 2016; 215: 13-20.

Lionetti V, Cervone F, Bellincampi D. Methyl esterification of pectin plays a role during plant–pathogen interactions and affects plant resistance to diseases. *J Plant Physiol*. 2012; 169: 1623-1630.

Lipinsky ES. Chemicals from biomass: petrochemical substitution options. *Science*. 1981; 212: 1465–1471.

Lippits MJ, Nieuwenhuys BE. Direct conversion of ethanol into ethylene oxide on copper and silver nanoparticles Effect of addition of CeOx and Li2O. *Catal Today*. 2010; 154:127-132.

Liu Y, Yu P, Song X, Qu YB. Hydrogen production from cellulose by coculture of Clostridium thermocellum JN4 and Thermoanaerobacterium thermosaccharolyticum GD17. *Int J Hydrogen Energ*. 2008; 33: 2927-2933.

Lloyd TA, Wyman CE. Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. *Bioresource Technol*. 2005; 96(18): 1967–77.

Loh YR, Sujana D, Rahman ME, Das CA. Sugarcane bagasse-The future composite material: A literature review. *Resour Conserv Recy*. 2013; 75: 14-22. DOI: 10.1016/j.resconrec.2013.03.002.

Lotfi G. Lignin-degrading bacteria. *J Agroaliment Proc Technol*. 2014; 20(1): 64-68.

Lu J, Weerasiri RR, Liu Y, Wang W, Ji S, Lee I. Enzyme production by the mixed fungal culture with nano-shear pretreated biomass and lignocellulose hydrolysis. *Biotechnol Bioeng*. 2013; 110(8): 2123-2130. DOI: 10.1002/bit.24883.

Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev*. 2002; 66: 506-577.

Madurwar MV, Mandavgane SA, Ralegaonkar RV. Use of sugarcane bagasse ash as brick material. *Curr Sci India*. 2014; 107(6): 1044-1051.

Mansfield SD, Meder R. Cellulose hydrolisi – the role of monocomponent cellulases in crystalline cellulose degradation. *Cellulose*. 2003; 10:159-169.

Mariani C, Cabrini R, Danin A, Piffanelli P, Fricano A, Gomarasca S, Dicandilo M, Grassi F, Soave C. Origin, diffusion and reproduction of the giant reed (*Arundo donax* L.): a promising weedy energy crop. *Ann Appl Biol*. 2010; 157: 191-202. DOI: 10.1111/j.1744-7348.2010.00419.x.

Mazzoli R, Bosco F, Mizrahi I, Bayer EA, Pessione E. Towards lactic acid bacteria-based biorefineries. *Biotechnol Adv*. 2014; 32: 1216–1236.

McNeil M, Darvill AG, Fry SC, Albersheim P. Structure and Function of the Primary Cell Walls of Plants. *Ann Rev Biochem*. 1984; 53: 625. DOI: 10.1146/annurev.bi.53.070184.003205.

Meinke A, Gilkes NR, Kilburn DG, Miller RC Jr, Warren RA. Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (CenD), a family A beta-1,4-glucanase. *J Bacteriol*. 1993; 175: 1910-1918.

Meinke A, Gilkes NR, Kwan E, Kilburn DG, Warren RA, Miller RC J. Cellobiohydrolase A (CbhA) from the cellulolytic bacterium Cellulomonas fimi is a beta-1,4-exocellobiohydrolase analogous to Trichoderma reesei CBH II. *Mol Microbiol.* 1994; 12: 413-442.

Melis A. Green alga hydrogen production: progress, challenges and prospects. *Int J Hydrogen Energ*. 2002; 27: 1217-28.

Miranda JR, Historia da cana de acucar—History of sugarcane, Komedi, Campinas, Brazil, 2008.

Modani PO, Vyawahare MR. Utilization of Bagasse Ash as a Partial Replacement of Fine Aggregate in Concrete. *Procedia Engineering*. 2013; 51: 25-29. DOI: 10.1016/j.proeng.2013.01.007.

Mohr H, Schopfer P. Plant Physiology. 1995. Springer, Berlin.

Mosier N, Wyman C, Dale B, Elander R, Hotzapple YYLM, Ladish M. Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technol*. 2005; 96:673-686.

Nair NU, Zhao H. Selective reduction of xylose to xylitol from a mixture of hemicellulosic sugars. *Metab Eng*. 2010; 12(5): 462-8. DOI: 10.1016/j.ymben.2010.04.005.

Ni K, Wanga Y, Cai Y, Pang H. Natural Lactic Acid Bacteria Population and Silage Fermentation of Whole-crop Wheat. *Asian Australas J Anim Sci*. 2015; 28: 1123-1132.

Nieves LM, Panyon LA, Wang X. Engineering Sugar Utilization and Microbial Tolerance toward Lignocellulose Conversion. 2015. DOI: 10.3389/fbioe.2015.00017.

O'Sullivan AC. Cellulose: the structure slowly unravels. *Cellulose*. 1997; 4:173–207.

Oakley JH, Hoadley AFA. Industrial scale steam reforming of bioethanol: A conceptual study. *Int J Hydrogen Energ*. 2010; 35:8472-8485.

Bajpai P. Pretreatment of Lignocellulosic Biomass for Biofuel Production. 2016. Springer Briefs in Green Chemistry for Sustainability, DOI 10.1007/978-981-10-0687-6_2.

Palin R, Geitmann A. The role of pectin in plant morphogenesis. *Biosystem.* 2012; 109: 397-402.

Palmqvist E, Hahn-Hägerdal B. Fermentation of lignocellulosic hydrolysates I: inhibition and detoxification. *Bioresource Technol*. 2000a; 74:17-24.

Palmqvist E, Hahn-Hägerdal B. Fermentation of lignocellulosic hydrolysates II: inhibitors and mechanisms of inhibition. *Bioresource Technol*. 2000b; 74:25-33.

Pandey A, Soccol CR, Nigam P, Soccol VT. Biotechnological potential of agro-industrial residues. I: sugarcane bagasse". *Bioresource Technol*. 2000; 74: 69-80. DOI: 10.1016/S0960-8524(99)00142-X.

Pauly M, Keegstra K. Cell wall carbohydrates and their modifications as a resource for biofuels. *Plant J*. 2008; 54 :559–68.

Pedersen M, Meyer AS. Lignocellulose pretreatment severity—relating pH to biomatrix opening. *New Biotechnol*. 2010; 27(6): 739–750.

Pérez J, Munoz-Dorado J, Dela Rubia T, Martinez J. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *Int Microbiol*. 2002; 5: 53-63.

Persson PV, Hafren J, Fogden A, Daniel G, Iversen T. Silica nanocasts of wood fibers: a study of cell-wall accessibility and structure. *Biomacromolecules*. 2004; 5:1097–1101.

Pessarakli M. Handbook of Photosynthesis, Marcel Dekker, New York, NY, USA, 1997.

Poopak S, Reza AR. Environmental Benefit of Using Bagasse in Paper Production-A Case Study of LCA in Iran. 2012. DOI: 10.5772/51553.

Pratt K, Lenaghan M, Mitchard ETA. Material flows accounting for Scotland shows the merits of a circular economy and the folly of territorial carbon reporting. *Carbon Balance Manage*. 2016; 11: 21.

Premjai P, Khammuang P, Somord K, Tawichai N, Intatha U, Soykeabkaew N. Bio-based composites of sugarcane bagasse: effect of bagasse particle size. T*he 26th Annual Meeting of the Thai Society for Biotechnology and International Conference*, 2014.

Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JC, Johansen KS, Krogh KB, Jørgensen CI, Tovborg M, Anthonsen A, Tryfona T, Walter CP, Dupree P, Xu F, Davies GJ, Walton PH. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci USA*. 2011; 108: 15079-15084.

Ratanakhanokchai K, Waeonukul R, Pason P, Tachaapaikoon C, Kyu KL, Sakka K, Kosugi A, Mori Y. Paenibacillus curdlanolyticus Strain B-6 Multienzyme Complex: A Novel System for Biomass Utilization. 2013; DOI: 10.5772/51820.

Ren HJ. The furfural industry in China future. *Chem Technol Mark*. 2001; 24: 12-5.

Ridley BL, O'Neil MA, Mohnen DA. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*. 2001; 57: 929–967.

Saha BC, Iten LB, Cotta MA, Wu YV. Dilute acid pretreatment, enzymatic saccharification and fermentation of wheat straw to ethanol. *Process Biochem*. 2005; 40(12):3693–700.

Saha BC. Hemicellulose bioconversion. *J Ind Microbiol Biot*. 2003; 30: 279-291.

Saini JK, Saini R, Tewari L. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech*. 2015; 5: 337-353. DOI: 10.1007/s13205-014-0246-5.

Sanchez OJ, Cardona CA. Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technol*. 2008; 99: 5270-5295.

Sandgren M, Stahlberg J, Mitchinson C. Structural and biochemical studies of GH family 12 cellulases: improved thermal stability, and ligand complexes. *Prog Biophys Mol Bio*. 2005; 89: 246–291.

Sannigrahi P, Ragauskas AJ, Tuskan GA. Poplar as a Feedstock for Biofuels: A Review of Compositional Characteristics. *Biofuels Bioprod Bior*. 2010; 4: 209-226. DOI: 10.1002/bbb.206.

Santos RB, Capanema EA, Balakshin MY, Chang H, Jameel H. Effect of hardwoods characteristics on kraft pulping process: emphasis on lignin structure. *BioResources*. 2011; 6(4): 3623-3637.

Sarkar N, Ghosh SK, Bannerjee S, Aikat K. Bioethanol production from agricultural waste: An overview. *Renew Energ*. 2011; 37: 19-27.

Schulze E. Zur Kenntnis der chemischen Zusammensetzung der pflanzlichen Zellmem. *Ber Dtsch Chem Ges*. 1891; 24: 2277-87.

Schwarz WH. The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol*. 2001; 56: 634-649.

Shallom D, Shoham Y. Microbial hemicellulases. *Curr Opin Microbiol*. 2003; 6:219-228.

Shatalov AA, Quilhó T, Pereira H. *Arundo donax L*. reed: new perspectives for pulping and bleaching 1. raw material characterization. *Tappi Journal Review Edpaper*. 2001; 84(1): 1-12.

Sheldon RA. Green and sustainable manufacture of chemicals from biomass: state of the art. *Green Chem*. 2014; 16: 950-63.

Shen H, Gilkes NR, Kilburn DG, Miller RC Jr, Warren RA. Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic bacterium Cellulomonas fimi. *Biochemistry*. 1995; 311:67-74.

Sirò I, Plackett D. Microfibrillated cellulose and new nanocomposite materials: a review. *Cellulose*. 2010; 17(3): 459–494.

Society JE. Biomass handbook. In: Shi ZP, Hua ZZ, editors. Beijing: Chemical Industry Press. 2002.

Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredez A, Persson S, Raab T, Vorwerk S, Youngs H. Toward a systems approach to understanding plant cell walls. *Science*. 2004; 306: 2206– 11.

Song ZX, Takahashi A, Nakamura I, Fujitani T. Phosphorus-modified ZSM-5 for conversion of ethanol to propylene. *Appl Catal A Gen*. 2010; 384:201-205.

Stackebrandt E, Schumann P, Prauser H. The family Cellulomonadaceae. *Prokaryotes.* 2006; 3: 983-1001.

Stoll D, Stalbrand H, Warren RAJ. Mannan-degrading enzymes from Cellulomonas fimi. *Appl Environ Microb*. 1999; 65: 2598-2605.

Su B, Wu M, Zhang Z, Lin J, Yang L. Efficient production of xylitol from hemicellulosic hydrolysate using engineered Escherichia coli. *Metab Eng*. 2015; 31: 112-22. DOI: 10.1016/j.ymben.2015.07.003.

Sun Y, Cheng J. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technol*. 2002; 83(1): 1-11.

Taherzadeh MJ, Karimi K. Enzyme-based hydrolysis processes for ethanol from lignocellulosic materials: a review. *Bioresources*. 2007; 2: 707-738.

Taylor LE, Henrissat B, Coutinho PM, Ekborg NA, Hutcheson SW, Weiner RM. Complete cellulase system in the marine bacterium Saccharophagus degradans strain 2-40T. *J Bacteriol*. 2006; 188: 3849-3861.

Timell TE. Recent progress in the chemistry of wood hemicellulose. *Wood Sci Technol*. 1967; 1: 45–70.

Tomme P, Kwan E, Gilkes NR, Kilburn DG, Warren RA. Characterization of CenC, an enzyme from Cellulomonas fimi with both endo- and exoglucanase activities. *J Bacteriol*. 1996; 178: 4216-4223.

Turner P, Mamo G, Nordberg Karlsson E. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microb Cell Fact*. 2007; 6: 9. DOI:10.1186/1475-2859-6-9.

Van Dyk JS, Pletschke BI. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. *Biotechnol Adv*. 2012; 30: 1458-1480.

Villares A, Moreau C, Bennati-Granier C, Garajova S, Foucat L, Falourd X, Saake B, Berrin JG, Cathala B. Lytic polysaccharide monooxygenases disrupt the cellulose fibers structure. *Sci Rep*. 2017; 7:40262. DOI: 10.1038/srep40262.

Villela Filho M, Araujo C, Bonfá A, Porto W. Chemistry Based on Renewable Raw Materials: Perspectives for a Sugar Cane-Based Biorefinery. *Enzyme Research*. 2011; DOI:10.4061/2011/654596.

Voragen AGJ, Coenen GJ, Verhoef RP, Schols HA. Pectin, a versatile polysaccharide present in plant cell walls. *Struct Chem*. 2009; 20: 263-275. DOI: 10.1007/s11224-009-9442-z.

Wakarchuk W, Kilburn DG, Miller RC Jr, Warren RAJ. The preliminary characterization of the β-glucosidases of Cellulomonas fimi. *J Gen Microbiol*. 1984; 130: 1385-89.

Wang C., Dong D., Wang H., Muller K., Qin Y., Wang H., Wu W. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocelluloses decomposition. *Biotechnol Biofuels*. 2016; 9: 22. DOI: 10.1186/s13068-016-0440-2.

Wang CC, Wang YQ, Chen JN, Li H, Zhang ZH. Research progress of technological processes in fuel ethanol production from lignocellulosic biomass. *Biotechnol Bull*. 2010; 2:51–7.

Wang RF, Shi WY. The production and application of furfural. *Henan Chem*. 2008; 25: 14-5.

Wang Y, Chen R, Cai J, Liu Z, Zheng Y, Wang H, Li Q, He N. Biosynthesis and Thermal Properties of PHBV Produced from Levulinic Acid by Ralstonia eutropha. *PLoS One*. 2013; 8(4): e60318. DOI:10.1371/journal.pone.0060318.

Weber S, Stubner S, Conrad R. Bacterial populations colonizing and degrading rice straw in anoxic paddy soil. *Appl Environ Microbiol*. 2001; 67:1318-1327.

Westereng B, Ishida T, Vaaje-Kolstad G, Wu M, Eijsink VGH, Igarashi K, Samejima M, Stahlberg J, Horn SJ, Sandgren M. The putative endoglucanase PcGH61D from Phanerochaete chrysosporium is a metal-dependent oxidative enzyme that cleaves cellulose. *PLoS One*. 2011; 16: e27807.

Wolf S, Mouille G, Pelloux J. Homogalacturonan methylesterification and plant development. *Mol Plant*. 2009; 2: 851-860.

Woo HL, Hazen TC, Simmons BA, DeAngelis KM. Enzyme activities of aerobic lignocellulolytic bacteria isolated from wet tropical forest soils. *Syst Appl Microbiol*. 2014; 37: 60-67.

World food situation: FAO cereal supply and demand brief. Rome, Italy: United Nations, Food and Agriculture Organization. 8 December 2016. Retrieved 14 December 2016.

Wu L, Moteki T, Gokhale AA, Flaherty DW, Toste FD. Production of Fuels and Chemicals from Biomass: Condensation Reactions and Beyond. *Chem*. 2016; 1: 32-58.

Xu H, Hong JH, Cai MY. Industry microbiology basis of its application. Beijing: Science Press; 1991.

Yang B, Wyman CE. Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels Bioprod Bior*. 2008; 2: 26–40.

Yang S, Fei Q, Zhang Y, Contreras LM, Utturkar SM, Brown SD, Himmel ME, Zhang M. Zymomonas mobilis as a model system for production of biofuels and biochemicals. *Microb Biotechnol*. 2016; 9(6): 699–717. DOI:10.1111/1751-7915.12408.

Yip VL, Withers SG. Breakdown of oligosaccharides by the process of elimination. *Curr Opin Chem Biol.* 2006; 10: 147-155.

Yu CY, Lee DW, Park SJ, Lee KY, Lee KH. Study on a catalytic membrane reactor for hydrogen production from ethanol steam reforming. *Int J Hydrogen Energ*. 2009; 34: 2947-2954.

Zeitsch KJ. The chemistry and technology of furfural and its many by-products. Amsterdam: Elsevier Science Ltd; 2000.

Zhang B, Huang HJ, Ramaswamy S. Reaction kinetics of the hydrothermal treatment of lignin. *Appl Biochem Biotechnol*. 2008; 147(1-3): 119-31. DOI: 10.1007/s12010-007-8070-6.

Zhang H, Dai CC, Zhu Q, Yang QY. Research advances in the biodegradation of lignin. *J Anhui Agric Sci*. 2006; 34:1780–4.

Zhang J, Shao X, Townsend OV, Lynd LR. Simultaneous saccharification and co-fermentation of paper sludge to ethanol by Saccharomyces cerevisiae RWB222—Part I: kinetic modeling and parameters. *Biotechnol Bioeng*. 2009;104(5):920–31.

Zhang JA, Zhang XY, Han RL, Liu DH, Li ZH. Cellulase to the fluid production of citric acid by Aspergillus niger fermentation. *Microbiology*. 2001; 21: 5-8.

Zhang Z, Hu S, Chen D, Zhu B. An analysis of an ethanol-based whole-crop refinery system in China. *Chin J Chem Eng*. 2016b; 24(11): 1609-1618.

Zhang Z, Song J, Han B. Catalytic transformation of lignocellulose into chemicals and fuel products in ionic liquids. *Chem Rev*. 2017; 117: 6834-6880.

Zhang Z, Song J, Han B. Catalytic transformation of lignocellulose into chemicals and fuel products in ionic liquids. *Chem Rev*. 2016a; 117: 6834-6880. DOI: 10.1021/acs.chemrev.6b00457.

"Polysaccharides", General Biology Hub: Learning Resource 3, 2010.

"World food situation: FAO cereal supply and demand brief". www.FAO.org, access 2016.

http://ec.europa.eu/environment/circular-economy/index_en.htm, access 31/07/2017.

http://www.cazy.org, access 30/09/2017.

https://www.ellenmacarthurfoundation.org/Circular Economy, access 31/07/2017.

Aim of work

Lignocellulose degradation nowadays represents a challenging step for value-added products and biofuels production, to face the inevitable depletion of non-renewable resources, even reducing harmful emissions. Several studies aim to improve the efficiency of renewable biomass conversion processes, reducing environmental impact and especially costs. A major obstacle to industrial-scale production of fermentable sugar monomers from lignocellulosic biomass lies in the inefficient deconstruction of plant material, due to its recalcitrance toward enzymatic breakdown, and in the relatively low activity of currently available hydrolytic enzymes. The search for new or more efficient cellulolytic enzymes and microorganisms is then very encouraged and represents the major aim of this thesis.

In this study, the lignocellulose degradation is investigated by different approaches. In particular, proteomic and metagenomic analyses were conducted, to highlight carbohydrate active enzymes and microbial communities most probably involved in lignocellulose decay. Preliminary *in silico* analysis was performed to a better knowledge of the potentially secreted proteins of the bacterium *Cellulomonas fimi*, subject of the proteomic study. The identification of the secreted proteins obtained after its growth on different lignocellulosic substrates could provide new insights about its lignocellulose degrading machinery, even highlighting possible relations between enzyme secretion and substrate of growth.

It's well known that lignocellulose decay is a complex process involving several microorganisms acting in synergy in natural environments that constitute microbial communities so the application of a metagenomic approach to identify microbial communities inhabiting decaying wood environments could provide interesting and useful information. Anyway, in the following three chapters, the aim of each analysis will be further explained.

Chapter I

In silico secretome of *Cellulomonas fimi*

The term "secretome" refers to the set of proteins secreted by a cell or an organism at a given time (Tjalsma *et al*., 2000; Alfaro *et al*., 2014), comprehending the proteins released into the extracellular medium and anchored to the membrane or cell wall, including integral membrane proteins (Alfaro *et al*., 2016). Protein secretion in bacteria plays an important role in the interaction of microbes with each other and with their environments (Song *et al*., 2009). The secretome of a microorganism varies depending on several parameters, like growth substrate, temperature and other environmental changes, so the prediction of the potentially secreted proteins, encoded in the genome, could allow to hypothesize microbial role in an environmental niche and how bacteria and fungi may perform peculiar biological processes. The development of genomic techniques and informatic technologies permitted to elaborate bioinfosecretomes, also known as "*in silico*-predicted secretomes" or "*in silico* secretomes", based on the identification of secretion signals in the predicted proteins, using the automatically annotated gene models of a genome sequence (Alfaro *et al*., 2014). Consequently, the quality of an *in silico* secretome depends on the quality of the genome sequence and annotation. Furthermore, several proteins with a correctly predicted secretion peptide are not secreted, for instance because they are resident proteins of the endoplasmatic reticulum (Scott *et al*., 2004). Thus, in the absence of direct experimental proof of secretion, an *in silico* predicted secretome does not correctly represent the real one, so a combination of *in silico* and proteomic techniques is still considered the best approach to characterize microbial secretomes (Braaksma *et al*., 2010). Several studies regarding *in silico* secretomes, both bacterial and fungal, were already published, improving the knowledge of their roles and capabilities (Falb *et al*., 2005; Boekhorst *et al*., 2006; Braaksma *et al*., 2010; Indrelid *et al*., 2014; Alfaro *et al*., 2016).

1 Aim of study

The study of bacterial secretome represents a key to understand how bacteria interact with their environment. As previously reported, *Actinobacteria* in the genus *Cellulomonas* are the only known cellulose degrading facultative anaerobe microorganisms and a better knowledge of their cellulolytic strategy could provide new insights for lignocellulose degradation studies. Since *Cellulomonas fimi* genome was sequenced and annotated (Christopherson *et al*., 2013), it can be analyzed for the identification of potentially secreted proteins, so an *in silico* secretome analysis was conducted by means of two different secretion prediction tools (SignalP and SecretomeP). The proteins identified as secreted were used to elaborate an *in silico* 2DE map and functionally annotated by the bioinformatic tool Blast2GO. Moreover, a CAZymes 2DE map was obtained, for a better analysis of the secreted proteins involved in polysaccharide metabolism.

2 Materials and methods

2.1 Bioinformatic analysis

Cellulomonas fimi ATCC 484 protein sequences were downloaded from UNIPROT database (www.uniprot.org) and submitted to two different secretion prediction software: SignalP 4.1, that predicts the presence of signal peptidase I cleavage sites (Petersen *et al*., 2011) and SecretomeP 2.0, that highlights proteins secreted by non-classical systems, like the Sec-dependent pathway (Bendtsen *et al*., 2005). Secreted proteins were grouped in three different categories: only predicted by SignalP 4.1, only predicted by SecretomeP 2.0 and commons (predicted by both the tools). Each category was then analyzed with Blast2GO ((http://www.blast2go.com/b2ghome) for Gene Ontology (GO) annotation using the standard parameters (Conesa *et al*., 2005). Data of biological process and molecular function were presented with pie charts. For each predicted protein, theoretical mass weight (Mw) and isoelectric point (pI) were calculated by Compute pI/Mw tool (www.expasy.org). Protein ID of *C. fimi* CAZymes were downloaded from the CAZy database (www.cazy.org) (Levasseur *et al*., 2013) and finally, *in silico* 2DE maps of the total secretome and of the CAZymes were plotted with Microsoft Excel.

3 Results and discussion

3.1 *In silico* secretome and Gene Ontology annotation

The *in silico* analysis permitted to identify the whole secreted proteome of *Cellulomonas fimi* ATCC 484. For a more detailed evaluation, two prediction tools were used as different secretion pathways were considered. In fact, only some proteins are secreted employing signal peptide pathway. *C. fimi* non-redundant protein sequences were downloaded from UniProt database and amounted to 3784. Downloaded sequences were submitted to SignalP 4.1 and SecretomeP 2.0, that highlighted 369 and 956 sequences, respectively. Comparing the results, 194 proteins were predicted by both SignalP 4.1 and SecretomeP 2.0 (named common proteins), while 175 showed classical N-terminal signal peptides and 758 were found to be secreted only by non-classical secretory pathway, accounting for a total of 1127 secreted proteins. Thus, about 25% of *C. fimi* protein sequences are proteins destined to extracellular environment and only about 9% of secreted proteins showed signal peptide.

The analysis with Blast2GO for Gene Ontology (GO) functional annotation was conducted on the three groups of proteins: those provided exclusively by SignalP 4.1 or SecretomeP 2.0 and common proteins. In particular, GO analysis describes biological processes and molecular functions (Fig. 1). About the 50% of secreted proteins was highlighted as "annotated seqs" by Blast2GO and assigned to a specific GO term. The analysis of SignalP 4.1 group showed that among biological process the most highly represented categories were: macromolecule metabolic process (GO:0043170), carbohydrate metabolic process (GO:0005975) and transport (GO:0006810). Regarding molecular function, indeed, the majority of protein sequences was associated with GO term related to hydrolase activity (GO:0016787). Moreover, most of proteins were associated with miscellaneous binding activities, including organic cyclic compound binding (GO:0097159), heterocyclic compound binding (GO:1901363), ion binding (GO:0043167) and

small molecule binding (GO:0036094). The two largest categories of biological process highlighted in SecretomeP 2.0 group were: establishment of localization (GO:0051234) and single-organism cellular process (GO:0044763). Others included cellular metabolic process (GO:0044237), organic substance metabolic process (GO:0071704), primary metabolic process (GO:0044238) and single-organism metabolic process (GO:0044710). As regard molecular function, Blast2GO analysis revealed that the majority of proteins were involved in catalytic activity (GO:0016787 - hydrolase activity), binding (GO:1901363 - heterocyclic compound binding, GO:0043167 - ion binding, GO:00306094 - small molecule binding, GO:0097367 - carbohydrate derivative binding, GO:0097159 - organic cyclic compound binding) and transport (GO:0022857 - transmembrane transporter activity, GO:0022892 - substrate-specific transporter activity). Common proteins were related to biomolecules metabolic processes, in particular polysaccharides (GO:0044260 - cellular macromolecule metabolic process, GO:0005976 - polysaccharide metabolic process, GO:0016052 - carbohydrate catabolic process, GO:0044262 - cellular carbohydrate metabolic process, GO:0009057 - macromolecule catabolic process, GO:0044723 - single-organism carbohydrate metabolic process). Molecular function was, indeed, clearly divided between hydrolase activity (GO:0016787) and binding (GO:0030246 - carbohydrate binding, GO:0005515 - protein binding, GO:0001871 - pattern binding, GO:0043167 - ion binding). The Gene Ontology analysis permitted to state that most of secretion proteins belong to “hydrolase activity” GO term, and in particular are cellulases, hemicellulases and other enzymes involved in polysaccharides metabolism. Moreover, several protein sequences were associated with binding functions.

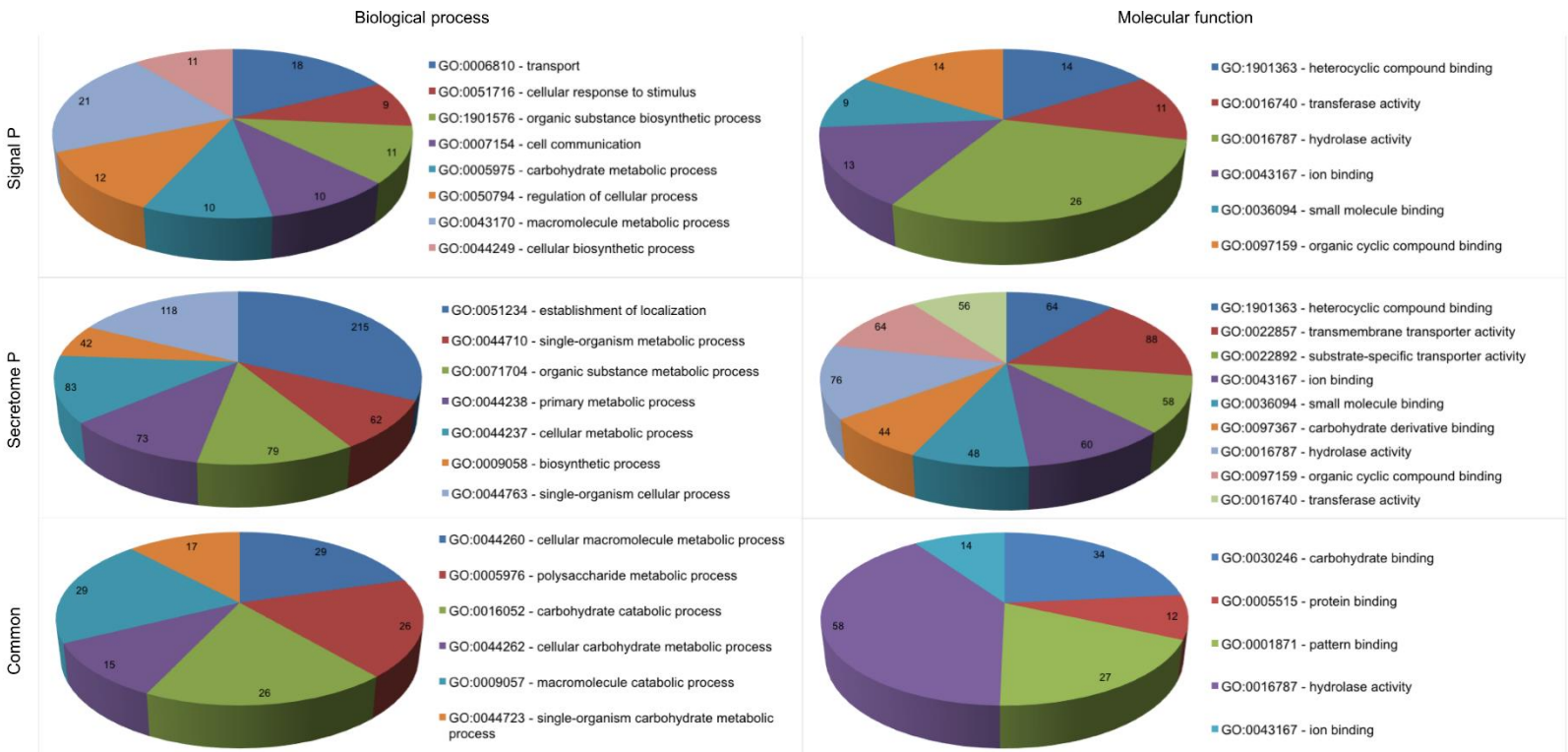


Figure 1: graphical representation of GO analysis results

3.2 *In silico* analysis of CAZymes

As already reported, proteins and enzymes involved in lignocellulose metabolism are collectively named CAZymes and their classification is available on CAZy database. In particular, 175 different CAZymes sequences of *C. fimi* were obtained from the database. Catalytic domain evaluation permitted to classify: 108 Glycoside Hydrolases (GHs), 8 Carbohydrate Esterases (CEs), 2 enzymes with both GH and CE domain, 6 Polysaccharide Lyases (PLs), 45 Glycosyl Transferases (GTs), one enzyme of Auxiliary Activity (AA) and 5 proteins with exclusively Carbohydrate Binding Module domain (CBM). Secretion prediction analysis showed that 40% of CAZymes is constituted by extracellular proteins. Further, 46 sequences were predicted as secreted by both tools, while 11 and 13 sequences were predicted only by SignalP 4.1 or SecretomeP 2.0, respectively. Intracellular CAZymes account for the 60% of the total.

C. fimi GH sequences belong to 41 different GH families. In particular, observing each family it can be stated that *C. fimi* holds the enzymatic activities necessary for an efficient lignocellulose degradation. GH families associated to endoglucanase (GH5, GH6 and GH9), cellobiohydrolase (GH6 and GH48) and several hemicellulase (GH10, GH11, GH26 and GH74) activities are secreted. Nevertheless, *C. fimi* secretome lacks in β-glucosidase and xylosidase activities as proteins belonging to GH1 and GH3 families are mainly found in intracellular compartment, as demonstrated by SignalP 4.1 and SecretomeP 2.0. Other CAZymes classes (CE, PL and AA) were predicted as secreted.

3.3 *In silico* secretome and CAZymes 2DE maps

In silico 2DE maps can represent a helpful instrument for visualize the potentially secreted proteins. In fact, typical experimental 2DE maps of *in vivo* secretome permit to analyze only those proteins with a Mw of 15-250 KDa and a pI 3-10. Therefore, some proteins cannot be represented in *in vivo* 2DE maps. Yet, post-translational modification cannot be shown by *in silico* maps, hence the *in vivo* approach remains of fundamental importance. Observing the theoretical pI, the majority of *C. fimi* secretory proteins predicted by both SignalP 4.1 and SecretomeP 2.0 shared a pI between 4 and 7. Thus, in order to achieve a good spots resolution in *in vivo* maps, IEF should be performed in the 4-7 pH range. The *in silico* 2DE map of the *C. fimi* total secretome was plotted and the proteins with pI 4-7 were highlighted with a yellow area (Fig. 2). The 2DE map of CAZymes was represented in Fig. 3. About 80% of CAZymes had a pI 4-7 and no one had pI lower than 4, so a good resolution of protein spots in the pH range 4-7 is advisable. Nevertheless, about 20% of CAZymes (both secreted and intracellular) had pI higher than 7 and cannot be detected.

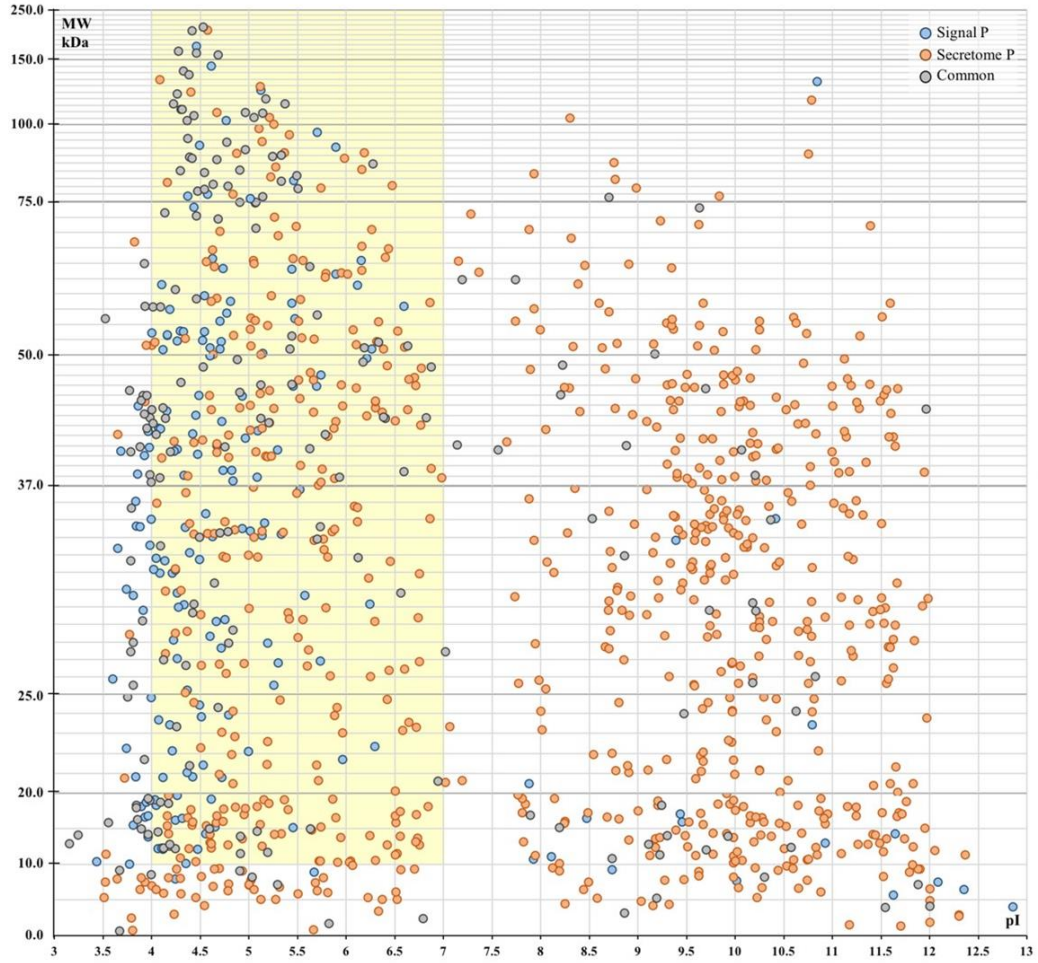


Figure 2: *Cellulomonas fimi in silico* secretome 2DE map

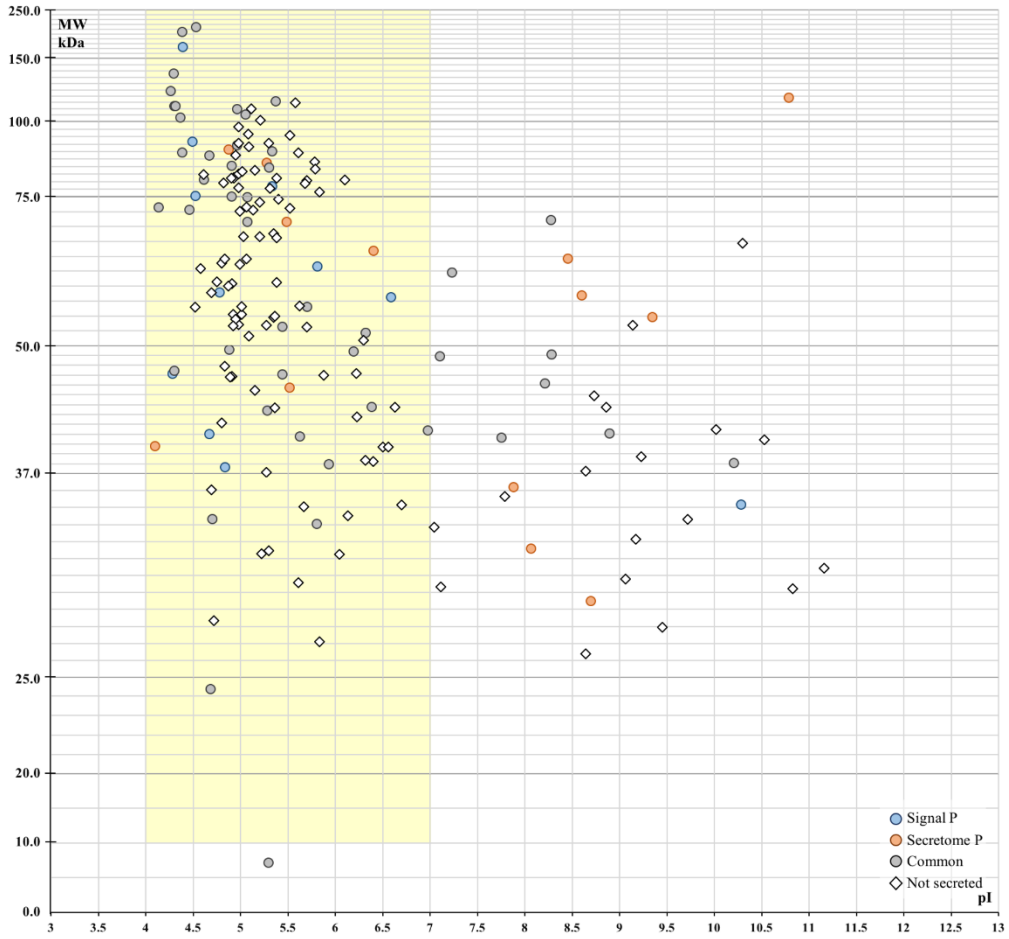


Figure 3: *Cellulomonas fimi* CAZymes 2DE map

4 Conclusions

The conducted analyses confirmed that *Cellulomonas fimi* ATCC484 can secrete carbohydrate active enzymes for lignocellulose degradation. However, the *in silico* secretome is a static list of potentially secreted proteins that can't provide a complete overview of how these proteins are actually used, so proteomic analyses of real secretomes, derived from a real biological activity, will always be necessary for a better knowledge of microorganisms activities. The *in silico* analysis of *C. fimi* secretome may represent a helpful tool for the choose of an appropriate proteomic method for the secretome characterization. In fact, the in silico 2DE map of CAZymes showed that most of them share a pI between 4 and 7, but some CAZymes have a pI higher than 10 and cannot be detected by 2DE analysis with IPG strips pH 3-10 (see the recently accepted manuscript, Spertino *et al.*, 2018).

References

Alfaro M, Castanera R, Lavìn JL, Grigoriev IV, Oguiza JA, Ramìrez L, Pisabarro AG. Comparative and transcriptional analysis of the predicted secretome in the lignocellulose-degrading basidiomycete fungus Pleurotus ostreatus. *Environ Microbiol*. 2016; 18(12): 4710-4726. DOI:10.1111/1462-2920.13360.

Alfaro M, Oguiza JA, Ramìrez L, Pisabarro AG. Comparative analysis of secretomes in basidiomycete fungi. *J Proteomics*; 102: 28-43.

Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. *BMC Microb*. 2005; 5: 58.

Boekhorst J, Wels M, Kleerebezem M, Siezen RJ. The predicted secretome of Lactobacillus plantarum WCFS1 sheds light on the interaction with its environment. *Microbiology*. 2006; 152: 3175-3183.

Braaksma M, Martens-Uzunova ES, Punt PJ, Schaap PJ. An inventory of the Aspergillus niger secretome by combining in silico predictions with shotgun proteomics data. *BMC Genomics*. 2010; 11:584.

Conesa A, Gotz S, Garzia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21: 3674-3676.

Falb M, Pfeiffer F, Palm P, Rodewald K, Hickmann V, Tittor J, Oesterhelt D. Living with two extremes: conclusions from the genome sequence of Natronomonas pharaonis. *Genome Res*. 2005; 15: 1336-1343.

Indrelid S, Mathiesen G, Jacobsen M, Lea T, Kleiveland CR. Computational and Experimental Analysis of the Secretome of Methylococcus capsulatus (Bath). *PLoS One*. 2014; 9(12): e114476. DOI:10.1371/journal.pone.0114476.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels*. 2013; 6:41.

Petersen TN, Brunak S, Heijne von G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011; 8: 785-786.

Scott M, Lu G, Hallett M, Thomas DY. The Hera database and its use in the characterization of endoplasmic reticulum proteins. *Bioinformatics*. 2004; 20: 937-944.

Song C, Kumar A, Saleh M. Bioinformatic Comparison of Bacterial Secretomes. *Genomics Proteomics Bioinformatics*. 2009; 7(1-2): 37-46. DOI: 10.1016/S1672-0229(08)60031-5.

Spertino S, Boatti L, Icardi S, Manfredi M, Cattaneo C, Marengo E, Cavaletto M. Cellulomonas fimi secretomes: in vivo and in silico approaches for the lignocellulose bioconversion. *J Biotechnol*. 2018. DOI: 10.1016/j.jbiotec.2018.01.018.

Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM. Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome. *Microbiol Mol Biol Rev*. 2000; 64: 515-547.

http://www.blast2go.com/b2ghome

www.cazy.org

www.expasy.org

www.uniprot.org

Chapter II

Cellulomonas fimi secretome characterization

The ability to degrade lignocellulose represents a very important subject of study and microorganisms like *Cellulomonas fimi* can constitute simple models, useful to reach a better understanding of the degradation process. As previously described, *C. fimi* can degrade LCB both in aerobic and anaerobic conditions, exploiting a "secreted enzymes" approach. Since this microorganism is easily cultivable, even in liquid growth media, the study of its secreted proteins, collectively referred as "secretome", is possible and very interesting. In the last years, many secretome studies were conducted, supported by the great development of high-throughput proteomic tools (Adav *et al*., 2012; Fernandes *et al*., 2014; Schneider *et al*., 2016; Wakarchuk *et al*., 2016; Zhu *et al*., 2016). Bacteria, yeast and fungi secretomes are easy to study as these microorganisms do not require any exogenous protein addition in their culture media, so the conditioned media collected after their growth contain mainly their own released proteins. Secretome studies of different microorganisms provide information about their roles in infectious diseases, food industry and biotechnology as well as agriculture and environment (Brown *et al*., 2012).

Recently, Wakarchuk *et al*. (Wakarchuk *et al*., 2016) have characterized the proteins secreted by *C. fimi* and *C. flavigena* grown on carboxymethylcellulose (CMC) or a soluble xylan fraction, detecting many known *C. fimi* CAZymes, in particular cellulases, operating with either hydrolytic and oxidative mechanism, and xylanases. Moreover, the study confirmed once again that the growth medium influences secreted protein expression, since xylanases were more represented in secretomes obtained after growth on xylan. The bacterial protein secretion variations depending on substrate of growth suggests the possibility to elaborate enzymatic cocktails *ad hoc* for the degradation of several lignocellulosic biomasses. Biorefineries in fact can provide value-added products starting from many non-food raw materials, such as cereal straw, sugarcane bagasse, perennial grasses, energy crops like *Arundo donax*, corn stover, agricultural, forest, municipal and industrial biomass wastes (Nizami *et al*., 2017). In a market

economy, the maximization of profit is fundamental, as well as product quality, operational safety and respect for the environment. For this purpose, researches to seek the most competitive feedstocks are always in act (Villela Filho *et al.*, 2011).

1 Aim of work

The aim of this work was to characterize *Cellulomans fimi* secretome, paying particular attention to CAZymes and enzymatic activities, after its growth (for 24 or 48 hours) on carboxymethyl-cellulose (CMC) or three different lignocellulosic biomasses (LCBs): wheat straw, sugar cane bagasse and *Arundo donax*. All of them are globally available in large quantities and represent attractive substrates for second generation biofuels production, avoiding food competition.

2 Materials and methods

2.1 Production of secretomes

Frozen strains of *C. fimi* ATCC 484 were streaked on Nutrient Agar plates (beef extract 3 g L^{-1}, peptone 5 g L^{-1}, agar 20 g L^{-1}, pH 6.8) at 30 °C for 48 h. Cells were inoculated in flasks containing 100 mL of Broth Medium (BD cat 234000) and incubated at 30 °C with shaking (150 rpm) until reaching OD_{600} ~0.7 (24 h of growth). The pre-culture was then inoculated in Mineral Medium ($NaNO_3$ 1 g L^{-1}, K_2HPO_4 1 g L^{-1}, KCl 0.5 g L^{-1}, $MgSO_4.7H_2O$ 0.5 g L^{-1}, yeast extract 0.5 g L^{-1}, pH 7) obtaining a 2% (v/v) final concentration. As unique carbon source was added 0.1% (w/v) carboxy-methylcellulose (CMC) medium viscosity (Sigma-Aldrich Co.) or one of the following steam-explosion pretreated biomasses: wheat straw (WS), *Arundo donax* (AD) and sugar cane bagasse (SCB). All cultivations were performed in biological triplicates and were incubated at 30 °C, 150 rpm for 24 or 48 h. The samples were centrifuged at 16000 x *g* for 30 min at 4 °C and the supernatants were collected. Replicates of the same culture condition were pooled and these secretomes were filtered with 0.22 μm PES filters (Sigma-Aldrich Co.), concentrated using Vivaspin 20 (Sartorius Stedim Biotech GmbH, Goettingen, Germany) with molecular weight cutoff 5 KDa, centrifuging at 5400 x *g* at 4° C in a swinging-bucket rotor (Allegra 25-R, Beckman Coulter Inc.) and used for enzymatic activity assays and proteomic analysis. The protein content of each sample was estimated by FluoroProfile Protein Quantification Kit (Sigma-Aldrich Co.), according to manufacturer's guide.

2.2 Zymographic analysis

In-gel activities of cellulases and xylanases were tested separating 1 μg of each native protein sample with 10% SDS-PAGE added with 0.4% AZO-CMC (Sigma-Aldrich Co.) or 0.1% Remazol Brilliant Blue R-D-Xylan (Sigma-Aldrich Co.) respectively, as described by Cattaneo *et al*. (Cattaneo *et al*., 2014). Enzyme activity was observed as a clear zone contrasting on a dark blue background. The images were scanned using an Image Scanner III (GE Healthcare).

2.3 Enzyme activity assays

Native concentrated secretomes were used to assess endoglucanase, xylanase and β-glucosidase activities. Each activity was assayed in 50 mM sodium-acetate buffer (pH 5.8) for 10 min at 50°C. The substrates used were: 2% CMC- medium viscosity, 2% birch wood xylan and 1% salicin (Sigma-Aldrich Co.) for endoglucanase, xylanase and β-glucosidase activity, respectively. Sugar release was evaluated according to the Nelson-Somogyi method (Nelson,1944). All samples were analyzed at least in triplicate and mean values were calculated. The activity is expressed in U/mg of protein. One unit of activity (U) is defined as the amount of enzyme which catalyzes the conversion of 1 μmol of substrate/min under standard conditions. Data were analyzed by a one-way ANOVA followed by Fisher's test with cut-off significance at $p = 0.05$ using Stat View 4.5 (Abacus Concepts) software.

2.4 Sample preparation for SDS-PAGE

Before proceeding to the protein precipitation, each sample was subjected to phenol extraction, as described by Mangiapane *et al*. (Mangiapane *et al*., 2014). Briefly, aliquots of the concentrated secretomes were mixed 1:1 with saturated phenol, incubated 10 min at 70 °C and 5 min in ice and then centrifuged (10.000 *x g*, 10 min at 4 °C). The phenol phase was transferred in a clean Eppendorf tube

and one volume of milliQ was added. Samples were incubated and centrifuged as previously described and finally the phenol phase was recovered and precipitated as described by Chen *et al*. (Chen *et al*., 2009), adding 5 volumes of cold methanol plus 0.1 M ammonium acetate and incubated overnight at -20 °C. Protein samples were then recovered by centrifugation (18.000 *x g*, 20 min at 4 °C) and washed twice with ammonium acetate 0.1 M in cold methanol for 30 min at -20 °C and twice with 80% cold acetone for 15 min at 4 °C. After each incubation, proteins were pelleted centrifuging samples (15.000 *x g*, 20 min at 4 °C). Finally, pellets were washed with 70% cold ethanol for 15 min at 4 °C, centrifuged (18.000 *x g*, 20 min at 4 °C), nitrogen-dried and stored at -80 °C.

2.5 Bidimensional electrophoresis (2DE)

C. fimi secretome collected after its growth on pretreated wheat straw was submitted to bidimensional electrophoresis. Precipitated samples were re-suspended in solubilization buffer (7 M Urea, 2 M Thiourea, 4% w/v CHAPS, 50 mM DTT) and the 2DE was performed as described by Spertino *et al*. (Spertino *et al*., 2012). Briefly, first dimension isoelectric focusing (IEF) was conducted using 7 cm immobilized linear pH range 3-10 strips (GE Healthcare, Milan, Italy), on an IPG-Phor unit (GE Healthcare Bio-Sciences); the protein samples were mixed with a rehydration buffer (7 M Urea, 2 M Thiourea, 4% w/v CHAPS, 50 mM DTT, 5% Triton X100, 5% 4–7 IPG Buffer or 3–10 IPG Buffer (GE Healthcare), and traces of Bromophenol Blue (BBF)) and the final volume was adjusted to 125 µL with solubilization buffer. For the second dimension, the focused strips were first reduced in SDS equilibration buffer (50 mM Tris–HCl pH 8.8, 6 M Urea, 30% v/v Glycerol, 2% w/v SDS) containing 10 mg mL^{-1} DTT for 15 min, rinsed with distilled water and then alkylated in SDS equilibration buffer with 45 mg mL^{-1} Iodoacetamide for 15 min. Finally, a 10% SDS-PAGE was performed and 2DE gels were stained with Colloidal Coomassie Brilliant

Blue G250 (Bio-Rad Laboratories) in accordance with Neuhoff *et al*. (1988) and scanned using an Image Scanner III (GE Healthcare).

2.6 SDS-PAGE

Precipitated samples were solubilized in 0.5 M Tris-HCl buffer pH 6.8. Aliquots corresponding to 20 µg of protein were mixed with one volume of 2X loading buffer (125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 10% β-mercaptoethanol, BBF in trace) and separated by 10% SDS-PAGE as described by Peila *et al*. (Peila *et al*., 2016). Protein bands were visualized using Coomassie Blue G-250 staining and minigels were scanned using an Image Scanner III (GE Healthcare). Three technical replicates were done.

2.7 In-gel digestion and protein identification

Each lane corresponding to a different secretome was excised from the gel using a sterile scalpel, cut into eight fractions (gel slices) based on molecular mass and each slice was cut into small pieces (Jain *et al*., 2010; Bastida *et al*., 2014; Ternan *et al*., 2014; Zhu *et al*., 2016). An in-gel trypsin digestion was carried out as described by Boatti *et al*. (Boatti *et al*., 2012). Before liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis, the tryptic mixtures were submitted to solid phase extraction using C18 Discovery® DSC-18 SPE Tube, according to manufacturer's protocol.

LC-MS/MS analyses were performed by a micro-LC Eksigent Technologies (Dublin, USA) system, as described by Manfredi *et al*. (Manfredi *et al*., 2016). The mass spectrometer worked in information dependent acquisition (IDA) mode. MS data were acquired with Analyst TF 1.7 (AB SCIEX, Concord, Canada). The mass spectrometry files obtained from gel slices of the same lane were taken together and were searched using Mascot v. 2.4 (Matrix Science Inc., Boston, MA, USA). Trypsin as digestion enzyme and 2 missed cleavages were specified. The instrument was set to ESI-QUAD-TOF and the following modifications were

specified for the search: carbamidomethylated cysteins, oxidized methionines and deamidated asparagines as variable modifications. A search tolerance of 0.1 Da was specified for the peptide mass tolerance, and 50 ppm for the MS/MS tolerance. False discovery rate (FDR) was set at 1%, so 99% of the proteins identified should be correct (Jain *et al*., 2010). The charges of the peptides to search for were set to 2+, 3+ and 4+, and the search was set on monoisotopic mass. To perform the protein identification using Mascot, we have created an in-house database downloading all the protein fasta sequences of *C. fimi* from http://www.ncbi.nlm.nih.gov.

2.8 Bioinformatic analysis

Cellulomonas fimi ATCC 484 protein sequences were downloaded from UNIPROT database (www.uniprot.org) and analyzed with two different secretion prediction software: SignalP 4.1 (Petersen *et al*., 2011) and SecretomeP 2.0 (Bendtsen *et al*., 2005), that evaluate two different secretory pathways, as explained in chapter I.

The Search Tool for the Retrieval of Interacting Genes (STRING) database (Version 10.0, http://string-db.org) was recruited to predict the potential interactions between the identified proteins. Such interactions include physical and functional associations and are derived from five main sources: genomic context predictions, high-throughput lab experiments, co-expression, automated textmining and previous knowledge in databases coverage (Szklarczyk *et al*., 2015).

Venn's diagrams were elaborate using the online tool available at: http://bioinformatics.psb.ugent.be/webtools/Venn/.

3 Results and discussion

3.1 Zymografic analysis

The native secretomes obtained after *C. fimi* growth on different substrates for 24 or 48 hours were analyzed by zymography to assess the endoglucanase and xylanase activities (Fig. 1).

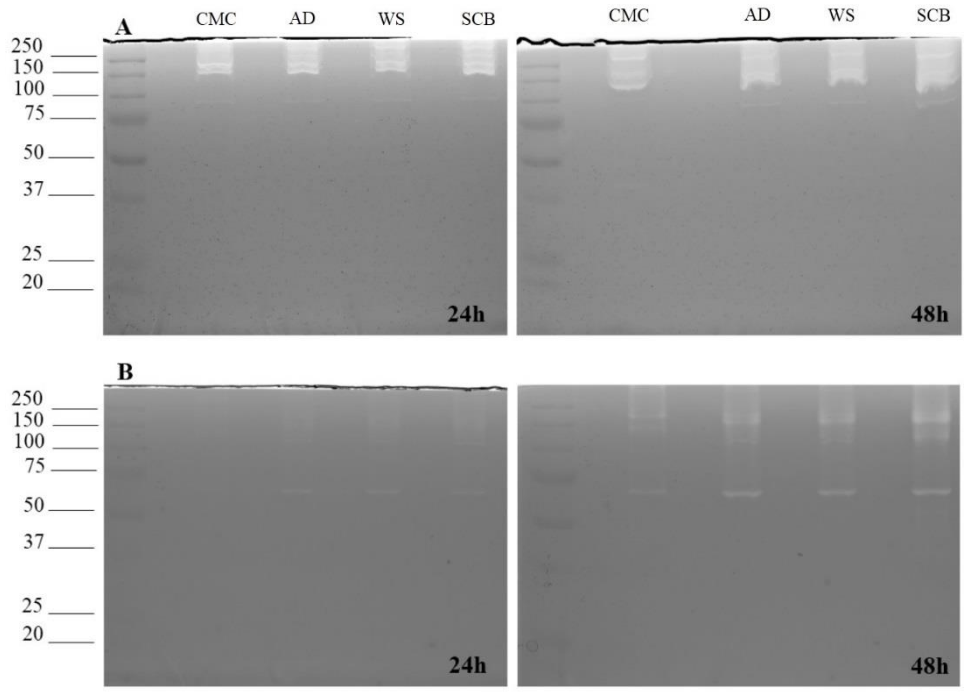


Figure 1: Endoglucanase (A) and xylanase (B) activity detection by zymography.

Degradation halos appeared for all the analyzed samples, confirming that *C. fimi* secretomes hold both these activities and the halos corresponding to the secretomes obtained after 48 hours of growth were more evident than those observed for the secretomes collected after 24 hours.

Regarding endoglucanase activity, the highest levels were detected in the secretomes collected after growth on LCBs. In particular, considering secretomes at 24 hours, within the degradation halo two hydrolysis bands were visible for the CMC secretome, between 150 and 250 KDa, while in LCBs secretomes four bands were detected, with a molecular weight higher than 150 KDa. The largest halo of degradation was attributed to SCB secretome. Finally, a hydrolysis band was revealed in all the samples, between 100 and 75 KDa. The secretomes collected after 48 hours of growth showed most evident hydrolysis bands. Three bands over 100 KDa were observable for the CMC secretome, while secretomes obtained after *C. fimi* growth on wheat straw, *A. donax* or sugar cane bagasse showed a further band between 100 and 75 KDa.

Regarding xylanase activity, the secretomes collected after 24 hours of growth on LCBs were very similar each other, showing two defined hydrolysis bands: between 75 and 50 KDa and between 150 and 100 KDa. Moreover, two bands at about 150 KDa and one over 250 KDa were detected, although less defined. In the secretome obtained after growth on CMC the degradation bands at molecular weight >250 KDa and between 75 and 50 KDa, already observed in the LCBs secretomes, were less evident, while between 150 and 100 KDa a very weak halo was detectable. The secretomes collected after 48 hours showed the same pattern of degradation but higher intensity, especially SCB secretome. Moreover, in the CMC secretome two bands (not revealed at 24 hours) were detected between 250 and 150 KDa and between 150 and 100 KDa.

3.2 Enzymatic activity assays

Glycoside hydrolases activities from *C. fimi* secretomes were then quantitatively evaluated by reducing sugars assays. Endoglucanase, xylanase and β-glucosidase activities were assayed in 50 mM sodium-acetate buffer (pH 5.8) for 10 min at 50 °C. Reducing sugar release was evaluated by Nelson-Somogyi method and the activities were calculated and expressed in U/mg. The results are shown in Table 1.

Table 1: **Enzymatic activities of *C. fimi* secretomes.** Mean values of enzymatic activities followed by standard errors, at 24 and 48 hours of growth. Specific enzymatic activity was expressed as IU mg^{-1} of protein. ND: not detected.

	Endoglucanase		**Xylanase**		**β-glucosidase**	
Growth substrate	**24 h**	**48 h**	**24 h**	**48 h**	**24 h**	**48 h**
CMC	0.92±0.14	6.29±0.94	1.42±0.25	4.56±0.88	0.49±0.07	1.62±0.44
AD	4.93±0.76	7.83±0.84	6.19±1.03	12.20±0.10	ND	0.09±0.04
WS	4.31±0.49	7.47±0.82	9.42±1.09	13.87±0.92	ND	0.13±0.04
SCB	4.44±0.51	11.14±0.92	8.94±1.19	19.02±1.07	0.82±0.17	0.25±0.04

The enzymatic assays confirmed that the highest values of endoglucanase and xylanase activities were detected in the secretomes collected after 48 hours of growth. The secretomes produced by *C. fimi* grown on LCBs showed higher endoglucanase and xylanase activities, compared to CMC and among LCBs, the substrate associated to the highest activities was sugar cane bagasse (Fig. 2).

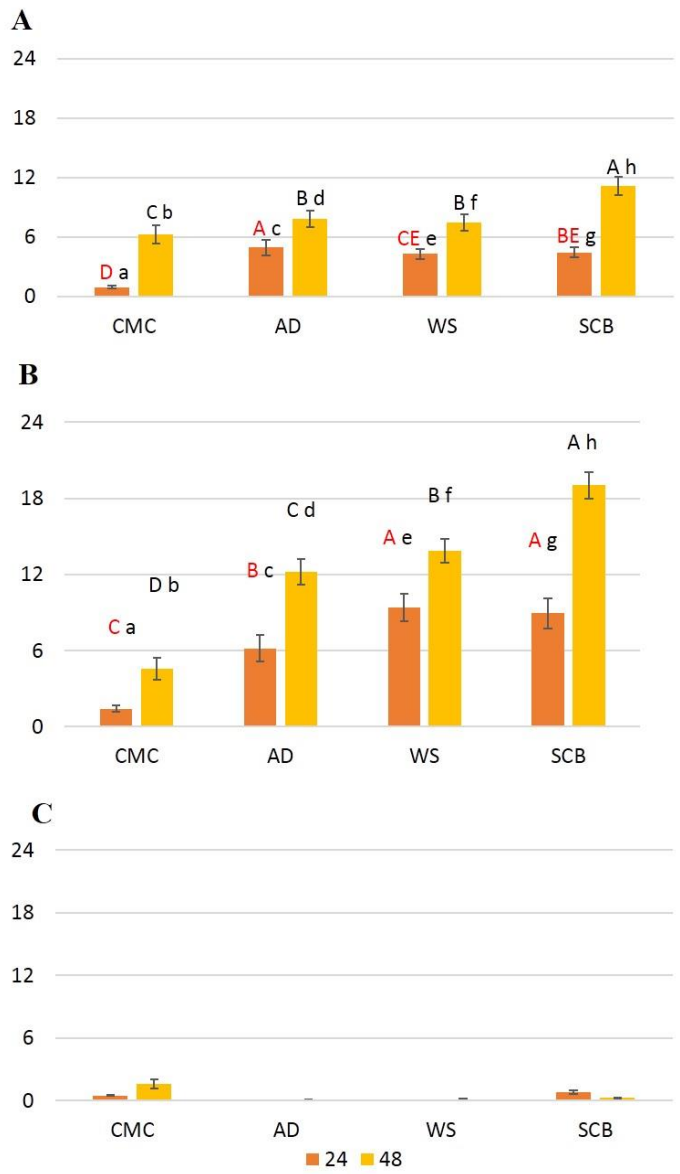


Figure 2: Endoglucanase (A), xylanase (B) and β-glucosidase (C) activity (U/mg) of *C. fimi* secretomes, collected after 24 or 48 hours of growth in presence of CMC or a pretreated LCB: wheat straw (WS), *A. donax* (AD) or sugar cane bagasse (SCB). Uppercase letters in black and red highlight statistical differences between all the samples at 48 and 24 hours, respectively. Tiny letters refer to significant differences between samples collected after 24 and 48 hours of growth on the same substrate.

In particular, except for the secretome obtained after *C. fimi* growth on CMC, it can be stated that, in every culture condition, the highest enzymatic activity was the xylanase, followed by endoglucanase and β-glucosidase. The secretome produced in presence of CMC showed indeed higher values of endoglucanase than xilanase activity (Fig.3). Such observation can be reasonably explained with the different nature of the substrate.

In agreement with the well-known *C. fimi* inability to secrete β-glucosidases, only very low values of this activity were detected in a few secretomes, and especially in the secretome collected after 48 hours of growth on CMC. In those obtained in presence of AD and WS, the β-glucosidase activity was undetectable. The low values recorded could be due to enzymatic side-activities.

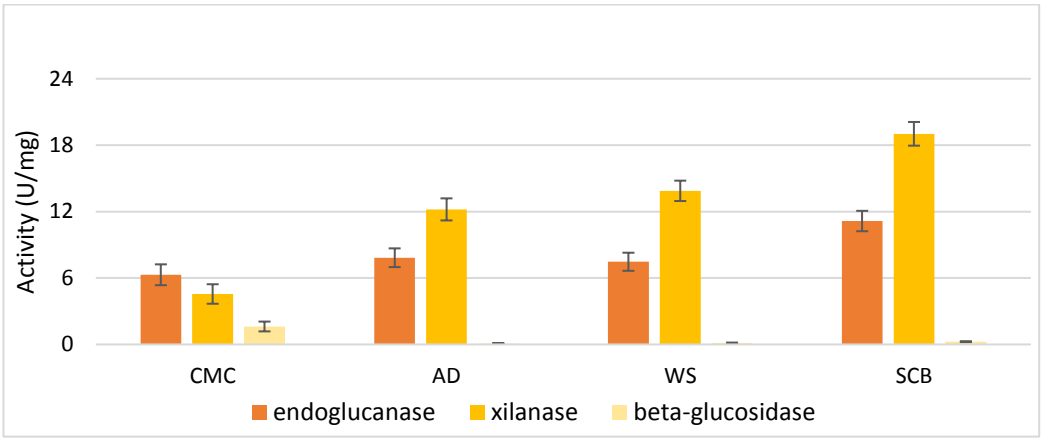


Figure 3: Overview of endoglucanase, xylanase and β-glucosidase activity values (U/mg) of each analyzed secretome, collected after 48 hours of growth.

3.3 2DE map

In Figure 4 is reported a typical 2DE map of a secretome obtained after *C. fimi* growth on wheat straw. From its observation, it can be stated that only a few spots were visualized, with respect to the *in silico* map (pag. 79). Moreover, the focalization is not optimal, probably for the presence of impurities and the formation of protein aggregates.

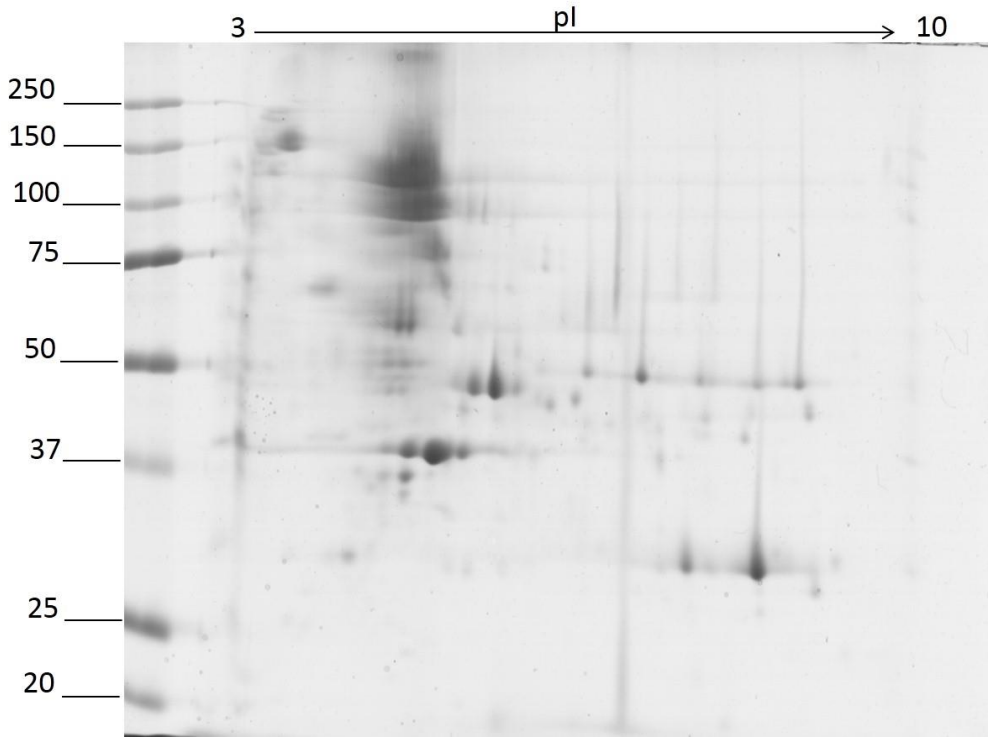


Figure 4: *C. fimi* secretome 2DE map.

3.4 Proteomic analysis

Since 2DE cannot be considered a robust method for secreted protein identification, an SDS-PAGE was performed (Fig. 5). Each lane, corresponding to a different secretome, was then excised (Fig. 6) and digested for the following MS/MS analysis. The identified proteins were reported in Tables S1-S8 (Appendix A).

For greater simplicity and clarity, samples will be indicated with an acronym, constituted by the abbreviation of the substrate used for their production and the time of growth to which they were collected. So, secretomes are named as follow:

CMC24/CMC48: obtained after *C. fimi* growth on CMC for 24/48 hours

AD24/AD48: obtained after *C. fimi* growth on *A. donax* for 24/48 hours

WS24/WS48: obtained after *C. fimi* growth on wheat straw for 24/48 hours

SCB24/SCB48: obtained after *C. fimi* growth on sugar cane bagasse for 24/48 hours

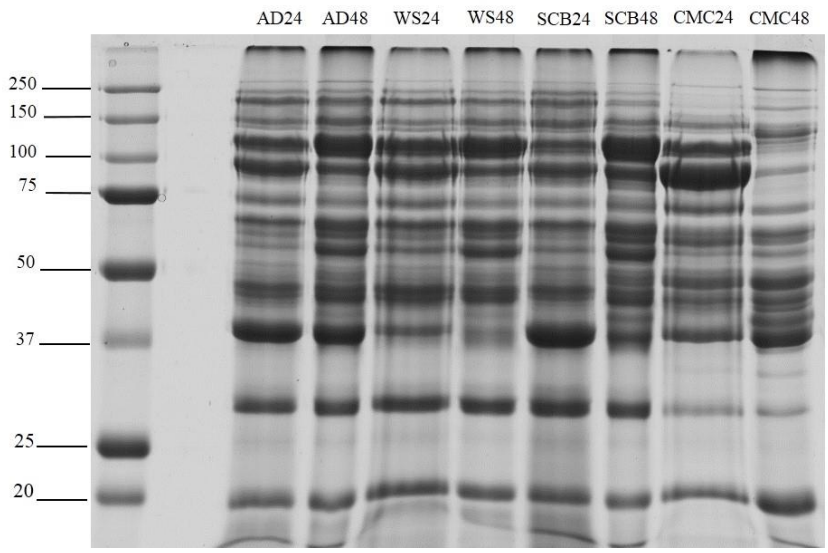


Figure 5: SDS-PAGE of *C. fimi* secretomes.

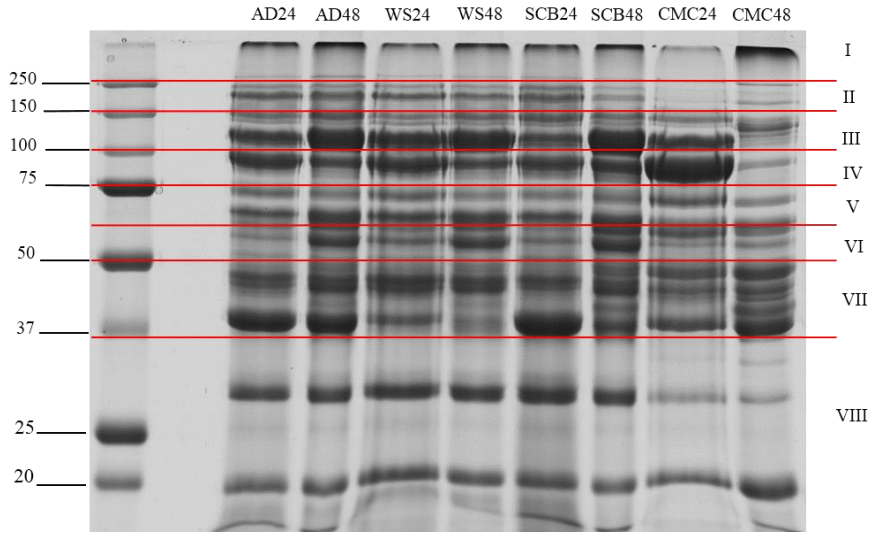


Figure 6: SDS-PAGE and gel slices (eight for each lane, indicated by roman numbers).

The MS/MS analysis permitted to identify 91 proteins in the secretome CMC24, 84 in CMC48, 80 in AD24, 93 in AD48, 100 in WS24, 74 in WS48, 123 in SCB24 e 76 in SCB48. In all the analyzed secretomes, CAZymes represented a percentage between 28% and 37% (Fig. 7).

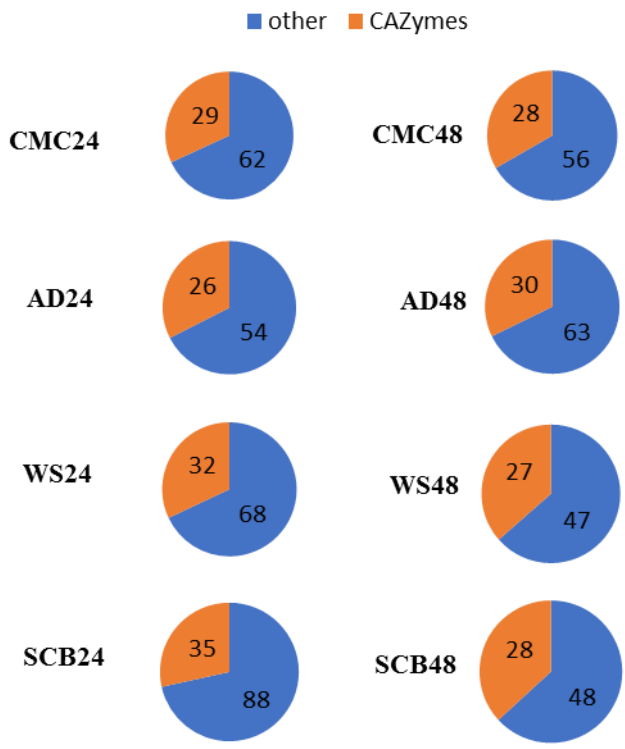


Figure 7: Total proteins identified in each secretome analyzed.

The detected CAZymes belong to several classes, according to the CAZy database. Most of them were found in all secretomes, such as GH, CE, PL, AA. GT class was revealed only in secretomes obtained after 24 hours of growth, while CBM class was detected only in CMC24 and WS24. GH class was the most represented in every secretome, counting several families (Figg. 8-9).

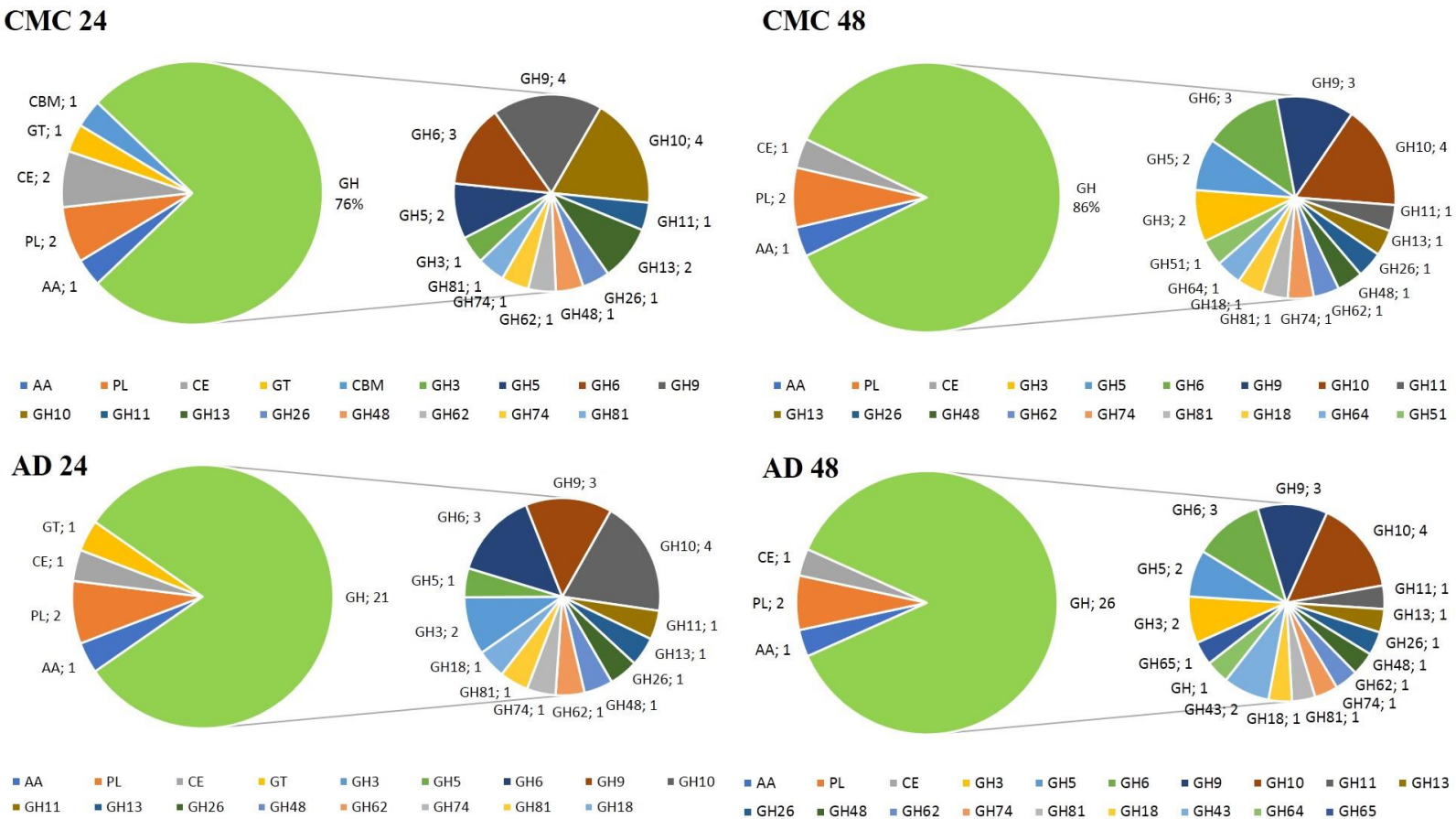


Figure 8: CAZyme classes and GH families identified in CMC and AD secretomes.

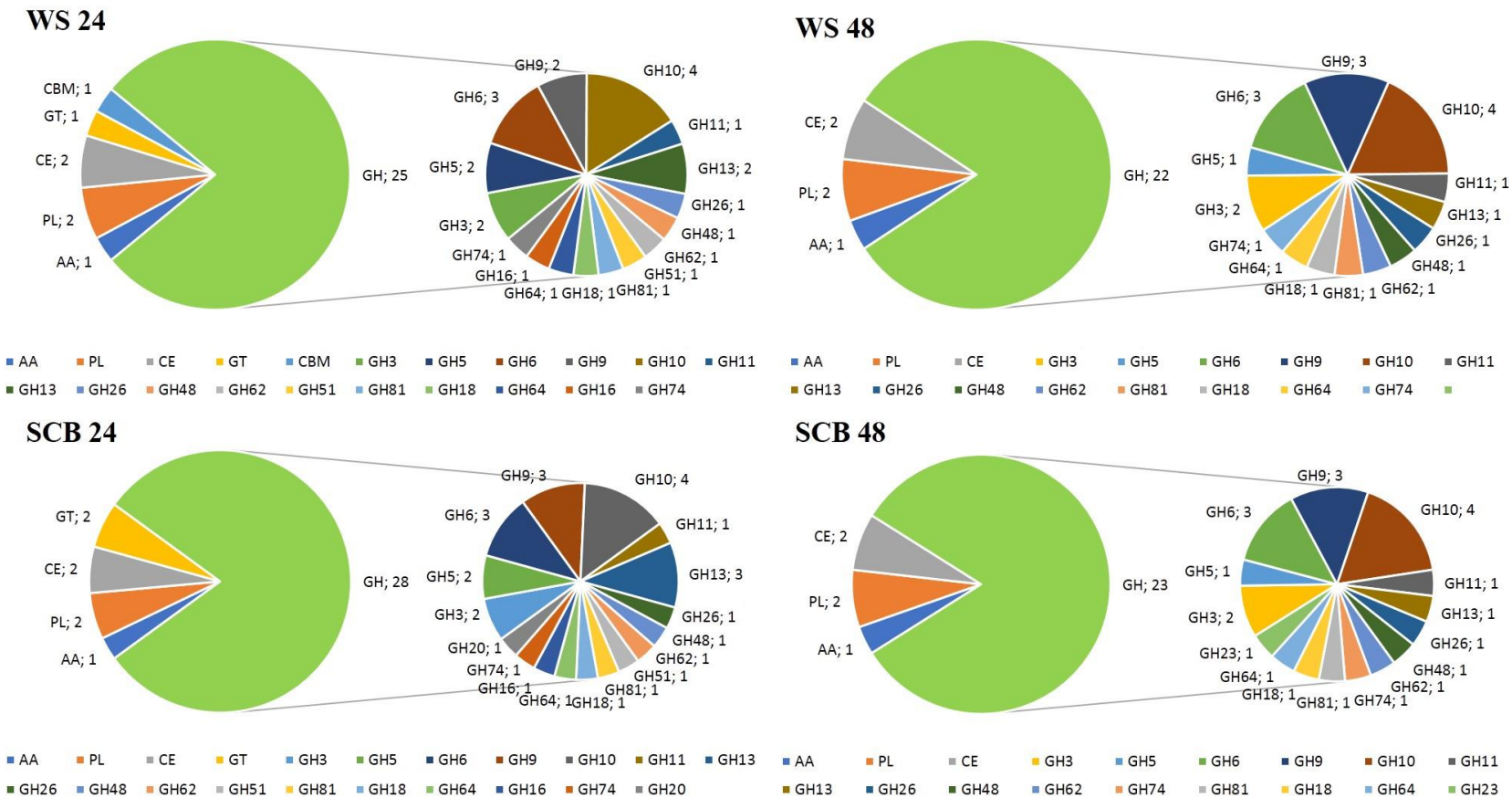


Figure 9: CAZyme classes and GH families identified in WS and SCB secretomes.

As shown by Venn's diagrams in Figure 10, 51 and 50 identified proteins were detected in all the secretomes collected after 24 and 48 hours of growth, respectively. Moreover, a deeper analysis highlighted that 40 proteins were always expressed in all the samples, collected after both 24 and 48 hours. Among them, 21, corresponding to 50%, were CAZymes.

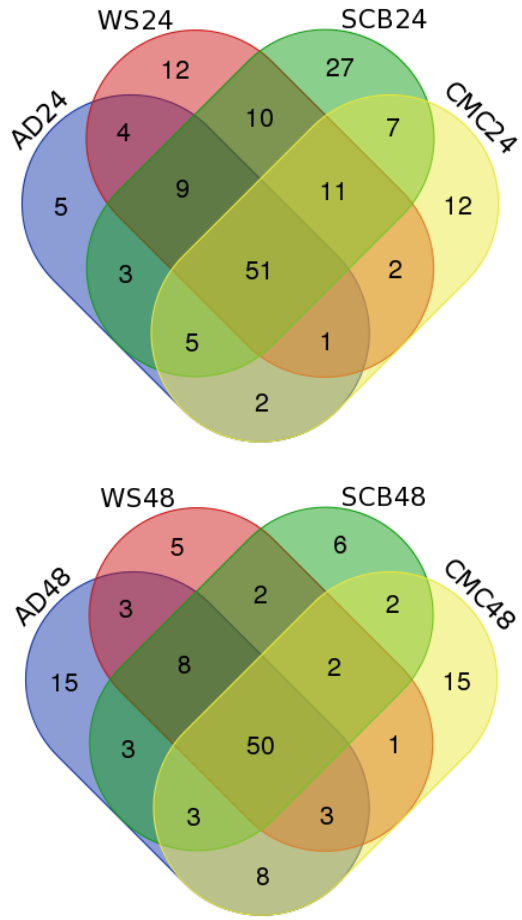


Figure 10: Venn's diagrams referred to the identified proteins in the secretomes collected after 24 (top) and 48 (bottom) hours of bacterial growth.

3.4.1 All the secretomes contained a same group of CAZymes secreted after both 24 and 48 hours of growth in presence of CMC or LCB

The proteomic analysis highlighted 21 CAZymes commonly secreted among samples. This observation may suggest that such CAZymes hold an important role in polysaccharide degradation, both in the initial and progressed phase.
Most of them belong to several families of the glycoside hydrolase class. This class includes CAZymes involved both in the degradation of cellulose and hemicellulose. The commonly known *C. fimi* cellulases that were identified in all the secretomes belong to the following families:

- GH6: 1,4-beta cellobiohydrolase, an endoglucanase known as CenA (gi|332340715), 1,4-beta cellobiohydrolase (gi|332338782) and Exoglucanase A, known as CbhA (gi|1708083).
- GH9: Glycoside hydrolase family 9, an endoglucanase known as CenB (gi|332337588) and Endoglucanase C, known as CenC (gi|121819).
- GH5: Glycoside hydrolase family 5, an endoglucanase known as CenD (gi|332339471).
- GH48: Exoglucanase B, known as CbhB (gi|1708084).

Other additional glucanases, not yet characterized in *C. fimi*, but already identified in *C. fimi* secretomes (Wakarchuk *et al*., 2016) belong to families:

- GH3: Glycoside hydrolase family 3 domain protein (gi|332338140).
- GH81: Endo-1,3(4)-beta-glucanase (gi|332341296).
- GH74: Cellulose-binding family II (gi|32339460). This family include endoglucanase, oligoxyloglucan reducing end-specific cellobiohydrolase and xyloglucanase.

Several identified glycosyl hydrolases are involved in hemicellulose degradation and are included in family:

- GH10: Endo-1,4-beta-xylanase (gi|332340687), Endo-1,4-beta-xylanase (gi|332337655) and Endo-1,4-beta-xylanase (gi|332338131). All of them are related to xylan degradation. Endo-1,4-beta-xylanase (gi|332338131) holds a Carbohydrate Esterase (CE4) domain, for the removal of acetylic substituents of the hemicellulose chain.
- GH11: Glycoside hydrolase family 11, formerly known as XynD (gi|332337936), involved in xylan degradation and deacetylation, as it brings a CE4 domain.
- GH62: Alpha-N-arabinofuranosidase (gi|332340686), that releases arabinose units from hemicellulose ramifications.
- GH26: Man26A (gi|5359710), involved in mannan degradation.
- GH13: Alpha-1,6-glucosidase, pullulanase-type (gi|332338678). Pullulanase are bacterial debranching enzymes, acting on pullulan, amylopectin and glycogen (Manners *et al*., 1997).

Other CAZymes commonly found in all the secretomes and not belonging to GH class were ascribed to family:

- AA10: Chitin-binding domain 3 protein (gi|332337832), a LPMO involved in cellulose degradation by an oxidative mechanism.
- PL1 and PL3: Pectate lyase/Amb allergen (gi|332341298) and Pectate lyase (gi|332337958), respectively. Pectate lyase cleave uronic acid-containing polysaccharide chains by a β-elimination mechanism.
- CE2: Cellulose-binding family II (gi|332339303), an acetyl xylan esterase.

3.4.2 Other proteins secreted after both 24 and 48 hours of growth in presence of CMC or LCB

Besides CAZymes, other proteins were revealed in all the secretomes.

Flagellin domain protein (gi|332338221), the main constituent of bacterial flagellum, was very abundant. It is frequently found in the growth media, as confirmed even by the analysis with SecretomeP; furthermore, its presence can be due to flagellum disruption caused by the agitation of the cultures (Komoriya *et al*., 1999; Kazemi-Pour *et al*., 2004).

Fibronectin type III domain protein (gi|332339883), a highly conserved actinobacterial protein already found in other secretomes of *C. fimi* and *C. flavigena* (Wakarchuk *et al*., 2016) was detected. Many secreted hydrolases contain Fibronectin type 3 (FN3) domains and in *Clostridium thermocellum* tandem FN3 domains from a cellulase (Cbh9A) have been shown to disrupt the surface of cellulose fibres (Kataeva *et al*., 2002). It has also been proposed that such a large protein could act as a scaffold for other secreted proteins, or could interact with carbohydrate substrates by means of FN3 domains (Wakarchuk *et al*., 2016).

Several membrane proteins were always found:

- Cell surface receptor IPT/TIG domain protein (gi|332341050): IPT (Ig-like, plexins, transcription factors) domains are found on the plasma membrane acting like receptors for transcription factors (Collesi *et al*., 1996).
- PKD domain containing protein (gi|332339822): PKD are extracellular domains involved in protein-protein and protein-carbohydrates interaction (Joachimiak *et al*., 2002).
- Extracellular ligand-binding receptor (gi|332339377), a membrane receptor (Rothberg *et al*., 1990).

- Extracellular solute binding proteins of families 1 (gi|332340270, gi|332339297, gi|332338842) and 5 (gi|332340445). They are typically membrane anchored lipoproteins related to different classes of nutrient uptake (Tam & Saier, 1993). Depending on sequence similarity, the extracellular solute binding proteins are grouped in eight families (Saurin & Dassa, 1994).

Among the proteins identified in all secretomes, there were even three peptidases and two nucleotides related proteins:

- Peptidase S1 and S6 chymotrypsin/Hap (gi|332338679), a secreted serine protease (Polgár, 1987);
- Peptidase S8 and S53 subtilisin kexin sedolisin (gi|332338096), a secreted endopeptidase (Siezen & Leunissen, 1997);
- Aminopeptidase Y (gi|332337698);
- TAP domain protein (gi|332337857): a receptor for mRNA nuclear exportation belonging to NXF1 (Nuclear RNA export factor 1) family (Grant *et al*., 2002);
- 5'-Nucleotidase domain-containing protein (gi|332339202): an important enzyme for nucleotides metabolism.

Finally, five proteins were not predicted as secreted but detected in all secretomes: Xylose isomerase (gi|332338137), Aminopeptidase N (gi|332338951), Alkaline phosphatase (gi|332340091), Enolase (gi|332340559) and Glutamine synthetase, type I (gi|332339768). Since it's known that some proteins, called "moonlighting proteins" may sometimes have a secondary function, not related to the main one, that can be explicated even in the extracellular environment (Huberts & van der Klei, 2010), we performed a search in the database MoonProt (http://moonlightingproteins.org/).

Enolase is a known moonlighting protein, as it is an important enzyme involved in the glycolysis process, for the conversion of 2-phosphoglycerate in

phosphoenolpyruvate and when secreted (by *Candida* and *Lactobacillus* strains, for example) is indeed involved in plasminogen and fibronectin binding (Jong *et al*., 2003; Antikainen *et al*., 2007).

3.4.3 Commonly identified proteins in all the secretomes at 24 hours

Besides intracellular proteins, not predicted as secreted by both the secretion prediction tools and most probably released after cellular lysis, in all the secretomes collected after 24 hours were exclusively identified:

- a peptidoglycan active enzyme, responsible for transglycosilation: Peptidoglycan glycosyltransferase (gi|332340923), belonging to GT51 family. The reason of its detection may reside in the fact that during growth phase or enhanced secretory activity, peptydoglycan remodelling enzymes promote the membran integration of transport and secretion systems (Madigan & Martinko, 2012);
- ligand binding proteins, in particular two Extracellular solute-binding protein family 1 (gi|332340306; gi|332340856) and Periplasmic binding protein/LacI transcriptional regulator (gi|332341019), a member of a family of primary receptors for chemotaxis and transport of many sugar based solutes (Saurin & Dassa, 1994);

Three moonlighting proteins were even detected:

- Glyceraldehyde-3-phosphate dehydrogenase, type I (gi|332339411) that just like enolase beside its role in the glycolysis process hold plasminogen binding capabilities highlighted in the secretome of *Bacillus anthracis* (Matta *et al*., 2010);
- Catalase (gi|332339192) that in the cytoplasm convert hydrogen peroxide, prortecting the cell from oxidative damage and when secreted by *Candida albicans* binds plasminogen (Crowe *et al*., 2003);

- Glucose-6-phosphate isomerase (gi|332338893) that is involved in the glycolysis process in the cytosol. Studies on the secretome *Lactobacillus crispatus* highlighted its secondary function, that is the binding of laminin and collagen type I (Kainulainen *et al*., 2012).

3.4.4 CMC, WS and SCB induced the expression of specific proteins at 24 hours

The proteins identified exclusively after 24 hours of growth on *A. donax* were intracellular or not involved in polysaccharide degradation. On the contrary, the other substrates stimulated the secretion of CAZymes or extracellular solute binding proteins, not revealed at 48 hours. Only in CMC24 was detected a protein belonging to CBM2 family: Cellulose-binding family II (gi|332337963). Two proteins were indeed identified only in WS24 secretome: Ricin B lectin (gi|332339947), belonging to CBM13 family and Extracellular solute-binding protein family 1 (gi|332338899). In the secretome SCB24, the exclusive proteins were six:

- two proteins named Extracellular solute-binding protein family 1 (gi|332337651, gi|332337695);
- Glycosyl transferase family 2 (gi|332338940), family GT2;
- Glycogen debranching enzyme GlgX (gi|332339655), family GH13;
- Alpha-L-arabinofuranosidase domain protein (gi|332340780), family GH51;
- Glycoside hydrolase, family 20, catalytic core (gi|332340615), family GH20.

All the revealed CAZymes were not predicted by both the secretion prediction tools.

3.4.5 Commonly identified proteins in all the secretomes at 48 hours

Three CAZymes, in particular glycoside hydrolases, were detected in all the secretomes collected after 48 hours of *C. fimi* growth:

- Glucan endo-1,3-beta-D-glucosidase (gi|332340860), family GH64;
- Glycoside hydrolase family 3 domain protein (gi|332339598), family GH3;
- Glycoside hydrolase family 18 (gi|332340692), family GH18.

Besides them, a protein called Extracellular solute-binding protein family 3 (gi|332339124) was revealed.
Other common proteins detected at 48 hours were intracellular and their presence in the secretome was probably due to cellular lysis.
The identified CAZymes are not characterized yet so it's impossible to state their function. GH3 family in fact includes several enzymatic activities, like glucosidase, xylan 1,4-β-xylosidase, β-glucosylceramidase, β-N-acetylhexosaminidase, α-L-arabinofuranosidase, glucan 1,3-β-glucosidase and glucan 1,4-β-glucosidase, while GH18 includes chitinase, lysozyme, endo-β-N-acetylglucosaminidase and peptidoglycan hydrolase with endo-β-N-acetylglucosaminidase specificity. However, their detection in all the secretomes collected at 48 hours, related to the highest enzymatic activities, suggested that they may have a possible synergistic role for lignocellulose degradation.

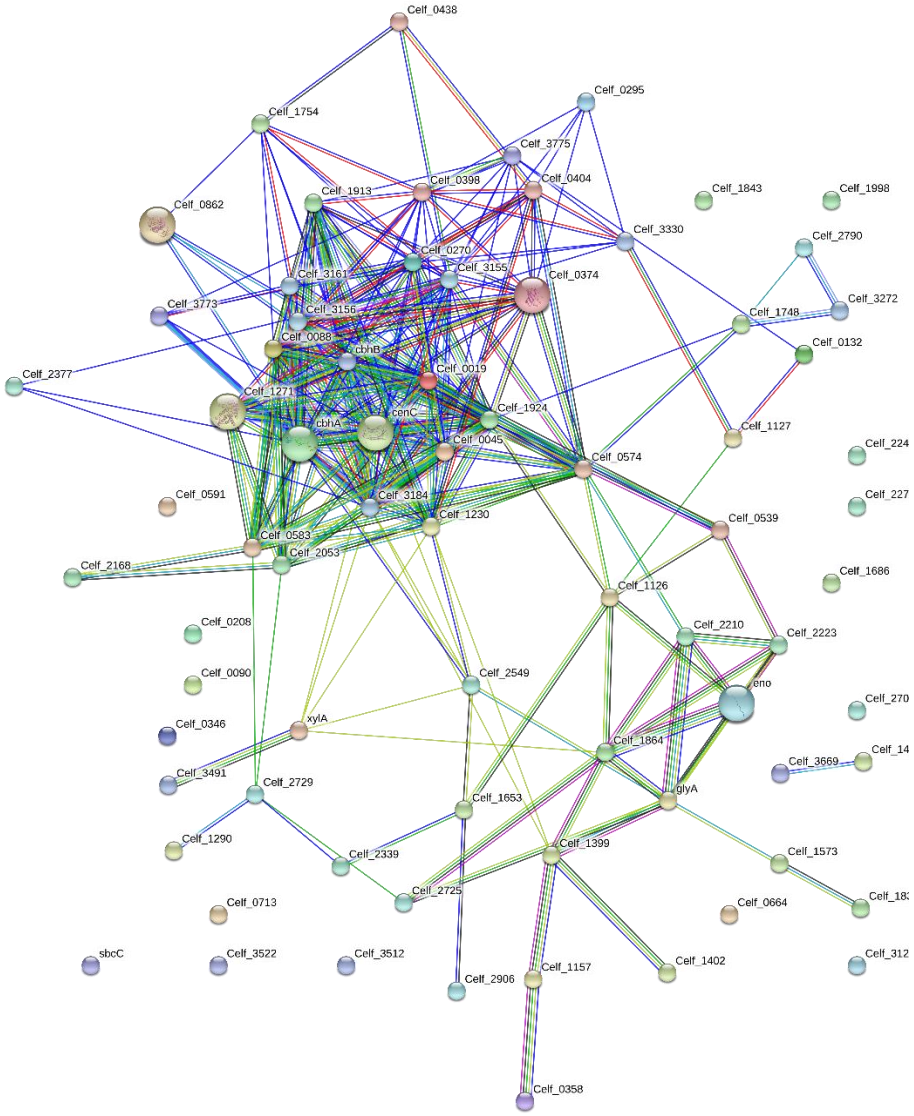
3.4.6 SCB and AD induced the expression of specific CAZymes at 48 hours

The previous observations suggested that *C. fimi* can degrade lignocellulose by means of a set of CAZymes, that is almost unchanged at substrate and time variation. However, two LCBs induced the expression of specific CAZymes, not detected in other secretomes. In particular, in the secretome collected after 48 hours of growth in presence of *A. donax* were identified:

- two CAZymes belonging to GH43 family: Glycoside hydrolase family 43 (gi|332340688) and Glycoside hydrolase family 43 (gi|332338345). This family includes several enzymatic activities, all involved in hemicellulose degradation.
- a Beta-phosphoglucomutase family hydrolase (gi|332339433), belonging to GH65 family but not predicted as secreted by both the tools utilized. Anyhow, STRING analysis highlighted its interaction with the CAZyme Alpha-1,6-glucosidase, pullulanase-type, as well as with other intracellular proteins.

The only peculiar CAZyme exclusively found in SCB48 secretome was indeed the Lytic transglycosylase catalytic (gi|332339713), belonging to GH23 family. STRING analyses showed its interaction with two proteins called Glycoside hydrolase family 3 domain protein (gi|332338140 and gi|332339598). The enzymes belonging to this family act on peptidoglycan and were already found in *C. fimi* secretomes obtained after growth on CMC (Wakarchuk *et al*, 2016).
Figure 11 shows a representative result of the analyses made by STRING. The network is referred to the secretome SCB48, that showed the highest enzymatic activity and only one additional CAZyme, so it can be considered a good baseline reference. Briefly, secreted proteins were grouped in two main clusters: the bigger one was constituted by CAZymes, with strong interactions among each other, while the second comprised proteins involved in intracellular processes. The observation of the revealed interactions may constitute a helpful tool for a better understanding of *C. fimi* LCB degrading machinery working. The protein Alpha-

1,6-glucosidase, pullulanase-type seems to connect the intracellular network with the secreted CAZymes network, suggesting that it may have an interesting role, that could represent an interesting subject of further studies. STRING results were very similar among secretomes (data not shown) and were associated to the same GO processes, except for CMC24 and CMC48 that lacked in "endo-1,4-beta-xylanase activity" (GO.0031176), "xylan catabolic process" (GO.0045493) and "cell wall organization or biogenesis" (GO.0071554).



Biological Process (GO)	
pathway ID	*pathway description*
GO:0044260	cellular macromolecule metabolic process
GO:0000272	polysaccharide catabolic process
GO:0030245	cellulose catabolic process
GO:0045493	xylan catabolic process
GO:0071554	cell wall organization or biogenesis

Molecular Function (GO)	
pathway ID	*pathway description*
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds
GO:0030247	polysaccharide binding
GO:0008810	cellulase activity
GO:0016162	cellulose 1,4-beta-cellobiosidase activity
GO:0031176	endo-1,4-beta-xylanase activity

Figure 11: STRING analysis output (legend in Table S9, Appendix A).

3.4.7 The emPAI evaluation showed changes in CAZymes secretion depending on time and substrate of growth

In order to evaluate possible changes in CAZymes secretion related to the substrate or the time of growth, emPAI values were analyzed. For each CAZyme, the protein content was calculated as the ratio between emPAI and ∑(emPAI), where ∑ (emPAI) is the summation of emPAI values for all the identified proteins (Ternan *et al*., 2014).

Since Flagellin domain protein was very abundant, the final emPAI value of each protein was re-calculated without take it in account. For each CAZyme, the corresponding emPAI values, calculated in each secretome, were reported in Table 2. emPAI values variations equal to or greater than 40% were considered as significant, for the detection of CAZymes secretion variability.

The analyses highlighted CAZymes always more expressed at 48 than 24 hours, like Endo-1,4-beta-xylanase (gi|332338131), Alpha-1,6-glucosidase, pullulanase-type (gi|332338678), Man26A (gi|5359710) and Glycoside hydrolase family 11, known as XynD (gi|332337936). Other CAZymes indeed increased at 48 hours only in all the secretomes obtained using LCB as substrate of growth: CenD, CenB, Glycoside hydrolase family 3 domain protein (gi|332339598) and Pectate lyase (gi|332337958). Only in SCB48 were observed increments of the secretion of CbhB, Exoglucanase Cex (gi|327179208), Chitin binding domain 3 protein (gi|332337832), CenA (gi|332340715) and Glucan-endo-1,3-beta-D-glucosidase (gi|332340860). CbhA and Endo-1,4-betaxylanase (gi|332340687) increased only in WS48. CenC and especially Cellulose binding family II (gi|332339460) increased in all secretomes, but in WS48 the variation was not significant; the same observation was made for Glycoside hydrolase family 3 domain protein (gi|332338140), that was not significant in AD48. 1,4-beta-cellobiohydrolase (gi|332338782) was the only CAZyme increased exclusively in AD48. Endo-1,4-betaxylanase (gi|332337655) was always found more secreted at 24 hours, like

Pectate lyase/Amb allergen (gi|332341298), Cellulose binding family II (gi|332339303) and Alpha-N-arabinofuranosidase (gi|332340686). Finally, some CAZymes were detected as "significantly oversecreted" at 48 hours in two culture conditions, like Endo-1,3-(4)-betaglucanase (gi|332341296) (CMC48 and WS48), Glycoside hydrolase family 18 (gi|332340692) (AD48 and SCB48) and Glycoside hydrolase family 9 (gi|332337613) (AD48 and SCB48).

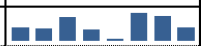
The influence of the substrate on CAZymes secretion was evaluated comparing for each protein the emPAI values obtained in the secretomes derived from the growth on LCBs with the one obtained in the CMC secretome. The analysis focused on the secretomes collected after 48 hours of growth, since they showed the highest enzymatic activities. It can be stated that after 48 hours of culture, LCBs induced a strong secretion of some CAZymes, with respect to CMC. CbhB, Endo-1,4-beta-xylanase (gi|332338131), CenB, Endo-1,4-betaxylanase (gi|332340687) and Endo-1,3-(4)-betaglucanase (gi|332341296) resulted more abundant in each LCB secretome. Other CAZymes indeed were overexpressed in the secretomes derived from one or two LCBs. Wheat straw stimulated the secretion of CbhA, Pectate lyase/Amb allergen (gi|332341298) and Pectate lyase (gi|332337958), while SCB stimulated the secretion of XynD. AD and especially SCB secretomes collected at 48 hours showed higher emPAI values for Glucan-endo-1,3-beta-D-glucosidase (gi|332340860) and Endo-1,4-betaxylanase (gi|332337655), while CenD was more secreted in presence of AD and WS.

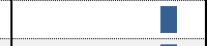
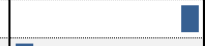
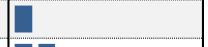
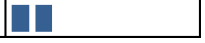
Glycoside hydrolase family 3 domain protein (gi|332338140) and Cellulose binding family II (gi|332339460) were more secreted in CMC48 and CMC24 respectively.

The emPAI analysis validated enzymatic activity assays results. For example, the higher xylanase activity observed in all the samples at 48 hours was in agreement with the increased secretion of hemicellulolytic enzymes, like Alpha-1,6-glucosidase, pullulanase-type, xylanase XynD and Endo-1,4-beta-xylanase. Moreover, the higher activities observed in the secretomes obtained after growth

on LCBs could be related to the increment of emPAI values of cellulolytic and hemicellulolytic enzymes, like endoglucanases CenB and CenD, Glycoside hydrolase family 3 domain protein (gi|332339598) and Pectate lyase (gi|332337958). Only in the secretome obtained after 48 hours of growth on SCB, that showed the highest enzymatic activities, five additional CAZymes shared increased emPAI values: cellobiohydrolase CbhB, Exoglucanase Cex, the oxidative enzyme Chitin binding domain 3 protein, Endoglucanase CenA and Glucan-endo-1,3-beta-D-glucosidase. The association of these data with enzymatic activity evaluation, by zymography and enzymatic assays, even suggested that some CAZymes may be secreted in different phases of the lignocellulolytic process. In each culture condition, for example, the protein Endo-1,4-beta-xylanase (gi|332337655) was found more secreted at 24 hours. So, it can be hypothesized that different xylanases, sharing the same enzymatic activity, could attend to different stages of the hydrolysis. Similarly, emPAI analyses highlighted that some CAZymes, mainly involved in hemicellulose degradation, were less secreted at 48 than 24 hours. Such enzymes, like Endo-1,4-beta-xylanase (gi|332337655), Pectate lyase/Amb allergen (gi|332341298), Cellulose binding family II (gi|332339303) and Alpha-N-arabinofuranosidase (gi|332340686) could degrade hemicellulose to increase cellulose accessibility, promoting the activity of further CAZymes. Hence, such analyses permitted not only to compare different secretomes but even to formulate hypothesis about *C. fimi* lignocellulose degrading approach.

Table 2: CAZymes emPAI values detected in each secretome.

Protein		CAZy classification	Secretomes (emPAI value)								Graphs
Definition	NCBI ID		CMC24	CMC48	AD24	AD48	WS24	WS48	SCB24	SCB48	
Exoglucanase A [CbhA]	gi|1708083	GH6	6.98	6.72	9.11	6.49	4.51	9.84	9.21	6.96	
Exoglucanase B [CbhB]	gi|1708084	GH48	2.08	2.26	3.65	4.50	2.97	4.14	1.92	7.86	
Endoglucanase C [CenC]	gi|121819	GH9	1.82	4.06	2.33	3.52	1.89	2.41	1.69	2.95	
Glycoside hydrolase family 5 [CenD]	gi|332339471	GH5	2.38	3.17	3.20	5.19	3.63	7.26	2.57	4.16	
Endo-1,4-betaxylanase	gi|332338131	GH10, CE4	0.45	0.73	2.26	3.20	1.16	3.76	1.11	3.42	
Glycoside hydrolase family 9 [CenB]	gi|332337588	GH9	2.55	2.93	2.90	6.33	3.23	7.26	3.06	9.19	
Endo-1,4-betaxylanase	gi|332337655	GH10	3.35	2.49	5.53	3.93	5.05	3.44	3.43	4.22	
Endo-1,4-betaxylanase	gi|332340687	GH10	2.02	1.24	1.90	2.18	1.81	2.67	2.69	3.74	
Pectate lyase/Amb allergen	gi|332341298	PL1	11.22	6.13	10.29	7.05	11.11	10.28	6.13	3.79	
Exoglucanase [Cex]	gi|327179208	GH10	7.46	7.91	5.27	4.98	5.31	0.82	2.88	6.22	
Glycoside hydrolase family 3 domain protein	gi|332338140	GH3	0.62	1.34	1.30	1.59	1.06	1.83	1.15	1.68	
Cellulose binding family II	gi|332339303	CE2	3.43	2.00	3.48	2.24	2.15	1.76	2.48	1.68	
Endo-1,3-(4)-betaglucanase	gi|332341296	GH81	0.17	0.35	1.30	1.31	0.59	1.66	0.91	1.12	
Chitin binding domain 3 protein	gi|332337832	AA10	2.12	1.06	0.85	0.69	1.68	1.09	0.60	1.05	
Alpha-N-arabinofuranosidase	gi|332340686	GH62	1.53	0.24	0.36	0.21	0.85	0.32	0.52	0.31	
1, 4-beta-cellobiohydrolase [CenA]	gi|332340715	GH6	4.69	5.67	0.92	1.27	1.53	1.68	0.92	3.11	
Alpha-1,6-glucosidase, pullulanase type	gi|332338678	GH13	0.82	2.47	0.38	0.80	0.31	0.69	0.08	0.98	
Man26A	gi|5359710	GH26	0.98	1.70	0.47	0.90	0.23	0.61	0.08	0.68	
1, 4-beta-cellobiohydrolase	gi|332338782	GH6	1.83	1.24	0.43	0.69	0.76	0.84	1.12	0.57	
Peptidoglycan glycosyltransferase	gi|332340923	GT51	0.07	-	0.34	-	0.09	-	0.17	-	
Glycoside hydrolase family 11 [XynD]	gi|332337936	GH11, CE4	0.42	0.66	0.30	0.56	0.36	0.73	0.22	1.42	
Glycoside hydrolase family 18	gi|332340692	GH18	-	0.94	0.51	0.81	0.28	0.15	0.25	1.01	

Protein		CAZy classification	Secretomes (emPAI value)								Graphs
Definition	NCBI ID		CMC24	CMC48	AD24	AD48	WS24	WS48	SCB24	SCB48	
Glycoside hydrolase family 3 domain protein	gi|332339598	GH3	1.31	1.31	0.09	1.57	0.07	0.94	0.10	0.26	
Glycoside hydrolase family 9	gi|332337613	GH9	0.07	-	0.11	0.69	-	0.10	0.05	0.92	
Cellulose binding family II	gi|332339460	GH74	0.07	1.09	0.11	1.48	0.16	0.17	0.15	1.29	
Pectate lyase	gi|332337958	PL3	0.88	1.06	0.21	1.13	0.80	3.02	0.35	1.40	
Glycoside hydrolase family 43	gi|332340688	GH43	-	-	-	0.15	-	-	-	-	
Cellulase	gi|332337938	GH5	1.24	1.59	-	0.28	0.12	-	0.16	-	
Glucan-endo-1,3-beta-D-glucosidase	gi|332340860	GH64	-	0.11	-	0.19	0.28	0.15	0.49	0.85	
Glycoside hydrolase family 43	gi|332338345	GH43	-	-	-	0.04	-	-	-	-	
Betaphosphoglucomutase family hydrolase	gi|332339433	GH65	-	-	-	0.05	-	-	-	-	
Cellulose binding family II	gi|332337964	CE1	0.14	-	-	-	0.17	0.19	0.22	0.18	
Extracellular solute binding protein family 1	gi|332339858	GH13	0.14	-	-	-	0.17	-	0.36	-	
Ricin B lectin	gi|332339947	CBM13	-	-	-	-	0.16	-	-	-	
Alpha-L-arabinofuranosidase domain protein	gi|332340852	GH51	-	0.11	-	-	0.14	-	-	-	
Glycoside hydrolase family 16	gi|332340645	GH16	-	-	-	-	0.17	-	0.21	-	
Glycosyl transferase family 2	gi|332338940	GT2	-	-	-	-	-	-	0.15	-	
Glycogen debranching enzyme GlgX	gi|332339655	GH13	-	-	-	-	-	-	0.05	-	
Alpha-L-arabinofuranosidase domain protein	gi|332340780	GH51	-	-	-	-	-	-	0.08	-	
Glycoside hydrolase, family 20, catalytic core	gi|332340615	GH20	-	-	-	-	-	-	0.08	-	
Lytic transglycosylase catalytic	gi|332338270	GH23	-	-	-	-	-	-	-	0.20	
Cellulose binding family II	gi|332337963	CBM2	0.52	-	-	-	-	-	-	-	
Glycoside hydrolase family 9	gi|332339254	GH9	0.06	0.06	-	-	-	-	-	-	

4 Conclusions

The characterization of *Cellulomonas fimi* secretome highlighted that this microorganism with lignocellulose-degrading capabilities can be cultured on pretreated lignocellulosic biomass. The analysis of the secretomes obtained on three different LCBs showed that 21 CAZymes were always secreted and maybe represent the basic degrading apparatus used by *C. fimi*. The enzymatic activity assays confirmed the presence of endoglucanase and xylanase activities, while β-glucosidase was undetectable or very low, in according to the lack of secreted betaglucosidases. Xylanase activity was the highest, followed by endoglucanase and both these activities were higher at 48 hours than at 24. This data can be confirmed by the emPAI values analysis. In fact, in all the secretomes, at 48 hours, it was observed an increase of hemicellulose-degrading CAZymes (Endo-1,4-beta-xylanase, Alpha-1,6-glucosidase, pullulanase-type, Man26A and XynD). Furthermore, the higher enzymatic activities detected in the secretomes obtained on LCBs reflect the increment of cellulolytic CAZymes, like CenD, CenB, CenC and Glycoside hydrolase family 3 domain protein.

The highest enzymatic activities were associated to SCB48, where increments in the secretion of CbhB, Exoglucanase Cex, Chitin binding domain 3 protein, CenA and Glucan-endo-1,3-beta-D-glucosidase were observed.

Finally, some enzymes, like Endo-1,4-betaxylanase (gi|332337655) (GH10), Pectate lyase/Amb allergen (PL1), Cellulose binding family II (CE2) and Alpha-N-arabinofuranosidase (GH62) were always found more secreted at 24 hours, suggesting their roles in the primary phases of LCB degradation.

References

Adav SS, Cheow ESH, Ravindran A, Dutta B, Sze SK. Label free quantitative proteomic analysis of secretome by Thermobifida fusca on different lignocellulosic biomass. *J Proteomics*. 2012; 75: 3694- 3706.

Antikainen J, Kuparinen V, Lahteenmaki K, Korhonen TK. Enolases from Gram-positive bacterial pathogens and commensal lactobacilli share functional similarity in virulence-associated traits. *FEMS Immunol Med Microbiol*. 2007; 51(3): 526-34.

Bastida F, Hernández T, García C. Metaproteomics of soils from semiarid environment: Functional and phylogenetic information obtained with different protein extraction methods. *J Proteomics*. 2014; 101: 31- 42.

Bendtsen JD, Kiemer LA, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. *BMC Microb*. 2005; 5: 58.

Boatti L, Robotti E, Marengo E, Viarengo A, Marsano F. Effects of nickel, chlorpyrifos and their mixture on the Dictyostelium discoideum proteome. *Int J Mol Sci*. 2012; 13: 15679-15705.

Brown KJ, Formolo CA, Seol H, Marathi RL, Duguez S, An E, Pillai D, Nazarian J, Rood BR, Hathout Y. Advances in the proteomic investigation of the cell secretome. *Expert Rev Proteomics*. 2012; 9(3): 337- 345.

Cattaneo C, Spertino S, Boatti L, Icardi S, Cavaletto M. Protein fingerprinting in the choice of cellulase cocktails for the conversion of lignocellulosic biomass. *Anal Methods-UK*. 2014; 6: 4046-4055.

Chen S, Rillig MC, Wang W. Improving soil protein extraction for metaproteome analysis and glomalin-related soil protein detection. *Proteomics*. 2009; 9: 4970–4973.

Collesi C, Santoro MM, Gaudino G, Comoglio PM. A splicing variant of the RON transcript induces constitutive tyrosine kinase activity and an invasive phenotype. *Mol Cell Biol*. 1996; 16(10): 5518-5526.

Crowe JD, Sievwright IK, Auld GC, Moore NR, Gow NA, Booth NA. Candida albicans binds human plasminogen: identification of eight plasminogen-binding proteins. *Mol Microbiol*. 2003; 47(6): 1637-51.

Fernandes I, Alves A, Correia A, Devreese B, Esteves AC. Secretome analysis identifies potential virulence factors of Diplodia corticola, a fungal pathogen involved in cork oak (Quercus suber) decline. *Fungal Biol- UK*. 2014; 118: 516- 523.

Grant RP, Hurt E, Neuhaus D, Stewart M. Structure of the C-terminal FG-nucleoporin binding domain of Tap/NXF1. *Nat Struct Biol*. 2002; 9: 247-51.

Huberts DHEW, van der Klei IJ. Moonlighting proteins: An intriguing mode of multitasking. *Biochim Biophys Acta*. 2010; 1803: 520-525.

Jain S, Graham RLJ, McMullan G, Ternan NG. Proteomic analysis of the insoluble subproteome of Clostridium difficile strain 630. *FEMS Microbiol Lett*. 2010; doi:10.1111/j.1574-6968.2010.02111.x.

Joachimiak A, Springer TA, Zhang RG, Wang JH, Liu JH, Jing H, Takagi J, Lindgren S. Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins. *Structure*. 2002; 10: 1453-1464.

Jong AY, Chen SH, Stins MF, Kim KS, Tuan TL, Huang SH. Binding of Candida albicans enolase to plasmin(ogen) results in enhanced invasion of human brain microvascular endothelial cells. *J Med Microbiol*. 2003; 52(8): 615-22.

Kainulainen V, Loimaranta V, Pekkala A, Edelman S, Antikainen J, Kylvj R, Laaksonen M, Laakkonen L, Finne J, Korhonen TK. Glutamine synthetase and glucose-6-phosphate isomerase are adhesive moonlighting proteins of Lactobacillus crispatus released by epithelial cathelicidin LL-37. *J Bacteriol*. 2012; 194(10): 2509-19. DOI: 10.1128/JB.06704-11.

Kataeva IA, Seidel RD 3rd, Shah A, West LT, Li XL, Ljungdahl LG. The fibronectin type 3-like repeat from the Clostridium thermocellum cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. *Appl Environ Microbiol*. 2002; 68(9): 4292-300.

Kazemi-Pour N, Condemine G, Huouvieux-Cotte-Pattat N. The secretome of the plant pathogenic bacterium Erwinia chrysanthemi. *Proteomics*. 2004; 4: 3177-3186.

Komoriya K, Shibano N, Higano T, Azuma N, Yamaguchi S, Aizawa SI. Flagellar proteins and type III-exported virulence factors are the predominant proteins secreted into the culture media of Salmonella typhimurium. *Mol Microbiol.* 1999; 34: 767-779.

Madigan MT, Martinko JM. 2012. Biologia dei microorganismi, Pearson.

Manfredi M, Martinotti S, Gosetti F, Ranzato E, Marengo E. The secretome signature of malignant mesothelioma cell lines. *J Proteomics*. 2016; 145: 3-10.

Mangiapane E, Lamberti C, Pessione A, Galano E, Amoresano A, Pessione E. Selenium effects on the metabolism of a Se-metabolizing Lactobacillus reuteri: analysis of envelope-enriched and extracellular proteomes. *Mol Biosyst*. 2014; 10: 1272-1280.

Manners DJ. Observations on the specificity and nomenclature of starch debranching enzymes. *J Appl Glycosci*. 1997; 44: 83–85.

Matta SK, Agarwal S, Bhatnagar R. Surface localized and extracellular Glyceraldehyde-3-phosphate dehydrogenase of Bacillus anthracis is a plasminogen binding protein. *Biochim Biophys Acta*. 2010; 1804(11): 2111-20. DOI: 10.1016/j.bbapap.2010.08.004.

Nelson N. A photometric adaption of the Somogyi method for the determination of glucose. *J Biol Chem*. 1944; 153: 375-380.

Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*. 1988; 9: 255-262.

Nizami AS, Rehan M, Waqas M, Naqvi M, Ouda OKM, Shahzad K, et al. Waste biorefineries: Enabling circular economies in developing countries. *Bioresource Technol*. 2017; 241: 1101-1117.

Peila C, Coscia A, Bertino E, Cavaletto M, Spertino S, Icardi S, Tortone C, Visser GHA, Gazzolo D. Effects of Holder pasteurization on the protein profile of human milk. *Italian Journal of Pediatrics*. 2016; 42: 36. DOI 10.1186/s13052-016-0248-5.

Petersen TN, Brunak S, Heijne von G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011; 8: 785-786.

Polgár L. The mechanism of action of aspartic proteases involves 'push-pull' catalysis. *FEBS Lett*. 1987; 13: 219(1): 1-4.

Rothberg KG, Ying YS, Kamen BA, Anderson RG. Cholesterol controls the clustering of the glycophospholipid-anchored membrane receptor for 5-methyltetrahydrofolate. *J Cell Biol*. 1990; 111: 2931–2938.

Saurin W, Dassa E. Sequence relationships between integral inner membrane proteins of binding protein-dependent transport systems: evolution by recurrent gene duplications. *Protein Sci*. 1994; 3: 325-344.

Schneider WDH, Gonçalves TA, Uchima CA, Couger MB, Prade R, Squina FM, Dillon AJP, Camassola M. Penicillium echinulatum secretome analysis reveals the fungi potential for degradation of lignocellulosic biomass. *Biotechnol Biofuels*. 2016; 9: 66.

Siezen RJ, Leunissen JA. The superfamily of subtilisin-like serine proteases. *Protein sci*. 1997; 6: 501-523.

Spertino S, Cipriani V, De Angelis C, Giuffrida MG, Marsano F, Cavaletto M. Proteome profile and biological activity of caprine, bovine and human milk fat globules. *Mol Biosyst*. 2012; 8: 967–974.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015; 43(1): D447-D452.

Tam R, Saier MH. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev*. 1993; 57(2): 320-46.

Ternan NG, Jain S, Graham RLJ, McMullan G. Semiquantitative Analysis of Clinical Heat Stress in Clostridium difficile Strain 630 Using a GeLC/MS Workflow with emPAI Quantitation. *PLoS One*. 2014; doi:10.1371/journal.pone.0088960.

Villela Filho M, Araujo C, Bonfá A, Porto W. Chemistry Based on Renewable Raw Materials: Perspectives for a Sugar Cane-Based Biorefinery. *Enzyme Research*. 2011; DOI:10.4061/2011/654596.

Wakarchuk WW, Brochu D, Foote S, Robotham A, Saxena H, Erak T, Kelly J. Proteomic analysis of the secretome of Cellulomonas fimi ATCC 484 and Cellulomonas flavigena ATCC 482. *PLoS One*. 2016; 11: e0151186.

Zhu N, Liu J, Yang J, Lin Y, Yang Y, Jil L, Li M, Yuan H. Comparative analysis of the secretomes of Schizophyllum commune and other wood-decay basidiomycetes during solid-state fermentation reveals its unique lignocellulose-degrading enzyme system. *Biotechnol Biofuels.* 2016; 9: 42.

Chapter III
Metagenomic analysis

In the last years, the search for alternative energy sources has obtained a growing interest, because of the dramatic rate of fossil fuels depletion and the drastic changes in the climatic conditions. Therefore, an innovative and appropriate use of naturally abundant and renewable sources is important to reach a sustainable future. The sustainability concept represents the driving force behind biorefineries, an emerging and promising sector with a great potential to process several feedstocks into value-added products (Arevalo-Gallegos *et al*., 2017). The effective conversion of lignocellulosic biomass (LCB) in fermentable monomers still represents an important target that can be reached using microbial enzymes. As the enzymatic hydrolysis represents the most expensive step of biorefineries processes, discovering new effective enzymatic activities or improving them is of primary importance to contain costs. In nature, the enzymatic hydrolysis of plant matter is operated by fungi and bacteria, that constitute communities of degrading microorganisms, acting in synergy (Wang *et al*., 2016). Nowadays, the leading industrial source of hydrolytic cocktails is the fungus *Trichoderma reesei*, but the complexity of the required enzymatic activities to reach better levels of lignocellulose degradation suggests the use of microbial consortia, instead of single microorganism strains (Cheng & Zhu, 2012). The recent development of next generation sequencing technologies has permitted to investigate complex environmental samples, like soil, using a metagenomic approach. The most important advantage of this cloning-independent method is the avoidance of bias derived from cloning and PCR amplification (Xu *et al*., 2015). Metagenomic approach has been used so far to study microbial communities in a wide range of environments (Iliev *et al*., 2017; Araujo *et al*., 2017a; Araujo *et al*., 2017b; Al-Hunaiti *et al*., 2017) and represents a promising opportunity for better understanding the structure of microbial niches of an environmental sample. The detected microorganisms are then considered the most probably responsible for the biological processes observed in that habitat. Soil is probably the most challenging of all natural environments for microbiologists, with respect to the

microbial community size and the diversity of species present (Mocali & Benedetti, 2010). The reassociation kinetics of DNA isolated from various soil samples suggest that the number of distinct prokaryotic genomes could vary from 2000 to 18000 genomes per gram of soil. These numbers might be underestimated because of the possible exclusion of genomes representing rare and unrecovered species (Torsvik *et al*., 2002). Soil environments heterogeneity, in terms of chemical and biological properties, probably contributes to the large microbial biodiversity (Daniel, 2005). Moreover, methodological biases remain an enormous challenge for microbial community characterization (Nesme *et al*., 2016).

1 Aim of work

The application of a metagenomic approach to decaying wood samples could provide a useful overview on microbial communities structure, highlighting microorganisms that could represent interesting object for lignocellulose degradation studies. Therefore, a metagenomic analyses was conducted on two different decaying woods, compared to two control soil samples. Bacterial and fungal genomic DNA was extracted and amplified before the taxonomic assignment for the detection of the most represented microorganisms in decaying woods. This study was conducted in collaboration with SmartSeq, a spin-off of our university.

2 Materials and Methods

2.1 Sample collection

Decaying wood samples (DW2 and DW4, each derived from five different specimens) were collected from parts of decaying plane trees and soil samples (S1 and S3, each derived from five different specimens) were collected from flowerbeds near healthy plane trees, from 10 cm below the surface and sieved (2 mm mesh) to remove large particles. All samples were collected on April 2015 in Salassa, Turin (Italy), kept in ice during transport and stored at -20°C until the analysis.

2.2 Physicochemical analysis

In order to estimate the percentage of humidity, aliquots of each sample were dried at 105 °C until reach a constant dry weight. The ratio between the weight loss and the starting weight represents the humidity content, according to Hausenbuiller (Hausenbuiller, 1975). The loss of ignition method described by Dean (Dean, 1974) was used for the organic matter content evaluation. Briefly, aliquots of dried samples were weighted, heated at 550 °C for 8 hours and kept in drier with silica gel, until they reach room temperature. The organic matter content of dried samples was calculated as the ratio between their weight loss and the starting weight. To determine sample pH, one part of soil or decaying wood was mixed with 5 parts of 0.01 M $CaCl_2$ solution. After agitation, pH was measured using a pHmeter (Schofield & Taylor, 1955).

2.3 Soil DNA extraction

Genomic DNA was extracted from each sample using the PowerSoil® DNA Isolation Kit according to the manufacturer's instructions (MO BIO Laboratories, Inc., Carlsbad, CA, United States). A 0.25 g soil sample was added to each tube with three replicates, and the DNA was dissolved in 70 µL elution buffer and stored at -20 °C before use. The purity and concentration of extracted nucleic acids was determined spectrophotometrically, by taking absorbance at 230, 260, and 280 nm.

2.4 DNA amplification and 454 sequencing

Polymerase chain reaction (PCR) was used to amplify DNA for analysis of bacterial and fungal communities. The V1-V4 hypervariable regions of bacterial 16S rRNA genes were amplified using primers 8F and b758 (Klindworth *et al*., 2013) and fungal internal transcribed spacer (ITS) 1-2 loci were amplified using primers ITS1-F and ITS4 (Toju *et al*., 2012). Primer pairs were modified with the 454 pyrosequencing (Roche diagnostics) adapters and Multiplex Identifier (MID). PCR reactions contained 5-20 ng DNA template were performed in a total volume of 20 µl (100 µM of dNTPs, 1.5 mM $MgCl_2$, 1× Buffer [67 mM Tris-HCl pH 8.8; 16.6 mM $(NH_4)_2$ SO_4; 0.01% Tween-20; $MgCl_2$ 5 mM], and 0.08 U of Taq DNA Polymerase (Thermo Fisher Scientific)) in a Bio-Rad thermocycler (CX1000, Bio-Rad Laboratories Srl) following specific conditions as recommended by Roche Diagnostics. Briefly, an initial denaturation at 94 °C for 5 min followed by 34 cycles of 94 °C for 1 min, 60 °C for 1 min, and 72 °C for 5 min, and a final extension of 72 °C for 10 min.

Amplicons obtained for the evaluation of bacterial 16S and fungal ITS were fragments of about 550 bp and 650 bp respectively.

The PCR products were purified using the Agencourt® AMPure® Kit (Beckman Coulter, CA, USA), quantified using QuantiT PicoGreen® kit (Thermo Scientific, Wilmington, DE) and pooled for pyrosequencing. Pyrosequencing using 454/Roche GS Junior was carried out according to the manufacturer's instructions.

2.5 Data analysis

Pyrosequencing raw data were processed using a custom bioinformatics pipeline as described in Novello *et al*. (Novello *et al*., 2017).

For bacterial communities, the taxonomic assignment was performed using RDP database (Cole *et al*., 2014) and for fungal communities the taxonomy classification was performed using UNITE fungal ITS database (release 7.2) (Abarenkov *et al*., 2010).

The data obtained was graphically represented by Microsoft Excel (Office - Microsoft Package).

Descriptive statistical analysis was made with RAM package of R statistical software 3.1.34 to obtain: (i) rarefaction curve graph, (ii) biodiversity indices (Shannon-Wiener Index, Simpson Index, Observed species), (iii) Beta diversity graph and (iv) Principal Component Analysis graph.

3 Results and Discussion

3.1 Physical-chemical properties of samples

The evaluation of physical-chemical properties highlighted important differences among samples. Decaying woods (sample DW2 and DW4) showed higher percentages of humidity and organic matter content than soils (sample S1 and S3) and alkaline pH (Tab. 1).

Table 1: Physical-chemical properties of samples.

Sample	pH	Organic matter (%)	Humidity (%)
S1	6.1	7.6	22.7
S3	4.6	7.2	19.0
DW2	7.6	91.8	85.2
DW4	8.3	88.7	84.5

3.2 Metagenomic analysis

The association of bacteria with wood has been recognized since the 1950s and 1960s but their role in wood deterioration received little attention compared to fungal one. In 1971, Greaves (Greaves, 1971) classified bacteria into groups, depending on their role in wood decay, distinguishing bacteria that affect wood permeability, attack wood structure, breakdown wood in synergy with other bacteria or represent passive colonizers. The initial stages of wood decay are related to bacteria succession, before fungal predominance (Clausen, 1995) and this observation is confirmed even by our data.

The number of reads obtained after PCR amplification are shown in Table 2.

Table 2: Fungal and bacterial reads (sequences) obtained after sequencing (S1-S3: soils; DW2-DW4: decaying woods)

Sample	Bacterial	Fungal
S1	13008	6938
S3	9540	7443
DW2	10893	11179
DW4	9106	10586

The fungal and bacterial diversity was performed analyzing the rarefaction curves (Fig. 1) that provide a measure of the depth of our experiments. In ecology, rarefaction permits to assess species richness from the results of sampling and a rarefaction curve is a plot of the number of species as a function of the number of samples. The Figure 1 shows that the number of observations was sufficient to evaluate the species richness in samples, since rarefaction curves reach the plateau.

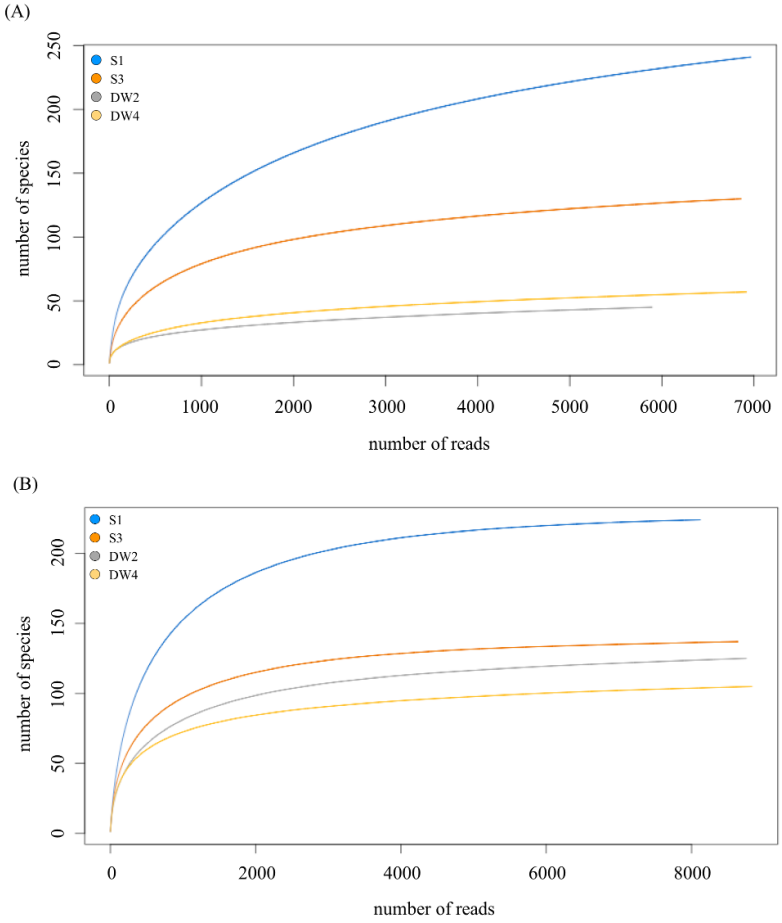


Figure 1: Rarefaction curves for each sample in (A) fungal and (B) bacterial communities.

To evaluate the local diversity of community (alpha diversity parameter), three different biological indices were used: (i) observed species, (ii) Shannon and (iii) Simpson indices showing that soils and decaying woods are inhabited by different fungal and bacterial communities. Decaying woods (DW2 and DW4) shared a very similar population, while soil samples (S1 and S3) have a greater variability, both regarding Fungi and Bacteria (Tab. 3).

Table 3: Fungal and bacterial biological indexes.

Sample	Observed species Fungi/bacteria	Simpson index Fungi/bacteria	Shannon index Fungi/bacteria
S1	241 / 224	0.97/ 0.96	4.10/ 4.15
S3	130/ 137	0.93/ 0.95	3.29/ 3.67
DW2	45/ 125	0.77/ 0.93	1.95/ 3.34
DW4	57/ 105	0.81/ 0.94	2.07/ 3.38

Beta diversity, the change in diversity of species from different environments, was calculated using the Bray-Curtis algorithm. As shown in Figure 2, the samples were clustered in two distinct groups, soils and decaying woods.

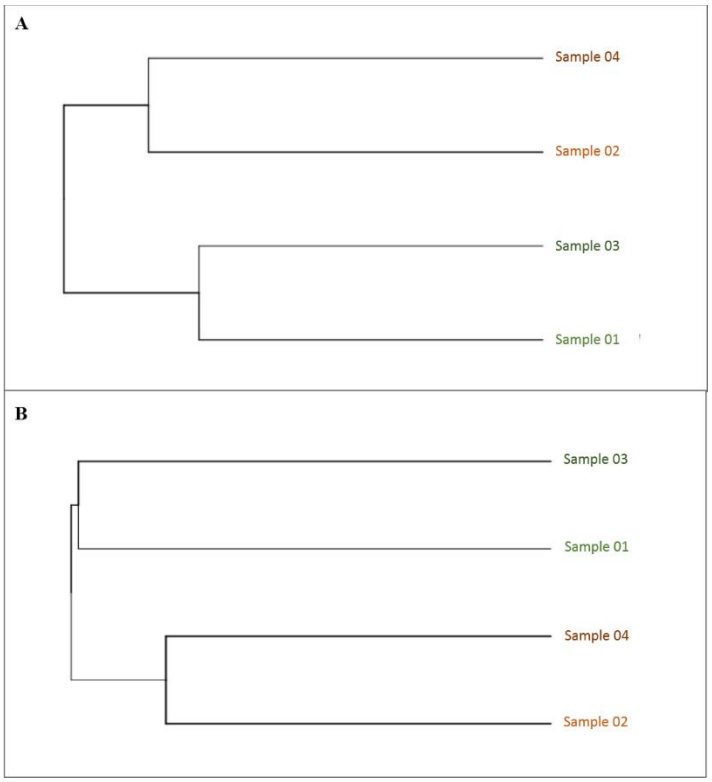


Figure 2: Beta diversity analysis and sample clusterization.

Moreover, the principal component analysis, referred both to fungal and bacterial communities, highlighted one more time that microbial communities are different among samples. In particular, the major variability in decaying wood samples was observed for fungal communities (Fig. 3), while soils were characterized by a higher bacterial variability (Fig.4).

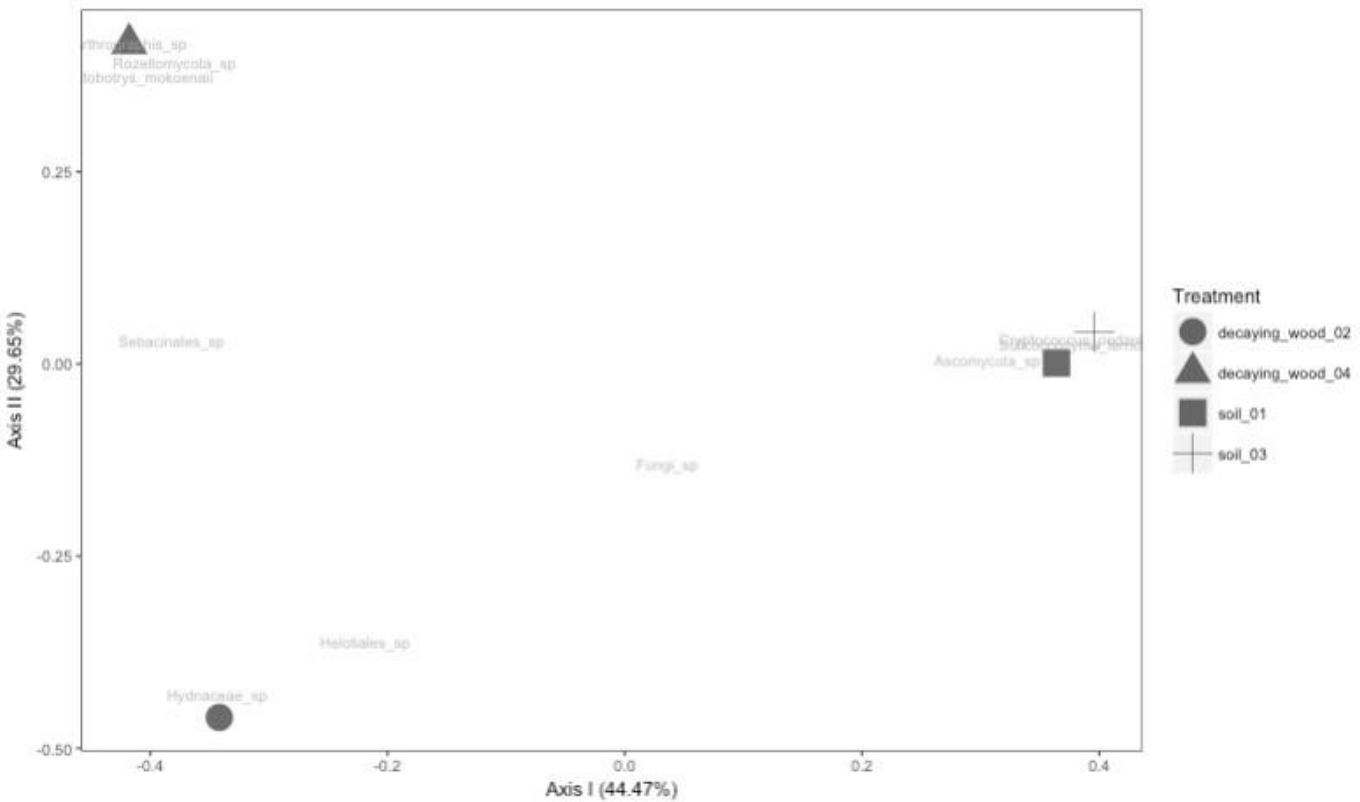


Figure 3: Principal Component Analysis – comparison of the fungal community at species level. Soil and decaying wood samples are well-grouped by the axis 1 (axis 1 - 44.47%, axis 2 – 29.65%).

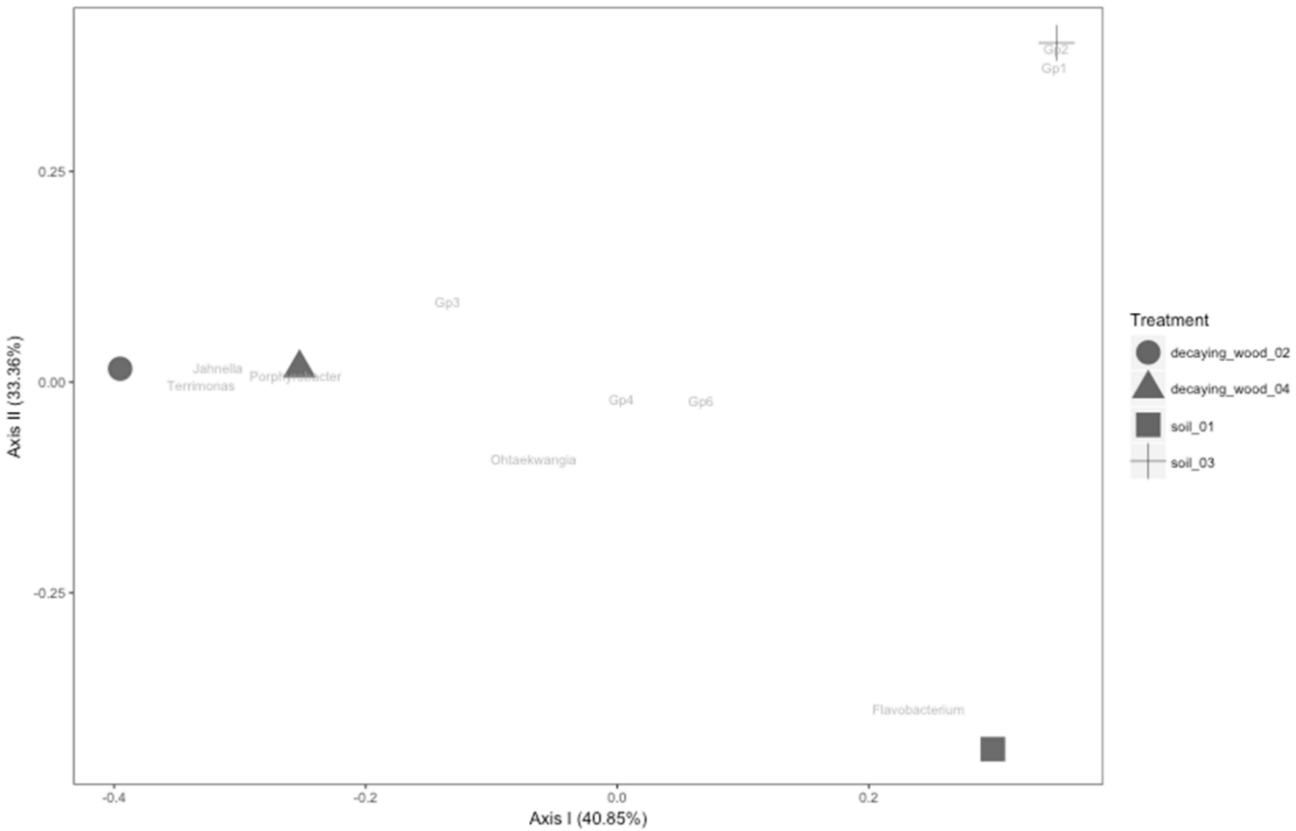


Figure 4: Principal Component Analysis – comparison of the bacterial community at genus level. Soil and decaying wood samples are well-grouped by the axis 1 (axis 1 - 40.85%, axis 2 – 33.36%).

Metagenomic data were analyzed at all taxonomic levels (data reported in Tabb. S1 and S2, Appendix B). The organisms, at different taxonomic levels, showing at least a 40% increase of the abundance value in DW samples compared to S samples were considered predominant in decaying woods.

2.2.1. Fungal communities

Regarding the fungal communities, Ascomycota and Basidiomycota were the most represented phyla, while Zygomycota were more abundant in soils than in decaying woods and Rozellomycota were revealed especially in one decaying wood sample (DW4) (Fig. 5A). Ascomycota are known for producing enzymes involved in the degradation of biopolymers, like cellulose and hemicellulose (Ma *et al*., 2013). Basidiomycetes survive better in presence of high organic matter content and are involved in C cycling in several environments, even acting as wood decomposers (Castro *et al*., 2016). Rozellomycota have an almost ubiquitous distribution in all habitat types (Grossart *et al*., 2016), since they have been isolated from several environments, such as soil, freshwater, marine, and anoxic habitats. Phylogenetic studies have suggested that their specialization for physical habitat is relatively limited, but distribution of these groups may be influenced by substrate pH. All known members of Rozellomycota are obligate pathogens of other eukaryotes, such as amoebae, algae, and other fungi so their distribution may depend indirectly on interaction specificity and habitat preference of host organisms (Tedersoo *et al*., 2017). Such information may justify the revelation of this phylum almost exclusively in one of the two decaying wood samples analyzed.

Regarding the class taxonomic level (Fig. 5B), the most represented was Agaricomycetes, followed by Saccharomycetes. Genomic and metagenomic studies on wood degrading fungi highlighted that most of them belong to the class Agaricomycetes (phylum Basidiomycota), even if not all Agaricomycetes are wood decayers (Ohm *et al*., 2014). Agaricomycetes in fact could also be

pathogens, parasites and symbionts (Araujo *et al*., 2017b). Dothideomycetes, Eurotiomycetes and Leotiomycetes were differently distributed among decaying wood samples. Dothideomycetes and Eurotiomycetes were more abundant in sample DW4, while Leotiomycetes in DW2. Phylogenetic analyses of species found during a survey of saprophytic microfungi on decomposing woody, herbaceous debris and soil from different regions in Southern Europe proved that most of them were related to Dothideomycetes and to lesser extent to Leotiomycetes (Hernandez-Restrepo *et al*., 2017), giving support to our observations.

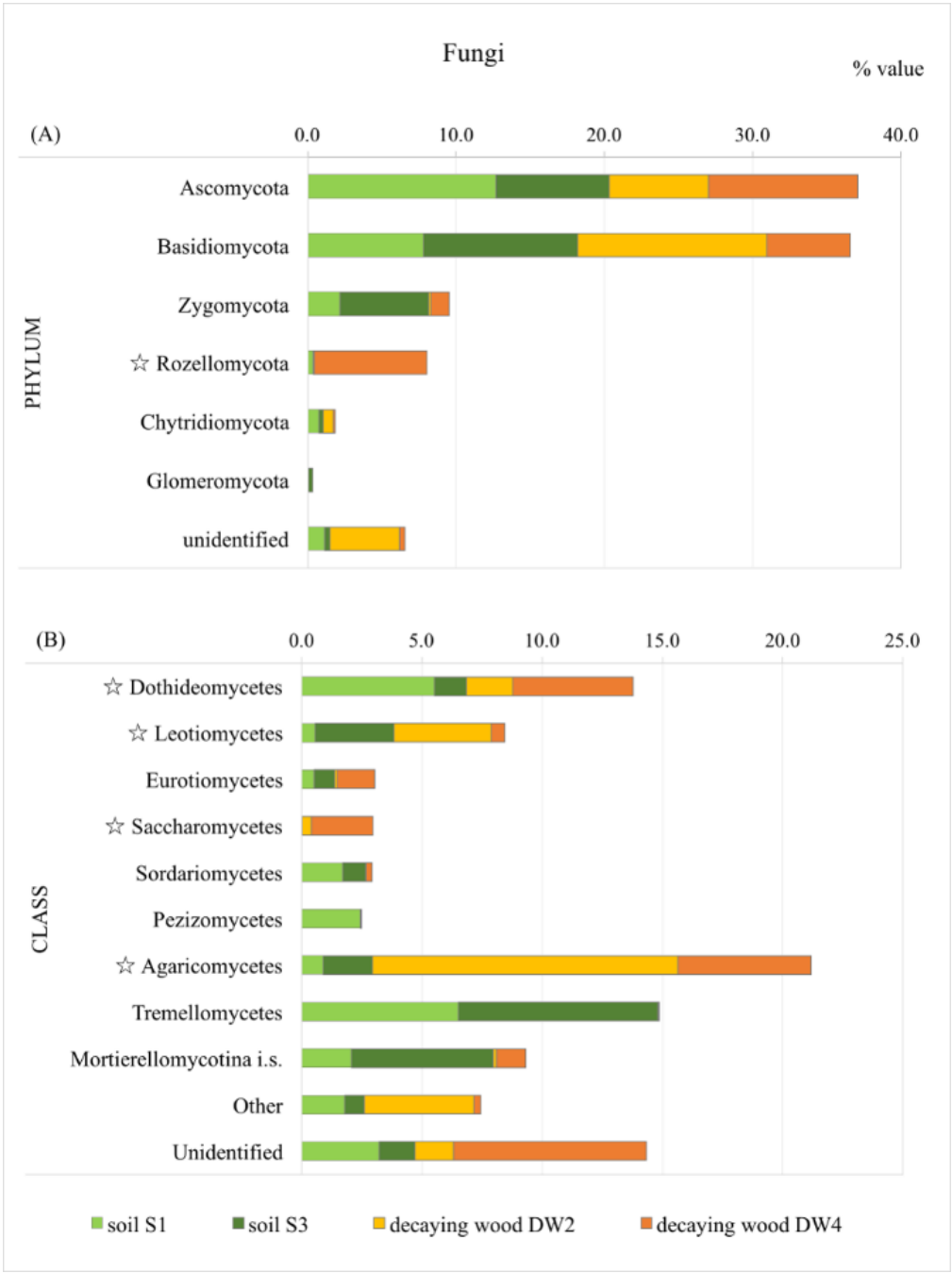


Figure 5: Fungal community in soil and decaying wood samples at (A) phylum and (B) class level. The organisms tagged with ☆ are prevalent in decaying wood.

At order level, Sebacinales (class Agaricomycetes) was more represented in both the decaying wood samples than in soils. Although Sebacinales presence in soil around roots is well known, our work could represent the first evidence of the colonization of dead plant material in soil by Sebacinales, since it was not explored *in situ* yet (Weiß *et al*., 2016). Indeed, *in vitro* culture and genomic studies support their saprotrophic abilities. For example, it was demonstrated that *Piriformospora indica* grows on dead roots (Zuccaro *et al*., 2011) and Sebacinales have enzymes allowing the saprotrophic processing of complex organic substrates (Basiewicz *et al*., 2012). Other orders predominant in decaying woods were Saccharomycetales, Eurotiales and Cantharellales (Fig. 6A).

Among families (Fig. 6B), Hydnaceae, Eremomycetaceae and Trichomonascaceae were exclusively found in decaying wood samples. *Hydnaceae* family was revealed almost exclusively in DW2, while Eremomycetaceae family was only found in DW4. Trichomonascaceae indeed were totally absent in soil samples and more abundant in DW4 than in DW2.

Six genera belonging to the phylum Ascomycota were exclusive or predominant in decaying wood samples: *Arthrographis*, *Blastobotrys*, *Lophiostoma*, *Cladosporium*, *Scytalidium* and *Aspergillus*. Only the genus *Perenniporia*, exclusively found in DW2, belong to Basidiomycota phylum (Fig. 6C). The analysis permitted to highlight even some fungal species predominant in decaying wood samples (Tab. 4). Most of them belong to the phylum Ascomycota and were more abundant in DW4. *Perenniporia fraxinea* and *Pleurotus dryinus* were the only Basidiomycetes and were mostly identified in DW2. Moreover, two species in the genus *Mortierella* (Zygomycota phylum), were revealed in decaying woods. Several studies have revealed that Basidiomycetes are dominant in the earlier stages of wood decomposition, while many Ascomycetes and Zygomycetes occur in the later stages. Species in the genera *Mortierella* have no ability to decompose LCB and are named secondary sugar fungi as they use simple sugars produced by holocellulose decomposers (Fukasawa *et al*., 2017).

Observing the species distribution, we can suppose that our decaying wood samples refer to different stages of wood decomposition, since Ascomycota are more abundant in decaying wood 4, while in decaying wood 2 are almost absent, and Basidiomycota predominate.

Moreover, even if fungal communities of soil and decaying wood have often been studied separately, several studies have shown that wood decaying fungi may occur in soil and soil-inhabiting fungal species can colonize decaying wood. Furthermore, many mycorrhizal fungi and soil saprotrophs can colonize heavily decayed woods, in contact with soil surface. As a consequence, fungal communities related to the late stages of wood decay may share similarity with soil-inhabiting ones (Mäkipää *et al*., 2017).

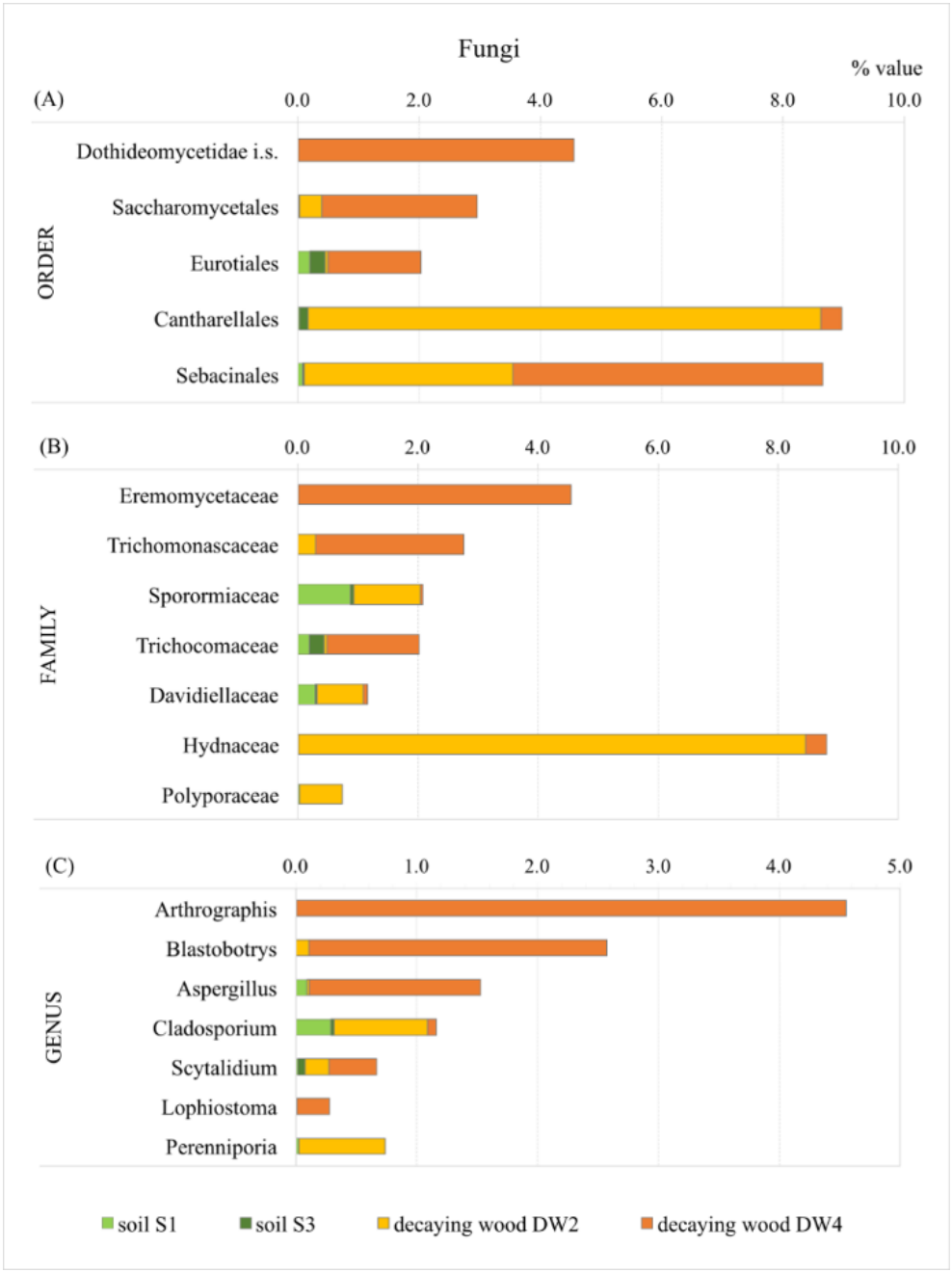


Figure 6: Fungal community in soil and decaying wood samples at (A) order, (B) family and (C) genus level. Only the organisms prevalent in decaying wood are shown.

Table 4: Fungal species predominant in decaying wood samples

Phylum	Fungal species predominant in DW
Ascomycota	*Blastobotrys mokoenaii*
	Aspergillus subversicolor
	Cladosporium delicatulum
	Scytalidium lignicola
	Sugiyamaella novakii
	Scedosporium minutisporum
	Leptodontidium trabinellum
	Mycosphaerella tassiana alias Cladosporium herbarum
Basidiomycota	*Perenniporia fraxinea*
	Pleurotus dryinus
Zygomycota	*Mortierella fimbricystis*
	Mortierella hyalina

2.2.2 Bacterial communities

Regarding bacterial communities, soils and decaying woods shared a great similarity and hosted most of the abundant phyla commonly found in soils (Araujo *et al*., 2017a; Janssen, 2006) (Fig. 7A). The most represented ones were Proteobacteria and Bacteroidetes. Bacteroidetes are Gram-negative bacteria widely distributed in soil and sediments (Gupta, 2004) and can secrete lignocellulose degrading enzymes, such as endo-β-1,4-xylanases, α-L-arabinofuranosidases, β-glucosidases and others, when grown in the presence of wheat straw (Jiménez *et al*., 2015). Actinobacteria were almost equally distributed among samples. Representatives of this phylum are among the most important decomposers in soil and rich actinobacterial communities are associated to sites with high organic matter content (Kopecky *et al*., 2011). Planctomycetes were indeed more abundant in decaying woods than in soils. Members of this phylum inhabit a wide range of aquatic and terrestrial environments with diverse environmental conditions (Ivanova *et al*., 2016). Acidobacteria, Verrucomicrobia and Chloroflexi prevailed in soils. Their tolerance to acid pH and low moisture is well-known (Quirino *et al*., 2012; Araujo *et al*., 2012).

Bacteria were described at genus level, because most of our reads were aligned with the RDP database sequences annotated - at species level - as "unclassified". The most represented bacterial classes in decaying woods were α-, γ- and δ-Proteobacteria. γ - and δ- Proteobacteria were almost exclusively found in decaying woods, while Planctomycetia (phylum Planctomycetes) were more abundant than in soils and Sphingobacteriia (phylum Bacteroidetes) predominated in only one decaying wood sample (DW2) (Fig. 7B).

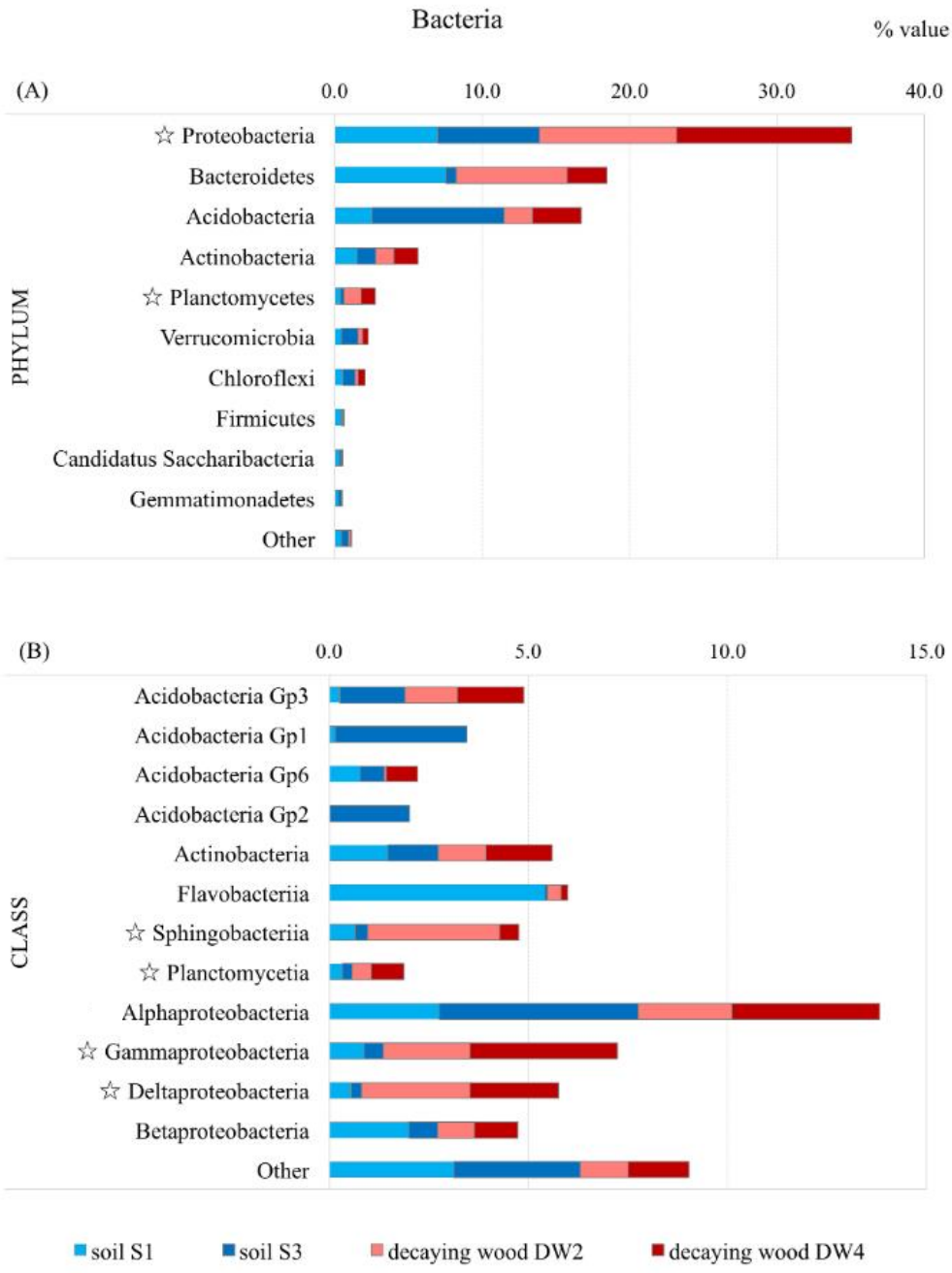


Figure 7: Bacterial community in soil and decaying wood samples at (A) phylum and (B) class level. The organisms tagged with ☆ are prevalent in decaying wood.

Myxococcales was the only bacterial order predominant in decaying woods (Fig. 8A). Myxobacteria are soil inhabiting, Gram-negative δ-Proteobacteria, distributed all over the world (Dawid, 2000) and can be divided in predators and cellulose-degraders (Mohr *et al*., 2017). Planctomycetales were weakly more represented in decaying wood samples, while Sphingobacteriiales were almost exclusive of DW2. Ohtaekwangia and Sphingomonadales were indeed almost exclusively detected in DW4.

At family level (Fig. 8B), Polyangiaceae (class δ-Proteobacteria) were almost equally divided into wood samples. Even Acidimicrobiaceae (phylum Actinobacteria) were exclusive of decaying woods despite their low frequency. Other families were especially detected in just one wood sample, as Chitinophagaceae (Sphingobacteriiales order), Phycisphaeraceae (phylum Planctomycetes) and Sinobacteraceae (class γ-Proteobacteria) in DW2 and Ohtaekwangia (phylum Bacteroidetes) and Erythrobacteraceae (class α-Proteobacteria) in DW4. Planctomycetaceae (phylum Planctomycetes) were slightly more expressed in decaying woods and prevailed in DW4.

The most represented genera in decaying wood samples are shown in Figure 8C. *Jahnella* (class δ-Proteobacteria) was the only bacterial genus identified exclusively in decaying woods. *Ilumatobacter* (phylum Actinobacteria) and *Phycisphaera* (phylum Planctomycetes) had a very low frequency in soils and prevailed in decaying woods. So far, type species of *Ilumatobacter* were previously detected in marine environments such as sediment of the estuaries, sand of the coastal beach and marine sponges (Matsumoto *et al*., 2009; Khan *et al*., 2012), lake water (Gugliandolo *et al*., 2016; Fang *et al*., 2015) and soil (Xiong *et al*., 2017), so this work could represent the first detection of this genus in decaying plant matter. *Terrimonas* (phylum Bacteroidetes, class Sphingobacteriia) and *Steroidobacter* (class γ-Proteobacteria) were more abundant in DW2 and were already detected in soils (Correa-Galeote *et al*., 2016; Chen *et al*., 2017). Indeed, *Ohtaekwangia* (phylum Bacteroidetes),

Porphyrobacter and *Thermovum* (class α-Proteobacteria) were more represented in DW4. *Porphyrobacter* has been proposed as a genus with four *Porphyrobacter* strains being isolated from a eutrophic freshwater pond in Australia (Fuerst *et al*., 1993). They are obligate aerobes and some species are able to synthesize bacteriochlorophyll (Liu *et al*., 2017).

Previous studies have clearly shown that the presence of wood degrading fungi strongly affected the composition of the bacterial community (Folman *et al*., 2008; Hervé *et al*., 2014; Hoppe *et al*., 2014). Bacterial community composition is clearly influenced by physical and chemical properties of the substrate, indicating that they are an active component of the wood-colonizing biota. Interactions with the wood degrading fungal community likely play a role in shaping the bacterial communities (Rinta-Kanto *et al*., 2016). Moreover, different plant litters provide different carbon sources to soil microbes and can select a subset of the bacterial community (Cassman *et al*., 2016).

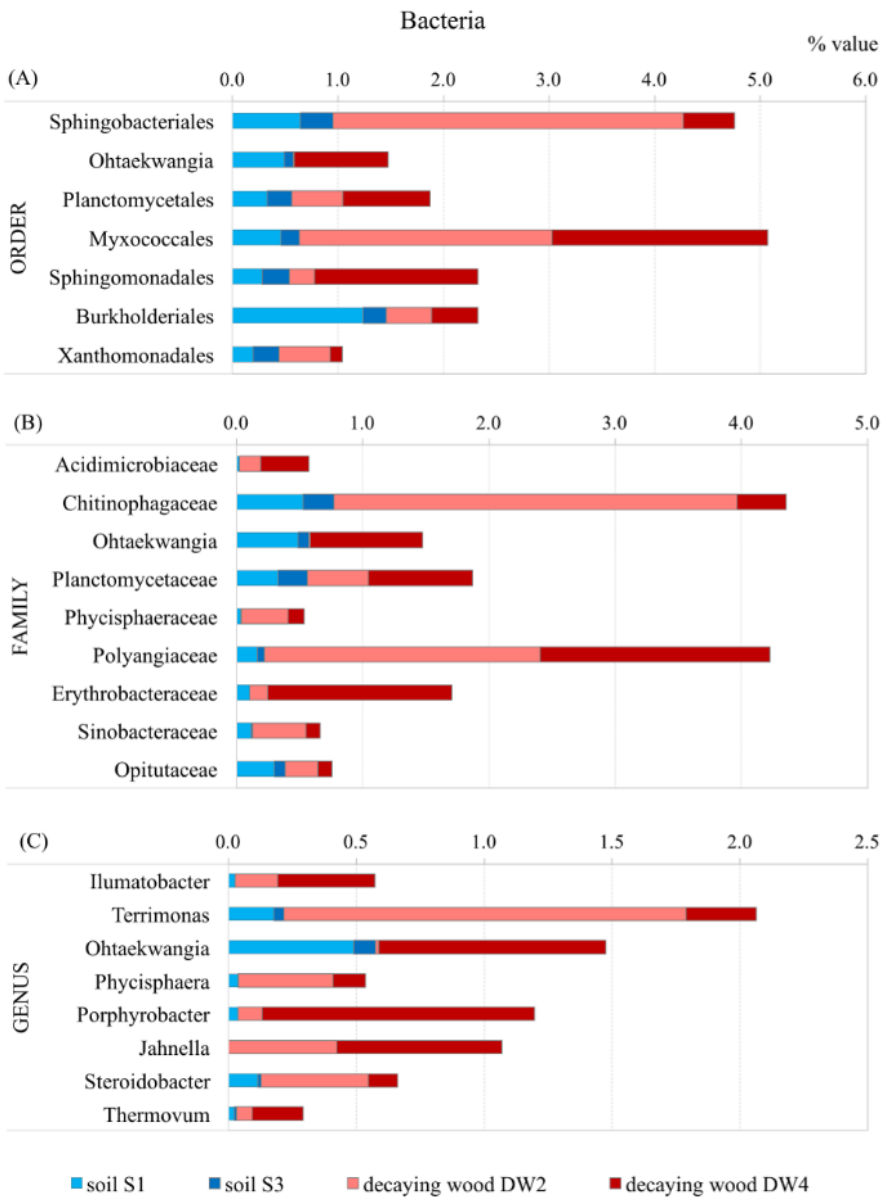


Figure 8: Bacterial community in soil and decaying wood samples at (A) order, (B) family and (C) genus level. Only the organisms prevalent in decaying wood are shown.

4 Conclusions

Wood decay processes have recently attracted so much attention, as lignocellulose represents the most abundant renewable resource on the Earth and can provide fermentable sugar monomers to produce value-added products. In order to improve the efficiency and ecological sustainability of the process, new insights about lignocellulosic biomass microbial degradation could be of fundamental importance. The advent of next generation sequencing and bioinformatics allow to get and elaborate a great number of data, even from environmental samples. This work provides a snapshot of fungal and bacterial communities that dominate decaying wood samples and that consequently are most probably involved in wood degradation. A deeper knowledge of the most represented microbial genera and species in decaying wood samples can highlight new LCB degraders, that could provide novel or more efficient enzymes, acting in synergy with the already used enzymatic cocktails. Moreover, our metagenomic analyses provided the first evidence of Sebacinales colonization of dead plant material.

References

Abarenkov K, Henrik Nilsson R, Larsson KH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AF, Tedersoo L, Ursing BM, Vrålstad T, Liimatainen K, Peintner U, Kõljalg U. The UNITE database for molecular identification of fungi— recent updates and future perspectives. *New Phytol*. 2010; 186(2): 281-285. DOI: 10.1111/j.1469-8137.2009.03160.x.

Al-Hunaiti A, Arar S, Täubel M, Wraith D, Maragkidou A, Hyvärinen A, Hussein T. Floor dust bacteria and fungi and their coexistence with PAHs in Jordanian indoor environments. *Sci Total Environ*. 2017; 601-602: 940-945. DOI: 10.1016/j.scitotenv.2017.05.211.

Amidon TE, Liu S. Water-based woody biorefinery. *Biotechnol Adv*. 2009; 27: 542-550. DOI: 10.1016/j.biotechadv.2009.04.012.

Araujo AS, Bezerra WM, dos Santos VM, Barbosa Rocha SM, da Silva Carvalho N, do Carmo Catanho Pereira de Lyra M, do Vale Barreto Figueiredo M, de Almeida Lopes AC, Maciel Melo VM. Distinct bacterial communities across a gradient of vegetation from a preserved Brazilian Cerrado. *Antonie van Leeuwenhoek*. 2017; 110:457-469. DOI: 10.1007/s10482-016-0815-1.

Araujo AS, Bezerra WM, dos Santos VM, Leal Nunes LAP, do Carmo Catanho Pereira de Lyra M, do Vale Barreto Figueiredo M, Maciel Melo VM. Fungal diversity in soils across a gradient of preserved Brazilian Cerrado. *J Microbiol*. 2017b; 55 (4): 273-279. DOI: 10.1007/s12275-017-6350-6.

Araujo JF, Castro AP, Costa MMC, Togawa RC, Pappas Junior GJ, Quirino BF, Bustamante MMC, Williamson L, Handelsman J, Kruger RH. Characterization of soil bacterial assemblies in Brazilian savanna-like vegetation reveals acidobacteria dominance. *Microb Ecol*. 2012; 64: 760-770. DOI: 10.1007/s00248-012-0057-3.

Arevalo-Gallegos A, Ahmad Z, Asgher M, Parra-Saldivar R, Iqbal HMN. Lignocellulose: A sustainable material to produce value-added products with a zero waste approach- A review. *Int J Biol Macromol*. 2017; 99: 308-318. DOI: 10.1016/j.ijbiomac.2017.02.097.

Basiewicz M, Weiß M, Kogel K-H, Langen G, Zorn H, Zuccaro A. Molecular and phenotypic characterization of a range of Sebacina vermifera strains associated with orchids, and the description of Piriformospora williamsii sp. nov. *Fungal Biol-UK*. 2012; 116: 204-213. DOI: 10.1016/j.funbio.2011.11.003.

Cassman NA, Leite MFA, Pan Y, de Hollander M, van Veen JA, Kuramae EE. Plant and soil fungal but not soil bacterial communities are linked in long-term fertilized grassland. *Sci Rep.* 2016; 6: 23680. DOI: 10.1038/srep23680.

Castro AP, Silva MRSS, Quirino BF, Bustamante MMC, Krüger RH. Microbial diversity in Cerrado biome (Neotropical Savanna) soils. *PLoS One.* 2016; 11: e0148785. DOI: 10.1371/journal.pone.0148785.

Chatterjee C, Pong F, Sen A. Chemical conversion pathways for carbohydrates. *Green Chem*. 2015; 17: 40−71. DOI: 10.1039/C4GC01062K.

Chen Q, Zang X, Hang X, Chen J, Wang H, Li S, Jiang J, Chen K. Terrimonas suqianensis sp. nov., isolated from a tetrabromobisphenol A-contaminated soil. *Antonie van Leeuwenhoek*. 2017; DOI:10.1007/s10482-017-0879-6.

Cheng JR, Zhu MJ. A novel co-culture strategy for lignocellulosic bioenergy production: a systematic review. *Int J Mod Biol Med*. 2012; 1: 266-193.

Clausen CA. Bacterial Associations with Decaying Wood: a Review. *Int Biodeter Biodegr*. 1995; 1996: 101-107.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014; 42: D633-D642. DOI: 10.1093/nar/gkt1244.

Correa-Galeote D, Bedmar EJ, Fernández-González AJ, Fernández-López M, Arone GJ. Bacterial Communities in the Rhizosphere of Amilaceous Maize (Zea mays L.) as Assessed by Pyrosequencing. *Front Plant Sci*. 2016; 7: 1016. DOI: 10.3389/fpls.2016.01016.

Daniel R. The metagenomics of soil. *Nat Rev Microbiol*. 2005; 3: 470-478. DOI:10.1038/nrmicro1160.

Dawid W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol Rev*. 2000; 24: 403-427. DOI: 10.1111/j.1574-6976.2000.tb00548.x.

Dean WE. Determination of carbonate and organic matter in calcareous sediments and sedimentary rocks by loss on ignition: comparison with other methods. *J Sediment Petrol*. 1974; 44 (1): 242-248.

Fang L, Chen L, Liu Y, Tao W, Zhang Z, Liu H, Tang Y. Planktonic and sedimentary bacterial diversity of Lake Sayram in summer. *MicrobiologyOpen*. 2015; 4(5): 814-825. DOI: 10.1002/mbo3.281.

Folman LB, Gunnewiek P, Boddy L, de Boer W. Impact of white-rot fungi on numbers and community composition of bacteria colonizing beech wood from forest soil. *FEMS Microbiol Ecol*. 2008; 63: 181-91. DOI: 10.1111/j.1574-6941.2007.00425.x.

Fuerst JA, Hawkins JA, Holmes A, Sly LI, Moore CJ, Stackebrandt E. Porphyrobacter neustonensis gen. nov., sp. nov., an aerobic bacteriochlorophyll-synthesizing budding bacterium from fresh water. *Int J Syst Bacteriol*. 1993; 43(1): 125-34. DOI: 10.1099/00207713-43-1-125.

Fukasawa Y, Osono T, Takeda H. Wood decomposing abilities of diverse lignicolous fungi on nondecayed and decayed beech wood. *Mycologia*. 2017; 103 (3): 474-482. DOI: 10.3852/10-246.

Greaves H. The bacterial factor in wood decay. *Wood Sci Technol*. 1971; 5: 616.

Grossart H, Wurzbacher C, James TY, Kagami M. Discovery of dark matter fungi in aquatic ecosystems demands a reappraisal of the phylogeny and ecology of zoosporic fungi. *Fungal Ecol*. 2016; 19: 28-38. DOI: 10.1016/j.funeco.2015.06.004.

Gugliandolo C, Michaud L, Lo Giudice A, Lentini V, Rochera C, Camacho A, Maugeri TL. Prokaryotic Community in Lacustrine Sediments of Byers Peninsula (Livingston Island, Maritime Antarctica). *Microb Ecol*. 2016; 71: 387-400. DOI: 10.1007/s00248-015-0666-8.

Gupta RS. The phylogeny and signature sequences characteristics of fibrobacteres, chlorobi, and bacteroidetes. *Critical Rev Microbiol*. 2004; 30: 123-140. DOI: 10.1080/10408410490435133.

Hausenbuiller RL. Soil Science Principles and Practice. 1975. 4th printing 1975, Wm. C. Brown Co., Dubuque, Iowa: 90.

Hernandez-Restrepo M, Gene J, Castaneda-Ruiz RF, Mena-Portales J, Crous PW, Guarro J. Phylogeny of saprobic microfungi from Southern Europe. *Stud Mycol*. 2017; 86: 53-97. DOI: 10.1016/j.simyco.2017.05.002.

Hervé V, Le Roux X, Uroz S, Gelhaye E, Frey-Klett P. Diversity and structure of bacterial communities associated with Phanerochaete chrysosporium during wood decay. *Environ Microbiol*. 2014; 16 (7): 2238-52. DOI:10.1111/1462-2920.12347.

Hoppe B, Kahl T, Karasch P, Wubet T, Bauhus J, Buscot F, Kruger D. Network analysis reveals ecological links between N-fixing bacteria and wood-decaying fungi. *PLoS One*. 2014; 9: e88141. DOI: 10.1371/journal.pone.0088141.

Iliev I, Yahubyan G, Marhova M, Apostolova E, Gozmanova M, Gecheva G, Kostadinova S, Ivanova A, Baev V. Metagenomic profiling of the microbial freshwater communities in two Bulgarian reservoirs. *J Basic Microbiol*. 2017; 9999: 1-11. DOI: 10.1002/jobm.201700137.

Ivanova AA, Kulichevskaya IS, Merkel AY, Toshchakov SV, Dedysh SN. High Diversity of Planctomycetes in Soils of Two Lichen-Dominated Sub-Arctic Ecosystems of Northwestern Siberia. *Front Microbiol*. 2016; 7: 2065. DOI: 10.3389/fmicb.2016.02065.

Janssen PH. Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. *Appl Environ Microb*. 2006; 72: 1719-1728. DOI: 10.1128/AEM.72.3.1719–1728.2006.

Jiménez DH, Maruthamuthu M, van Elsas JD. Metasecretome analysis of a lignocellulolytic microbial consortium grown on wheat straw, xylan and xylose. *Biotechnol Biofuels*. 2015; 8: 199. DOI: 10.1186/s13068-015-0387-8.

Khan ST, Takagi M, Shin-ya K. Actinobacteria associated with the marine sponges Cinachyra sp., Petrosia sp., and Ulosa sp. and their culturability. *Microbes Environ*. 2012; 27 (1): 99-104. DOI: 10.1264/jsme2.ME11270.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013; 41(1): e1. DOI: 10.1093/nar/gks808.

Kopecky J, Kyselkova M, Omelka M, Cermak L, Novotna J, Grundmann GL, Moenne-Loccoz Y, Sagova-Mareckova M. Actinobacterial community dominated by a distinct clade in acidic soil of a waterlogged deciduous forest. *FEMS Microbiol Ecol*. 2011; 78: 386-394. DOI: 10.1111/j.1574-6941.2011.01173.x.

Liu Q, Wu Y, Cheng H, Xu L, Wang C, Xu X. Complete genome sequence of bacteriochlorophyll-synthesizing bacterium Porphyrobacter neustonensis DSM 9434. *Stand Genomic Sci*. 2017; 12: 32. DOI 10.1186/s40793-017-0243-5.

Ma A, Zhuang X, Wu J, Cui M, Lv D, Liu C. Ascomycota members dominate fungal communities during straw residue decomposition in arable soil. *PLoS One*. 2013; 8: e66146. DOI: 10.1371/journal.pone.0066146.

Mäkipää R, Rajala T, Schigel D, Rinne KT, Pennanen T, Abrego N, Ovaskainen O. Interactions between soil- and dead wood-inhabiting fungal communities during the decay of Norway spruce logs. *ISME J*. 2017; 11: 1964-1974. DOI:10.1038/ismej.2017.57.

Matsumoto A, Kasai H, Matsuo Y, Omura S, Shizuri Y, Takahashi Y. Ilumatobacter fluminis gen. nov., sp. nov., a novel actinobacterium isolated from the sediment of an estuary. *J Gen Appl Microbiol*. 2009; 55 (3): 201-05.

Mocali S, Benedetti A. Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. *Res Microbiol*. 2010; 161: 497-505. DOI: 10.1016/j.resmic.2010.04.010.

Mohr KI, Zindler T, Wink J, Wilharm E, Stadler M. Myxobacteria in high moor and fen: An astonishing diversity in a neglected extreme habitat. *MicrobiologyOpen*. 2017; 6: e464. DOI:10.1002/mbo3.464O

Nesme J, Achouak W, Agathos SN, Bailey M, Baldrian P, Brunel D, Frostegård A, Heulin T, Jansson JK, Jurkevitch E, Kruus KL, Kowalchuk GA, Lagares A, Lappin-Scott HM, Lemanceau P, Le Paslier D, Mandic-Mulec I, Murrell JC, Myrold DD, Nalin R, Nannipieri P, Neufeld JD, O'Gara F, Parnell JJ, Pühler A, Pylro V, Ramos JL, Roesch LFW, Schloter M, Schleper C, Sczyrba A, Sessitsch A, Sjöling S, Sørensen J, Sørensen SJ, Tebbe CC, Topp E, Tsiamis G, van Elsas JD, van Keulen G, Widmer F, Wagner M, Zhang T, Zhang X, Zhao L, Zhu Y, Vogel TM, Simonet P. Back to the Future of Soil Metagenomics. *Front Microbiol*. 2016; 7: 73. DOI: 10.3389/fmicb.2016.00073.

Novello G, Gamalero E, Bona E, Boatti L, Mignone F, Massa N, Cesaro P, Lingua G, Berta G. The Rhizosphere Bacterial Microbiota of Vitis vinifera cv. Pinot Noir in an Integrated Pest Management Vineyard. *Front Microbiol*. 2017; 8: 1528. DOI: 10.3389/fmicb.2017.01528.

Ohm RA, Riley R, Salamov A, Min B, Choi I, Grigoriev I. Genomics of wood-degrading fungi. *Fungal Genet Biol*. 2014; 72: 82-90. DOI: 10.1016/j.fgb.2014.05.001.

Quirino BF, Pappas GJ, Tagliaferro AC, Collevatti RG, Neto EL, Silva MR, Bustamante MM, Kruger RH. Molecular phylogenetic diversity of bacteria associated with soil of savanna-like Cerrado vegetation. *Microbiol Res*. 2009; 164: 59–70. DOI: 10.1016/j.micres.2006.12.001.

Rinta-Kanto JM, Sinkko H, Rajala T, Al-Soud WA, Sørensen SJ, Tamminen MV, Timonen S. Natural decay process affects the abundance and community structure of Bacteria and Archaea in Picea abies logs. *FEMS Microbiol Ecol*. 2016; DOI: 10.1093/femsec/fiw087.

Schofield RK, Taylor AW. The Measurement of Soil pH. *Soil Sci Soc Am J*. 1955; 19 (2): 164-167. DOI:10.2136/sssaj1955.03615995001900020013x.

Sista Kameshwar AK, Qin W. Recent developments in using advanced sequencing technologies for the genomic studies of lignin and cellulose degrading microorganisms. *Int J Biol Sci*. 2016; 12 (2): 156-171. DOI: 10.7150/ijbs.13537.

Tedersoo L, Bahram M, Puusepp R, Nilsson RH, James TY. Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome*. 2017; DOI 10.1186/s40168-017-0259-5.

Toju H, Tanabe AS, Yamamoto S, Sato H. High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples. *PLoS One* 2012; 7(7): e40863. DOI: 10.1371/journal.pone.0040863.

Torsvik V, Daae FL, Sandaa,R-A, Øvreås L. Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol*. 2002; 5: 240-245.

Tuck CO, Perez E, Horvath IT, Sheldon RA, Poliakoff M. Valorization of biomass: Deriving more value from waste. *Science*. 2012; 337: 695-699. DOI: 10.1126/science.1218930.

Wang C, Dong D, Wang H, Muller K, Qin Y, Wang H, Wu W. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocelluloses decomposition. *Biotechnol Biofuels*. 2016; 9: 22. DOI: 10.1186/s13068-016-0440-2.

Wang H, Ruan H, Pei H, Wang H, Chen X, Tucker MP, Cort JR, Yang B. Biomass-derived lignin to jet fuel range hydrocarbons via aqueous phase hydrodeoxygenation. *Green Chem*. 2015; 17: 5131-5135. DOI: 10.1039/C5GC01534K.

Weiß M, Waller F, Zuccaro A, Selosse M. Sebacinales – one thousand and one interactions with land plants. *New Phytol*. 2016; 211: 20–40. DOI: 10.1111/nph.13977.

Xiong W, Li R, Ren Y, Liu C, Zhao Q, Wu H, Jousset A, Shen Q. Distinct roles for soil fungal and bacterial communities associated with the suppression of vanilla Fusarium wilt disease. *Soil Biol Biochem*. 2017; 107: 198-207. DOI: 10.1016/j.soilbio.2017.01.010.

Xu B, Xu W, Li J, Dai L, Xiong C, Tang X, Yang Y, Mu Y, Zhou J, Ding J, Wu Q, Huang Z. Metagenomic analysis of the *Rhinopithecus bieti* fecal microbiome reveals a broad diversity of bacterial and glycoside hydrolase profiles related to lignocellulose degradation. *BMC Genomics*. 2015; 16: 174. DOI: 10.1186/s12864-015-1378-7.

Zhang Z, Song J, Han B. Catalytic transformation of lignocellulose into chemicals and fuel products in ionic liquids. *Chem Rev*. 2016; 117: 6834-6880. DOI: 10.1021/acs.chemrev.6b00457.

Zuccaro A, Lahrmann U, Guldener U, Langen G, Pfiffi S, Biedenkopf D, Wong P, Samans B, Grimm C, Basiewicz M, Murat C, Martin F, Kogel K. Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont Piriformospora indica. *PLoS Pathog*. 2011; 7: e1002290. DOI: 10.1371/journal.ppat.1002290.

Discussion

Biorefineries and renewable biomass sources have attracted growing attention since the last years, as fossil fuels reserves are decreasing, with a negative effect on prices. In addition, the growing consumption of fossil fuels leads to several negative environmental impacts, such as global warming, ozone depletion and greenhouse gasses emission. Biorefineries aim is analogous to conventional refineries, that is the production of biofuels, bioenergy and chemical building blocks for value-added products. The difference relies in the starting feedstocks, as biorefineries utilize renewable sources, like biomass, instead of fossil fuels (Gavrilescu, 2014). Among renewable resources, lignocellulosic biomass is certainly the most abundant. It represents the major structural component of plants and can be found in large amounts even as a waste, generated for example from forestry and agricultural practices, paper and pulp industries and many agro-industries (Sanchez & Cardona, 2008). Moreover, the advent of second generation bioethanol, obtained from lignocellulose, gave more importance to dedicated cultures, known as energy crops. Lignocellulose consists of a complex matrix containing cellulose, hemicellulose and lignin (Sandgren *et al*., 2005). Each polymer can be degraded and converted in different value-added products, including not only biofuels but even chemicals, like organic acids, used for several applications, such as plastic production (Chen, 2014). In the light of a circular economy, the recycling of resources, the minimization of energy consumption and wastes is of fundamental importance and the passage from a "take, make, dispose" system to a circular economy implicates the capability to convert wastes into useful products, like chemical building blocks or fuels. One of the most significant obstacles still to be overcome is the lignocellulose recalcitrance. Currently, several studies are focused on pretreatment methods improvement, to reach a better deconstruction of lignocellulose, increasing enzymatic cellulose accessibility, and a consistent lignin removal. However, the hydrolysis step remains the major responsible for biofuels high costs. In particular, hydrolytic

enzymes are preferred to acid or alkaline processes since they are environmental friendly specific biocatalysts, that can operate under much milder reaction conditions, avoiding equipment corrosion and undesirable products release (Kumar & Sharma, 2017). Regarding enzymes, the search for new or more efficient enzymatic species is always in act. In particular, commercial enzymatic cocktails are mainly derived from fungi, like *Trichoderma reesei*, that can secrete large amounts of free enzymes. However, the implementation of such enzymatic cocktails can improve hydrolytic yields (Bussamra *et al*., 2015; Peciulyte *et al*., 2017). Since in nature several degrading systems, both fungal and bacterial, have evolved to overcome lignocellulose recalcitrance, the investigation of microbial communities involved in the decay process can provide important information. Such investigations are encouraged by the development of bioinformatics and next-generation sequencing technologies, representing a very helpful tool for microorganism genome annotation and metagenomic analysis. The genome sequencing permits a deeper knowledge of the microorganisms, giving the capability to derive *in silico* secretomes, that represent important starting points for further analysis. Metagenomic analyses indeed can reveal non-cultivable microorganisms and provide an interesting snapshot of microbial communities inhabiting environmental samples, even suggesting their possible relation. Moreover, since the industrial lignocellulose conversion to valuable products is costly and requires harsh conditions, like high temperature and acid environment, the study of microorganisms that degrade lignocellulose under natural conditions should contribute to the development of eco-friendly biomass degrading processes (Oh *et al*., 2017).

Lignocellulose degradation represents the subject of this study and it was investigated with different approaches. Briefly, a preliminary *in silico* analysis permitted to highlight lignocellulose degrading capabilities of *Cellulomonas fimi*, a Gram-positive mesophilic bacterium, and provided helpful insights about the

choice of the following sample analysis method. Then, the secretome of *C. fimi*, grown for different periods (24/48 h) with carboxymethyl-cellulose (CMC) or three different pretreated lignocellulosic biomasses (LCBs) (wheat straw, *A. donax*, sugar cane bagasse) as unique carbon source, was characterized. Protein samples were separated with SDS-PAGE and then submitted to trypsin digestion and MS/MS analysis for protein identification. Several CAZymes were detected and emPAI analysis allowed a quantitative evaluation of their secretion. Bacterial secretomes were also characterized in terms of enzymatic activities. In particular, xylanase, endoglucanase and β-glucosidase activities were evaluated. Finally, wood decay process was analyzed with a metagenomic approach in order to identify the major microorganisms responsible for lignocellulose degradation. Fungal and bacterial genomic DNA was extracted from environmental samples (soils and decaying woods), amplified with specific primers and used for the taxonomic assignment. The comparison between microbial communities revealed in soil samples and in decaying wood samples highlighted fungal and bacterial microorganisms most probably involved in the decay process. Such results may have an important resonance for the development of new strategies for lignocellulose degradation. The *in silico* analysis is resulted very important for the choice of the most appropriate proteomic method. Before trypsin digestion, in fact, a protein separation is advisable, not only for the visualization of the secretome, but also because samples, especially those obtained after growth on lignocellulosic biomasses, contain a lot of impurities, very difficult to eliminate, such as plant-derived colored compounds or biomass residues, that could remain, even after filtration. The 2DE map of CAZymes showed that most of them share a pI between 4 and 7. However, our analysis confirmed that only a few spots can be visualized in the 2DE maps (even in the pI range 3-10), with respect to the *in silico* map, as some CAZymes were not detected because of complications in the isoelectrophocusing process. It's well known, in fact, that many cellulolytic

enzymes show the tendency to strongly associate, forming high molecular weight complexes that difficultly enter the gel for the second-dimension separation. In order to perform a more comprehensive analysis, avoiding protein loss, the samples were separated with SDS-PAGE before the digestion and MS analysis. Such approach, combined with the gel slices analysis, attempted to simulate an in-liquid digestion, since the abundance of impurities and flagellin would have negatively affected the results. The MS/MS analysis highlighted several proteins that were commonly secreted among all samples, suggesting that *C. fimi* exploits a highly conserved pool of CAZymes, even when changing the substrate and the time of culture. However, a few proteins were detected only at 24 or 48 hours of growth and others were related to a specific substrate. A better understanding of this topic derived from emPAI analysis, that provided information about CAZymes abundance variation depending on time and substrate of growth. These data were consistent with the enzymatic activity evaluation by zymography and enzymatic assays. The emPAI analysis in fact validated enzymatic activity assays results. For example, the higher xylanase activity observed in all the samples at 48 hours was in agreement with the increased secretion of hemicellulolytic enzymes and the higher activities observed in the secretomes obtained after growth on LCBs could be related to the increment of emPAI values of cellulolytic and hemicellulolytic enzymes. The association of emPAI data with enzymatic activity evaluation, by zymography and enzymatic assays, even suggested that some CAZymes, even with the same enzymatic activity, may be secreted in different phases of the lignocellulolytic process. In each culture condition, for example, an Endo-1,4-beta-xylanase was found more secreted at 24 hours than at 48. Such observation permits to hypothesize that different xylanases, sharing the same enzymatic activity, could attend to different stages of the hydrolysis. Similarly, emPAI analyses highlighted that some CAZymes, mainly involved in hemicellulose degradation, were less secreted at 48 than 24 hours. Such enzymes,

like Endo-1,4-beta-xylanase, Pectate lyase/Amb allergen, Cellulose binding family II and Alpha-N-arabinofuranosidase could degrade hemicellulose to increase cellulose accessibility, promoting the activity of further CAZymes. Hence, such analyses permitted not only to compare different secretomes but even to formulate hypothesis about *C. fimi* lignocellulose degrading machinery. Despite few experimental data are available, STRING analysis permitted even to investigate protein-protein relations, highlighting some proteins, like Enolase or Alpha-1,6-glucosidase, pullulanase-type, that may have an interesting role in lignocellulose degradation. Hence, further investigations could provide useful information. It could be very interesting to evaluate the enzymatic activities of secretomes from *C. fimi* strains obtained after gene silencing. Since lignocellulose degradation is due to the synergistic action of several CAZymes, knock out experiments can provide important information to clarify the mechanism of the process. Part of this study was conducted within the project SPRITE, in collaboration with C5-6 Italy, Biochemtex group. In order to improve the cellulose hydrolysis efficiency, *C. fimi* secretome was added to industrial commercial cocktails and several hydrolysis tests were conducted. Yields in oligosaccharides and monomers were evaluated, proving even on a pilot scale that *C. fimi* secretome can increase enzymatic yields, acting in synergy with commercial enzymes. Since lignin represents a waste, that in most biorefineries is currently burned to be converted in energy, the detection of new ligninolytic microorganisms and their enzymes can be very important. Metagenomic analyses from this side can suggest microbial consortia that could synergistically cooperate for lignocellulose degradation. Further, not only lignin degradation but also its conversion in value-added products is becoming more and more important. The optimization of the hydrolytic step in fact is fundamental and could be improved from the most simple details, since lignocellulose degradation is a complex process involving several factors, acting in synergy.

References

Bussamra BC, Freitas S, Costa AC. Improvement on sugar cane bagasse hydrolysis using enzymatic mixture designed cocktail. *Bioresour Technol*. 2015;1 87: 173-81. DOI: 10.1016/j.biortech.2015.03.117.

Chen H. Biotechnology of Lignocellulose - Theory and Practice. 2014. Springer Dordrecht Heidelberg New York London (Ed.), ISBN 978-94-007-6897-0. DOI: 10.1007/978-94-007-6898-7.

Gavrilescu M. Biorefinery systems: an overview. In: Gupta VG, Tuohy M, Kubicek CP, Saddler J, Xu F, editors. Bioenergy Research: Advances and Applications. 2014; 219-241.

Kumar AK, Sharma S. Recent updates on different methods of pretreatment of lignocellulosic feedstocks: a review. *Bioresour Bioprocess*. 2017; 4: 7. DOI 10.1186/s40643-017-0137-9.

Oh HN, Lee TK, Park JW, No JH, Kim D, Sul WJ. Metagenomic SMRT sequencing-based exploration of novel lignocellulose-degrading capability in wood detritus from Torreya nucifera in Bija Forest on Jeju Island. *J Microbiol Biotechnol*. 2017; 27(9): 1670-1680. DOI: 10.4014/jmb.1705.05008.

Peciulyte A, Pisano M, de Vries RP, Olsson L. Hydrolytic potential of five fungal supernatants to enhance a commercial enzyme cocktail. *Biotechnol Lett*. 2017; 39: 1403-1411. DOI 10.1007/s10529-017-2371-9.

Sanchez OJ, Cardona CA. Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technol*. 2008; 99: 5270-5295.

Sandgren M, Stahlberg J, Mitchinson C. Structural and biochemical studies of GH family 12 cellulases: improved thermal stability, and ligand complexes. *Prog Biophys Mol Bio*. 2005; 89: 246–291.

Appendix A

Table S1: secreted proteins collected after 24 h of *C. fimi* growth on *A. donax*.

	Protein Identification [common name] (gi NCBI)	CAZy classification	MW (kDa)	Peptide sequences (significant sequences)	emPAI	SignalP	SecretomeP
1	Exoglucanase A [CbhA] (1708083)	GH6 [Cel6B]	89588	27 (24)	4.27	YES	YES
2	Exoglucanase B [CbhB] (1708084)	GH48 [Cel48A]	115101	27 (21)	1.71	YES	YES
3	Flagellin domain protein (332338221)	-	39886	20 (16)	16.54	NO	YES
4	Endoglucanase C [CenC] (121819)	GH9 [Cel9B]	115658	19 (16)	1.09	YES	YES
5	Peptidase S8 and S53 subtilisin kexin sedolisin (332338096)	-	138613	24 (18)	0.85	YES	YES
6	Glycoside hydrolase family 5 [Cen D] (332339471)	GH5 [Cel5A]	79059	20 (14)	1.5	YES	YES
7	Endo-1,4-beta-xylanase (332338131)	GH10, CE4	141693	26 (22)	1.06	YES	YES
8	Glycoside hydrolase family 9 [CenB] (332337588)	GH9 [Cel9A]	109266	31 (20)	1.36	YES	YES
9	TAP domain protein (332337857)	-	63142	18 (16)	2.59	YES	YES

10	Endo-1,4-beta-xylanase (332337655)	GH10	49754	15 (12)	2.59	YES	YES
11	Endo-1,4-beta-xylanase (332340687)	GH10	53458	13 (7)	0.89	YES	YES
12	Pectate lyase/Amb allergen (332341298)	PL1	52912	23 (19)	4.82	YES	YES
13	Exoglucanase [Cex] 327179208	GH10 [Xyn10A]	54499	22(14)	2.47	YES	YES
14	Glycoside hydrolase family 3 domain protein (332338140)	GH3	178143	27 (20)	0.61	YES	YES
15	Cellulose-binding family II (332339303)	CE2	52461	14 (12)	1.63	YES	YES
16	Extracellular solute-binding protein family 5 (332340445)	-	67393	13 (11)	1.13	YES	YES
17	Cell surface receptor IPT/TIG domain protein (332341050)	-	99332	17 (13)	0.75	YES	YES
18	Aminopeptidase Y (332337698)	-	53985	9 (5)	0.48	YES	YES
19	Endo-1,3(4)-beta-glucanase (332341296)	GH81	116431	21 (13)	0.61	NO	YES
20	Alkaline phosphatase (332340091)	-	44035	14 (8)	1.61	NO	NO
21	Chaperonin GroEL (332338193)	-	56352	20 (14)	1.87	NO	NO
22	PKD domain containing protein (332339822)	-	87063	15 (9)	0.55	YES	YES
23	Extracellular solute-binding protein family 1 (332340803)	-	60706	7 (5)	0.42	YES	YES

24	Extracellular solute-binding protein family 1 (332340270)	-	45120	12 (7)	0.93	YES	NO
25	5'-Nucleotidase domain-containing protein (332339202)	-	167817	14 (8)	0.23	YES	YES
26	Chitin-binding domain 3 protein (332337832)	AA10	37478	6 (3)	0.4	YES	YES
27	Fibronectin type III domain protein (332339883)	-	210933	26 (8)	0.18	NO	YES
28	Glutamine synthetase, type I (332339768)	-	53327	14 (8)	0.89	NO	NO
29	Alpha-N-arabinofuranosidase (332340686)	GH62	53079	5 (2)	0.17	YES	YES
30	Peptidase S1 and S6 chymotrypsin/Hap (332338679)	-	49521	8 (3)	0.29	YES	YES
31	Glucose-6-phosphate isomerase (332338893)	-	60314	14 (7)	0.64	NO	NO
32	1, 4-beta cellobiohydrolase [CenA] (332340715)	GH6 [Cel6A]	47018	7 (4)	0.43	YES	YES
33	Alpha-1,6-glucosidase, pullulanase-type (332338678)	GH13	209811	21 (8)	0.18	YES	YES
34	Man26A (5359710)	GH26 [Man26A]	107064	11 (5)	0.22	YES	YES
35	Aminopeptidase N (332338951)	-	94435	14 (2)	0.09	NO	NO
36	Stress protein (332337771)	-	20150	4 (2)	0.51	NO	NO
37	1, 4-beta cellobiohydrolase (332338782)	GH6	46631	7 (2)	0.2	YES	YES

38	Extracellular solute-binding protein family 1 (332339297)	-	63225	12 (6)	0.5	YES	YES
39	Peptidoglycan glycosyltransferase (332340923)	GT51	86700	10 (3)	0.16	NO	YES
40	Glycoside hydrolase family 11 [XynD] (332337936)	GH11, CE4 [Xyn11A]	66855	4 (2)	0.14	YES	YES
41	Glycoside hydrolase family 18 (332340692)	GH18	58146	12 (3)	0.24	YES	YES
42	Extracellular ligand-binding receptor (332339377)		43218	2 (2)	0.22	YES	NO
43	Glycoside hydrolase family 3 domain protein (332339598)	GH3	96753	11 (1)	0.04	YES	NO
44	Xylose isomerase (332338137)	-	43425	6 (3)	0.34	NO	NO
45	Glycogen/starch/alpha-glucan phosphorylase [maltodextrin phosphorylase] (332341073)	-	92938	12 (1)	0.05	NO	NO
46	L-arabinose isomerase (332338427)	-	55111	10 (4)	0.36	NO	NO
47	Glycoside hydrolase family 9 (332337613)	GH9	94310	8(1)	0.05	YES	YES
48	Aconitate hydratase 1 (332339684)	-	100704	14 (3)	0.14	NO	NO
49	LPXTG-motif cell wall anchor domain protein (332341025)	-	129721	8 (1)	0.03	NO	YES
50	Extracellular solute-binding protein family 1 (332340856)	-	48074	6 (2)	0.19	YES	YES

51	Phosphoglucomutase, alpha-D-glucose phosphate-specific (332341135)	-	58852	6 (1)	0.07	NO	NO
52	Catalase (332339192)	-	54817	7 (2)	0.17	NO	NO
53	Extracellular solute-binding protein family 1 (332338842)	-	46266	6 (2)	0.2	YES	YES
54	Basic membrane lipoprotein (332340404)	-	38019	7 (4)	0.56	YES	YES
55	Peptidase M24 (332338709)	-	55886	6 (1)	0.08	NO	NO
56	Periplasmic binding protein/LacI transcriptional regulator (332341019)	-	40664	1 (1)	0.11	YES	YES
57	Ribosomal protein L5 (332338567)	-	21150	4 (1)	0.22	NO	NO
58	Hypothetical protein Celf_2513 (332340055)	-	82303	9 (1)	0.05	YES	YES
59	Glyceraldehyde-3-phosphate dehydrogenase, type I (332339411)	-	35807	11 (1)	0.13	NO	NO
60	Cellulose-binding family II (332339460)	GH74, CBM2	95879	14(1)	0.05	YES	YES
61	Pectate lyase (332337958)	PL3	44887	10(1)	0.1	YES	YES
62	Protein of unknown function DUF124 (332341292)	-	27337	7(1)	0.17	NO	NO
63	Lipoprotein LpqB, beta-propeller domain-like protein	-	57776	7(1)	0.08	YES	NO

	(332340014)						
64	Dihydroxy-acid dehydratase (332340266)	-	66305	6(1)	0.07	YES	YES
65	ATP synthase F1, alpha subunit (332340342)	-	58082	4(1)	0.08	NO	NO
66	TAP domain protein (332338042)	-	54198	2(2)	0.17	YES	NO
67	N-acetyl-gamma-glutamyl-phosphate reductase (332339210)	-	36207	1(1)	0.11	NO	NO
68	SCP-like extracellular (332338299)	-	36228	7(1)	0.12	NO	NO
69	Hypothetical protein Celf_3047 (332340581)	-	55195	8(1)	0.08	YES	YES
70	Dihydrolipoamide dehydrogenase (332339755)	-	48441	6(1)	0.09	NO	NO
71	Extracellular solute-binding protein family 1 (332340306)	-	47972	4(1)	0.09	YES	YES
72	Enolase (332340559)	-	45287	6(2)	0.21	NO	NO
73	Domain of unknown function DUF2394 (332338139)		35622	2(1)	0.13	NO	NO
74	Hypothetical protein Celf_3380 (332340910)	-	12743	1(1)	0.38	NO	YES
75	Transcriptional regulator, winged helix family (332339235)	-	117884	9(1)	0.04	NO	NO
76	Guanosine pentaphosphate synthetase I/polyribonucleotide nucleotidyltransferase (332339091)	-	79580	9(1)	0.05	NO	NO
77	Signal transduction histidine kinase	-	55289	12(1)	0.08	NO	NO

	(332338844)						
78	Extracellular solute-binding protein family 1 (332340331)	-	59378	7(1)	0.07	NO	YES
79	Regulatory protein TetR (332337908)	-	24168	6(1)	0.19	NO	NO
80	Stress protein (332337770)	-	20550	2 (1)	0.23	NO	NO

Table S2: secreted proteins collected after 48 h of *C. fimi* growth on *A. donax*.

	Protein Identification [common name] (gi NCBI)	**CAZy classification**	**MW (kDa)**	**Peptide sequences (significant sequences)**	**emPAI**	**SignalP**	**SecretomeP**
1	Exoglucanase B [CbhB] (1708084)	GH48 [Cel48A]	115101	33 (31)	3.72	YES	YES
2	Exoglucanase A [CbhA] (1708083)	GH6 [Cel6B]	89588	29 (27)	5.36	YES	YES
3	Flagellin domain protein (332338221)	-	39886	24 (17)	16.48	NO	YES
4	Glycoside hydrolase family 9 [CenB] (332337588)	GH9 [Cel9A]	109266	43 (38)	5.23	YES	YES
5	Glycoside hydrolase family 5 [Cen D] (332339471)	GH5 [Cel5A]	79059	28 (22)	4.29	YES	YES

6	Endoglucanase C [CenC] (121819)	GH9 [Cel9B]	115658	30 (27)	2.91	YES	YES
7	Peptidase S8 and S53 subtilisin kexin sedolisin (332338096)	-	138613	34 (30)	2.02	YES	YES
8	Endo-1,4-beta-xylanase (332338131)	GH10, CE4	141693	42 (37)	2.64	YES	YES
9	Exoglucanase [Cex] (327179208)	GH10 [Xyn10A]	54499	26 (17)	4.11	YES	YES
10	Endo-1,4-beta-xylanase (332337655)	GH10	49754	17 (12)	3.25	YES	YES
11	Endo-1,4-beta-xylanase (332340687)	GH10	53458	14 (10)	1.8	YES	YES
12	Glycoside hydrolase family 3 domain protein (332338140)	GH3	178143	41 (33)	1.31	YES	YES
13	Aminopeptidase Y (332337698)	-	53985	18 (13)	2.8	YES	YES
14	5'-Nucleotidase domain-containing protein (332339202)	-	167817	25 (21)	0.79	YES	YES
15	PKD domain containing protein (332339822)	-	87063	24 (22)	2.56	YES	YES
16	Pectate lyase/Amb allergen (332341298)	PL1	52912	22 (21)	5.82	YES	YES
17	TAP domain protein (332337857)	-	63142	18 (17)	2.83	YES	YES
18	Cellulose-binding family II (332339303)	CE2	52461	11 (11)	1.85	YES	YES

19	Extracellular solute-binding protein family 5 (332340445)	-	67393	17 (15)	2.52	YES	YES
20	Alpha-1,6-glucosidase, pullulanase-type (332338678)	GH13	209811	37 (25)	0.66	YES	YES
21	Glycoside hydrolase family 3 domain protein (332339598)	GH3	96753	20 (17)	1.3	YES	NO
22	Endo-1,3(4)-beta-glucanase (332341296)	GH81	116431	21 (18)	1.08	NO	YES
23	Cellulose-binding family II (32339460)	GH74, CBM2	95879	26 (17)	1.22	YES	YES
24	Pectate lyase (332337958)	PL3	44887	16 (7)	0.93	YES	YES
25	Chitin-binding domain 3 protein (332337832)	AA10	37478	8 (4)	0.57	YES	YES
26	1, 4-beta cellobiohydrolase [CenA] (332340715)	GH6 [Cel6A]	47018	8 (7)	1.05	YES	YES
27	Glycoside hydrolase family 11 [XynD] (332337936)	GH11, CE4 [Xyn11A]	66855	13 (5)	0.46	YES	YES
28	Man26A (5359710)	GH26 [Man26A]	107064	18 (14)	0.74	YES	YES
29	Cell surface receptor IPT/TIG domain protein (332341050)	-	99332	16 (13)	0.74	YES	YES
30	Glycoside hydrolase family 9 (332337613)	GH9	94310	18 (10)	0.57	YES	YES
31	Alkaline phosphatase (332340091)	-	44035	11 (7)	1.15	NO	NO
32	1, 4-beta cellobiohydrolase	GH6	46631	7 (5)	0.57	YES	YES

	(332338782)						
33	Fibronectin type III domain protein (332339883)	-	210933	27 (10)	0.22	NO	YES
34	Peptidase S1 and S6 chymotrypsin/Hap (332338679)	-	49521	5 (3)	0.29	YES	YES
35	Extracellular solute-binding protein family 1 (332340803)	-	60706	9 (6)	0.52	YES	YES
36	Aminopeptidase N (332338951)	-	94435	19 (12)	0.72	NO	NO
37	Alpha-N-arabinofuranosidase (332340686)	GH62	53079	4 (2)	0.17	YES	YES
38	Glycoside hydrolase family 43 (332340688)	GH43, CBM13	76536	8 (2)	0.12	YES	YES
39	Glycoside hydrolase family 18 (332340692)	GH18	58146	12 (7)	0.67	YES	YES
40	Extracellular solute-binding protein family 1 (332340270)	-	45120	9 (6)	0.76	YES	NO
41	Extracellular solute-binding protein family 1 (332339297)	-	63225	17 (9)	0.83	YES	YES
42	Enolase (332340559)	-	45287	9 (4)	0.45	NO	NO
43	Periplasmic binding protein (332341193)	-	34098	4 (3)	0.45	YES	NO
44	Extracellular solute-binding protein family 5 (332339390)	-	60117	7 (3)	0.24	YES	YES
45	Glutamine synthetase, type I (332339768)	-	53327	16 (6)	0.61	NO	NO
46	Glucose-6-phosphate isomerase (332338893)	-	60314	11 (5)	0.42	NO	NO

47	NLPA lipoprotein (332338762)	-	33494	4 (1)	0.13	YES	YES
48	Extracellular solute-binding protein family 1 (332338842)	-	46266	8 (3)	0.32	YES	YES
49	Stress protein (332337771)	-	20150	5 (4)	1.29	NO	NO
50	Flagellar hook-basal body protein (332338236)	-	39855	8 (3)	0.37	NO	YES
51	Extracellular ligand-binding receptor (332339377)		43218	2 (2)	0.22	YES	NO
52	Hypothetical protein Celf_2704 (332340245)	-	24970	8 (3)	0.66	NO	NO
53	Phosphate ABC transporter, periplasmic phosphate-binding protein (332338148)	-	37419	3 (1)	0.12	YES	NO
54	Peptidase M24 (332338709)	-	55886	11 (2)	0.16	NO	NO
55	Xylose isomerase (332338137)	-	43425	3 (2)	0.22	NO	NO
56	Glyceraldehyde-3-phosphate dehydrogenase, type I (332339411)	-	35807	10 (2)	0.27	NO	NO
57	Cellulase (332337938)	GH5, CBM46	61605	7 (3)	0.23	YES	NO
58	Glycerophosphoryl diester phosphodiesterase (332338855)	-	79002	11 (3)	0.17	YES	NO
59	Glucan endo-1,3-beta-D-glucosidase (332340860)	GH64	55696	6 (2)	0.16	YES	YES

60	Extracellular solute-binding protein family 1 (332340856)	-	48074	3 (2)	0.19	YES	YES
61	Periplasmic binding protein (332339005)	-	34886	2 (1)	0.13	YES	NO
62	Ribosomal protein L18 (332338571)	-	12954	2 (1)	0.37	NO	NO
63	Lipoprotein LpqB, beta-propeller domain-like protein (332340014)	-	57776	7(3)	0.25	YES	NO
64	Extracellular solute-binding protein family 3 (332339124)	-	30041	2 (1)	0.15	YES	YES
65	Extracellular solute-binding protein family 1 (332340331)	-	59378	10 (2)	0.15	NO	YES
66	WD40-like beta Propeller containing protein (332337747)	-	116717	8 (1)	0.04	YES	YES
67	Cobalamin synthesis CobW domain protein (332340651)	-	39903	8 (1)	0.11	NO	NO
68	Chaperonin GroEL (332338193)	-	56352	9 (2)	0.16	NO	NO
69	Neprilysin (332338954)	-	72498	11 (1)	0.06	NO	NO
70	Basic membrane lipoprotein (332340404)	-	38019	4 (1)	0.12	YES	YES
71	Catalase (332339192)	-	54817	13 (1)	0.08	NO	NO
72	Ricin B lectin (332337998)	-	60246	13 (1)	0.07	YES	YES
73	Peptidyl-dipeptidase Dcp (332339321)	-	76280	6 (1)	0.06	NO	NO

74	Periplasmic binding protein/LacI transcriptional regulator (332341019)	-	40664	2 (1)	0.11	YES	YES
75	Hypothetical protein Celf_3353 (332340883)	-	37795	5 (2)	0.25	YES	NO
76	Regulatory protein TetR (332337908)	-	24168	8 (1)	0.19	NO	NO
77	Alpha/beta hydrolase fold protein (332338325)	-	30206	3 (1)	0.15	NO	NO
78	Glycoside hydrolase family 43 (332338345)	GH43	127647	15 (1)	0.03	YES	YES
79	GrpE protein (332338016)	-	23333	1 (1)	0.2	NO	NO
80	Glycine hydroxymethyltransferase (332338687)	-	45466	5 (1)	0.1	NO	NO
81	Hypothetical protein Celf_3380 (332340910)	-	12743	1(1)	0.38	NO	YES
82	TAP domain protein (332338042)	-	54198	3 (1)	0.08	YES	NO
83	3-oxoacyl-(acyl-carrier-protein) reductase (332339263)	-	25808	8 (1)	0.18	NO	NO
84	Tail sheath protein (332337732)	-	56259	13 (1)	0.08	NO	NO
85	Peptide deformylase (332339815)	-	18873	3 (1)	0.25	NO	NO
86	Hypothetical protein Celf_1758 (332339307)	-	122020	28 (1)	0.04	NO	NO
87	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	-	29484	5 (1)	0.15	NO	NO

	(332340482)						
88	Dihydroxy-acid dehydratase (332340266)	-	66305	12 (1)	0.07	NO	NO
89	Hypothetical protein Celf_2249 (332339794)	-	14782	3 (1)	0.32	YES	YES
90	Methylcrotonoyl-CoA carboxylase (332340101)	-	57448	12 (1)	0.08	NO	NO
91	Beta-phosphoglucomutase family hydrolase (332339433)	GH65	114851	21 (1)	0.04	NO	NO
92	Hypothetical protein Celf_0974 (332338528)	-	49735	7 (1)	0.09	YES	NO
93	Stress protein (332337770)	-	20550	8 (1)	0.23	NO	NO

Table S3: secreted proteins collected after 24 h of *C. fimi* growth on wheat straw.

	Protein Identification [common name] (gi NCBI)	**CAZy classification**	**MW (kDa)**	**Peptide sequences (significant sequences)**	**emPAI**	**SignalP**	**SecretomeP**
1	Exoglucanase B [CbhB] (1708084)	GH48 [Cel48A]	115101	27 (21)	1.71	YES	YES
2	Exoglucanase A [CbhA] (1708083)	GH6 [Cel6B]	89588	26 (20)	2.6	YES	YES
3	Peptidase S8 and S53 subtilisin kexin sedolisin (332338096)	-	138613	32 (28)	1.67	YES	YES
4	Flagellin domain protein	-	39886	20 (14)	11.76	NO	YES

	(332338221)						
5	Glycoside hydrolase family 9 [CenB] (332337588)	GH9 [Cel9A]	109266	34 (24)	1.86	YES	YES
6	Glycoside hydrolase family 5 [Cen D] (332339471)	GH5 [Cel5A]	79059	20 (15)	2.09	YES	YES
7	Endoglucanase C [CenC] (121819)	GH9 [Cel9B]	115658	18 (16)	1.09	YES	YES
8	Exoglucanase [Cex] (327179208)	GH10 [Xyn10A]	54499	26 (16)	3.06	YES	YES
9	Endo-1,4-beta-xylanase (332338131)	GH10, CE4	141693	29 (17)	0.67	YES	YES
10	Endo-1,4-beta-xylanase (332337655)	GH10	49754	16 (12)	2.91	YES	YES
11	Endo-1,4-beta-xylanase (332340687)	GH10	53458	17 (9)	1.04	YES	YES
12	5'-Nucleotidase domain-containing protein (332339202)	-	167817	30 (26)	0.99	YES	YES
13	Pectate lyase/Amb allergen (332341298)	PL1	52912	26 (22)	6.4	YES	YES
14	TAP domain protein (332337857)	-	63142	19 (15)	2.13	YES	YES
15	Glycoside hydrolase family 3 domain protein (332338140)	GH3	178143	28 (20)	0.61	YES	YES
16	Cellulose-binding family II (332339303)	CE2	52461	11 (10)	1.24	YES	YES
17	Aminopeptidase Y (332337698)	-	53985	8 (5)	0.6	YES	YES

18	Chitin-binding domain 3 protein (332337832)	AA10	37478	6 (5)	0.97	YES	YES
19	Extracellular solute-binding protein family 5 (332340445)	-	67393	8 (6)	0.55	YES	YES
20	Peptidase S1 and S6 chymotrypsin/Hap (332338679)	-	49521	11 (8)	0.98	YES	YES
21	Alpha-N-arabinofuranosidase (332340686)	GH62	53079	8 (4)	0.49	YES	YES
22	Extracellular solute-binding protein family 1 (332340803)	-	60706	11 (5)	0.42	YES	YES
23	PKD domain containing protein (332339822)	-	87063	17 (12)	0.8	YES	YES
24	Chaperonin GroEL (332338193)	-	56352	20 (13)	1.66	NO	NO
25	Cell surface receptor IPT/TIG domain protein (332341050)	-	99332	16 (13)	0.74	YES	YES
26	Endo-1,3(4)-beta-glucanase (332341296)	GH81	116431	12 (8)	0.34	NO	YES
27	Enolase (332340559)	-	45287	10 (7)	1.11	NO	NO
28	Extracellular solute-binding protein family 1 (332340270)	-	45120	9 (7)	0.93	YES	NO
29	Glyceraldehyde-3-phosphate dehydrogenase, type I (332339411)	-	35807	14 (8)	1.57	NO	NO
30	1, 4-beta cellobiohydrolase (332338782)	GH6	46631	6 (3)	0.44	YES	YES
31	1, 4-beta cellobiohydrolase [CenA] (332340715)	GH6 [Cel6A]	47018	7 (5)	0.88	YES	YES

32	Fibronectin type III domain protein (332339883)	-	210933	34 (7)	0.15	NO	YES
33	Alkaline phosphatase (332340091)	-	44035	9 (5)	0.78	NO	NO
34	Glutamine synthetase, type I (332339768)	-	53327	17 (5)	0.49	NO	NO
35	Extracellular solute-binding protein family 1 (332338842)	-	46266	8 (4)	0.44	YES	YES
36	Xylose isomerase (332338137)	-	43425	7 (3)	0.34	NO	NO
37	Glucose-6-phosphate isomerase (332338893)	-	60314	10 (6)	0.53	NO	NO
38	Extracellular solute-binding protein family 1 (332339297)	-	63225	15 (8)	0.71	YES	YES
39	Periplasmic binding protein (332341193)	-	34098	3 (3)	0.45	YES	NO
40	Glycoside hydrolase family 11 [XynD] (332337936)	GH11, CE4 [Xyn11A]	66855	9 (3)	0.21	YES	YES
41	Alpha-1,6-glucosidase, pullulanase-type (332338678)	GH13	209811	25 (8)	0.18	YES	YES
42	Periplasmic binding protein (332339005)	-	34886	3 (3)	0.62	YES	NO
43	Extracellular ligand-binding receptor (332339377)	-	43218	2 (2)	0.34	YES	NO
44	Phosphate ABC transporter, periplasmic phosphate-binding protein (332338148)	-	37419	3 (2)	0.25	YES	NO
45	Periplasmic binding protein/LacI transcriptional regulator	-	40664	7 (3)	0.37	YES	YES

	(332341019)						
46	Dihydrolipoamide dehydrogenase (332339755)	-	48441	8 (4)	0.42	NO	NO
47	Catalase (332339192)	-	54817	15 (6)	0.59	NO	NO
48	L-arabinose isomerase (332338427)	-	55111	15 (5)	0.47	NO	NO
49	Man26A (5359710)	GH26 [Man26A]	107064	9 (3)	0.13	YES	YES
50	NLPA lipoprotein (332338762)	-	33494	4 (2)	0.29	YES	YES
51	Ribosomal protein L20 (332339196)	-	14351	6 (1)	0.34	NO	NO
52	Glycoside hydrolase family 3 domain protein (332339598)	GH3	96753	8 (1)	0.04	YES	NO
53	Glycoside hydrolase family 18 (332340692)	GH18	58146	13 (2)	0.16	YES	YES
54	Glucan endo-1,3-beta-D-glucosidase (332340860)	GH64	55696	7 (2)	0.16	YES	YES
55	Glycogen/starch/alpha-glucan phosphorylase [maltodextrin phosphorylase] (332341073)	-	92938	13 (1)	0.05	NO	NO
56	Extracellular solute-binding protein family 1 (332340856)	-	48074	3 (2)	0.19	YES	YES
57	Aminopeptidase N (332338951)	-	94435	21 (3)	0.14	NO	NO
58	Pectate lyase (332337958)	PL3	44887	14 (3)	0.46	YES	YES
59	Ribosomal protein S4 (332339561)	-	23485	4 (2)	0.43	NO	NO
60	Cellulose-binding family II	GH74, CBM2	95879	17 (2)	0.09	YES	YES

	(32339460)						
61	Ribosomal protein L2 (332338558)	-	30456	14 (2)	0.32	NO	YES
62	Cellulose-binding family II (332337964)	CE1	44990	3 (1)	0.1	YES	YES
63	Extracellular solute-binding protein family 1 (332339858)	GH13	43054	2 (1)	0.1	YES	YES
64	Aconitate hydratase 1 (332339684)	-	100704	17 (3)	0.14	NO	NO
65	Ricin B lectin (332339947)	GHnc, CBM13	50141	5 (1)	0.09	YES	YES
66	Hypothetical protein Celf_2249 (332339794)	-	14782	2 (1)	0.32	YES	YES
67	Ricin B lectin (332337998)	-	60246	11 (1)	0.07	YES	YES
68	Thimet oligopeptidase (332337920)	-	70844	11 (1)	0.06	NO	NO
69	Cobalamin synthesis CobW domain protein (332340651)	-	39903	9 (1)	0.11	NO	NO
70	Domain of unknown function DUF2394 (332338139)	-	35622	1 (1)	0.13	NO	NO
71	Basic membrane lipoprotein (332340404)	-	38019	4 (1)	0.12	YES	YES
72	Regulatory protein TetR (332337908)	-	24168	8 (1)	0.19	NO	NO
73	Peptidyl-dipeptidase Dcp (332339321)	-	76280	9 (1)	0.06	NO	NO
74	Transaldolase (332339419)	-	39758	6 (2)	0.24	NO	NO
75	Extracellular solute-binding protein family 1 (332338899)	-	47100	7 (1)	0.09	YES	YES

76	Phosphoglucomutase, alpha-D-glucose phosphate-specific (332341135)	-	58852	9 (1)	0.07	NO	NO
77	Cellulase (332337938)	GH5, CBM46	61605	9 (1)	0.07	YES	NO
78	Extracellular solute-binding protein family 1 (332340331)	-	59378	9 (1)	0.07	NO	YES
79	Transcriptional regulator, winged helix family (332339235)	-	117884	14 (1)	0.04	NO	NO
80	Alpha-L-arabinofuranosidase domain protein (332340852)	GH51	55481	7 (1)	0.08	NO	NO
81	Extracellular solute-binding protein family 1 (332340306)	-	47972	5 (1)	0.09	YES	YES
82	Hypothetical protein Celf_1998 (332339545)	-	54461	6 (1)	0.08	NO	YES
83	Ribosomal protein S3 (332338561)	-	30844	15 (1)	0.15	NO	NO
84	Flagellar hook-basal body protein (332338236)	-	39855	7 (2)	0.24	NO	YES
85	Dihydroxy-acid dehydratase (332340266)	-	66305	10 (1)	0.07	NO	NO
86	Xylulokinase (332338136)	-	49036	10 (1)	0.09	NO	NO
87	Ribosomal protein L18 (332338571)	-	12954	3 (1)	0.38	NO	NO
88	Glycine hydroxymethyltransferase (332338687)	-	45466	8 (1)	0.1	NO	NO
89	GrpE protein (332338016)	-	23333	3 (1)	0.2	NO	NO
90	Guanosine pentaphosphate synthetase I/polyribonucleotide nucleotidyltransferase	-	79580	13 (1)	0.05	NO	NO

	(332339091)						
91	Transketolase (332339420)	-	76587	10 (1)	0.06	NO	NO
92	Ribosomal protein S12 (332338549)	-	13633	6 (1)	0.35	NO	YES
93	Peptidoglycan glycosyltransferase (332340923)	GT51	86700	17 (1)	0.05	NO	YES
94	Hypothetical protein Celf_3152 (332340683)	-	38242	4 (1)	0.12	NO	NO
95	Ribosomal protein L5 (332338567)	-	21150	5 (1)	0.22	NO	NO
96	Hypothetical protein Celf_0653 (332338210)	-	91828	12 (1)	0.05	NO	NO
97	Hypothetical protein (753797931)	-	47461	6 (1)	0.09	NO	NO
98	Extracellular solute-binding protein family 5 (332339390)	-	60117	4 (1)	0.07	YES	YES
99	Glycoside hydrolase family 16 (332340645)	GH16	45423	7 (1)	0.1	YES	YES
100	Methylcrotonoyl-CoA carboxylase (332340101)	-	57448	8 (1)	0.08	NO	NO

Table S4: secreted proteins collected after 48 h of *C. fimi* growth on wheat straw.

	Protein Identification [common name] (gi NCBI)	CAZy classification	MW (kDa)	Peptide sequences (significant sequences)	emPAI	SignalP	SecretomeP
1	Exoglucanase B [CbhB] (1708084)	GH48 [Cel48A]	115101	29 (22)	2.17	YES	YES
2	Exoglucanase A [CbhA] (1708083)	GH6 [Cel6B]	89588	26 (24)	5.15	YES	YES
3	Glycoside hydrolase family 5 [Cen D] (332339471)	GH5 [Cel5A]	79059	26 (22)	3.8	YES	YES
4	Glycoside hydrolase family 9 [CenB] (332337588)	GH9 [Cel9A]	109266	40 (34)	3.8	YES	YES
5	Peptidase S8 and S53 subtilisin kexin sedolisin (332338096)	-	138613	33 (28)	1.95	YES	YES
6	Endoglucanase C [CenC] (121819)	GH9 [Cel9B]	115658	19 (17)	1.26	YES	YES
7	Endo-1,4-beta-xylanase (332338131)	GH10, CE4	141693	40 (34)	1.97	YES	YES
8	Exoglucanase [Cex] (327179208)	GH10 [Xyn10A]	54499	26 (14)	0.43	YES	YES

9	Flagellin domain protein (332338221)	-	39886	19 (13)	5.83	NO	YES
10	Endo-1,4-beta-xylanase (332337655)	GH10	49754	18 (10)	1.8	YES	YES
11	Endo-1,4-beta-xylanase (332340687)	GH10	53458	13 (9)	1.4	YES	YES
12	5'-Nucleotidase domain-containing protein (332339202)	-	167817	21 (19)	0.67	YES	YES
13	Glycoside hydrolase family 3 domain protein (332338140)	GH3	178143	38 (28)	0.96	YES	YES
14	Pectate lyase/Amb allergen (332341298)	PL1	52912	26 (21)	5.38	YES	YES
15	Aminopeptidase Y (332337698)	-	53985	10 (5)	0.48	YES	YES
16	Pectate lyase (332337958)	PL3	44887	11 (8)	1.58	YES	YES
17	PKD domain containing protein (332339822)	-	87063	19 (17)	1.42	YES	YES
18	Cellulose-binding family II (332339303)	CE2	52461	8 (7)	0.92	YES	YES
19	Extracellular solute-binding protein family 5 (332340445)	-	67393	14 (13)	1.28	YES	YES
20	Endo-1,3(4)-beta-glucanase (332341296)	GH81	116431	22 (16)	0.87	NO	YES
21	TAP domain protein (332337857)	-	63142	17 (12)	1.41	YES	YES
22	1, 4-beta cellobiohydrolase [CenA]	GH6 [Cel6A]	47018	7 (5)	0.88	YES	YES

	(332340715)						
23	Alpha-1,6-glucosidase, pullulanase-type (332338678)	GH13	209811	32 (15)	0.36	YES	YES
24	Glycoside hydrolase family 3 domain protein (332339598)	GH3	96753	13 (9)	0.49	YES	NO
25	Peptidase S1 and S6 chymotrypsin/Hap (332338679)	-	49521	9 (3)	0.29	YES	YES
26	Man26A (5359710)	GH26 [Man26A]	107064	14 (7)	0.32	YES	YES
27	Extracellular solute-binding protein family 1 (332340803)	-	60706	9 (3)	0.23	YES	YES
28	Glycoside hydrolase family 11 [XynD] (332337936)	GH11, CE4 [Xyn11A]	66855	11 (5)	0.38	YES	YES
29	Chitin-binding domain 3 protein (332337832)	AA10	37478	5 (3)	0.57	YES	YES
30	Alpha-N-arabinofuranosidase (332340686)	GH62	53079	4 (2)	0.17	YES	YES
31	Fibronectin type III domain protein (332339883)	-	210933	27 (4)	0.08	NO	YES
32	Periplasmic binding protein (332341193)	-	34098	3 (3)	0.45	YES	NO
33	1, 4-beta cellobiohydrolase (332338782)	GH6	46631	5 (4)	0.44	YES	YES
34	Extracellular solute-binding protein family 1 (332339297)	-	63225	10 (5)	0.4	YES	YES
35	Stress protein (332337771)	-	20150	13 (4)	0.2	NO	NO
36	Cellulose-binding family II (332339460)	GH74, CBM2	95879	17 (2)	0.09	YES	YES

37	Aminopeptidase N (332338951)	-	94435	14 (7)	0.37	NO	NO
38	Cell surface receptor IPT/TIG domain protein (332341050)	-	99332	7 (5)	0.24	YES	YES
39	Glycoside hydrolase family 18 (332340692)	GH18	58146	8 (1)	0.08	YES	YES
40	Extracellular solute-binding protein family 1 (332338842)	-	46266	5 (4)	0.45	YES	YES
41	Extracellular ligand-binding receptor (332339377)	-	43218	2 (2)	0.22	YES	NO
42	Glutamine synthetase, type I (332339768)	-	53327	7 (3)	0.27	NO	NO
43	NLPA lipoprotein (332338762)	-	33494	7 (2)	0.29	YES	YES
44	Alkaline phosphatase (332340091)	-	44035	8 (3)	0.34	NO	NO
45	Peptidase M24 (332338709)	-	55886	8 (3)	0.26	NO	NO
46	Enolase (332340559)	-	45287	8 (3)	0.33	NO	NO
47	Glyceraldehyde-3-phosphate dehydrogenase, type I (332339411)	-	35807	9 (5)	0.81	NO	NO
48	Cellulose-binding family II (332337964)	CE1	44990	3 (1)	0.1	YES	YES
49	Glycoside hydrolase family 9 (332337613)	GH9	94310	7 (1)	0.05	YES	YES
50	Glucose-6-phosphate isomerase (332338893)	-	60314	8 (2)	0.15	NO	NO
51	Xylose isomerase	-	43425	4 (2)	0.22	NO	NO

	(332338137)						
52	Extracellular solute-binding protein family 1 (332340270)	-	45120	6 (2)	0.21	YES	NO
53	Ribosomal protein L27 (332339968)	-	45120	7 (1)	0.58	NO	NO
54	Cobalamin synthesis CobW domain protein (332340651)	-	39903	3 (1)	0.11	NO	NO
55	Hypothetical protein Celf_2704 (332340245)	-	24970	5 (1)	0.18	NO	NO
56	Ribosomal protein L18 (332338571)	-	12954	1 (1)	0.38	NO	NO
57	1-deoxy-D-xylulose 5-phosphate reductoisomerase (332339073)	-	40195	4 (1)	0.11	NO	NO
58	Glucan endo-1,3-beta-D-glucosidase (332340860)	GH64	55696	7 (1)	0.08	YES	YES
59	Extracellular solute-binding protein family 3 (332339124)	-	30041	1 (1)	0.15	YES	YES
60	Regulatory protein TetR (332337908)	-	24168	5 (1)	0.19	NO	NO
61	Dihydrolipoamide dehydrogenase (332339755)	-	48441	7 (1)	0.09	NO	NO
62	Phosphate ABC transporter, periplasmic phosphate-binding protein (332338148)	-	37419	2 (1)	0.12	YES	NO
63	Transaldolase (332339419)	-	39758	6 (1)	0.11	NO	NO
64	Extracellular solute-binding protein family 1 (332340331)	-	59378	6 (1)	0.07	NO	YES
65	Transcriptional regulator, winged helix family	-	117884	18 (1)	0.04	NO	NO

	(332339235)						
66	Dihydroxy-acid dehydratase (332340266)	-	66305	9 (1)	0.07	NO	NO
67	Peptide deformylase (332339815)	-	18873	8 (1)	0.25	NO	NO
68	Hypothetical protein (753797949)	-	39470	3 (1)	0.11	YES	NO
69	Zn-dependent hydrolase, glyoxylase (332339402)	-	24782	5 (1)	0.19	NO	NO
70	Hypothetical protein Celf_2249 (332339794)	-	14782	3 (1)	0.33	YES	YES
71	GrpE protein (332338016)	-	23333	4 (1)	0.2	NO	NO
72	Hypothetical protein Celf_3380 (332340910)	-	12743	1 (1)	0.38	NO	YES
73	3-dehydroquinate synthase (332339538)	-	39463	3 (1)	0.11	NO	NO
74	AMP-dependent synthetase and ligase (332338991)	-	64475	10 (1)	0.07	NO	NO

Table S5: secreted proteins collected after 24 h of *C. fimi* growth on sugar cane bagasse.

	Protein Identification [common name] (gi NCBI)	**CAZy classification**	**MW (kDa)**	**Peptide sequences (significant sequences)**	**emPAI**	**SignalP**	**SecretomeP**
1	Flagellin domain protein (332338221)	-	39886	31 (23)	75.49	NO	YES
2	Exoglucanase A [CbhA]	GH6 [Cel6B]	89588	36 (29)	8.7	YES	YES

	(1708083)						
3	Exoglucanase B [CbhB] (1708084)	GH48 [Cel48A]	115101	30 (22)	1.81	YES	YES
4	Endoglucanase C [CenC] (121819)	GH9 [Cel9B]	115658	23 (19)	1.6	YES	YES
5	Peptidase S8 and S53 subtilisin kexin sedolisin (332338096)	-	138613	34 (29)	2.01	YES	YES
6	Glycoside hydrolase family 5 [Cen D] (332339471)	GH5 [Cel5A]	79059	24 (19)	2.43	YES	YES
7	Glycoside hydrolase family 9 [CenB] (332337588)	GH9 [Cel9A]	109266	37 (30)	2.89	YES	YES
8	Endo-1,4-beta-xylanase (332338131)	GH10, CE4	141693	31 (23)	1.05	YES	YES
9	Endo-1,4-beta-xylanase (332337655)	GH10	49754	18 (14)	3.24	YES	YES
10	Endo-1,4-beta-xylanase (332340687)	GH10	53458	18 (12)	2.54	YES	YES
11	Exoglucanase [Cex] (327179208)	GH10 [Xyn10A]	54499	24 (14)	2.72	YES	YES
12	Glycoside hydrolase family 3 domain protein (332338140)	GH3	178143	34 (30)	1.09	YES	YES
13	5’-Nucleotidase domain-containing protein (332339202)	-	167817	31 (28)	1.14	YES	YES
14	Chaperonin GroEL	-	56352	35 (32)	10.88	NO	NO

	(332338193)						
15	Pectate lyase/Amb allergen (332341298)	PL1	52912	24 (20)	5.79	YES	YES
16	Cell surface receptor IPT/TIG domain protein (332341050)	-	99332	19 (15)	0.9	YES	YES
17	TAP domain protein (332337857)	-	63142	22 (17)	3.08	YES	YES
18	Aminopeptidase Y (332337698)	-	53985	12 (6)	0.6	YES	YES
19	Extracellular solute-binding protein family 5 (332340445)	-	67393	17 (13)	1.73	YES	YES
20	Enolase (332340559)	-	45287	12 (8)	1.31	NO	NO
21	Fibronectin type III domain protein (332339883)	-	210933	40 (15)	0.38	NO	YES
22	Peptidase S1 and S6 chymotrypsin/Hap (332338679)	-	49521	9 (5)	0.67	YES	YES
23	Cellulose-binding family II (332339303)	CE2	52461	16 (13)	2.34	YES	YES
24	Endo-1,3(4)-beta-glucanase (332341296)	GH81	116431	25 (17)	0.86	NO	YES
25	1, 4-beta cellobiohydrolase (332338782)	GH6	46631	11 (8)	1.06	YES	YES
26	Alpha-N-arabinofuranosidase (332340686)	GH62	53079	9 (5)	0.49	YES	YES
27	Chitin-binding domain 3 protein (332337832)	AA10	37478	6 (4)	0.57	YES	YES
28	Alkaline phosphatase (332340091)	-	44035	13 (8)	1.37	NO	NO
29	Extracellular solute-binding protein family 1 (332340803)	-	60706	9 (6)	0.52	YES	YES

30	Aminopeptidase N (332338951)	-	94435	28 (16)	1.05	NO	NO
31	Glutamine synthetase, type I (332339768)	-	53327	21 (13)	1.8	NO	NO
32	Extracellular solute-binding protein family 1 (332338842)	-	46266	8 (4)	0.44	YES	YES
33	Glyceraldehyde-3-phosphate dehydrogenase, type I (332339411)	-	35807	12 (7)	1.28	NO	NO
34	PKD domain containing protein (332339822)	-	87063	20 (12)	0.79	YES	YES
35	Extracellular solute-binding protein family 1 (332340270)	-	45120	11 (9)	1.55	YES	NO
36	1, 4-beta cellobiohydrolase [CenA] (332340715)	GH6 [Cel6A]	47018	8 (7)	0.87	YES	YES
37	Extracellular solute-binding protein family 1 (332339297)	-	63225	16 (10)	0.95	YES	YES
38	NLPA lipoprotein (332338762)	-	33494	5 (3)	0.46	YES	YES
39	Glycoside hydrolase family 11 [XynD] (332337936)	GH11, CE4 [Xyn11A]	66855	8 (3)	0.21	YES	YES
40	L-arabinose isomerase (332338427)	-	55111	17 (8)	0.85	NO	NO
41	Xylose isomerase (332338137)	-	43425	5 (3)	0.34	NO	NO
42	Periplasmic binding protein/LacI transcriptional regulator (332341019)	-	40664	4 (4)	0.51	YES	YES

43	ATP synthase F1, alpha subunit (332340342)	-	58082	14 (9)	0.93	NO	NO
44	Pectate lyase (332337958)	PL3	44887	9 (2)	0.33	YES	YES
45	Glucose-6-phosphate isomerase (332338893)	-	60314	13 (7)	0.63	NO	NO
46	Periplasmic binding protein (332339005)	-	34886	4 (3)	0.44	YES	NO
47	Lipoprotein LpqB, beta-propeller domain-like protein (332340014)	-	57776	15 (7)	0.67	YES	NO
48	Peptidase M24 (332338709)	-	55886	12 (6)	0.58	NO	NO
49	Cellulose-binding family II (332337964)	CE1	44990	2 (2)	0.21	YES	YES
50	Extracellular solute-binding protein family 1 (332340331)	-	59378	11 (6)	0.65	NO	YES
51	Catalase (332339192)	-	54817	10 (6)	0.59	NO	NO
52	Periplasmic binding protein (332341193)	-	34098	3 (3)	0.45	YES	NO
53	Extracellular solute-binding protein family 1 (332340856)	-	48074	8 (5)	0.55	YES	YES
54	Aconitate hydratase 1 (332339684)	-	100704	21 (5)	0.23	NO	NO
55	Phosphate ABC transporter, periplasmic phosphate-binding protein (332338148)	-	37419	2 (2)	0.25	YES	NO

56	Basic membrane lipoprotein (332340404)	-	38019	5 (2)	0.25	YES	YES
57	Dihydrolipoamide dehydrogenase (332339755)	-	48441	9 (5)	0.55	NO	NO
58	Extracellular ligand-binding receptor (332339377)	-	43218	5 (3)	0.34	YES	NO
59	Stress protein (332337771)	-	20150	5 (3)	0.86	NO	NO
60	Glycogen/starch/alpha-glucan phosphorylase [maltodextrin phosphorylase] (332341073)	-	92938	21 (4)	0.2	NO	NO
61	Putative F420-dependent oxidoreductase (332338604)	-	36769	7 (2)	0.26	NO	NO
62	Glycoside hydrolase family 9 (332337613)	GH9	94310	8 (1)	0.05	YES	YES
63	Alpha-1,6-glucosidase, pullulanase-type (332338678)	GH13	209811	24 (4)	0.08	YES	YES
64	Flagellar hook-basal body protein (332338236)	-	39855	5 (2)	0.24	NO	YES
65	Succinyl-CoA synthetase, beta subunit (332338665)	-	40965	10 (2)	0.23	NO	NO
66	Ricin B lectin (332337998)	-	60246	12 (3)	0.23	YES	YES
67	Amidohydrolase (332338891)	-	44181	11 (1)	0.1	NO	NO
68	Extracellular solute-binding protein family 1 (332339858)	GH13	43054	5 (2)	0.34	YES	YES
69	Glycoside hydrolase family 3 domain protein (332339598)	GH3	96753	12 (2)	0.09	YES	NO
70	N-acetylmuramoyl-L-alanine amidase family 2	-	94631	9 (3)	0.14	YES	YES

	(332340054)						
71	Glycoside hydrolase family 18 (332340692)	GH18	58146	11 (3)	0.24	YES	YES
72	TAP domain protein (332338042)	-	54198	4 (4)	0.37	YES	NO
73	Curculin domain protein (mannose-binding) lectin (332339921)	-	53997	4 (2)	0.17	YES	YES
74	Phosphoglucomutase, alpha-D-glucose phosphate-specific (332341135)	-	58852	11 (4)	0.33	NO	NO
75	Guanosine pentaphosphate synthetase I/polyribonucleotide nucleotidyltransferase (332339091)	-	79580	11 (3)	0.17	NO	NO
76	Glucan endo-1,3-beta-D-glucosidase (332340860)	GH64	55696	11 (5)	0.46	YES	YES
77	Ribosomal protein L18 (332338571)	-	12954	6 (2)	0.89	NO	NO
78	Ribosomal protein L4/L1e (332338556)	-	24253	7 (1)	0.19	NO	YES
79	Cellulose-binding family II (32339460)	GH74, CBM2	95879	12 (3)	0.14	YES	YES
80	Glycoside hydrolase family 16 (332340645)	GH16	45423	4 (2)	0.2	YES	YES
81	Peptidoglycan glycosyltransferase (332340923)	GT51	86700	12 (3)	0.16	NO	YES
82	Ribosomal protein S4 (332339561)	-	23485	8 (1)	0.19	NO	NO
83	Cellulase (332337938)	GH5, CBM46	61605	9 (2)	0.15	YES	NO
84	Extracellular solute-binding protein family 1	-	47972	8 (2)	0.19	YES	YES

	(332340306)						
85	Hypothetical protein Celf_2249 (332339794)	-	14782	2 (1)	0.32	YES	YES
86	Leucyl aminopeptidase (332339753)	-	49979	7 (1)	0.09	NO	YES
87	Cell envelope-related transcriptional attenuator (332340061)	-	41524	4 (1)	0.11	YES	YES
88	Man26A (5359710)	GH26 Man26A]	107064	11 (2)	0.08	YES	YES
89	Glucose-1-phosphate adenylyltransferase (332339267)	-	44979	6 (1)	0.1	NO	NO
90	WD40-like beta Propeller containing protein (332337747)	-	116717	1 (1)	0.04	YES	YES
91	Ribosomal protein L24 (332338566)	-	12549	6 (1)	0.39	NO	NO
92	Glycine hydroxymethyltransferase (332338687)	-	45466	8 (1)	0.1	NO	NO
93	Carbon starvation protein CstA (332337945)	-	79679	11 (1)	0.05	NO	YES
94	Extracellular solute-binding protein family 1 (332337651)	-	45283	6 (1)	0.1	YES	NO
95	Transketolase (332339420)	-	76587	7 (1)	0.06	NO	NO
96	Transaldolase (332339419)	-	39758	13 (1)	0.11	NO	NO
97	Extracellular solute-binding protein family 1 (332337695)	-	47942	3 (1)	0.09	NO	YES
98	DNA-directed RNA polymerase, beta' subunit (332338548)	-	143483	28 (1)	0.03	NO	NO
99	Flagellar hook-associated protein FlgK (332338218)	-	48839	5 (1)	0.09	NO	NO

100	Hypothetical protein Celf_1679 (332339228)	-	34792	2 (1)	0.13	NO	NO
101	DNA-directed RNA polymerase, beta subunit (332338547)	-	128655	23 (1)	0.03	NO	NO
102	Protein of unknown function DUF124 (332341292)	-	27337	5 (1)	0.17	NO	NO
103	Ribosomal protein L17 (332338584)	-	18585	5 (1)	0.25	NO	NO
104	Putative exonuclease (332341232)	-	109116	21 (1)	0.04	NO	NO
105	Ribosomal protein S9 (332338595)	-	17515	4 (1)	0.27	NO	NO
106	Two component transcriptional regulator, winged helix family (332338869)	-	23189	3 (1)	0.2	NO	NO
107	Ribosomal protein L2 (332338558)	-	30456	9 (1)	0.15	NO	YES
108	Ribosomal protein S3 (332338561)	-	30844	9 (1)	0.15	NO	NO
109	Glycosyl transferase family 2 (332338940)	GT2	31569	6 (1)	0.14	NO	NO
110	Glycogen debranching enzyme GlgX (332339655)	GH13	83431	3 (1)	0.05	NO	NO
111	Ribosomal protein L14 (332338565)	-	13435	14 (1)	0.36	NO	NO
112	Peptidase M20 (332339322)	-	48556	6 (1)	0.09	NO	NO
113	Hypothetical protein Celf_3047 (332340581)	-	55195	6 (2)	0.17	YES	YES
114	Oxidoreductase domain protein (332341060)	-	36691	9 (1)	0.12	NO	NO

115	Tail sheath protein (332337732)	-	56259	10 (1)	0.08	NO	NO
116	Alpha/beta hydrolase fold protein (332338325)	-	30206	2 (1)	0.15	NO	NO
117	Alpha-L-arabinofuranosidase domain protein (332340780)	GH51	56678	8 (1)	0.08	NO	NO
118	Ribosomal protein L5 (332338567)	-	21150	5 (1)	0.22	NO	NO
119	Glycoside hydrolase, family 20, catalytic core (332340615)	GH20	53273	6 (1)	0.08	NO	NO
120	Thimet oligopeptidase (332337920)	-	70844	7 (1)	0.06	NO	NO
121	Protein of unknown function DUF839 (332337822)	-	76485	9 (1)	0.06	NO	NO
122	Xylulokinase (332338136)	-	49036	7 (1)	0.09	NO	NO
123	SMF family protein (332339063)	-	41861	5 (1)	0.11	NO	NO

Table S6: secreted proteins collected after 48 h of *C. fimi* growth on sugar cane bagasse.

	Protein Identification [common name] (gi NCBI)	**CAZy classification**	**MW (kDa)**	**Peptide sequences (significant sequences)**	**emPAI**	**SignalP**	**SecretomeP**
1	Exoglucanase B [CbhB] (1708084)	GH48 [Cel48A]	115101	39 (31)	4.27	YES	YES
2	Exoglucanase A [CbhA]	GH6 [Cel6B]	89588	28 (24)	3.78	YES	YES

	(1708083)						
3	Glycoside hydrolase family 9 [CenB] (332337588)	GH9 [Cel9A]	109266	47 (37)	4.99	YES	YES
4	Endoglucanase C [CenC] (121819)	GH9 [Cel9B]	115658	23 (20)	1.6	YES	YES
5	Glycoside hydrolase family 5 [Cen D] (332339471)	GH5 [Cel5A]	79059	23 (16)	2.26	YES	YES
6	Flagellin domain protein (332338221)	-	39886	21 (15)	13.14	NO	YES
7	Exoglucanase [Cex] (327179208)	GH10 [Xyn10A]	54499	25 (17)	3.38	YES	YES
8	Peptidase S8 and S53 subtilisin kexin sedolisin (332338096)	-	138613	31 (27)	1.59	YES	YES
9	5'-Nucleotidase domain-containing protein (332339202)	-	167817	32 (24)	0.98	YES	YES
10	Endo-1,4-beta-xylanase (332338131)	GH10, CE4	141693	41 (32)	1.86	YES	YES
11	Glycoside hydrolase family 3 domain protein (332338140)	GH3	178143	39 (27)	0.91	YES	YES
12	PKD domain containing protein (332339822)	-	87063	21 (13)	0.89	YES	YES
13	Endo-1,4-beta-xylanase (332337655)	GH10	49754	16 (11)	2.29	YES	YES
14	Endo-1,4-beta-xylanase	GH10	53458	14 (10)	2.03	YES	YES

	(332340687)						
15	TAP domain protein (332337857)	-	63142	13 (9)	0.96	YES	YES
16	1, 4-beta cellobiohydrolase [CenA] (332340715)	GH6 [Cel6A]	47018	15 (10)	1.69	YES	YES
17	Cellulose-binding family II (332339303)	CE2	52461	8 (8)	0.91	YES	YES
18	Pectate lyase/Amb allergen (332341298)	PL1	52912	16 (12)	2.06	YES	YES
19	Alpha-1,6-glucosidase, pullulanase-type (332338678)	GH13	209811	41 (21)	0.53	YES	YES
20	Glycoside hydrolase family 11 [XynD] (332337936)	GH11, CE4 [Xyn11A]	66855	15 (6)	0.77	YES	YES
21	Extracellular solute-binding protein family 5 (332340445)	-	67393	9 (8)	0.76	YES	YES
22	Aminopeptidase Y (332337698)	-	53985	17 (10)	1.37	YES	YES
23	Endo-1,3(4)-beta-glucanase (332341296)	GH81	116431	19 (12)	0.61	NO	YES
24	Pectate lyase (332337958)	PL3	44887	12 (6)	0.76	YES	YES
25	Cellulose-binding family II (32339460)	GH74, CBM2	95879	24 (12)	0.7	YES	YES
26	Peptidase S1 and S6 chymotrypsin/Hap (332338679)	-	49521	12 (4)	0.41	YES	YES
27	Glycoside hydrolase family 9 (332337613)	GH9	94310	16 (9)	0.5	YES	YES
28	Chitin-binding domain 3 protein (332337832)	AA10	37478	6 (4)	0.57	YES	YES

29	Fibronectin type III domain protein (332339883)	-	210933	28 (9)	0.2	NO	YES
30	Glycoside hydrolase family 18 (332340692)	GH18	58146	11 (6)	0.55	YES	YES
31	Man26A (5359710)	GH26 [Man26A]	107064	15 (8)	0.37	YES	YES
32	1, 4-beta cellobiohydrolase (332338782)	GH6	46631	5 (3)	0.31	YES	YES
33	Extracellular solute-binding protein family 5 (332339390)	-	60117	8 (5)	0.42	YES	YES
34	Periplasmic binding protein (332341193)	-	34098	4 (4)	0.64	YES	NO
35	Enolase (332340559)	-	45287	11 (5)	0.59	NO	NO
36	Glycoside hydrolase family 3 domain protein (332339598)	GH3	96753	12 (3)	0.14	YES	NO
37	Alpha-N-arabinofuranosidase (332340686)	GH62	53079	7 (2)	0.17	YES	YES
38	Glyceraldehyde-3-phosphate dehydrogenase, type I (332339411)	-	35807	17 (4)	0.6	NO	NO
39	Extracellular solute-binding protein family 1 (332340803)	-	60706	6 (1)	0.07	YES	YES
40	Hypothetical protein Celf_2704 (332340245)	-	24970	5 (2)	0.4	NO	NO
41	Extracellular solute-binding protein family 1 (332338842)	-	46266	5 (3)	0.32	YES	YES
42	Periplasmic binding protein (332339005)	-	34886	4 (3)	0.44	YES	NO
43	Glucan endo-1,3-beta-D-glucosidase (332340860)	GH64	55696	10 (5)	0.46	YES	YES

44	Ricin B lectin (332337998)	-	60246	17 (3)	0.23	YES	YES
45	Cell surface receptor IPT/TIG domain protein (332341050)	-	99332	8 (2)	0.09	YES	YES
46	Alkaline phosphatase (332340091)	-	44035	7 (3)	0.33	NO	NO
47	Aminopeptidase N (332338951)	-	94435	14 (4)	0.2	NO	NO
48	Stress protein (332337771)	-	20150	5 (2)	0.51	NO	NO
49	Periplasmic binding protein/LacI transcriptional regulator (332341019)	-	40664	6 (2)	0.23	YES	YES
50	Phosphate ABC transporter, periplasmic phosphate-binding protein (332338148)	-	37419	3 (2)	0.25	YES	NO
51	Hypothetical protein Celf_2249 (332339794)	-	14782	4 (1)	0.32	YES	YES
52	Extracellular ligand-binding receptor (332339377)	-	43218	2 (2)	0.22	YES	NO
53	Cellulose-binding family II (332337964)	CE1	44990	2 (1)	0.1	YES	YES
54	Neprilysin (332338954)	-	72498	14 (1)	0.06	NO	NO
55	Cobalamin synthesis CobW domain protein (332340651)	-	39903	6 (1)	0.11	NO	NO
56	Extracellular solute-binding protein family 3 (332339124)	-	30041	1 (1)	0.15	YES	YES
57	Extracellular solute-binding protein family 1 (332340270)	-	45120	3 (2)	0.21	YES	NO
58	Thimet oligopeptidase	-	70844	10 (1)	0.06	NO	NO

	(332337920)						
59	Extracellular solute-binding protein family 1 (332340331)	-	59378	8 (2)	0.15	NO	YES
60	Glutamine synthetase, type I (332339768)	-	53327	12 (1)	0.08	NO	NO
61	Dihydrolipoamide dehydrogenase (332339755)	-	48441	9 (1)	0.09	NO	NO
62	Extracellular solute-binding protein family 1 (332339297)	-	63225	11 (1)	0.07	YES	YES
63	Curculin domain protein (mannose-binding) lectin (332339921)	-	53997	6 (1)	0.08	YES	YES
64	Regulatory protein TetR (332337908)	-	24168	12 (1)	0.19	NO	NO
65	Hypothetical protein Celf_3512 (332341040)	-	21621	2 (1)	0.21	YES	YES
66	Putative exonuclease (332341232)	-	109116	22 (1)	0.04	NO	NO
67	Hypothetical protein Celf_1998 (332339545)	-	54461	7 (1)	0.08	NO	YES
68	Xylose isomerase (332338137)	-	43425	4 (1)	0.1	NO	NO
69	Glycine hydroxymethyltransferase (332338687)	-	45466	5 (1)	0.1	NO	NO
70	Hypothetical protein Celf_0090 (332337657)	-	19014	4 (1)	0.25	NO	NO
71	Transcriptional regulator, winged helix family (332339235)	-	117884	13 (1)	0.04	NO	NO
72	Hypothetical protein (753797949)	-	39470	5 (1)	0.11	YES	NO

73	Dihydroxy-acid dehydratase (332340266)	-	66305	14 (1)	0.07	NO	NO
74	Peptidase M24 (332338709)	-	55886	10 (1)	0.08	NO	NO
75	Lytic transglycosylase catalytic (332339713)	GH23	38842	8 (1)	0.11	YES	YES
76	Hypothetical protein Celf_0713 (332338270)	-	16730	3 (1)	0.28	NO	YES

Table S7: secreted proteins collected after 24 h of *C. fimi* growth on carboxymethyl-cellulose.

	Protein Identification [common name] (gi NCBI)	**CAZy classification**	**MW (kDa)**	**Peptide sequences (significant sequences)**	**emPAI**	**SignalP**	**SecretomeP**
1	Exoglucanase A [CbhA] (1708083)	GH6 [Cel6B]	89588	30 (25)	4.84	YES	YES
2	Flagellin domain protein (332338221)	-	39886	27 (23)	62.9	NO	YES
3	Exoglucanase B [CbhB] (1708084)	GH48 [Cel48A]	115101	29 (19)	1.44	YES	YES
4	Endoglucanase C [CenC] (121819)	GH9 [Cel9B]	115658	23 (17)	1.26	YES	YES
5	Peptidase S8 and S53 subtilisin kexin sedolisin (332338096)	-	138613	26 (25)	1.37	YES	YES
6	Glycoside hydrolase family 5 [Cen D]	GH5 [Cel5A]	79059	22 (14)	1.65	YES	YES

	(332339471)						
7	TAP domain protein (332337857)	-	63142	26 (21)	5.19	YES	YES
8	Glycoside hydrolase family 9 [CenB] (332337588)	GH9 [Cel9A]	109266	37 (25)	1.77	YES	YES
9	Pectate lyase/Amb allergen (332341298)	PL1	52912	25 (22)	7.78	YES	YES
10	Exoglucanase [Cex] (327179208)	GH10 [Xyn10A]	54499	26 (18)	5.17	YES	YES
11	1, 4-beta cellobiohydrolase [CenA] (332340715)	GH6 [Cel6A]	47018	17 (13)	3.25	YES	YES
12	Cellulose-binding family II (332339303)	CE2	52461	15 (12)	2.38	YES	YES
13	Peptidase S1 and S6 chymotrypsin/Hap (332338679)	-	49521	12 (9)	1.58	YES	YES
14	Chitin-binding domain 3 protein (332337832)	AA10	37478	9 (7)	1.47	YES	YES
15	Alpha-1,6-glucosidase, pullulanase-type (332338678)	GH13	209811	35 (22)	0.57	YES	YES
16	Glycoside hydrolase family 3 domain protein (332338140)	GH3	178143	24 (15)	0.43	YES	YES
17	Man26A (5359710)	GH26 [Man26A]	107064	15 (13)	0.68	YES	YES
18	Endo-1,4-beta-xylanase (332340687)	GH10	53458	24 (9)	1.4	YES	YES
19	Endo-1,4-beta-xylanase	GH10	49754	20 (12)	2.32	YES	YES

	(332337655)						
20	Enolase (332340559)	-	45287	17 (13)	2.39	NO	NO
21	Cellulase (332337938)	GH5, CBM46	61605	16 (9)	0.86	YES	NO
22	PKD domain containing protein (332339822)	-	87063	14 (9)	0.56	YES	YES
23	5'-Nucleotidase domain-containing protein (332339202)	-	167817	21 (9)	0.29	YES	YES
24	Aminopeptidase Y (332337698)	-	53985	10 (3)	0.27	YES	YES
25	1, 4-beta cellobiohydrolase (332338782)	GH6	46631	12 (8)	1.27	YES	YES
26	Hypothetical protein Celf_3047 (332340581)	-	55195	16 (12)	1.53	YES	YES
27	Extracellular ligand-binding receptor (332339377)		43218	6 (4)	0.64	YES	NO
28	Extracellular solute-binding protein family 1 (332338842)	-	46266	10 (7)	1.09	YES	YES
29	Endo-1,4-beta-xylanase (332338131)	GH10, CE4	141693	11 (8)	0.31	YES	YES
30	Periplasmic binding protein/LacI transcriptional regulator (332341019)	-	40664	9 (5)	0.69	YES	YES
31	Extracellular solute-binding protein family 5 (332340445)	-	67393	6 (6)	0.46	YES	YES
32	Alkaline phosphatase (332340091)	-	44035	9 (6)	0.96	NO	NO
33	Glycoside hydrolase family 11 [XynD] (332337936)	GH11, CE4 [Xyn11A]	66855	6 (3)	0.29	YES	YES

34	Pectate lyase (332337958)	PL3	44887	11 (5)	0.61	YES	YES
35	Alpha-N-arabinofuranosidase (332340686)	GH62	53079	10 (7)	1.06	YES	YES
36	Fibronectin type III domain protein (332339883)	-	210933	30 (7)	0.15	NO	YES
37	Dihydrolipoamide dehydrogenase (332339755)	-	48441	11 (7)	0.85	NO	NO
38	Glutamine synthetase, type I (332339768)	-	53327	12 (6)	0.62	NO	NO
39	Cell surface receptor IPT/TIG domain protein (332341050)	-	99332	6 (4)	0.19	YES	YES
40	Xylose isomerase (332338137)	-	43425	7 (4)	0.48	NO	NO
41	Glucose-6-phosphate isomerase (332338893)	-	60314	12 (4)	0.33	NO	NO
42	Stress protein (332337771)	-	20150	5 (4)	1.3	NO	NO
43	Endo-1,3(4)-beta-glucanase (332341296)	GH81	116431	5 (2)	0.12	NO	YES
44	Extracellular solute-binding protein family 1 (332340306)	-	47972	7 (3)	0.31	YES	YES
45	Curculin domain protein (mannose-binding) lectin (332339921)	-	53997	8 (4)	0.37	YES	NO
46	Flagellar hook-basal body protein (332338236)	-	39855	5 (2)	0.24	NO	YES
47	Periplasmic binding protein (332341193)	-	34098	2 (1)	0.13	YES	NO
48	Glycoside hydrolase family 9 (332337613)	GH9	94310	9 (1)	0.05	YES	YES

49	Cellulose-binding family II (332337963)	CBM2	41163	5 (2)	0.36	YES	YES
50	Phosphate ABC transporter, periplasmic phosphate-binding protein (332338148)	-	37419	3 (3)	0.41	YES	YES
51	Glyceraldehyde-3-phosphate dehydrogenase, type I (332339411)	-	35807	10 (2)	0.27	NO	NO
52	Transaldolase (332339419)	-	39758	10 (2)	0.24	NO	NO
53	Hypothetical protein Celf_1998 (332339545)	-	54461	9 (2)	0.17	NO	YES
54	N-acetylmuramoyl-L-alanine amidase family 2 (332340054)	-	94631	7 (1)	0.05	YES	YES
55	Cellulose-binding family II (332337964)	CE1	44990	2 (1)	0.1	YES	YES
56	Cell envelope-related transcriptional attenuator (332340061)	-	41524	5 (1)	0.11	YES	YES
57	NLPA lipoprotein (332338762)	-	33494	6 (2)	0.29	YES	YES
58	Hypothetical protein Celf_2249 (332339794)	-	14782	2 (1)	0.33	YES	YES
59	Extracellular solute-binding protein family 1 (332339858)	GH13	43054	5 (1)	0.1	YES	YES
60	SCP-like extracellular (332338299)	-	36228	6 (1)	0.12	NO	NO
61	Protein of unknown function DUF124 (332341292)	-	27337	6 (1)	0.17	NO	NO
62	Aminopeptidase N (332338951)	-	94435	15 (2)	0.09	NO	NO
63	Glycoside hydrolase family 9	GH9	115872	9 (1)	0.04	YES	YES

	(332339254)						
64	Ribosomal protein S4 (332339561)	-	23485	3 (1)	0.2	NO	NO
65	Peptidyl-dipeptidase Dcp (332339321)	-	76280	4 (1)	0.06	NO	NO
66	Hypothetical protein (753797949)	-	39470	5 (1)	0.11	YES	NO
67	GCN5-related N-acetyltransferase (332339991)	-	19511	6 (1)	0.24	NO	NO
68	Catalase (332339192)	-	54817	8 (2)	0.17	NO	NO
69	Extracellular solute-binding protein family 1 (332340856)	-	48074	4 (1)	0.09	YES	YES
70	Periplasmic binding protein (332339005)	-	34886	2 (1)	0.13	YES	NO
71	Extracellular solute-binding protein family 1 (332340270)	-	45120	4 (1)	0.1	YES	NO
72	Cysteine synthase A (332338971)	-	32249	2 (1)	0.14	NO	NO
73	Extracellular solute-binding protein family 1 (332339297)	-	63225	7 (1)	0.07	YES	YES
74	L-arabinose isomerase (332338427)	-	55111	8 (1)	0.08	NO	NO
75	Alpha/beta hydrolase fold protein (332338325)	-	30206	2 (1)	0.15	NO	NO
76	3-isopropylmalate dehydrogenase (332340239)	-	37108	2 (1)	0.12	NO	NO
77	Hypothetical protein Celf_3380 (332340910)	-	12743	1(1)	0.38	NO	YES
78	Peptidoglycan glycosyltransferase (332340923)	GT51	86700	16 (1)	0.05	NO	YES

79	Amidohydrolase (332338891)	-	44181	5 (1)	0.1	NO	NO
80	Peptidase M24 (332338709)	-	55886	8 (2)	0.16	NO	NO
81	Hypothetical protein Celf_2499 (332340042)	-	83764	4 (1)	0.05	YES	YES
82	Carbon starvation protein CstA (332337945)	-	79679	11 (1)	0.06	NO	YES
83	NAD-dependent epimerase/dehydratase (332340040)	-	33174	5 (1)	0.14	NO	NO
84	Hypothetical protein Celf_2000 (332339547)	-	62014	9 (1)	0.07	NO	YES
85	Regulatory protein TetR (332337908)	-	24168	12 (1)	0.19	NO	NO
86	Basic membrane lipoprotein (332340404)	-	38019	1 (1)	0.12	YES	YES
87	Protein of unknown function DUF885 (332340382)	-	62265	5 (1)	0.07	NO	NO
88	Flagellar M-ring protein FliF (332338228)	-	54876	2 (1)	0.08	NO	NO
89	Guanosine pentaphosphate synthetase I/polyribonucleotide nucleotidyltransferase (332339091)	-	79580	8 (1)	0.06	NO	NO
90	Hypothetical protein Celf_0713 (332338270)	-	16730	3 (1)	0.28	NO	YES
91	Cellulose-binding family II (32339460)	GH74, CBM2	95879	10 (1)	0.05	YES	YES

Table S8: secreted proteins collected after 48 h of *C. fimi* growth on carboxymethyl-cellulose.

	Protein Identification [common name] (gi NCBI)	CAZy classification	MW (kDa)	Peptide sequences (significant sequences)	emPAI	SignalP	SecretomeP
1	Flagellin domain protein (332338221)	-	39886	28 (23)	146.61	NO	YES
2	Endoglucanase C [CenC] (121819)	GH9 [Cel9B]	115658	30 (27)	2.91	YES	YES
3	Exoglucanase A [CbhA] (1708083)	GH6 [Cel6B]	89588	28 (24)	4.81	YES	YES
4	Exoglucanase B [CbhB] (1708084)	GH48 [Cel48A]	115101	27 (19)	1.62	YES	YES
5	Glycoside hydrolase family 5 [Cen D] (332339471)	GH5 [Cel5A]	79059	22 (17)	2.27	YES	YES
6	Peptidase S8 and S53 subtilisin kexin sedolisin (332338096)	-	138613	26 (25)	1.37	YES	YES
7	1, 4-beta cellobiohydrolase [CenA] (332340715)	GH6 [Cel6A]	47018	15 (14)	4.06	YES	YES
8	Exoglucanase [Cex] (327179208)	GH10 [Xyn10A]	54499	25 (18)	5.66	YES	YES
9	Alpha-1,6-glucosidase, pullulanase-type (332338678)	GH13	209811	63 (49)	1.77	YES	YES

10	Glycoside hydrolase family 9 [CenB] (332337588)	GH9 [Cel9A]	109266	38 (27)	2.1	YES	YES
11	Man26A (5359710)	GH26 [Man26A]	107064	25 (19)	1.22	YES	YES
12	Aminopeptidase Y (332337698)	-	53985	16 (12)	2.52	YES	YES
13	Peptidase S1 and S6 chymotrypsin/Hap (332338679)	-	49521	9 (5)	0.67	YES	YES
14	TAP domain protein (332337857)	-	63142	19 (15)	2.84	YES	YES
15	Pectate lyase/Amb allergen (332341298)	PL1	52912	26 (18)	4.39	YES	YES
16	PKD domain containing protein (332339822)	-	87063	21 (13)	0.89	YES	YES
17	Glycoside hydrolase family 3 domain protein (332338140)	GH3	178143	35 (27)	0.96	YES	YES
18	Cellulose-binding family II (332339303)	CE2	52461	12 (10)	1.43	YES	YES
19	Chitin-binding domain 3 protein (332337832)	AA10	37478	6 (4)	0.76	YES	YES
20	Endo-1,4-beta-xylanase (332337655)	GH10	49754	16 (11)	1.78	YES	YES
21	Endo-1,4-beta-xylanase (332340687)	GH10	53458	14 (8)	0.89	YES	YES
22	1, 4-beta cellobiohydrolase (332338782)	GH6	46631	10 (6)	0.89	YES	YES
23	5’-Nucleotidase domain-containing protein (332339202)	-	167817	19 (11)	0.32	YES	YES
24	Endo-1,4-beta-xylanase	GH10, CE4	141693	18 (14)	0.52	YES	YES

	(332338131)						
25	Glycoside hydrolase family 11 [XynD] (332337936)	GH11, CE4 [Xyn11A]	66855	7 (4)	0.47	YES	YES
26	Enolase (332340559)	-	45287	12 (11)	1.8	NO	NO
27	Extracellular ligand-binding receptor (332339377)		43218	6 (5)	0.99	YES	NO
28	Extracellular solute-binding protein family 1 (332338842)	-	46266	10 (7)	1.08	YES	YES
29	Extracellular solute-binding protein family 5 (332340445)	-	67393	13 (11)	1.27	YES	YES
30	Glycoside hydrolase family 3 domain protein (332339598)	GH3	96753	18 (15)	0.94	YES	NO
31	Cellulase (332337938)	GH5, CBM46	61605	18 (11)	1.14	YES	NO
32	Endo-1,3(4)-beta-glucanase (332341296)	GH81	116431	15 (6)	0.25	NO	YES
33	Glycoside hydrolase family 18 (332340692)	GH18	58146	13 (6)	0.67	YES	YES
34	Periplasmic binding protein/LacI transcriptional regulator (332341019)	-	40664	8 (7)	1.07	YES	YES
35	Cellulose-binding family II (32339460)	GH74, CBM2	95879	20 (13)	0.78	YES	YES
36	Hypothetical protein (753797949)	-	39470	7 (5)	0.71	YES	NO
37	Alkaline phosphatase (332340091)	-	44035	10 (6)	0.96	NO	NO
38	Aminopeptidase N	-	94435	21 (10)	0.57	NO	NO

	(332338951)						
39	Fibronectin type III domain protein (332339883)	-	210933	27 (5)	0.11	NO	YES
40	Pectate lyase (332337958)	PL3	44887	15 (5)	0.76	YES	YES
41	Stress protein (332337771)	-	20150	5 (4)	1.29	NO	NO
42	Extracellular solute-binding protein family 3 (332339124)	-	30041	1 (1)	0.15	YES	YES
43	Dihydrolipoamide dehydrogenase (332339755)	-	48441	11 (8)	1.02	NO	NO
44	Glutamine synthetase, type I (332339768)	-	53327	14 (7)	0.75	NO	NO
45	Extracellular solute-binding protein family 5 (332339390)	-	60117	6 (6)	0.53	YES	YES
46	Extracellular solute-binding protein family 1 (332340306)	-	47972	8 (4)	0.42	YES	YES
47	Alpha-N-arabinofuranosidase (332340686)	GH62	53079	6 (2)	0.17	YES	YES
48	Glucose-6-phosphate isomerase (332338893)	-	60314	13 (4)	0.33	NO	NO
49	Flagellar hook-basal body protein (332338236)	-	39855	9 (5)	0.7	NO	YES
50	Thimet oligopeptidase (332337920)	-	70844	11 (3)	0.2	NO	NO
51	PKD domain containing protein (332340573)	-	120006	15 (4)	0.15	YES	YES
52	Stress protein (332337770)	-	20550	5 (2)	0.5	NO	NO
53	Xylose isomerase (332338137)	-	43425	4 (3)	0.34	NO	NO

54	Phosphate ABC transporter, periplasmic phosphate-binding protein (332338148)	-	37419	4 (3)	0.4	YES	NO
55	Extracellular solute-binding protein family 1 (332339297)	-	63225	9 (2)	0.14	YES	YES
56	Extracellular repeat protein, HAF family (332340974)	-	40811	9 (2)	0.23	YES	NO
57	Transaldolase (332339419)	-	39758	7 (2)	0.24	NO	NO
58	Hypothetical protein Celf_0769 (332338326)	-	36708	3 (2)	0.26	YES	YES
59	Peptidase M24 (332338709)	-	55886	10 (2)	0.16	NO	NO
60	Neprilysin (332338954)	-	72498	8 (2)	0.12	NO	NO
61	Periplasmic binding protein (332341193)	-	34098	1 (1)	0.13	YES	NO
62	TAP domain protein (332338042)	-	54198	3 (2)	0.17	YES	NO
63	Cysteine synthase A (332338971)	-	32249	4 (1)	0.14	NO	NO
64	Glycoside hydrolase family 9 (332339254)	GH9	115872	9 (1)	0.04	YES	YES
65	Ribosomal protein L5 (332338567)	-	21150	6 (1)	0.22	NO	NO
66	Hypothetical protein Celf_1998 (332339545)	-	54461	6 (1)	0.08	NO	YES
67	Hypothetical protein Celf_2704 (332340245)	-	24970	3 (2)	0.4	NO	NO
68	Cell surface receptor IPT/TIG domain protein	-	99332	5 (3)	0.14	YES	YES

	(332341050)						
69	Glucan endo-1,3-beta-D-glucosidase (332340860)	GH64	55696	6 (1)	0.08	YES	YES
70	Basic membrane lipoprotein (332340404)	-	38019	3 (1)	0.12	YES	YES
71	Hypothetical protein Celf_2703 (332340244)	-	19277	1 (1)	0.24	YES	NO
72	Cobalamin synthesis CobW domain protein (332340651)	-	39903	5 (1)	0.11	NO	NO
73	Alpha/beta hydrolase fold protein (332338325)	-	30206	3 (1)	0.15	NO	NO
74	Extracellular solute-binding protein family 1 (332340856)	-	48074	5 (1)	0.09	YES	YES
75	Transcriptional regulator, LacI family (332337650)	-	36317	6 (1)	0.12	NO	NO
76	Transcriptional regulator, LuxR family (332338067)	-	103247	8 (1)	0.04	NO	NO
77	UvrD/REP helicase (332338723)	-	73474	6 (1)	0.06	NO	NO
78	Catalase (332339192)	-	54817	6 (1)	0.08	NO	NO
79	Protein of unknown function DUF1212 (332341097)	-	44760	6 (1)	0.1	NO	YES
80	Hypothetical protein Celf_3380 (332340910)	-	12743	1(1)	0.38	NO	YES
81	NLPA lipoprotein (332338762)	-	33494	9 (1)	0.13	YES	YES
82	Extracellular solute-binding protein family 1 (332340270)	-	45120	4 (1)	0.1	YES	NO
83	Alpha-L-arabinofuranosidase domain protein	GH51	55481	6 (1)	0.08	NO	NO

	(332340852)						
84	Anthranilate synthase component I (332339348)	-	55878	7 (1)	0.08	NO	NO

Table S9: STRING legend.

Celf_0019	Glycoside hydrolase family 9 (1045 aa)
Celf_0045	Glycoside hydrolase family 9 (912 aa)
Celf_0088	Endo-1,4-beta-xylanase (464 aa)
Celf_0090	Hypothetical protein (179 aa)
Celf_0132	Aminopeptidase Y (517 aa)
Celf_0208	Stress protein (191 aa)
Celf_0270	Chitin-binding domain 3 protein (355 aa)
Celf_0295	TAP domain-containing protein (598 aa)
Celf_0346	Regulatory protein TetR (224 aa)
Celf_0358	Thimet oligopeptidase (639 aa)
Celf_0374	Glycoside hydrolase family 11 (645 aa)
Celf_0398	Pectate lyase (423 aa)
Celf_0404	Cellulose-binding family II protein (433 aa)
Celf_0438	Ricin B lectin (561 aa)
Celf_0539	Peptidase S8 and S53 subtilisin kexin sedolisin (1383 aa)
Celf_0574	Endo-1,4-beta-xylanase (1350 aa)
xylA	Xylose isomerase (395 aa)
Celf_0583	Glycoside hydrolase family 3 domain-containing protein (1745 aa)
Celf_0591	Phosphate ABC transporter, periplasmic phosphate-binding protein; Part of the ABC transporter complex PstSACB involved in phosphate import (370 aa)
Celf_0664	Flagellin domain-containing protein; Flagellin is the subunit protein which polymerizes to form the filaments of bacterial flagella (391 aa)
Celf_0713	Hypothetical protein (162 aa)
Celf_0862	Mannan endo-1,4-beta-mannosidase (1010 aa)
Celf_1126	Pullulanase-type alpha-1,6-glucosidase (1973 aa)

	Celf_1127	Peptidase S1 and S6 chymotrypsin/Hap (492 aa)
	glyA	Glycine hydroxymethyltransferase; Catalyzes the reversible interconversion of serine and glycine with tetrahydrofolate (THF) serving as the one-carbon carrier. This reaction serves as the major source of one-carbon groups required for the biosynthesis of purines, thymidylate, methionine, and other important biomolecules. Also exhibits THF- independent aldolase activity toward beta-hydroxyamino acids, producing glycine and aldehydes, via a retro-aldol mechanism (429 aa)
	Celf_1157	Peptidase M24 (511 aa)
	Celf_1230	1, 4-beta cellobiohydrolase (433 aa)
	Celf_1271	Glycoside hydrolase family 10 (510 aa)
	Celf_1290	Family 1 extracellular solute-binding protein (434 aa)
	Celf_1399	Aminopeptidase N (866 aa)
	Celf_1402	Neprilysin (661 aa)
	Celf_1453	Periplasmic binding protein (336 aa)
	cenC	Glycoside hydrolase family 9; The biological conversion of cellulose to glucose generally requires three types of hydrolytic enzymes- (1) Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2) Exocellobiohydrolases that cut the dissaccharide cellobiose from the non-reducing end of the cellulose polymer chain; (3) Beta-1,4-glucosidases which hydrolyze the cellobiose and other short cello- oligosaccharides to glucose (1101 aa)
	Celf_1573	Family 3 extracellular solute-binding protein (288 aa)
	Celf_1653	5'-Nucleotidase domain-containing protein (1651 aa)
	Celf_1686	Winged helix family transcriptional regulator (1133 aa)
	Celf_1748	Family 1 extracellular solute-binding protein (570 aa)
	Celf_1754	Cellulose-binding family II protein (498 aa)
	Celf_1830	Extracellular ligand-binding receptor (422 aa)
	Celf_1843	Family 5 extracellular solute-binding protein (555 aa)
	Celf_1864	Glyceraldehyde-3-phosphate dehydrogenase, type I (334 aa)
	Celf_1913	Cellulose-binding family II protein (918 aa)
	Celf_1924	Glycoside hydrolase family 5 (747 aa)

	cbhA	1, 4-beta cellobiohydrolase; This enzyme hydrolyzes 1,4-beta-D-glucosidic linkages of cellulose. Weak activity against carboxymethylcellulose, bacterial microcrystalline cellulose and barley beta-glucan. Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds with inversion of anomeric configuration (872 aa)
	Celf_1998	Hypothetical protein (512 aa)
	Celf_2053	Glycoside hydrolase family 3 domain-containing protein (915 aa)
	Celf_2168	Lytic transglycosylase catalytic subunit (389 aa)
	Celf_2210	Dihydrolipoamide dehydrogenase (461 aa)
	Celf_2223	Glutamine synthetase, type I (474 aa)
	Celf_2249	Hypothetical protein (145 aa)
	Celf_2278	PKD domain-containing protein (836 aa)
	Celf_2339	Fibronectin type III domain-containing protein (2036 aa)
	Celf_2377	Curculin domain-containing protein (mannose-binding) lectin (526 aa)
	Celf_2549	Alkaline phosphatase (425 aa)
	Celf_2704	Hypothetical protein (231 aa)
	Celf_2725	Dihydroxy-acid dehydratase (620 aa)
	Celf_2729	Family 1 extracellular solute-binding protein (431 aa)
	Celf_2790	Family 1 extracellular solute-binding protein (552 aa)
	Celf_2906	Family 5 extracellular solute-binding protein (621 aa)
	eno	Enolase; Catalyzes the reversible conversion of 2- phosphoglycerate into phosphoenolpyruvate. It is essential for the degradation of carbohydrates via glycolysis (427 aa)
	Celf_3120	Cobalamin synthesis CobW domain-containing protein (366 aa)
	Celf_3155	alpha-N-arabinofuranosidase (500 aa)
	Celf_3156	Endo-1,4-beta-xylanase (493 aa)
	Celf_3161	Glycoside hydrolase family 18 (552 aa)
	Celf_3184	1, 4-beta cellobiohydrolase (449 aa)
	Celf_3272	Family 1 extracellular solute-binding protein (554 aa)

	Celf_3330	Glucan endo-1,3-beta-D-glucosidase (539 aa)
	cbhB	Glycoside hydrolase family 48; Hydrolyzes cellohexaose to a mixture of cellotetraose, cellotriose and cellobiose, with only a trace of glucose. It hydrolyzed cellopentaose to cellotriose and cellobiose, and cellotetraose to cellobiose, but it did not hydrolyze cellotriose. Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds with inversion of anomeric configuration (1090 aa)
	Celf_3491	Periplasmic binding protein/LacI transcriptional regulator (385 aa)
	Celf_3512	Hypothetical protein (211 aa)
	Celf_3522	Cell surface receptor IPT/TIG domain-containing protein (1006 aa)
	Celf_3669	Periplasmic binding protein (332 aa)
	sbcC	Putative exonuclease; SbcCD cleaves DNA hairpin structures. These structures can inhibit DNA replication and are intermediates in certain DNA recombination reactions. The complex acts as a 3'->5' double strand exonuclease that can open hairpins. It also has a 5' single-strand endonuclease activity (1044 aa)
	Celf_3773	Endo-1,3(4)-beta-glucanase (1118 aa)
	Celf_3775	Pectate lyase/Amb allergen (522 aa)

Appendix B

Table S1a: Fungal community - list of all Phyla identified in soil and decaying wood samples. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	TOT%	*soil*		*decaying wood*	
			S1	S3	DW2	DW4
Fungi	Ascomycota	37.10	12.70	7.63	6.68	10.09
Fungi	Basidiomycota	36.56	7.80	10.39	12.75	5.63
Fungi	Zygomycota	9.52	2.15	6.01	0.15	1.22
Fungi	**Rozellomycota**	8.01	0.37	0.02	0.03	7.60
Fungi	Chytridiomycota	1.83	0.75	0.29	0.68	0.10
Fungi	Glomeromycota	0.34	0.06	0.29	0.00	0.00
Fungi	unidentified	6.54	1.15	0.35	4.70	0.34

Table S1b: Fungal community – Class level. In table are reported the organisms with a total frequency greater than 1 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	TOT%	*soil*		*decaying wood*	
				S1	S3	DW2	DW4
Fungi	Basidiomycota	**Agaricomycetes**	21.18	0.89	2.05	12.71	5.53
Fungi	Basidiomycota	Tremellomycetes	14.88	6.52	8.29	0.02	0.05
Fungi	Ascomycota	**Dothideomycetes**	13.78	5.52	1.33	1.93	5.01
Fungi	Zygomycota	Mortierellomycotina cls i. s.	9.31	2.05	5.90	0.15	1.21
Fungi	Ascomycota	**Leotiomycetes**	8.44	0.56	3.26	4.06	0.56
Fungi	Ascomycota	Eurotiomycetes	3.04	0.51	0.85	0.11	1.58
Fungi	Ascomycota	**Saccharomycetes**	2.95	0.01	0.01	0.37	2.56
Fungi	Ascomycota	Sordariomycetes	2.92	1.70	0.96	0.01	0.24
Fungi	Ascomycota	Pezizomycetes	2.48	2.43	0.05	0.00	0.00
Fungi	Other	Other	7.45	1.79	0.80	4.57	0.29
Fungi	Unidentified	Unidentified	14.33	3.20	1.50	1.61	8.02

Table S1c: Fungal community – Order level. In table are reported the organisms with a total frequency greater than 1 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	order	TOT%	*soil*		*decaying wood*	
					S1	S3	DW2	DW4
Fungi	Zygomycota	Mortierellomycotina cls i. s.	Mortierellales	9.31	2.05	5.90	0.15	1.21
Fungi	Basidiomycota	Agaricomycetes	**Cantharellales**	8.97	0.03	0.14	8.46	0.35
Fungi	Basidiomycota	Agaricomycetes	**Sebacinales**	8.66	0.08	0.02	3.44	5.11
Fungi	Basidiomycota	Tremellomycetes	Tremellales	7.26	2.40	4.83	0.02	0.01
Fungi	Ascomycota	Leotiomycetes	Helotiales	7.15	0.23	2.39	4.04	0.49
Fungi	Ascomycota	Dothideomycetes	Pleosporales	6.36	3.80	1.14	1.12	0.30
Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	4.79	1.63	3.16	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	**Dothideomycetidae ord i. s.**	4.55	0.00	0.00	0.00	4.55
Fungi	Ascomycota	Saccharomycetes	**Saccharomycetales**	2.95	0.01	0.01	0.37	2.56
Fungi	Ascomycota	Pezizomycetes	Pezizales	2.48	2.43	0.05	0.00	0.00
Fungi	Ascomycota	Eurotiomycetes	**Eurotiales**	2.03	0.20	0.25	0.05	1.53
Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	1.53	1.44	0.09	0.00	0.00
Fungi	Ascomycota	Sordariomycetes	Sordariales	1.53	0.74	0.79	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Capnodiales	1.44	0.46	0.03	0.80	0.15
Fungi	Basidiomycota	Agaricomycetes	Trechisporales	1.26	0.00	1.26	0.00	0.00

Fungi	Ascomycota	Leotiomycetes	Leotiomycetes ord i. s.	1.26	0.32	0.86	0.02	0.06
Fungi	Basidiomycota	Agaricomycetes	Agaricales	1.09	0.67	0.28	0.09	0.05
Fungi	Other	Other	Other	12.64	4.48	2.21	5.35	0.60
Fungi	Unidentified	Unidentified	Unidentified	15.89	4.55	1.64	1.65	8.05

Table S1d: Fungal community – Family level. In table are reported the organisms with a total frequency greater than 0.5 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	order	family	TOT%	*soil*		*decaying wood*	
						S1	S3	DW2	DW4
Fungi	Basidiomycota	Agaricomycetes	Cantharellales	**Hydnaceae**	8.81	0.00	0.01	8.45	0.35
Fungi	Basidiomycota	Tremellomycetes	Tremellales	Tremellales fam I. s.	7.13	2.40	4.70	0.02	0.01
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	6.81	1.43	4.03	0.15	1.21
Fungi	Ascomycota	Dothideomycetes	Dothideomycetidae ord I. s.	**Eremomycetaceae**	4.55	0.00	0.00	0.00	4.55
Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	Piskurozymaceae	4.02	1.56	2.46	0.00	0.00
Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	**Trichomonascaceae**	2.76	0.00	0.00	0.30	2.46
Fungi	Ascomycota	Dothideomycetes	Pleosporales	**Sporormiaceae**	2.08	0.88	0.06	1.11	0.04
Fungi	Ascomycota	Eurotiomycetes	Eurotiales	**Trichocomaceae**	2.01	0.19	0.25	0.05	1.53
Fungi	Ascomycota	Pezizomycetes	Pezizales	Pyronemataceae	1.82	1.79	0.03	0.00	0.00
Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	1.53	1.44	0.09	0.00	0.00
Fungi	Ascomycota	Sordariomycetes	Sordariales	Chaetomiaceae	1.43	0.65	0.78	0.00	0.00
Fungi	Ascomycota	Leotiomycetes	Helotiales	Helotiaceae	1.20	0.03	1.18	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Capnodiales	**Davidiellaceae**	1.16	0.29	0.03	0.78	0.07
Fungi	Ascomycota	Leotiomycetes	Helotiales	Helotiales fam I. s.	1.16	0.16	0.29	0.29	0.41
Fungi	Ascomycota	Leotiomycetes	Leotiomycetes ord I. s.	Leotiomycetes fam I. s.	0.91	0.04	0.80	0.02	0.06

Fungi	Ascomycota	Dothideomycetes	Dothideomycetes ord I. s.	Dothideomycetes fam I. s.	0.87	0.87	0.00	0.00	0.00
Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	Filobasidiaceae	0.77	0.07	0.70	0.00	0.00
Fungi	Basidiomycota	Agaricomycetes	Polyporales	**Polyporaceae**	0.74	0.03	0.00	0.71	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	0.73	0.14	0.59	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporales fam I. s.	0.71	0.66	0.05	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Melanommataceae	0.70	0.70	0.00	0.00	0.00
Fungi	Ascomycota	Pezizomycetes	Pezizales	Pezizaceae	0.62	0.62	0.00	0.00	0.00
Fungi	Ascomycota	Sordariomycetes	Xylariales	Xylariales fam I. s.	0.58	0.54	0.03	0.00	0.00
Fungi	Other	Other	Other	Other	0.50	0.16	0.25	0.06	0.03
Fungi	Unidentified	Unidentified	Unidentified	Unidentified	9.54	3.82	2.75	1.79	1.19

Table S1e: Fungal community – Genus level. In table are reported the organisms with a total frequency greater than 0.2 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	order	family	genus	TOT%	*soil*		*decaying wood*	
							S1	S3	DW2	DW4
Fungi	Basidiomycota	Tremellomycetes	Tremellales	Tremellales fam I. s.	Cryptococcus	6.95	2.28	4.66	0.00	0.01
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	6.81	1.43	4.03	0.15	1.21
Fungi	Ascomycota	Dothideomycetes	Dothideomycetidae ord I. s.	Eremomycetaceae	**Arthrographis**	4.55	0.00	0.00	0.00	4.55
Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	Piskurozymaceae	Solicoccozyma	4.02	1.56	2.46	0.00	0.00
Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Trichomonascaceae	**Blastobotrys**	2.57	0.00	0.00	0.11	2.46
Fungi	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	**Aspergillus**	1.53	0.09	0.00	0.02	1.42
Fungi	Ascomycota	Sordariomycetes	Sordariales	Chaetomiaceae	Trichocladium	1.33	0.55	0.77	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	**Cladosporium**	1.16	0.29	0.03	0.78	0.07
Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Guehomyces	0.96	0.96	0.00	0.00	0.00
Fungi	Ascomycota	Pezizomycetes	Pezizales	Pyronemataceae	Pseudaleuria	0.96	0.96	0.00	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	Preussia	0.85	0.79	0.06	0.00	0.00
Fungi	Ascomycota	Leotiomycetes	Leotiomycetes ord I. s.	Leotiomycetes fam I. s.	Leohumicola	0.85	0.00	0.79	0.00	0.06
Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	Filobasidiaceae	Goffeauzyma	0.77	0.07	0.70	0.00	0.00
Fungi	Basidiomycota	Agaricomycetes	Polyporales	Polyporaceae	**Perenniporia**	0.74	0.03	0.00	0.71	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Melanommataceae	Karstenula	0.70	0.70	0.00	0.00	0.00

Fungi	Ascomycota	Leotiomycetes	Helotiales	Helotiales fam I. s.	**Scytalidium**	0.67	0.02	0.06	0.20	0.40
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporales fam I. s.	Didymella	0.60	0.60	0.00	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	Drechslera	0.59	0.00	0.59	0.00	0.00
Fungi	Ascomycota	Sordariomycetes	Xylariales	Xylariales fam I. s.	Microdochium	0.57	0.54	0.03	0.00	0.00
Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Cystofilobasidium	0.52	0.43	0.09	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Dothideomycetes ord I. s.	Dothideomycetes fam I. s.	Monodictys	0.51	0.51	0.00	0.00	0.00
Fungi	Basidiomycota	Agaricomycetes	Agaricales	Pluteaceae	Pluteus	0.49	0.48	0.00	0.01	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Cucurbitariaceae	Pyrenochaetopsis	0.40	0.36	0.04	0.00	0.00
Fungi	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	Penicillium	0.39	0.08	0.20	0.02	0.09
Fungi	Ascomycota	Dothideomycetes	Dothideomycetes ord I. s.	Dothideomycetes fam I. s.	Minutisphaera	0.36	0.36	0.00	0.00	0.00
Fungi	Ascomycota	Leotiomycetes	Leotiomycetes ord I. s.	Myxotrichaceae	Pseudogymnoascus	0.34	0.28	0.06	0.00	0.00
Fungi	Basidiomycota	Entorrhizomycetes	Entorrhizales	Entorrhizaceae	Entorrhiza	0.31	0.31	0.00	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Lophiostomataceae	**Lophiostoma**	0.28	0.00	0.00	0.00	0.27
Fungi	Ascomycota	Pezizomycotina cls I. s.	Pezizomycotina ord I. s.	Pezizomycotina fam I. s.	Slimacomyces	0.26	0.26	0.00	0.00	0.00
Fungi	Chytridiomycota	Chytridiomycetes	Rhizophlyctidales	Rhizophlyctidaceae	Rhizophlyctis	0.23	0.20	0.00	0.03	0.01
Fungi	Chytridiomycota	Chytridiomycetes	Rhizophydiales	Rhizophydiaceae	Rhizophydium	0.23	0.02	0.21	0.00	0.00
Fungi	Ascomycota	Leotiomycetes	Helotiales	Helotiales fam I. s.	Tetracladium	0.22	0.14	0.08	0.00	0.00
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Exophiala	0.21	0.04	0.16	0.00	0.00
Fungi	Other	Other	Other	Other	Other	10.41	3.11	2.05	4.52	0.73

Fungi	Unidentified	Unidentified	Unidentified	Unidentified	Unidentified	48.59	8.32	8.15	18.43	13.69

Table S1f: Fungal community – Species level. In table are reported the organisms with a total frequency greater than 0.1 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	order	family	genus	species	TOT%	*soil*		*decaying wood*	
								S1	S3	DW2	DW4
Fungi	Basidiomycota	Tremellomycetes	Tremellales	Tremellales fam I. s.	Cryptococcus	*Cryptococcus podzolicus*	5.83	1.43	4.40	0.00	0.01
Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	Piskurozymaceae	Solicoccozyma	*Solicoccozyma terricola*	3.91	1.49	2.41	0.00	0.00
Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Trichomonascaceae	Blastobotrys	***Blastobotrys mokoenaii***	2.57	0.00	0.00	0.11	2.46
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	*Mortierella rishikesha*	1.90	0.38	1.52	0.00	0.00
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	*Mortierella minutissima*	1.48	0.01	1.46	0.00	0.00
Fungi	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	Aspergillus	***Aspergillus subversicolor***	1.39	0.02	0.00	0.01	1.36
Fungi	Ascomycota	Sordariomycetes	Sordariales	Chaetomiaceae	Trichocladium	*Trichocladium opacum*	1.31	0.54	0.77	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Cladosporium	***Cladosporium delicatulum***	1.16	0.29	0.03	0.77	0.07
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	***Mortierella fimbricystis***	1.12	0.00	0.00	0.00	1.12
Fungi	Basidiomycota	Tremellomycetes	Tremellales	Tremellales fam I. s.	Cryptococcus	*Cryptococcus aerius*	0.98	0.78	0.20	0.00	0.00

Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Guehomyces	*Guehomyces pullulans*	0.96	0.96	0.00	0.00	0.00
Fungi	Ascomycota	Leotiomycetes	Leotiomycetes ord I. s.	Leotiomycetes fam I. s.	Leohumicola	*Leohumicola minima*	0.85	0.00	0.79	0.00	0.06
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	Preussia	*Preussia alloiomera*	0.80	0.75	0.06	0.00	0.00
Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	Filobasidiaceae	Goffeauzyma	*Goffeauzyma gastrica*	0.77	0.07	0.70	0.00	0.00
Fungi	Basidiomycota	Agaricomycetes	Polyporales	Polyporaceae	Perenniporia	***Perenniporia fraxinea***	0.74	0.03	0.00	0.71	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Melanommataceae	Karstenula	*Karstenula rhodostoma*	0.70	0.70	0.00	0.00	0.00
Fungi	Ascomycota	Leotiomycetes	Helotiales	Helotiales fam I. s.	Scytalidium	***Scytalidium lignicola***	0.66	0.01	0.06	0.20	0.40
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporales fam I. s.	Didymella	*Didymella calidophila*	0.60	0.60	0.00	0.00	0.00
Fungi	Ascomycota	Sordariomycetes	Xylariales	Xylariales fam I. s.	Microdochium	*Microdochium bolleyi*	0.51	0.48	0.03	0.00	0.00
Fungi	Basidiomycota	Agaricomycetes	Agaricales	Pluteaceae	Pluteus	*Pluteus hispidulus*	0.46	0.46	0.00	0.00	0.00
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	*Mortierella angusta*	0.41	0.23	0.18	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Cucurbitariaceae	Pyrenochaetopsis	*Pyrenochaetopsis leptospora*	0.38	0.34	0.04	0.00	0.00
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	*Mortierella exigua*	0.37	0.16	0.21	0.00	0.00
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	*Mortierella humilis*	0.36	0.00	0.36	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Dothideomycetes ord I. s.	Dothideomycetes fam I. s.	Minutisphaera	*Minutisphaera japonica*	0.35	0.35	0.00	0.00	0.00

Fungi	Ascomycota	Leotiomycetes	Leotiomycetes ord I. s.	Myxotrichaceae	Pseudogymnoascus	*Pseudogymnoascus roseus*	0.34	0.28	0.06	0.00	0.00
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	***Mortierella hyalina***	0.34	0.16	0.00	0.14	0.04
Fungi	Basidiomycota	Entorrhizomycetes	Entorrhizales	Entorrhizaceae	Entorrhiza	*Entorrhiza aschersoniana*	0.30	0.30	0.00	0.00	0.00
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	*Mortierella alpina*	0.28	0.25	0.03	0.00	0.00
Fungi	Ascomycota	Pezizomycotina cls I. s.	Pezizomycotina ord I. s.	Pezizomycotina fam I. s.	Slimacomyces	*Slimacomyces isiola*	0.26	0.26	0.00	0.00	0.00
Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Cystofilobasidium	*Cystofilobasidium infirmominiatum*	0.26	0.20	0.05	0.00	0.00
Fungi	Chytridiomycota	Chytridiomycetes	Rhizophlyctidales	Rhizophlyctidaceae	Rhizophlyctis	*Rhizophlyctis rosea*	0.23	0.20	0.00	0.03	0.01
Fungi	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	Penicillium	*Penicillium decumbens*	0.21	0.05	0.16	0.00	0.00
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Exophiala	*Exophiala equina*	0.21	0.04	0.16	0.00	0.00
Fungi	Zygomycota	Mortierellomycotina cls I. s.	Mortierellales	Mortierellaceae	Mortierella	*Mortierella polygonia*	0.19	0.16	0.00	0.00	0.04
Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Trichomonascaceae	Sugiyamaella	***Sugiyamaella novakii***	0.19	0.00	0.00	0.19	0.00
Fungi	Ascomycota	Pezizomycotina cls I. s.	Pezizomycotina ord I. s.	Pseudeurotiaceae	Pseudeurotium	*Pseudeurotium hygrophilum*	0.18	0.15	0.04	0.00	0.00
Fungi	Chytridiomycota	Chytridiomycetes	Rhizophydiales	Rhizophydiales fam I. s.	Operculomyces	*Operculomyces laminatus*	0.18	0.08	0.01	0.09	0.00
Fungi	Ascomycota	Sordariomycetes	Microascales	Microascaceae	Scedosporium	***Scedosporium minutisporum***	0.17	0.00	0.00	0.00	0.17

Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Cystofilobasidium	*Cystofilobasidium capitatum*	0.17	0.14	0.04	0.00	0.00
Fungi	Chytridiomycota	Blastocladiomycetes	Blastocladiales	Blastocladiaceae	Blastocladiella	*Blastocladiella britannica*	0.15	0.15	0.00	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Hysteriales	Gloniaceae	Cenococcum	*Cenococcum geophilum*	0.14	0.14	0.00	0.00	0.00
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Mycosphaerella	***Mycosphaerella tassiana***	0.14	0.05	0.00	0.02	0.08
Fungi	Ascomycota	Leotiomycetes	Helotiales	Helotiales fam I. s.	Tetracladium	*Tetracladium marchalianum*	0.14	0.11	0.03	0.00	0.00
Fungi	Basidiomycota	Tremellomycetes	Trichosporonales	Trichosporonaceae	Apiotrichum	*Apiotrichum gracile*	0.14	0.00	0.11	0.00	0.03
Fungi	Basidiomycota	Tremellomycetes	Tremellales	Sirobasidiaceae	Sirobasidium	*Sirobasidium magnum*	0.13	0.00	0.13	0.00	0.00
Fungi	Basidiomycota	Tremellomycetes	Holtermanniales	Holtermanniales fam I. s.	Holtermanniella	*Holtermanniella takashimae*	0.13	0.04	0.09	0.00	0.00
Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	Piskurozymaceae	Solicoccozyma	*Solicoccozyma terrea*	0.11	0.06	0.05	0.00	0.00
Fungi	Ascomycota	Eurotiomycetes	Onygenales	Onygenaceae	Chrysosporium	*Chrysosporium lobatum*	0.10	0.10	0.00	0.00	0.00
Fungi	Basidiomycota	Agaricomycetes	Agaricales	Pleurotaceae	Pleurotus	***Pleurotus dryinus***	0.10	0.00	0.00	0.07	0.03
Fungi	Ascomycota	Leotiomycetes	Helotiales	Helotiales fam i. s.	Leptodontidium	***Leptodontidium trabinellum***	0.10	0.00	0.00	0.09	0.01
Fungi	Other	Other	Other	Other	Other	*Other*	11.84	4.07	2.34	4.55	0.89
Fungi	Unidentified	Unidentified	Unidentified	Unidentified	Unidentified	*Unidentified*	56.84	10.20	9.66	18.46	18.52

Table S2a: Bacterial community - list of all Phyla identified in soil and decaying wood samples. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	TOT%	*soil*		*decaying wood*	
			S1	S3	DW2	DW4
Bacteria	**Proteobacteria**	35.06	6.97	6.92	9.28	11.89
Bacteria	Bacteroidetes	18.46	7.51	0.72	7.51	2.73
Bacteria	Acidobacteria	16.70	2.51	8.98	1.89	3.32
Bacteria	Actinobacteria	5.64	1.49	1.26	1.21	1.67
Bacteria	**Planctomycetes**	2.73	0.37	0.24	1.16	0.96
Bacteria	Verrucomicrobia	2.23	0.42	1.15	0.30	0.36
Bacteria	Chloroflexi	2.05	0.54	0.82	0.21	0.48
Bacteria	Firmicutes	0.61	0.45	0.06	0.04	0.06
Bacteria	Candidatus Saccharibacteria	0.53	0.29	0.13	0.08	0.02
Bacteria	Gemmatimonadetes	0.47	0.22	0.23	0.02	0.00
Bacteria	Nitrospirae	0.39	0.09	0.30	0.00	0.00
Bacteria	Cyanobacteria Chloroplast	0.20	0.16	0.02	0.03	0.00
Bacteria	candidate division WPS-2	0.16	0.04	0.02	0.09	0.01
Bacteria	Armatimonadetes	0.14	0.07	0.07	0.00	0.00
Bacteria	Parcubacteria	0.08	0.04	0.03	0.00	0.00

Bacteria	candidate division WPS-1	0.04	0.00	0.03	0.00	0.00
Bacteria	Latescibacteria	0.03	0.00	0.03	0.00	0.00
Bacteria	Hydrogenedentes	0.03	0.00	0.00	0.02	0.00
Bacteria	Spirochaetes	0.01	0.01	0.00	0.00	0.00
Bacteria	Lentisphaerae	0.01	0.01	0.00	0.00	0.00
Bacteria	Tenericutes	0.01	0.00	0.00	0.01	0.00
Bacteria	unclassified Bacteria	14.33	3.78	3.98	3.10	3.47

Table S2b: Bacterial community – Class level. In table are reported the organisms with a total frequency greater than 1 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	TOT%	*soil*		*decaying wood*	
				S1	S3	DW2	DW4
Bacteria	Proteobacteria	**Alphaproteobacteria**	13.82	2.75	5.01	2.36	3.71
Bacteria	Proteobacteria	**Gammaproteobacteria**	7.23	0.88	0.47	2.18	3.71
Bacteria	Bacteroidetes	Flavobacteriia	5.98	5.42	0.06	0.35	0.16
Bacteria	Proteobacteria	**Deltaproteobacteria**	5.76	0.54	0.26	2.73	2.22
Bacteria	Actinobacteria	Actinobacteria	5.59	1.46	1.26	1.21	1.65
Bacteria	Acidobacteria	Acidobacteria Gp3	4.88	0.26	1.64	1.31	1.67
Bacteria	Bacteroidetes	**Sphingobacteriia**	4.76	0.64	0.31	3.32	0.49
Bacteria	Proteobacteria	Betaproteobacteria	4.73	1.99	0.72	0.93	1.09
Bacteria	Acidobacteria	Acidobacteria Gp1	3.45	0.15	3.30	0.00	0.00
Bacteria	Acidobacteria	Acidobacteria Gp6	2.21	0.77	0.61	0.05	0.79
Bacteria	Acidobacteria	Acidobacteria Gp2	2.01	0.01	1.99	0.00	0.00
Bacteria	Planctomycetes	**Planctomycetia**	1.87	0.33	0.23	0.48	0.83
Bacteria	Bacteroidetes	Bacteroidetes i. s.	1.48	0.49	0.09	0.01	0.89
Bacteria	Acidobacteria	Acidobacteria Gp4	1.33	0.50	0.36	0.45	0.02
Bacteria	Other	Other	9.03	3.13	3.18	1.20	1.53
Bacteria	unclassified	unclassified	25.58	5.60	5.42	8.38	6.18

Table S2c: Bacterial community – Order level. In table are reported the organisms with a total frequency greater than 1 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	order	TOT%	*soil*		*decaying wood*	
					S1	S3	DW2	DW4
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	7.35	1.49	3.38	1.28	1.21
Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	5.98	5.42	0.06	0.35	0.16
Bacteria	Proteobacteria	Deltaproteobacteria	**Myxococcales**	5.07	0.46	0.18	2.39	2.04
Bacteria	Bacteroidetes	Sphingobacteriia	**Sphingobacteriales**	4.76	0.64	0.31	3.32	0.49
Bacteria	Acidobacteria	Acidobacteria Gp3	Gp3	4.03	0.18	1.23	1.11	1.51
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	2.94	1.01	0.86	0.83	0.24
Bacteria	Proteobacteria	Alphaproteobacteria	**Sphingomonadales**	2.33	0.28	0.26	0.24	1.55
Bacteria	Proteobacteria	Betaproteobacteria	**Burkholderiales**	2.32	1.24	0.22	0.43	0.44
Bacteria	Acidobacteria	Acidobacteria Gp6	Gp6	2.21	0.77	0.61	0.05	0.79
Bacteria	Acidobacteria	Acidobacteria Gp1	Gp1	2.21	0.09	2.12	0.00	0.00
Bacteria	Acidobacteria	Acidobacteria Gp2	Gp2	2.01	0.01	1.99	0.00	0.00
Bacteria	Planctomycetes	Planctomycetia	**Planctomycetales**	1.87	0.33	0.23	0.48	0.83
Bacteria	Bacteroidetes	Bacteroidetes i. s.	**Ohtaekwangia**	1.48	0.49	0.09	0.01	0.89
Bacteria	Acidobacteria	Acidobacteria Gp4	Gp4	1.19	0.44	0.35	0.37	0.02
Bacteria	Proteobacteria	Gammaproteobacteria	**Xanthomonadales**	1.04	0.19	0.25	0.48	0.12
Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	1.02	0.31	0.49	0.16	0.05

Bacteria	Other	Other	Other	11.79	3.84	3.96	1.70	2.29
Bacteria	unclassified	unclassified	unclassified	39.95	7.67	8.29	11.73	12.27

Table S2d: Bacterial community – Family level. In table are reported the organisms with a total frequency greater than 0.5 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	order	family	TOT%	*soil*		*decaying wood*	
						S1	S3	DW2	DW4
Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	5.88	5.34	0.06	0.33	0.15
Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	**Chitinophagaceae**	4.35	0.53	0.24	3.19	0.39
Bacteria	Proteobacteria	Deltaproteobacteria	Myxococcales	**Polyangiaceae**	4.23	0.17	0.06	2.18	1.82
Bacteria	Acidobacteria	Acidobacteria Gp3	Gp3	Gp3	4.03	0.18	1.23	1.11	1.51
Bacteria	Acidobacteria	Acidobacteria Gp6	Gp6	Gp6	2.21	0.77	0.61	0.05	0.79
Bacteria	Acidobacteria	Acidobacteria Gp1	Gp1	Gp1	2.21	0.09	2.12	0.00	0.00
Bacteria	Acidobacteria	Acidobacteria Gp2	Gp2	Gp2	2.01	0.01	1.99	0.00	0.00
Bacteria	Planctomycetes	Planctomycetia	Planctomycetales	**Planctomycetaceae**	1.87	0.33	0.23	0.48	0.83
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	1.80	0.31	1.28	0.10	0.11
Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	**Erythrobacteraceae**	1.71	0.11	0.00	0.14	1.46
Bacteria	Bacteroidetes	Bacteroidetes i. s.	Ohtaekwangia	**Ohtaekwangia**	1.48	0.49	0.09	0.01	0.89
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	1.46	0.33	0.62	0.19	0.33
Bacteria	Acidobacteria	Acidobacteria Gp4	Gp4	Gp4	1.19	0.44	0.35	0.37	0.02

Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	0.89	0.79	0.10	0.01	0.00
Bacteria	Acidobacteria	Acidobacteria Gp5	Gp5	Gp5	0.85	0.29	0.25	0.00	0.31
Bacteria	Verrucomicrobia	Subdivision3	Subdivision3 gen. i. s.	Subdivision3 gen. i. s.	0.85	0.07	0.66	0.01	0.11
Bacteria	Verrucomicrobia	Opitutae	Opitutales	**Opitutaceae**	0.76	0.30	0.09	0.26	0.11
Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	**Sinobacteraceae**	0.66	0.12	0.01	0.42	0.11
Bacteria	Acidobacteria	Acidobacteria Gp16	Gp16	Gp16	0.65	0.20	0.04	0.02	0.39
Bacteria	Actinobacteria	Actinobacteria	Acidimicrobiales	**Acidimicrobiaceae**	0.57	0.03	0.00	0.17	0.38
Bacteria	Planctomycetes	Phycisphaerae	Phycisphaerales	**Phycisphaeraceae**	0.54	0.04	0.00	0.37	0.13
Bacteria	Candidatus Saccharibacteria	Saccharibacteria gen. i. s.	Saccharibacteria gen. i. s.	Saccharibacteria gen. i. s.	0.53	0.29	0.13	0.08	0.02
Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	0.53	0.08	0.34	0.09	0.03
Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	0.50	0.16	0.25	0.06	0.03
Bacteria	Other	Other	Other	Other	9.54	3.82	2.75	1.79	1.19
Bacteria	unclassified	unclassified	unclassified	unclassified	47.89	9.47	11.28	13.44	13.70

Table S2d: Bacterial community – Genus level. In table are reported the organisms with a total frequency greater than 0.2 %. The highlighted organisms are prevalent in decaying wood samples.

kingdom	phylum	class	order	family	genus	TOT%	*soil*		*decaying wood*	
							S1	S3	DW2	DW4
Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	5.07	4.70	0.04	0.22	0.11
Bacteria	Acidobacteria	Acidobacteria Gp3	Gp3	Gp3	Gp3	4.03	0.18	1.23	1.11	1.51
Bacteria	Acidobacteria	Acidobacteria Gp6	Gp6	Gp6	Gp6	2.21	0.77	0.61	0.05	0.79
Bacteria	Acidobacteria	Acidobacteria Gp1	Gp1	Gp1	Gp1	2.21	0.09	2.12	0.00	0.00
Bacteria	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	**Terrimonas**	2.06	0.18	0.04	1.57	0.28
Bacteria	Acidobacteria	Acidobacteria Gp2	Gp2	Gp2	Gp2	2.01	0.01	1.99	0.00	0.00
Bacteria	Bacteroidetes	Bacteroidetes i.s.	Ohtaekwangia	Ohtaekwangia	**Ohtaekwangia**	1.48	0.49	0.09	0.01	0.89
Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Erythrobacteraceae	**Porphyrobacter**	1.20	0.04	0.00	0.09	1.07
Bacteria	Acidobacteria	Acidobacteria Gp4	Gp4	Gp4	Gp4	1.19	0.44	0.35	0.37	0.02
Bacteria	Proteobacteria	Deltaproteobacteria	Myxococcales	Polyangiaceae	**Jahnella**	1.07	0.00	0.00	0.42	0.65
Bacteria	Acidobacteria	Acidobacteria Gp5	Gp5	Gp5	Gp5	0.85	0.29	0.25	0.00	0.31
Bacteria	Verrucomicrobia	Subdivision3	Subdivision3 gen. i.s.	Subdivision3 gen. i.s.	Subdivision3 gen. i.s.	0.85	0.07	0.66	0.01	0.11
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Afipia	0.76	0.11	0.53	0.01	0.11
Bacteria	Verrucomicrobia	Opitutae	Opitutales	Opitutaceae	Opitutus	0.72	0.30	0.09	0.26	0.08
Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae	**Steroidobacter**	0.66	0.11	0.01	0.42	0.11

Bacteria	Acidobacteria	Acidobacteria Gp16	Gp16	Gp16	Gp16	0.65	0.20	0.04	0.02	0.39
Bacteria	Actinobacteria	Actinobacteria	Acidimicrobiales	Acidimicrobiaceae	**Ilumatobacter**	0.57	0.03	0.00	0.17	0.38
Bacteria	Planctomycetes	Phycisphaerae	Phycisphaerales	Phycisphaeraceae	**Phycisphaera**	0.54	0.04	0.00	0.37	0.13
Bacteria	Candidatus Saccharibacteria	Saccharibacteria gen. i.s.	Saccharibacteria gen. i.s.	Saccharibacteria gen. i.s.	Saccharibacteria gen. i.s.	0.53	0.29	0.13	0.08	0.02
Bacteria	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas	0.47	0.22	0.23	0.02	0.00
Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira	0.39	0.09	0.30	0.00	0.00
Bacteria	Actinobacteria	Actinobacteria	Gaiellales	Gaiellaceae	Gaiella	0.39	0.19	0.15	0.00	0.05
Bacteria	Verrucomicrobia	Spartobacteria	Spartobacteria gen. i.s.	Spartobacteria gen. i.s.	Spartobacteria gen. i.s.	0.37	0.02	0.32	0.01	0.03
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae	**Thermovum**	0.29	0.02	0.01	0.06	0.20
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bradyrhizobium	0.27	0.05	0.22	0.00	0.00
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Rugamonas	0.22	0.20	0.03	0.00	0.00
Bacteria	Acidobacteria	Acidobacteria Gp7	Gp7	Gp7	Gp7	0.21	0.15	0.06	0.00	0.00
Bacteria	Other	Other	Other	Other	Other	15.22	4.75	2.86	5.46	1.92
Bacteria	unclassified	unclassified	unclassified	unclassified	Unclassified	61.63	12.76	13.56	18.12	16.72

Aknowledgements

"Chi torna da un viaggio non è mai la stessa persona che è partita"

E anche questo viaggio sta volgendo al termine. Come quando si torna a casa da una vacanza, ripenso a questi tre anni passati e quello che scelgo come *souvenir* è la gratitudine, per aver conosciuto e a volte anche collaborato con tante persone che, in bene o in male, mi hanno trasmesso qualcosa.

Il primo ringraziamento va alla Prof.sa Maria Cavaletto, per avermi offerto questa opportunità di crescita, non solo formativa ma anche personale.

Grazie infinite ai miei colleghi e amici, in particolare Dr. Stefano Spertino, Dr.sa Lara Boatti, Dr.sa Stefania Ribulla e Dr.sa Romina Cannata, per tutti i consigli, gli insegnamenti, le esperienze e soprattutto le risate e l'affetto sincero.

Ultimo, ma non per importanza, GRAZIE a Manuel e a tutta la nostra grande famiglia, che sempre mi ha supportato, credendo in me e appoggiando le mie scelte, con tutta la pazienza e l'amore del mondo.

DECLARATION AND AUTHORISATION TO ANTIPLAGIARISM DETECTION

The undersigned Sara Icardi, student of Chemistry&Biology Ph.D course (XXX Cycle)

declares:

- to be aware that the University has adopted a web-based service to detect plagiarism through a software system called "Turnitin",
- her Ph.D. thesis was submitted to Turnitin scan and reasonably it resulted an original document, which correctly cites the literature

acknowledges:

- her Ph.D. thesis can be verified by her Ph.D. tutor and/or Ph.D Coordinator in order to confirm its originality.

Date: 09/11/2017 Signature: