UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

EDUARDO CRUZ MORAES

Enzyme and metabolic pathways prospection for the biosynthesis of vanilin and valorization of lignin from sugarcane bagasse

Prospecção de enzimas e vias metabólicas para a biossíntese de vanilina e valorização da lignina do bagaço de cana-de-açúcar

CAMPINAS
2017

EDUARDO CRUZ MORAES

Enzyme and metabolic pathways prospection for the biosynthesis of vanilin and valorization of lignin from sugarcane bagasse

Prospecção de enzimas e vias metabólicas para a biossíntese de vanilina e valorização da lignina do bagaço de cana-de-açúcar

Thesis presented to the Institute of Biology of the University of Campinas for partial fulfillment of the requirements for the degree of Doctor in FUNCTIONAL AND MOLECULAR BIOLOGY, in the area of BIOCHEMISTRY.

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em *BIOLOGIA FUNCIONAL E MOLECULAR,* na área de *BIOQUÍMICA*.

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO EDUARDO CRUZ MORAES E ORIENTADA PELO DR. FABIO MARCIO SQUINA.

Orientador: Dr. Fabio Marcio Squina
Co-Orientador: Dra. Thábata Maria Alvarez

CAMPINAS
2017

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2014/26152-4; CNPq, 149186/2014-2; CAPES

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

M791p

Moraes, Eduardo Cruz, 1986-
Prospecção de enzimas e vias metabólicas para a biossíntese de vanilina e valorização da lignina do bagaço de cana-de-açúcar / Eduardo Cruz Moraes. – Campinas, SP : [s.n.], 2017.

Orientador: Fabio Marcio Squina.
Coorientador: Thabata Maria Alvarez.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Lignina. 2. Compostos aromáticos - Biodegradação. 3. Metagenoma. 4. Vanilina. 5. Ácido ferúlico. I. Squina, Fabio Marcio. II. Alvarez, Thabata Maria, 1986-. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Enzyme and metabolic pathways prospection for the biosynthesis of vanilin and valorization of lignin from sugarcane bagasse
Palavras-chave em inglês:
Lignin
Aromatic compounds - Biodegradation
Metagenome
Vanillin
Ferulic acid
Área de concentração: Bioquímica
Titulação: Doutor em Biologia Funcional e Molecular
Banca examinadora:
Fabio Marcio Squina [Orientador]
Fabiano Jares Contesini
Richard John Ward
Wanius José Garcia da Silva
Fernanda Mandelli
Data de defesa: 14-12-2017
Programa de Pós-Graduação: Biologia Funcional e Molecular

Campinas, 14 de dezembro de 2017

COMISSÃO EXAMINADORA

Dr. Fabio Marcio Squina (Orientador)

Dr. Fabiano Jares Contesini

Dr. Richard John Ward

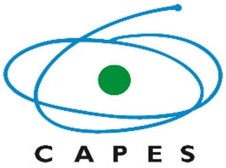
Dr. Wanius José Garcia da Silva

Dra. Fernanda Mandelli

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

AGRADECIMENTOS

Esta tese foi redigida a partir do trabalho realizado no projeto **"Prospecção de Enzimas e Otimização de Vias Metabólicas para Produção de Vanilina a partir de Lignina do Bagaço de Cana-de-Açúcar"**, sob orientação do Dr. Fabio Squina e da Dr.ª Thabata Maria Alvarez, financiado pelo **Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)** (processo nº 149186/2014-2, vigência de 01/08/2014 a 31/03/2015) e pela **Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)** (processo nº 2014/26152-4, vigência de 01/04/2015 a 31/07/2017), no âmbito do convênio desta com a **Coordenação de Aperfeiçoamento de Pessoal de Nível superior (CAPES)**. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do(s) autor(es) e não necessariamente refletem a visão do CNPq, FAPESP ou da CAPES.



RESUMO

Lignina, um dos materiais mais recalcitrantes da natureza, consiste em um polímero complexo composto de vários anéis aromáticos ligados por ligações do tipo éter, estando presente em combinação com celulose e hemicelulose na parede celular vegetal. A lignina se encontra fortemente associada a polissacarídeos da parede celular, o que dificulta seu aproveitamento e praticamente inviabiliza seu isolamento na forma não alterada. Vanilina é um dos produtos da oxidação de lignina, e é obtida na forma natural a partir da vagem de orquídeas tropicais do gênero *Vanilla*. Devido à sua baixa produtividade e à dificuldade de cultivo fora de seu habitat natural, a demanda mundial por este produto vem sendo atendida através de síntese química a partir de recursos não renováveis como guaicol e ácido glioxílico. Recentemente, rotas fermentativas naturais têm sido desenvolvidas para a produção de vanilina a partir de fontes renováveis.

Dada a importância comercial da molécula de vanilina, associada à demanda pela substituição de matérias primas não renováveis por matérias primas renováveis de baixo custo, é proposto neste trabalho o estabelecimento e caracterização de um consórcio de micro-organismos degradadores de lignina como estratégia para a valorização deste subproduto. Este consórcio foi criado através do cultivo de micro-organismos oriundos de solo de canavial em meio de cultura contendo uma solução de compostos provenientes de lignina. Esta solução é obtida através do processo de delignificação do bagaço de cana-de-açúcar, e é composta majoritariamente de compostos fenólicos de baixo peso molecular. Com o sequenciamento metagenômico do consórcio, foi possível ter acesso às enzimas e vias metabólicas envolvidas em processos de bioconversão destes compostos fenólicos. Como método para validação desta estratégia para a prospecção de enzimas lignolíticas, foram selecionadas as enzimas feruloil-CoA sintetase (fcs) e enoil-CoA hidratase/aldolase (ech), seguido de ensaios enzimáticos com estas enzimas para avaliar a capacidade de conversão de ácido ferúlico em vanilina.

A caracterização do consórcio revelou a capacidade dos micro-organismos presentes de degradar compostos fenólicos de baixo peso molecular do meio de cultura. Uma subsequente análise taxonômica indicou o enriquecimento do consórcio, com a diminuição da diversidade de micro-organismos em relação à amostra original de solo, e o aumento de

espécies descritas na literatura como relacionadas à degradação de compostos aromáticos. Desta maneira, pode se observar a especialização do consórcio através da seleção de micro-organismos aptos a degradarem a lignina presente no meio.

A análise do metagenoma do consórcio também permitiu a identificação de enzimas que participam na degradação de lignina e compostos aromáticos, tais como lacases, peroxidases e outras enzimas envolvidas no metabolismo de hidrocarbonetos policíclicos aromáticos (HPAs). A partir dos dados de sequenciamento metagenômico, realizou-se a seleção *in silico* por análise de similaridade de sequência três enzimas feruloil-CoA sintetase (fcs) e duas enoil-Coa hidratase/aldolase (ech). Através de análises de eletroforese capilar e espectrometria de massas, verificou-se a produção de vanilina a partir de ácido ferúlico em ensaios enzimáticos com as enzimas selecionadas.

Em paralelo, foram isolados micro-organismos capazes de crescer de forma autônoma em meios contendo lignina e com possível atividade lignolítica. Em especial, a cepa isolada de *Paenarthrobacter* sp. (denominada de HW13) se mostrou capaz de crescer em meio de cultura contendo apenas lignina kraft como fonte de carbono. Esta cepa foi também testada quanto a capacidade de crescimento/tolerância a diversos compostos fenólicos de reconhecida toxidade, bem como teve seu genoma sequenciado e comparado com um genoma de *Paenarthrobacter* obtido a partir dos dados de sequenciamento metagenômico do consórcio, revelando um alto grau de similaridade.

Desta maneira, foi mostrado neste trabalho que a estratégia de estabelecimento de um consórcio degradador de lignina, aliado ao acesso de enzimas lignolíticas dos micro-organismos pelo isolamento dos mesmos e por ferramentas de metagenômica, consistem em formas efetivas para revelar rotas de bioconversão de compostos aromáticos oriundos de lignina. A obtenção de enzimas presentes nestas rotas possui o potencial de valorização da lignina de bagaço de cana através da conversão desta em moléculas de maior valor agregado.

ABSTRACT

Lignin, one of nature's most recalcitrant materials, consists of a complex polymer composed of various aromatic rings bound by ester-like bonds, being present in combination with cellulose and hemicellulose in the plant cell wall. Lignin is strongly associated with polysaccharides of the cell wall, which makes difficult its use and isolation in an unchanged form. Vanillin is one of the products of lignin oxidation, and is obtained in the natural form from the pod of tropical orchids of the genus *Vanilla*. Due to its low productivity and the difficulty of growing outside its natural habitat, the world demand for this product has been met through chemical synthesis from non-renewable resources such as guaicol and glyoxylic acid. Recently, natural fermentative routes have been developed for the production of vanillin from renewable sources.

Given the commercial importance of the vanillin molecule, associated to the demand for the substitution of non-renewable raw materials for low cost renewable raw materials, it is proposed in this work the establishment and characterization of a consortium of lignin degrading microorganisms as a strategy for the valorization of this sub-product. This consortium was created through the cultivation of microorganisms from sugarcane fields in a culture medium containing a solution of compounds from lignin. This solution is obtained through the process of delignification of sugarcane bagasse, and is composed mainly of phenolic compounds of low molecular weight. With the metagenomic sequencing of the consortium, it was possible to have access to the enzymes and metabolic pathways involved in the bioconversion processes of these phenolic compounds. As a method to validate this strategy for the prospection of lignolytic enzymes, feruloil-CoA synthetase (fcs) and enoyl-CoA hydratase / aldolase (ech) enzymes were selected and evaluated for the conversion capacity of ferulic acid to vanillin.

The characterization of the consortium revealed the ability of the present microorganisms to degrade low molecular weight phenolic compounds from the culture medium. A subsequent taxonomic analysis indicated the enrichment of the consortium, with the decrease of the diversity of microorganisms in relation to the original soil sample, and the increase of species described in the literature as related to the degradation of aromatic

compounds. In this way, the specialization of the consortium can be observed through the selection of microorganisms able to degrade the lignin present in the medium.

The analysis of the metagenome of the consortium also allowed the identification of enzymes involved in the degradation of lignin and aromatic compounds, such as laccases, peroxidases and other enzymes involved in polycyclic aromatic hydrocarbon metabolism (PAHs). From the metagenomic sequencing data, three feruloil-CoA synthetase (fcs) and two enoyl-Coa hydratase / aldolase (ech) enzymes were selected *in silico* by sequence similarity analysis. Capillary electrophoresis and mass spectrometry analyzes showed the production of vanillin from ferulic acid in enzymatic assays with the selected enzymes.

In parallel, microorganisms capable of growing independently in media containing lignin and with possible lignolytic activity were isolated. In particular, the isolated strain of Paen*arthrobacter* sp. (referred to as HW13) was shown to be capable of growing in culture medium containing only kraft lignin as the carbon source. This strain was also tested for growth / tolerability to various phenolic compounds of recognized toxicity, as well as having its genome sequenced and compared to an *Paenarthrobacter* genome obtained from the metagenomic sequencing data of the consortium, revealing a high degree of similarity.

Therefore, it was shown in this work that the strategy of establishing a lignin degrading consortium, together with the access of lignolytic enzymes of the microorganisms by their isolation and by metagenomic tools, consist of effective ways to reveal bioconversion routes of aromatic compounds derived from lignin. The obtaining of enzymes present in these routes has the potential of valorization of the sugarcane bagasse lignin through the conversion of this into molecules of greater added value.

SUMÁRIO

CAPÍTULO 1 – INTRODUÇÃO

PREFÁCIO

1. Consórcios microbianos

Consórcios microbianos consistem em grupos de dois ou mais micro-organismos que se desenvolvem em conjunto em determinado nicho e que apresentam sinergia de processos bioquímicos (Kull, 2010). O termo foi introduzido em 1873 pelo botânico alemão Johannes Reinke para tipificar a relação simbiótica entre fungos e bactérias na qual consistem os líquens. De fato, o termo consórcio precede o termo simbiose para definir este tipo de interação biológica. Consórcios são estabelecidos naturalmente no meio ambiente, de acordo com a presença de determinados compostos que exercem uma pressão seletiva no meio, seja atuando como fonte limitante de carbono para o crescimento dos micro-organismos presentes ou selecionando aqueles que são resistentes ou capazes de degradar compostos tidos como tóxicos. Uma série de trabalhos descreve consórcios de micro-organismos capazes de degradar determinados corantes lançados ao meio ambiente como efluentes da indústria têxtil (Senan e Abraham, 2004; Tony et al., 2009; Phugare et al., 2011; Lade et al., 2015; Hamid et al., 2015; Joshi et al., 2015), consórcios capazes de degradar hidrocarbonetos oriundos do petróleo (Gargouri et al., 2014; Ichor et al., 2016; Patowary et al., 2016; Das e Chandran, 2011) e outros capazes de degradar diferentes tipos de plásticos (Shah et al., 2008; Skariyachan et al., 2016).

Fica evidente, portanto, a importância do estabelecimento e caracterização de consórcios visando desenvolver processos de biorremediação. Além disso, consórcios microbianos visando a degradação de biomassa têm sido estabelecidos a partir de comunidades microbianas presentes no solo (em geral, de locais onde há grande disponibilidade de matéria orgânica, tais como florestas e manguezais) (Gao et al., 2014; de Lima Brossi et al., 2016) ou de comunidades presentes no trato gastrointestinal de animais herbívoros (Fon et al., 2014). Estes consórcios podem ser enriquecidos através do fornecimento de uma fonte específica de carbono durante o cultivo, a fim de exercer uma pressão seletiva sobre os micro-organismos e criar uma comunidade especializada na

degradação desta fonte. Em um contexto bioeconômico, utiliza-se neste processo biomassas que são subproduto da agricultura, tais como palha de trigo ou bagaço de cana-de-açúcar, de forma a identificar micro-organismos capazes de degradar lignocelulose (de Lima Brossi et al., 2016; Korenblum et al., 2016; Jiménez et al., 2016). O acesso ao material genético destes consórcios possibilita a identificação de enzimas lignocelulolíticas, tais como celulases e hemicelulases, que podem ser empregadas em processos biotecnológicos para a produção de etanol de segunda geração (Mamo et al., 2013). O estudo do genoma de micro-organismos presentes em consórcios, enriquecidos ou não, pode ser feito com cepas isoladas (genômica), ou com comunidades microbianas presentes em amostras ambientais através de ferramentas de metagenômica.

2. Abordagem metagenômica aplicada ao estudo da diversidade microbiana e prospecção de enzimas

A grande maioria dos micro-organismos encontrados em amostras ambientais não são passíveis de serem isolados e cultivados em laboratório, impossibilitando o seu completo estudo e caracterização (Cowan, 2000). No entanto, o desenvolvimento de técnicas de sequenciamento e triagem em larga escala propiciaram o acesso ao genoma destes micro-organismos sem a necessidade do seu isolamento. A esta abordagem é atribuído o nome de "metagenômica". As diversas abordagens que compõem o que se chama hoje de metagenômica incluem a caracterização de comunidades microbianas por métodos independentes de cultivo, aplicando análises de ampla cobertura para o acesso ao genoma dos micro-organismos (Handelsman, 2004; Research Council (U.S.). Comittee on metagenomics, 2007). Denominamos o DNA extraído diretamente de amostras ambientais de "metagenoma". Duas estratégias são frequentemente utilizadas na busca de alvos funcionais em metagenomas, sendo elas: triagem funcional de bibliotecas metagenômicas e triagem dirigida por similaridade de sequência nucleotídica (Handelsman, 2004).

Triagens funcionais de bibliotecas metagenômicas consistem na extração do DNA metagenômico, seguido de fragmentação e inserção do DNA em um vetor adequado, transformação em um hospedeiro bacteriano e triagem das células recombinantes em substratos de interesse para a identificação de atividade enzimática dos fragmentos clonados (Handelsman, 2004) (Fig. 1). Os clones podem ser testados quanto à expressão de fenótipos ou características específicas, tais como atividades enzimáticas ou a produção de antibióticos.

Uma vez selecionado o clone, este segue para a etapa de sequenciamento, visando à identificação de genes responsáveis pelas características de interesse.

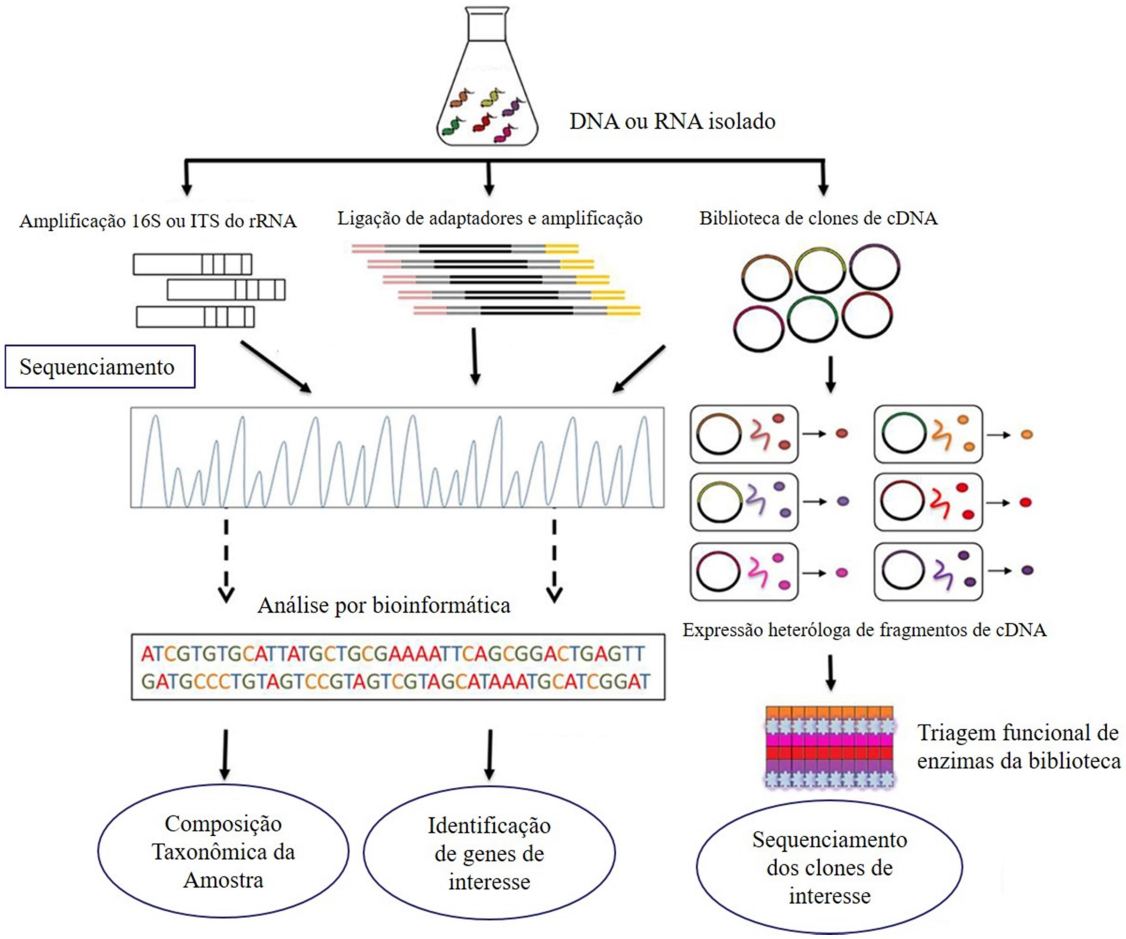


Figura 1: Esquema de abordagem metagenômica aplicada ao estudo de amostra de solo. Traduzido e adaptado de Coughlan et al. (2015).

Triagens baseadas em similaridade de sequência nucleotídica consistem na identificação e sequenciamento de genes no DNA total (metagenoma) ou RNA total (metatranscriptoma) de uma amostra de micro-organismos com base em análises comparativas (*in silico*) de sequências em bancos de dados. O material genético pode ser extraído diretamente da amostra para sequenciamento (por exemplo, uma amostra de solo) ou de comunidades microbianas submetidas a algum tipo de enriquecimento, como no caso dos consórcios microbianos. Os primeiros estudos que realizaram a triagem baseada em sequências fizeram uso de hibridização ou PCR com primers degenerados para a detecção de genes de interesse, tendo como referência sequências de DNA conservadas de famílias proteicas (Amann et al., 1995). Tal estratégia é limitada pelo número de alvos a serem

testados e por não haver garantia de obtenção de genes completos ou grupos gênicos pertencentes a determinada via de síntese de um produto desejado. Deste modo, a aplicação típica para esse tipo de abordagem acabou sendo a amplificação de regiões conservadas do genoma, tais como o rRNA 16S ou o gene *recA,* para estudos de diversidade filogenética (Yun e Ryu, 2005).

Recentemente, estudos baseados no sequenciamento massivo de DNAs metagenômicos têm gerado um grande volume de dados que permitem desde a avaliação de relações filogenéticas até a descoberta de novos genes e dedução de vias metabólicas em micro-organismos (Daniel, 2004). Atualmente, as plataformas de sequenciamento de última geração (NGS) Illumina, como MiSeq e HiSeq, são as mais utilizadas para sequenciamentos de ácidos nucleicos em larga escala. Esta tecnologia de sequenciamento de última geração, juntamente com ferramentas de bioinformática desenvolvidas para tratamento da grande quantidade de dados gerados, permitiram o acesso ao genoma de diversos seres vivos de maneira muito mais rápida, a um custo muito menor do que era necessário com as tecnologias anteriores. Apenas para fins de comparação, o projeto genoma humano, iniciado em 1990 e finalizado em 2001, teve um custo total estimado entre 500 milhões e 1 bilhão de dólares. Hoje, o custo do mapeamento de um genoma humano se encontra na casa dos 1 mil dólares, segundo dados do National Human Genome Research Institute (NIH) (https://www.genome.gov/sequencingcosts/). O acesso ao metagenoma de consórcios só foi possível graças ao desenvolvimento de tais tecnologias, possibilitando o sequenciamento de diversos genomas simultaneamente.

3. A lignina do bagaço de cana-de-açúcar

O bagaço de cana-de-açúcar, assim como todo material lignocelulósico, apresenta em sua composição majoritária celulose (35%), hemicelulose (24%) e lignina (22%) (Rezende et al., 2011) (Fig. 2). Trata-se de um subproduto abundante na agricultura e um substrato promissor para produção de etanol de segunda geração, além de uma matéria prima alternativa para a indústria química de ponta para a produção de moléculas de alto valor agregado (Goldenberg, 2008; Rocha et al., 2011, Zakzeski et al., 2010). Um dos seus componentes, a lignina, consiste em um polímero complexo composto de vários anéis aromáticos estavelmente ligados, sendo considerada um dos materiais mais recalcitrantes da natureza (Fengel e Weneger, 1989; Kerbauy, 2008). A forte associação entre a celulose,

hemicelulose e lignina é um dos principais fatores que dificultam o aproveitamento dos materiais lignocelulósicos. Desta forma, a deslignificação representa um dos passos mais importantes para a viabilização de processos industriais baseados na conversão de biomassa vegetal (Ogeda e Petri, 2010).

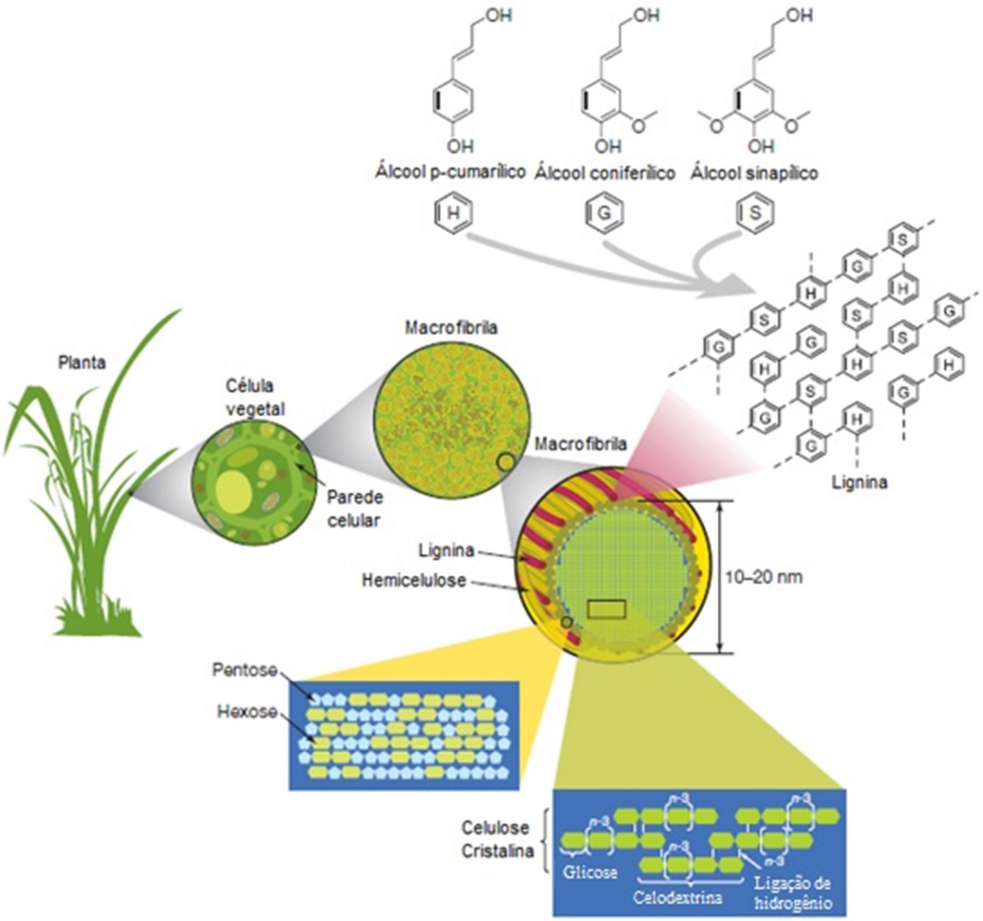


Figura 2: Representação geral da parede celular vegetal. Adaptado e traduzido de Rubin (2008).

A estrutura da lignina é extremamente complexa, e consiste de uma rede de polímeros composta por ligações covalentes entre átomos de carbono e ligações do tipo éter derivadas da polimerização desidrogenativa dos álcoois cinamílicos (monolignóis) álcool p-cumarílico, álcool coniferílico e álcool sinapílico (Freundenberg, 1965). A polimerização desidrogenativa destes álcoois resulta nas unidades hidroxifenil (H), guaiacil (G) e siringil (S), cuja composição é variável entre diferentes espécies de plantas. Por exemplo, a lignina de árvores coníferas (*softwood*) é composta principalmente por guaiacil, enquanto a lignina de árvores folhosas (*hardwood*) e de monocotiledôneas (como a cana-de-açúcar) é composta majoritariamente por guaiacil e siringil (Fengel e Wegener, 1989). A forte ligação entre lignina e os polissacarídeos da parede celular praticamente inviabiliza seu isolamento na

forma não alterada. Portanto, a maioria dos dados sobre a estrutura da lignina é baseada na análise de frações isoladas de monômeros e dímeros derivados da degradação química da lignina. Os produtos solúveis em água, resultantes da oxidação da lignina do bagaço de cana, são predominantemente ácido p-hidroxibenzóico, vanilina e siringaldeído. Durante a oxidação alcalina, quantidades consideráveis de ácido p-coumarico e ácidos ferúlicos são convertidos em p-hidroxibenzaldeído ou ácido p-hidroxibenzóico e vanilina ou ácido vanílico, respectivamente (Alves et al., 2012).

A lignina do bagaço de cana contém quantidades praticamente iguais de unidades não condensadas de guaiacil e siringil. De acordo com Kato et al. (1987), o ácido ferúlico e o p-cumárico estão esterificados, respectivamente, à hemicelulose e à lignina. Recentemente, os conteúdos de 11 híbridos experimentais de cana-de-açúcar selecionados pelo conteúdo de lignina variável foram avaliados, revelando que a porcentagem de ácidos hidroxicinâmicos esterificados à hemicelulose (ácidos ferúlico e cumárico liberados pelo tratamento químico moderado) variou entre 2,3% e 3,6%. Por outro lado, o conteúdo total de ácidos hidroxicinâmicos (liberados pelo tratamento alcalino severo) variou de 5,0% a 9,2%, variação esta que foi proporcional ao conteúdo de lignina destes híbridos (Masarin et al., 2011).

A lignina obtida de material lignocelulósico é uma alternativa renovável altamente promissora para produção de compostos fenólicos, como o siringaldeído, o p-hidroxibenzaldeído e polímeros, com as mais diversas aplicações industriais. Os compostos fenólicos derivados da lignina podem ser convertidos em éteres de arila para aplicação como aditivos em gasolina ou para a produção de solventes e ácidos orgânicos através de processos de craqueamento e dealquilação. A lignina pode também ser empregada como matéria-prima para produção de resinas fenólicas e poliuretânicas, usadas para a produção de tintas e vernizes (Lora e Glasser, 2002; Stewart, 2008). Além disso, diversos estudos reportam a capacidade antioxidante dos compostos fenólicos e polifenólicos na diminuição do risco de desenvolvimento de diversas desordens crônico-degenerativas, tais como câncer, inflamações, doenças cardiovasculares. Estes compostos podem atuar ainda contra a oxidação lipídica em alimentos pela inibição da atividade de enzimas relacionadas com a produção de radicais (Pan et al., 2006; Gülçin e Beydemir, 2013) e como ativos em formulações de cosméticos utilizados para combater o envelhecimento (Ribeiro, 2006).

Enzimas oxidativas para a despolimerização da lignina, tais como lignina peroxidases, Mn^{2+} peroxidases e lacases, foram estudadas em detalhe em fungos "white-rot" (fungo da podridão branca) (Wong, 2009). No entanto, até a presente data, um processo comercial para a valorização comercial da lignina não foi desenvolvido, em parte pela falta de enzimas recombinantes disponíveis. Diversas enzimas bacterianas degradadoras de lignina foram identificadas (Taylor et al., 2012), entre elas a peroxidase DypB, recentemente identificada em *Rhodococcus jostii* RHA1, que oxida Mn^{2+} e compostos modelos de lignina e apresenta elevado nível de expressão heteróloga em *E. coli* (Ahmad et al., 2011).

4. Vanilina: descrição, importância e aplicações

Vanilina (4-hidróxi-3-metoxibenzaldeído) é tradicionalmente conhecida como aroma de baunilha. Trata-se de um composto cristalino de cor branca, usado diretamente na indústria de alimentos, de cosméticos e na síntese de outros compostos (Figura 3). É tradicionalmente obtida da vagem de orquídeas tropicais do gênero *Vanilla,* fazendo desta planta a única orquídea de interesse comercial fora do contexto ornamental. Devido à baixa produtividade (na proporção de 2% de vanilina pura por peso total da vagem seca) e à dificuldade de cultivo fora de seu habitat, a produção total de vanilina natural supre apenas uma pequena parte da demanda mundial. Segundo dados da empresa Rhodia, em 2010, a demanda mundial total excedeu 15.000 Ton, das quais apenas 2.000 toneladas (cerca de 13%) foram produzidas de maneira natural (http://www.solvay.com/en/binaries/GPS_2011_12_v2_Vanillin_gb-139567.pdf).

Historicamente, esta grande demanda tem sido atendida através de síntese química, primeiramente a partir de eugenol (encontrado no óleo de cravo) e de compostos derivados da lignina encontrados no licor negro, resíduo da indústria de celulose (Kortekaas et al., 1998; Hocking, 1997). Posteriormente estes métodos foram substituídos pela síntese química da molécula de vanilina a partir de recursos não renováveis como guaicol e ácido glioxílico (Esposito et al., 2000). Por conta da baixa produtividade, extratos naturais de vanilina custam entre US$ 1.200 - US$ 4.000/Kg, enquanto a vanilina sintetizada a partir de matérias-prima derivadas do petróleo custa em torno de US$ 15/Kg (Walton et al., 2003).

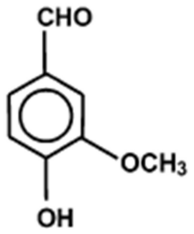


Figura 3: Estrutura da molécula de vanilina. Adaptado de Fitzgerald et. al (2005)

Aproximadamente 50% da produção mundial de vanilina sintética é usada como intermediária para síntese de herbicidas, agentes antiespumantes ou drogas tais como: papaverina, L-dopa, L-metildopa e agentes antimicrobianos. A vanilina também é utilizada na composição de desodorizantes de ambientes e ceras para pisos e assoalhos (Clark, 1990). Além destas características, a vanilina possui também propriedades antimicrobianas e antioxidantes. As propriedades antioxidantes da vanilina devem-se ao fato de esta interagir com espécies radicais, evitando, dessa forma, os processos oxidativos (Angelo e Jorge, 2007). Tais características justificam o amplo uso da vanilina em alimentos na condição de conservantes.

5. Produção de vanilina sintética (artificial)

Os custos elevados associados ao baixo rendimento da produção de vanilina a partir de fontes naturais estimularam a pesquisa por rotas de produção de vanilina sintética. Dessa forma, por volta de 1874-1875, menos de vinte anos após seu isolamento e caracterização, a molécula de vanilina foi sintetizada a partir de eugenol, tornando-se disponível comercialmente na França e Estados Unidos por US$ 176/Kg (Hocking, 1997). O eugenol continuou a ser utilizado como material de partida para a síntese de vanilina até 1920. Embora a produção de vanilina por essa rota sintética possuísse alto rendimento, esta consumia grandes quantidades de reagentes, o que desmotivou a produção da molécula por esse método.

Posteriormente, a vanilina passou a ser sintetizada a partir da lignina contida no licor negro (Kortekaas et al., 1998). O licor negro é um resíduo das indústrias de papel e celulose que possui grande quantidade de sólidos dissolvidos e uma grande diversidade de substâncias químicas oriundas da degradação de lignina (fenóis de baixa e alta massa molar e de toxicidade reconhecida) e polissacarídeos. A hipótese de que a vanilina poderia ser

sintetizada a partir de substratos constituídos de lignina foi realizada por meio de uma publicação anônima em 1875 (Hocking, 1997). A produção industrial de vanilina a partir da lignina contida no licor negro começou nos Estados Unidos em 1936 com as empresas Salvo Chemical Corp. e Marathon Paper Mills Co. de Wisconsin, utilizando a tecnologia desenvolvida por Tomlinson e Hibbert (Kortekaas et al., 1998). Um ano depois, a Howard Smith Paper Mills Ltd. desenvolveu a primeira planta de produção de vanilina em grande escala no Canadá. Estudos realizados pela Ontario Pulp and Paper, em 1940, levaram à construção da segunda planta de produção de vanilina, atingindo em 1945 a produção de 227 Ton/ano (Hocking, 1997).

Embora a produção de vanilina a partir do licor negro ter sido parcialmente motivada pelo interesse em reduzir o conteúdo de ligninas, açúcares e, consequentemente, a demanda bioquímica de oxigênio (DBO) desse subproduto, o processo possuía muitos inconvenientes como o fato de gerar um grande volume de resíduos tóxicos. Para produzir 1 kg de vanilina, eram utilizados 160 kg de líquidos cáusticos e gerados 150 kg de outras substâncias de reconhecido caráter tóxico e sem importância comercial (Hocking, 1997; Furukawa et al., 1998). Esse fator acabou levando ao fechamento das plantas no Canadá e nos EUA no fim da década de 1980 e início de 1990 (Hocking, 1997).

Atualmente, a maior parte da vanilina sintética é produzida por meio do uso de precursores derivados do petróleo. Os precursores mais utilizados são o guaiacol e o p-cresol, oriundos da indústria petroquímica (Mukhopadhyay, 2005). Este é o método corriqueiramente empregado pela indústria hoje, pois trata-se de um método mais econômico em relação ao que utiliza lignina do licor negro e também por gerar menos efluentes devido aos rendimentos das reações serem maiores (Hocking, 1997) (Figura 4).

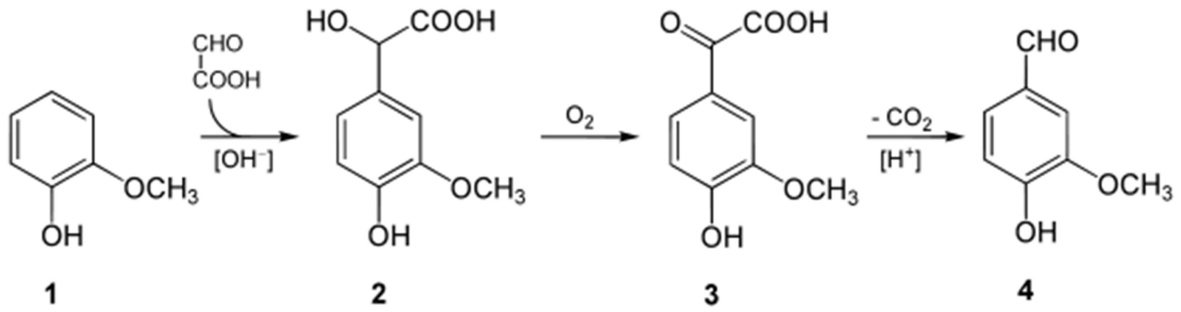


Figura 4: Via de síntese de vanilina (4) a partir de guaiacol (1). Guaiacol reage com o ácido hidroxílico através de uma substituição aromática eletrofílica. O ácido vanilmandélico (2) resultante é então convertido em vanilina através do intermediário ácido 4-hidroxu-3-metoxifenilglioxílico por descarboxilação oxidativa.

6. Produção de vanilina a partir de fontes renováveis: oportunidade biotecnológica

Recentemente, rotas fermentativas naturais foram desenvolvidas para a produção de vanilina a partir de fontes renováveis. Uma rota biossintética *de novo* foi desenvolvida em *Schizosaccharomyces pombe* através da via do chiquimato (Hansen et al., 2009), produzindo em torno de 45 mg/L de vanilina, utilizando glicose como matéria prima. Outras abordagens usaram como matéria prima compostos de relativo baixo custo, derivados de fontes naturais, tal como o ácido ferúlico (Lee et al., 2009; Yang et al., 2013). Nesta abordagem, foi possível a produção de vanilina em *E. coli* com a utilização das enzimas feruloil-CoA sintetase (fcs) e enoil-CoA hidratase/aldolase (ech) (Figura 5) a uma concentração de até 5,14 g/L, com rendimento de conversão molar de 86,6% (Lee et al., 2009). Em outra abordagem, isoeugenol foi convertido em vanilina utilizando a enzima isoeugenol monooxigenase (iem) de *Pseudomonas putida* IE27 (Yamada et al., 2008) (Figura 4) expressa em *E. coli***.** As células recombinantes, capazes de expressar iem, produziram 28,3 g de vanilina por litro a partir de uma solução com 230 mM de isoeugenol, com rendimento de conversão molar de 81% (Yamada et al., 2008).

A vanilina produzida por fermentação microbiana é vendida como vanilina natural pela empresa Rhodia com a marca Rhovanil Natural® à US$ 700/Kg. Desta maneira, produtos produzidos a partir de fontes renováveis por bioconversão/processos fermentativos podem carregar o selo “natural” nos mercados dos EUA e da União Européia. Tanto isoeugenol como ácido ferúlico podem ser obtidos de fontes naturais (renováveis), tais como

o óleo de cravo e o bagaço de cana-de-açúcar. Isto oferece boas oportunidades de valorização e uma não dependência de recursos finitos derivados do petróleo.

Reações enzimáticas complementares em rotas bioquímicas, como no caso das enzimas fcs e ech (Yang et al., 2013), demonstram que estes sistemas podem ser otimizados usando princípios de biologia sintética. O grupo de pesquisa do PhD Timothy D.H. Bugg (Departamento de Química da Universidade de Warwick) reportou recentemente que um mutante deletério de *Rhodococcus jostii* RHA1, cuja via catabólica da vanilina foi bloqueada, foi capaz de acumular até 96 mg/L de vanilina quando cultivado em meio mínimo contendo 2,5% (massa/volume) de lignocelulose extraída de palha de trigo (Bugg et al., 2011), demonstrando que uma abordagem de engenharia de rotas bioquímicas pode ser usada para a geração de compostos aromáticos úteis a partir de biomassa de plantas (Sainsbury et al., 2013). Além disso, muitas das enzimas identificadas associadas com a produção de vanilina estão protegidas por patentes, portanto a identificação de novas enzimas/processos catalíticos se mostra de grande importância. Existe um imenso campo a ser explorado, no que diz respeito à produção biotecnológica da vanilina: uma molécula de alto valor agregado obtida a partir de fontes abundantes, como resíduos agroindustriais. Além disso, há uma gama de micro-organismos potenciais a serem estudados, que somados a novas tecnologias e a engenharia genética, tendem a tornar o processo biotecnológico atraente para as indústrias.

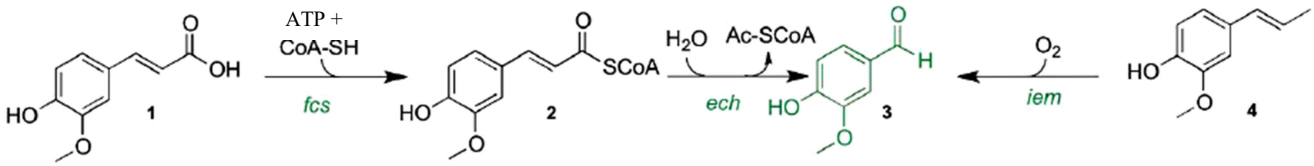


Figura 5: Rota biossintética convergente. Enzimas envolvidas: *fcs* (feruloil-CoA sintetase), *ech* (enoil-CoA hidratase/aldolase) e *iem* (isoeugenol monoxigenase); reagentes e produtos ácido ferúlico (1), feruloil-CoA (2), vanilina (3) e isoeugenol (4).

MOTIVAÇÃO E OBJETIVOS

Neste contexto, a utilização de uma abordagem metagenômica apresenta vantagens em relação aos métodos tradicionais de análise de genomas, pois possibilita a análise global da microbiota de uma determinada amostra, considerando, portanto, a interação destes micro-organismos em processos de bioconversão (Research Council (U.S.). Comittee on metagenomics, 2007). Conforme descrito no item 4 deste capítulo, compostos fenólicos derivados de lignina podem ser utilizados como precursores de moléculas de alto valor agregado, como a vanilina. No entanto, pela falta de metodologias para o aproveitamento destes compostos, eles são usualmente incinerados para geração de energia ou então descartados, gerando um agravante ambiental. Deste modo, faz se necessário o desenvolvimento de processos que permitam o aproveitamento da lignina gerada do pré-tratamento do bagaço de cana-de-açúcar. Com isto, seria possível um aumento na eficiência da cadeia de produção de cana-de-açúcar, sobretudo por tornar o processo de produção de etanol de 2ª geração mais atraente do ponto de vista econômico através da utilização de um co-produto para geração de moléculas de alto valor agregado. Deste modo, este processo se integra à tendência mundial de busca de matérias primas renováveis para substituição de combustíveis fósseis e de processos industriais ecologicamente corretos e pautados em princípios de sustentabilidade ambiental.

Portanto, este trabalho de doutorado teve como objetivo o estabelecimento e caracterização de um consórcio degradador de lignina (CLig) a partir de uma amostra de solo de canavial, visando sua caracterização quanto a capacidade de degradação de lignina, diversidade microbiana e presença de enzimas envolvidas em vias de catabolismo de compostos aromáticos. O projeto também visou a prospecção das enzimas feruloil-CoA sintetase e enoil-CoA hidratase/aldolase a partir dos dados de sequenciamento metagenômico do consórcio e o teste destas enzimas quanto à capacidade de conversão de ácido ferúlico em vanilina. Paralelamente, desenvolveu-se a estratégia de isolamento de micro-organismos do consórcio para o estudo dos mesmos quanto a capacidade de degradação de lignina e a presença de enzimas lignolíticas.

CAPÍTULO 2 – DOCUMENTOS EM PREPARAÇÃO

O documento a seguir corresponde à versão preliminar do artigo científico submetida ao periódico Biotechnology for Biofuels e contém os principais resultados obtidos no projeto de pesquisa. A versão final e aceita para publicação, intitulada "Lignolytic-consortium omics analyses reveal novel genomes and pathways involved in lignin modification and valorization", se encontra em https://doi.org/10.1186/s13068-018-1073-4.

DOCUMENTO 1. Targeted metagenomics identifies novel genomes and pathways involved in lignin conversion and valorization

*Eduardo C. Moraes[1], Thabata M. Alvarez[2], Gabriela F. Persinoti[1], Geizecler Tomazetto[1], Livia B. Brenelli[1], Douglas A. Paixão[1], Gabriela Cristina Ematsu[1], Juliana A. Aricetti[1], Camila Caldana[1], Neil Dixon[3], Timothy D. H. Bug[4], Fabio M. Squina[5]**

1. Laboratório de Ciência e Tecnologia do Bioetanol, Centro Nacional de Energia e Materiais, Campinas, Brazil
2. Universidade Positivo, Master in Industrial Biotechnology, Curitiba, Paraná, Brazil.
3. Department of Chemistry, University of Warwick, Coventry, United Kingdom.
4. Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, Manchester, United Kingdom.
5. Programa de Processos Tecnológicos e Ambientais, Universidade de Sorocaba, Sorocaba, Brazil

* *Correspondence to fabio.squina@prof.uniso.br; Rod. Raposo Tavares, km, 92.5, Sorocaba-SP, CEP 18023-000, Brasil*

Abstract

Background: Lignin is a heterogeneous polymer representing a renewable source of aromatic and phenolic bio-derived products for the chemical industry. However, the inherent structural complexity and recalcitrance of lignin makes its conversion into valuable chemicals a challenge. Natural microbial communities produce biocatalysts derived from a large number of inhabitants, which operate synergistically to successfully perform a variety of bioconversion processes. Thus, metagenomic approaches are a powerful tool to reveal novel optimized metabolic pathways for lignin conversion and valorization.

Results: The specialized lignin-degrading community (CLig) was obtained from a soil sample (ScMet) of a sugarcane plantation. The CLig taxonomical analyses (based on 16S rDNA) indicated prevalence of Proteobacteria, Actinobacteria and Firmicutes members, including the Alcaligenaceae and Micrococcaceae families which were enriched in the CLig compared to ScMet. *Candida* and *Rhodotorula* spp. were the dominant fungal genus in CLig, accounting for 64 % of all fungal sequences assigned (based on internal transcribed spacer region sequencing). Through global DNA sequencing analysis around 240 thousand gene models and 65 draft bacterial genomes were predicted and identified, respectively. Along with depicting several peroxidases, dye-decolorizing peroxidases, laccases, carbohydrate esterases and lignocellulosic auxiliary (redox) activities, the major pathways related to aromatic degradation were identified, including benzoate (or methylbenzoate) degradation to catechol (or methylcathecol), catechol ortho-cleavage, catechol meta-cleavage and phthalate degradation. A novel *Paenarthrobacter* strain harboring seven gene clusters related to aromatic degradation was isolated from CLig, and was able to grow on lignin as a sole carbon source. Furthermore, a recombinant pathway for vanillin production was assembled based on novel gene sequences coding for a feruloyl-CoA synthetase and an enoyl-CoA hydratase/aldolase retrieved from the metagenomic dataset.

Conclusion: The enrichment protocol described in the present study was successful for the establishment a microbial consortium towards the lignin and aromatic metabolism, providing pathways and enzyme sets for downstream synthetic biology strain engineering approaches. This work represents a pioneering study on lignin conversion and valorization strategies based on metagenomics, revealing several novel lignin-conversion enzymes, aromatic-degrading bacterial genomes, and a novel bacterial strain of potential biotechnological interest. The

validation of a biosynthetic route for vanillin synthesis confirmed the applicability of the targeted metagenome discovery approach for lignin valorization strategies.

Keywords: lignin, aromatic compound degradation, metagenome, vanillin, ferulic acid

Background

High global carbon dioxide emission levels and the need for renewable feedstocks for the energy and chemical industries represent the greatest challenges to humankind [1-3]. Lignocellulosic biomass is the most abundant carbon-based material in nature, representing an attractive renewable source to replace oil derivatives, and to provide aromatic building blocks for the chemical industry [4-6].

Lignin is one of the major components of plant cell walls, and consists of a highly recalcitrant and heterogeneous polymer network formed via radical coupling reactions involving the three major monolignols: p-coumaryl, coniferyl and sinapyl alcohol, linked by C-C and C-O bonds [7-9]. However, controlled deconstruction of the macromolecule to produce lower molecular weight compounds is critical for lignin valorization strategies [5,10]. The rapid expansion of cellulosic biorefineries has increased the production of lignin-rich streams [6], which at present are mainly burned for energy supply [11]. Nevertheless, examples of chemical depolymerization or acidification treatment of lignin to produce low-molecular-weight aromatic compounds have proven to be effective for lignin valorization strategies [12-14].

There is increasing interest in the development of biological lignin depolymerization routes using enzymes and microorganisms [15]. In this sense, intensive studies have been performed to elucidate the enzymatic repertoire involved in lignin depolymerization by fungi and bacteria [16,17]. Typically, enzyme-based technologies rely on sets of genes produced by a single cultivable organism, whereas natural microbial communities produce biocatalysts derived from a large number of inhabitants, which operate synergistically to perform a variety of bioconversion process. Therefore, metagenomic approaches show potential to gain further access to microorganisms, genes and metabolic pathways linked to lignin conversion and valorization [18-20].

This work describes a whole metagenomic-based approach to reveal genomes and metabolic pathways involved in lignin conversion. For this purpose, a specialized lignin-degrading community (CLig) was developed from a soil sample using a lignin-rich stream as the sole carbon source. The taxonomic assignment and functional annotation of metagenomic sequencing data were interpreted and applied to reconstruct bacterial draft genomes, depicting

lignin depolymerization and valorization pathways. Moreover, a novel strain of potential biotechnological interest was retrieved from CLig, *Paenarthrobacter* is capable of metabolizing lignin fragments and contains several aromatic degradation gene clusters in its genome. Finally, to validate the usefulness metagenomic characterization of CLig to provide lignin valorization pathways, the biosynthetic route of vanillin production from ferulic acid was confirmed using novel corresponding enzymes feruloyl-CoA synthetase (*FerA_B3*) and enoyl-coA hydratase/aldolase (*FerB_B11*), whose genes sequences were retrieved from the CLig dataset.

Methods

Development of the CLig

The sugarcane soil sample (ScMet) was collected from a sugarcane farm in the city of São Carlos (GPS coordinates: -22º 04' 77.98, -48º 63´ 52.99´´; Fazenda Tropical). The lignin-degrading consortium (CLig) was established by adding 1 g ScMet to a flask containing 100 ml of culture medium supplemented with a stock solution containing soluble lignin. The stock solution, referred to herein as low molecular weight (LW) lignin, consists of a supernatant obtained after the acidification of black liquor generated from the delignification of steam-exploded sugarcane bagasse [21]. The culture media consisted of a 1:1 dilution of the lignin stock solution in distilled water and addition of minimum nutrients from Bushnell Haas Broth (Sigma-Aldrich) and pH 7.0. The flask was incubated at 30 ºC and 150 rpm. Microbial enrichment was obtained by transferring aliquots (3 ml) of the microbial suspension to fresh medium weekly. This procedure was carried out during 50 weeks prior to the first CLig analysis. The DNS assay [22] was performed to determine the total reducing sugar consumed. CLig samples were centrifuged and 100 µl of the supernatant were mixed to 100 µl of the DNS reagent in triplicate as described by Miller [22].

Lignin degradation evaluation

The consumption of water-soluble lignin fragments present in the stock solution by the microbial community was evaluated by gel permeation chromatography (GPC) in a Superdex® 30 prep grade 127 ml column (GE Healthcare) coupled with an ÄKTA System (GE Healthcare). The GPC gradient was performed isocratically using NaOH 0.1 mol L^{-1} as the eluent at a 0.5 mL min^{-1} flow rate and 20 °C. For each supernatant, 500 µL was diluted in 500 µL of NaOH 0.1 mol L^{-1} prior to injection [12]. Detection of lignin fragments was obtained using a UV detector (280 nm) and the molecular weights were determined using different polyphenolic compounds as external standards.

Sequencing analysis

Procedures for amplicon and metagenomic library construction, sequencing and bioinformatics analyses are detailed in Additional file 1 (Methods section). In summary, total

microbial DNA was extracted using the FastDNA Spin Kit for soil (MP Biomedicals) and purified with Power Clean® DNA Clean-Up Kits (Mo Bio Laboratories) according to the manufacturer's instructions. The hypervariable V4 region of the 16S rRNA and internal transcribed spacer region 2 (ITS2) were amplified using universal primers [23-24]. PCR products were purified and pooled together in equimolar amounts for sequencing on an Illumina Miseq system according to standard procedures. The raw sequences were processed using the UPARSE pipeline [25] with default setting for merge, chimera detection (UCHIME method) and clustering into operation taxonomic units (OTU, sequence identity ≥ 97 %). Representative sequences were taxonomically classified by means of the USEARCH using Ribosomal Database Project (RDP) Classifier [26] and UNITE Fungal ITS training sets [27]. For statistical analyses, the alpha (Observed species, Chao1, Good's coverage) and beta diversities were calculated with the R statistical software system package version 3.2.

For metagenome sequencing, the library was prepared according the manufacturer's instructions and paired-end sequenced on the Illumina HiSeq 2500. Trimmed sequences were assembled *de novo* using IDBA-UD [28]. Genes were called using MetaGeneMark [29] and compared to SwissProt, and UniRef90 databases, Pfam [30], dbCAN [31] and EggNOG databases [32] using the HMMER3 (http:// hmmer.janelia.org/) package. Taxonomy assignment of genome bins was performed using Phyla-AMPHORA [33].

Cultivation, isolation and genomic-based characterization of microorganisms

The CLig culture broth was serially diluted and plated directly on agar media supplemented with 0.25 % (w/v) of low molecular weight (LW) soluble lignin and high molecular weight (HW) insoluble lignin, prepared as described by Rocha et al. [21]. Culture plates were then incubated at 30 ^{o}C for 48 hours. Obtained isolates were plated on agar supplemented with HW lignin and incubated again at 30 ^{o}C for 48 hours. Subsequently, the colonies were covered with a 1.5% (w/v) agar solution supplemented with 0.01% (w/v) Azure B dye and formation of decolorization halos around isolates indicated lignolytic activity. A bacterium colony was isolated and identified based on the 16S rRNA gene sequencing analysis using BLASTn search against the RDP database with 97 % sequence identity and E-value of 1e-10. For genome sequencing, genomic DNA of the isolates were extracted as previously described and sequenced by applying the paired-end and mate-pair library protocols on an Illumina MiSeq platform. Genome reads were first quality-filtered and

assembled using Velvet [34], and SSPACE [35] for scaffolding using the mate-pairs reads. Pilon [36] was used to further improve the genome assembly. The assembled genome sequence was imported into the annotation platform Integrated Microbial Genomes (IMG/ER, Markowitz et al., 2009) for automatic prediction of genes. Finally, the assembled genome of the isolate was compared to the genome bins using Mauve and Mummer (version 3.0, [38]) to determine their similarity.

Nucleotide sequence accession numbers

Raw sequencing datasets are available in the EBA database under study accession number PRJEB20169.

Cloning and expression of genes involved in vanillin biosynthesis

The biotransformation of ferulic acid into vanillin was based on expression of two genes, *FerA* and *FerB*, encoding feruloyl coenzyme A synthetase and enoyl-CoA hydratase, respectively. These genes sequences were obtained from the CLig metagenomic dataset and synthesized by Biomatik (Biomatik Corporation, Canada). Genes were inserted into the pET28a-vector and transformed into *Escherichia coli* BL21(DE3). A His6-tag was fused at the N-terminal to promote purification in a His-Trap-Ni-NTA column (GE Healthcare) for both proteins. Details and specifications of the expression step, strains and plasmids are described in detail in Additional file 1 (Methods section).

Enzymatic assays

The enzymatic assays were performed according to methodology described in Yang et al. (2013)[40]. Briefly, the first reaction, conversion of ferulic acid into feruloyl-CoA, consisted of 100 mM potassium phosphate buffer (pH 7.0), 2.5 mM $MgCl_2$, 1 mM ferulic acid, 2 mM ATP, 0.4 mM coenzyme A and 1.5 μg of the purified protein FerA. The mixture was incubated at 30 ºC for one hour. Next, the feruloyl-CoA was converted into vanillin by addition of 1.5 μg of the purified protein FerB and incubated for 24 hours at the same temperature. Samples were taken from last reaction for GC-MS quantitative analysis to detect the vanillin production. The substrates and cofactors were purchased from Sigma-Aldrich (St. Loius, MO, USA).

Analytical methods for vanillin production

To detect phenolics and vanillin production from enzymatic reactions, we used a qualitative GC-MS analysis. The resulting phenolics from enzymatic reactions were extracted by adjusting the pH of the samples to below 2 with 6 mol L-1 HCl and addition of butyl acetate (1:1, v:v) and then derivatized [41]. The derivatized samples (1 µL) were analyzed on an Combi-PAL autosampler (Agilent Technologies GmbH, Waldbronn, Germany) coupled to an Agilent 7890 gas chromatograph in split less mode coupled to a Leco Pegasus 2 time-of-flight mass spectrometer (LECO, St. Joseph, MI, USA) as described by Weckwerth et al. 2004 [42]. The chromatograms were exported from the Leco ChromaTOF software (version 3.25) to the R software. Peak detection, retention time alignment and library matching were performed using the Target Search R-package [43]. Metabolites were quantified by the peak intensity of a selective mass. The intensity normalization procedure was performed by dividing the fresh-weight, followed by sum of the total ion count and global outlier replacement [44,45].

Results and Discussion

The development of CLig

The establishment of microbial consortia is a powerful strategy to develop a specialized community with a particular metabolic ability. Aiming for enrichment of microorganisms able to degrade and modify lignin and/or lignin-derived aromatic compounds, we successfully established a lignin-degrading community (CLig) and maintained this for up to 50 consecutive weeks.

As illustrated in Fig. 1A, the liquid waste-stream from the CLig community contains lignin fragments ranging from 1,200 Da to 300 Da, including phenolic monomers such as ferulic and cinnamic acid (Additional file 1: Table S1), along with minimal sugar concentrations (450 µM L^{-1}). The CLig consumed compounds ranging from ~700 Da to ~300 Da during the first 34 hours of growth (Fig. 1B), indicating that the consortium was able to use the soluble lignin-fragments as a carbon source. Moreover, during the first 40 hours an increase in OD600 and consumption of sugars available in the medium was observed, indicating the growth of microbial community (Additional file 1: Fig. S1).

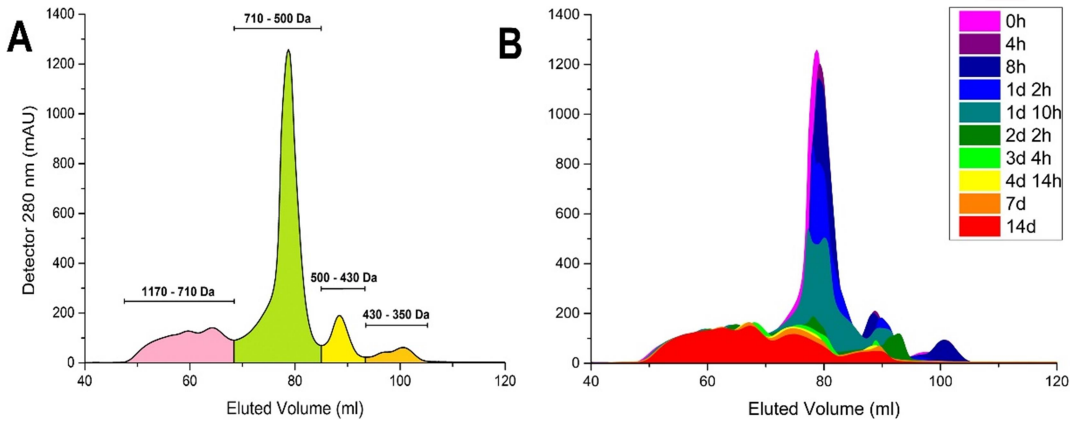


Fig 1. GPC chromatograms showing the molecular weight distribution of (A) the lignin-waste stream used as the sole carbon source for establishment of the lignin-degrading consortium (CLig) and (B) the high and low molecular weight fragments detected during 14 days of CLig cultivation. Above the peaks is the molecular weight distributions in Dalton (Da), which are inversely correlated with the elution time. Different phenolic compounds and polymers were used as internal standards.

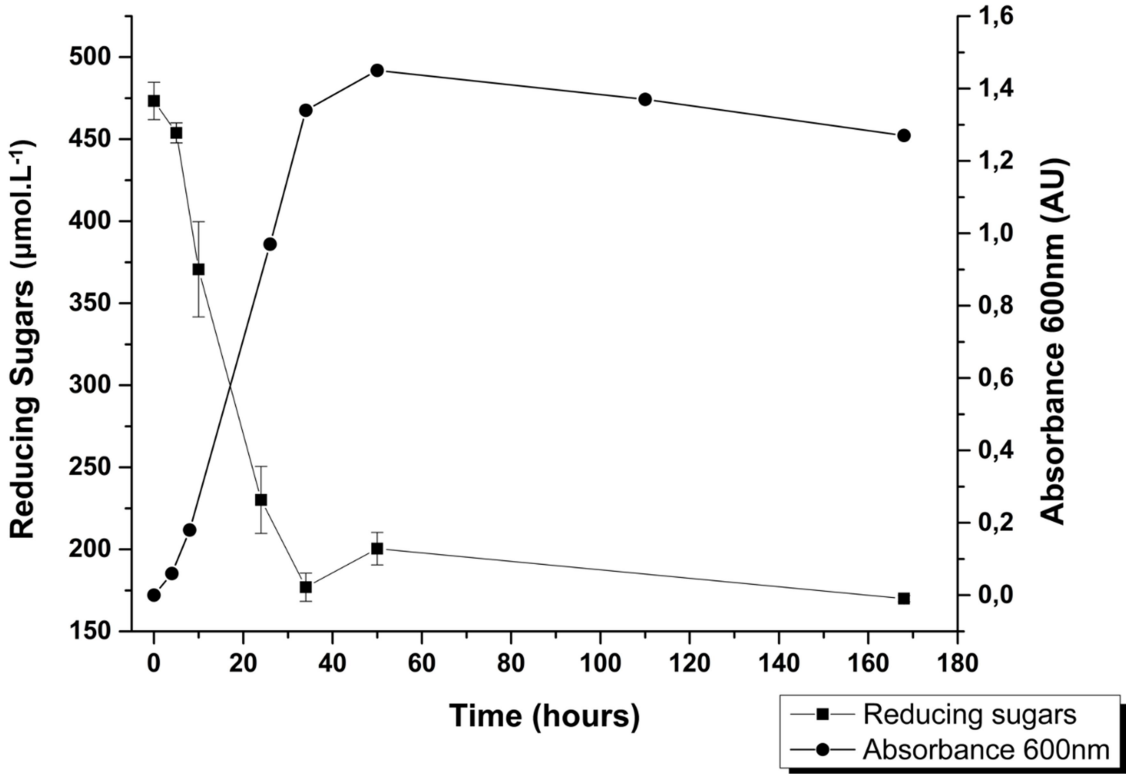


Fig S1. Monitoring of microbial growth by OD 600 nm along time in parallel to evaluation of reducing sugar consumption.

Table S1. Compounds identified by gas chromatography–mass spectrometry (GC-MS) in the lignin-waste stream used for establishment of lignin-degrading microbial community (CLig)[1].

Retention Time (min)	Match (%)	Compound
9.739	87.6	butanoic acid
10.241	96.1	latic acid
12.235	93.8	butanoic acid
12.995	94.3	3-Hydroxypropanoic acid
13.218	86.8	Butyric acid, 3-hydroxy
13.991	92	Oxalic acid
14.29	73.8	Pentenoic acid, 4-[(trimethylsilyl)oxy]-, trimethylsilyl ester
15.192	80.0	Pentenoic acid, 4-[(trimethylsilyl)oxy]-, trimethylsilyl ester
15.857	92.1	glycerol
17.511	79.7	Propanoic acid, 2-methyl-2,3-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester
18.366	81.2	Glyceric acid
18.78	75.5	Succinic acid
21.744	84.8	2(3H)-Furanone, dihydro-3-[(trimethylsilyl)oxy]-3-[[(trimethylsilyl)oxy]methyl]-
25.209	73.0	Butyric acid, 4-hydroxy
25.901	75.8	Butyric acid, 4-hydroxy
27.251	80.5	Arabinoic acid, 2,3,5-tris-O-(trimethylsilyl)-, ç-lactone, l-
28.221	81.6	D-Erythro-Pentonic acid, 3-deoxy-2,5-bis-O-(trimethylsilyl)-2-C-[[(trimethylsilyl)oxy]methyl]-, ç-lactone
28.621	82.6	D-Erythro-Pentonic acid, 3-deoxy-2,5-bis-O-(trimethylsilyl)-2-C-[[(trimethylsilyl)oxy]methyl]-, ç-lactone
31.212	61.0	Acetic acid, [4-methoxy-3-(trimethylsiloxy)phenyl]-, methyl ester
33.192	93.2	Cinnamic acid, 4-hydroxy-, trans-
36.393	76.6	trans-Ferulic acid

[1]The lignin source used in this study was a soluble stream generated by pilot-scale steam explosion and alkaline delignification of sugar cane bagasse [20]. The lignin profiling was performed with derivatized samples by a gas chromatography–mass spectrometry (GC-MS) system (Agilent GC 6890 and MSD 5973N series, Agilent, USA), according to Suguiyama et al [21]. The peaks were identified and quantified in comparison with authentic standards and the NIST Mass Spectral Library. The peak retention time and confidence level in sample identification with GC-MS are presented.

Diversity estimates

To evaluate the microbial richness of CLig, several indices were calculated based on sequencing of the 16S rDNA gene and ITS2 to analyze the prokaryotic and fungal diversity, respectively. The 16S rRNA diversity of ScMet was also performed for comparison purposes. Statistical data summarizing the sequencing results are shown in Additional file 1 (Table S2). Regarding the prokaryotic diversity, a total number of OTUs identified (based on

16S amplicons) in the ScMet was 1,558, far superior in comparison of the 355 OTUs identified in CLig (Additional file 1: Table S2). Alpha diversity, such as rarefaction curves, Chao1, Shannon and Simpson´s diversity indices are shown in Additional file 1 (Fig. S2, S3, S4 and S5). The Chao1 estimator predicted over 1,500 and 350 species-level OTUs for ScMet and CLig, respectively (Additional file 1: Fig. S4). The Good's coverage index was 0.99 for both CLig and ScMet, indicating that overall bacterial diversity has been observed and well represented in both samples. Moreover, Shannon and Simpson´s diversity indices were higher for ScMet (6.3 and 0.99, respectively) in comparison to CLig (3.4 and 0.92) (Additional file 1: Fig. S4). Regarding fungi diversity, CLig has a significantly less diverse community in comparison to the prokaryotic community (Additional file 1: Table S2). The observed species metrics and Chao1 estimator predicted up to 10 species-level fungal OTUs, and the Shannon and Simpson´s diversity indices calculated were 1.44 and 0.71, respectively (Additional file 1: Fig. S5). The reduction of the diversity of the microbial community in CLig sample was expected since only a small fraction of the natural microbial community from soil would be able to growth in a lignin rich-media.

Table S2. Sequencing statistics and data processing of amplicons libraries constructed for profiling of the CLig and ScMet analyzed.

Gene amplicon/Target primers/Replicate	Raw sequences[1]	Sequences Trimming/Merged/Chimeras[2]	Average amplicon size [bp]	OTUs[3]
CLig				
16S rRNA/*Bacteria*/A	744,353	310,123	292	353
16S rRNA/*Bacteria*/B	519,315	212,420	292	343
16S rRNA/*Bacteria*/C	519,302	206,284	292	355
ScMet				
16S rRNA/ *Bacteria /A*	42,349	20,193	292	1419
16S rRNA/ *Bacteria /A*	101,080	52,903	292	1551
16S rRNA/ *Bacteria /A*	137,133	77,397	292	1558
CLig				
ITS2/*Fungi*/A	51,188	26,601	337	9
ITS2/*Fungi*/B	40,950	21,844	331	10
ITS2/*Fungi*/C	52,795	17,838	334	7

[1] Total counting of paired-end reads.
[2] After quality filtering of reads, merged sequences and removal of chimeric sequences.
[3] Number of Operational Taxonomic Units (OTUs).

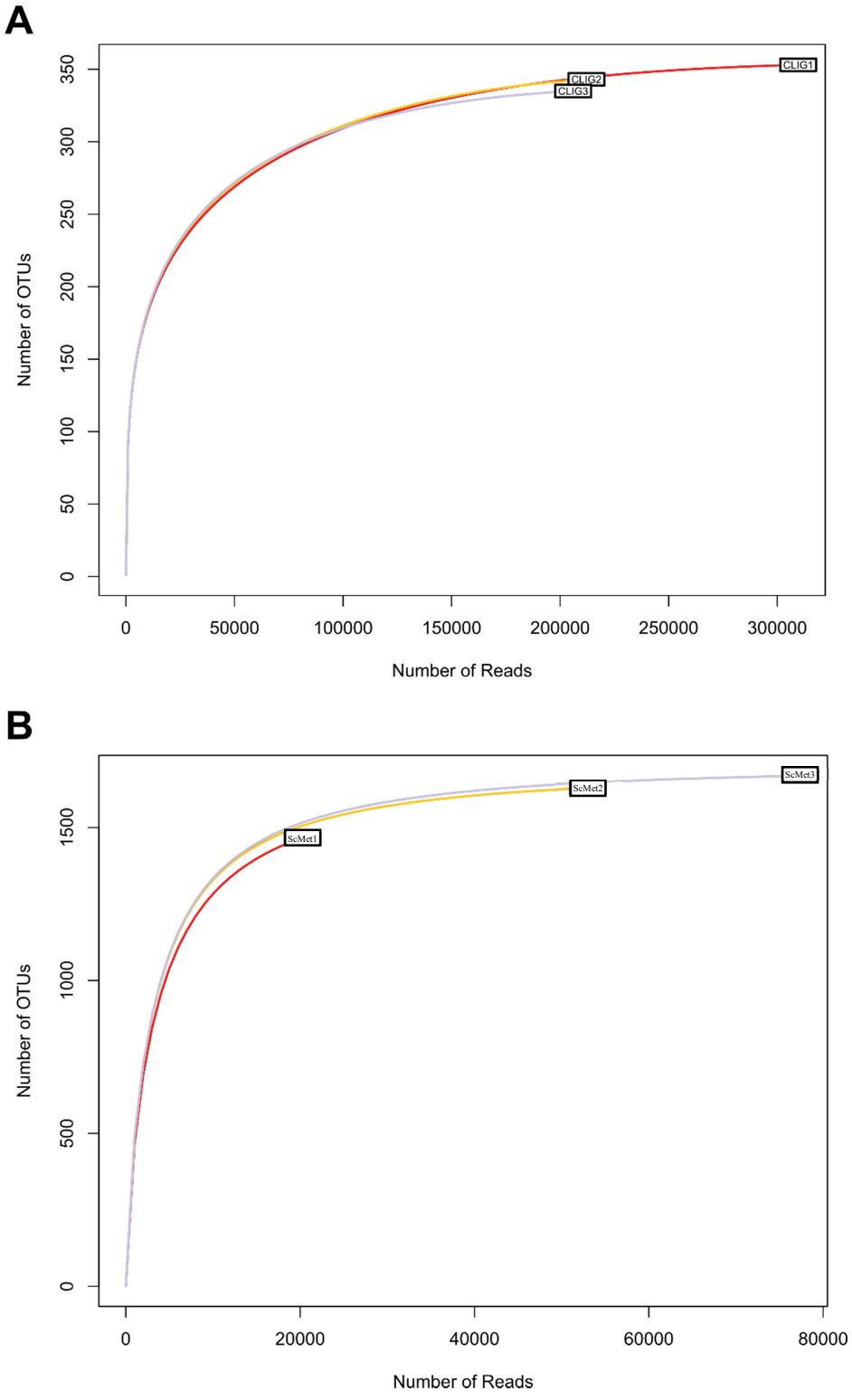


Fig. S2. Rarefaction curves based on V4 region from 16S rRNA gene amplicons derived from CLig (**A**) e ScMet (**B**).

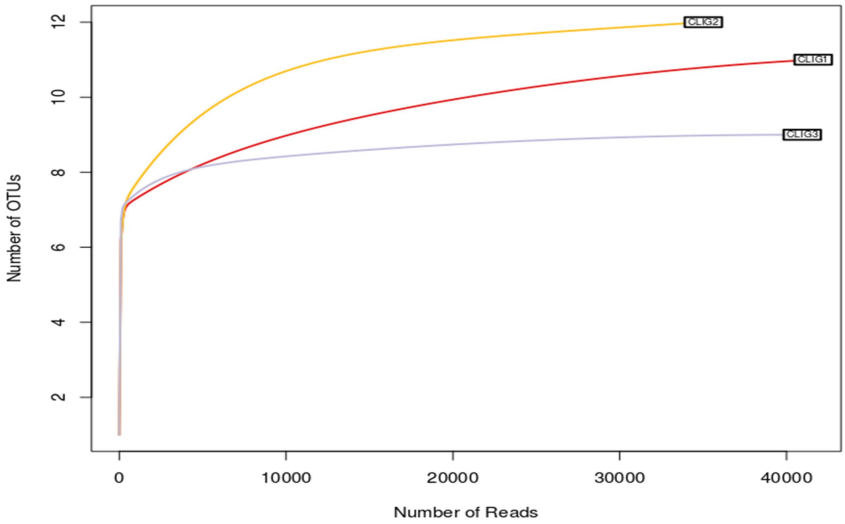


Fig. S3. Rarefaction curves based on ITS2 region sequencing derived from CLig.

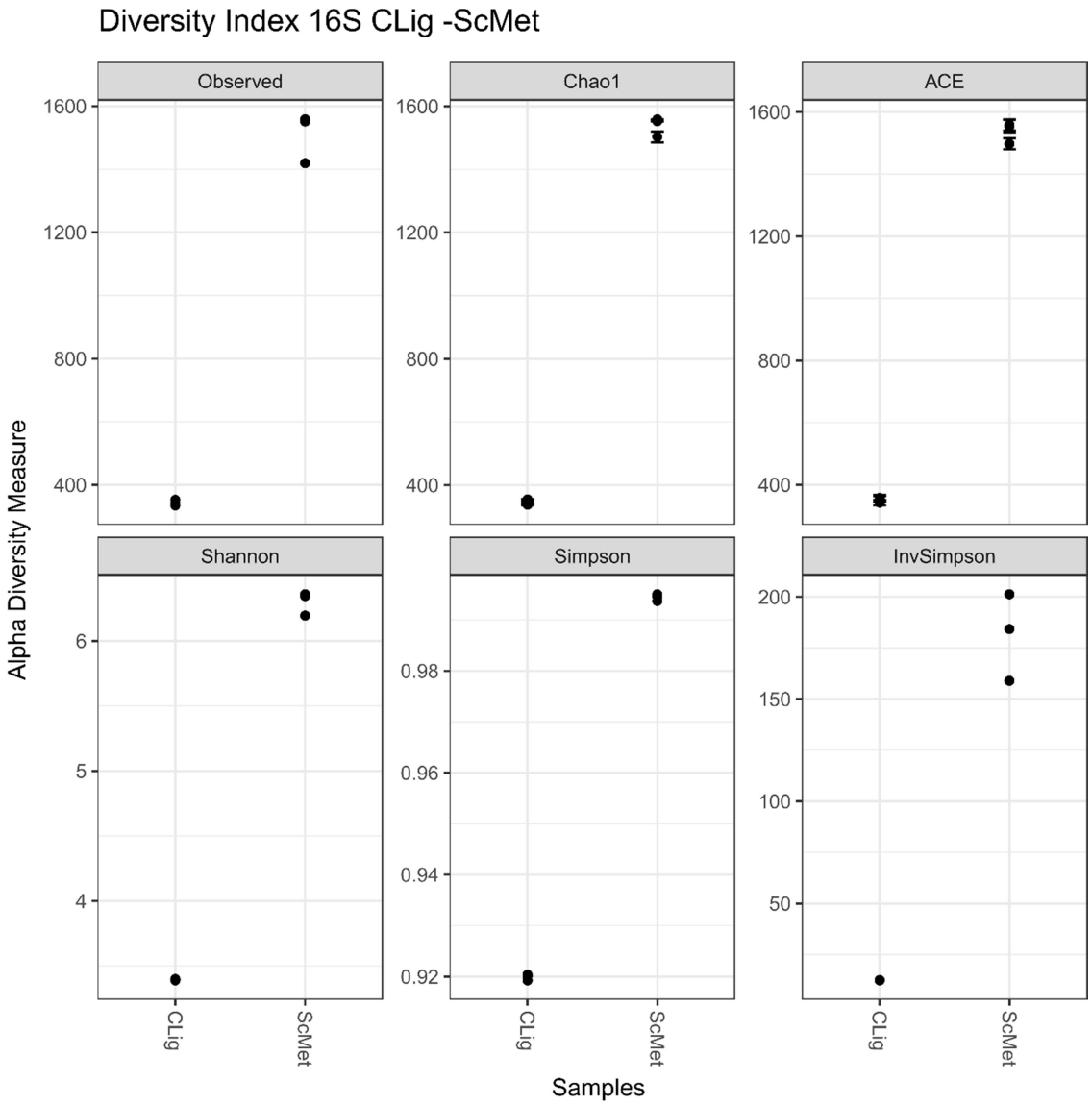


Fig. S4. The Community richness and diversity estimations expressed by the number of OTUs observed, Chao1, Shannon and Simpson indices. OTUs were clustered at 97% of identity using USEARCH pipeline and diversity index were calculated using phyloseq R package.

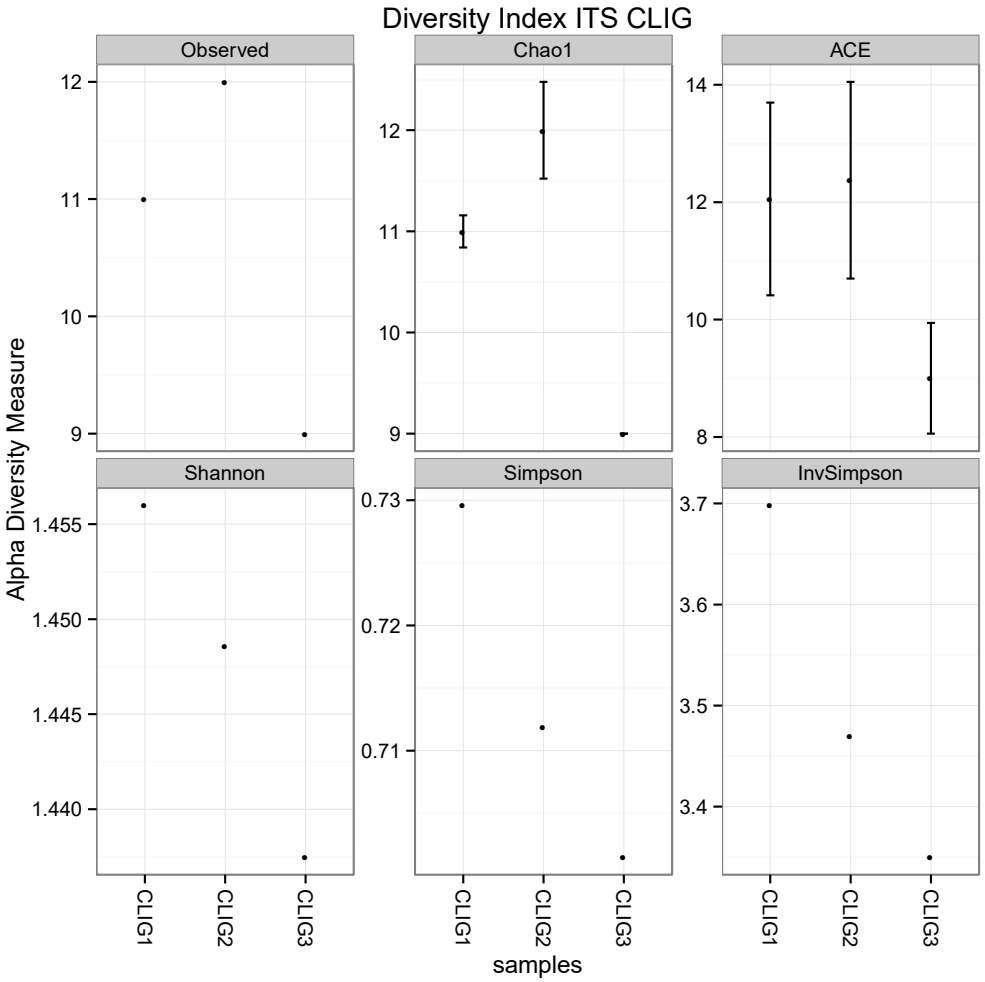


Fig. S5. Fungal community richness and diversity estimations expressed by the number of OTUs observed using Chao1, Shannon and Simpson indices were shown. OTUs were clustered at 97% of identity using USEARCH pipeline and diversity index were calculated using phyloseq R package.

Microbial community taxonomic composition

In order to evaluate the taxonomic composition of CLig and ScMet, the OTU representative sequences were assigned from the phylum to genus level. Figure 2 shows the ScMet and CLig bacterial phylogenetic profile. A total of 17 phyla were identified in ScMet amplicon data and the most abundant phyla were *Proteobacteria* (29 % of the total reads), *Acidobacteria* (22 %) and *Actinobacteria* (15 %) (Fig. 2A). At the class rank, *Alphaproteobacteria* (15 %), *Actinobacteria* (14 %) and *Acidobacteria* (14 %) were predominant in CLig (Additional file 1: Fig. S6). The overall taxonomic profile of ScMet corroborated with previous sequencing studies addressing soil microbial communities [47-49].

The CLig exhibited a significant difference in phyla relative abundance compared to ScMet. The CLig 16S sequences of *Proteobacteria* origin increased from 29% to 58%, and these were assigned mainly to the *Betaproteobacteria* (25 %) and *Alphaproteobacteria* (24 %) classes (Fig. 2A). On the other hand, the relative abundance of 16S sequences of the *Acidobacteria* class was less than 0.1% in the CLig and about 21% in the ScMet (Fig. 2A). Members of *Actinobacteria* and *Firmicutes* were the second and the third predominant phyla in CLig with 20% and 12% of the identified sequences, respectively. Among the bacterial families*, Alcaligenaceae* (23 %), *Micrococcaceae* (10 %), *Phyllobacteriaceae* (9 %), and *Paenibacillaceae* (6 %) were significantly enriched in the CLig compared to ScMet (Fig. 2B). At the genera level, the majority of sequences in the CLig were assigned to *Achromobacter*, *Paenarthrobacter*, *Pseudaminobacter* and *Paenibacillus*, representing almost half of all of the 16S sequences retrieved. Furthermore, eight families represented about 30% of all 16S reads of CLig: *Pseudomonadaceae, Xanthobacteraceae, Microbacteriaceae, Xanthomonadaceae, Mycobacteriaceae, Sphingobacteriaceae, Acetobacteraceae* and *Planococcaceae* (Fig. 2B). Interestingly, the predominant family in CLig (*Alcaligenaceae*) was virtually non-existent in the ScMet (less than 0.5 %). It should be noted that the major phylogenetic groups found in CLig, such as the *Betaproteobacteria*, *Alphaproteobacteria*, and *Actinobacteria* classes, were frequently reported as degraders of lignin and/or lignin-derived aromatic compounds [17][50-54].

The fungal kingdom was represented in CLig by *Basidiomycota* and *Ascomycota* members (Fig. 2C). On average, approximately 94% of all the analyzed sequences could be assigned at the genus level (Fig. 2D). Among them, *Candida* sp. (39%), followed by *Rhodosporidium* sp. (26%), *Trichosporon* sp. (20%), and *Cyberlindnera* sp. (9%) were those of relative dominance. All these genera were previously reported as being able to metabolize aromatic compounds [55-57]. For example, *Candida tropicalis* was described as being capable of degrading phenolic compounds [56,58,59], and *Trichosporon* and *Cyberlindnera* were reported as dye-decolourizing yeasts [57], as well as, displaying genes consistent with the lignin degradation system that was similar to the basidiomycete *Phanerochaete chrysosporium* [60].

The communities had few sequences assigned to the *Archaea* domain, such as OTUs assigned to the phylum *Thaumarchaeota*, which was identified in ScMet and CLig with 0.26 % and 0.02 % of total relative OTU abundance, respectively (data not shown). The phylum *Euryarchaeota* was identified only in the CLig (0.01 % relative abundance, data not shown). Moreover, sequences affiliated with the protozoan class *Litostomatea* were detected in CLig based on 18S rDNA analyses (data not shown). According to Simek et al. [61] and Jürgens et al. [62], the taxonomic structure of microbial communities can be shaped by protozoa due to its preferential predation of particular bacterial taxa.

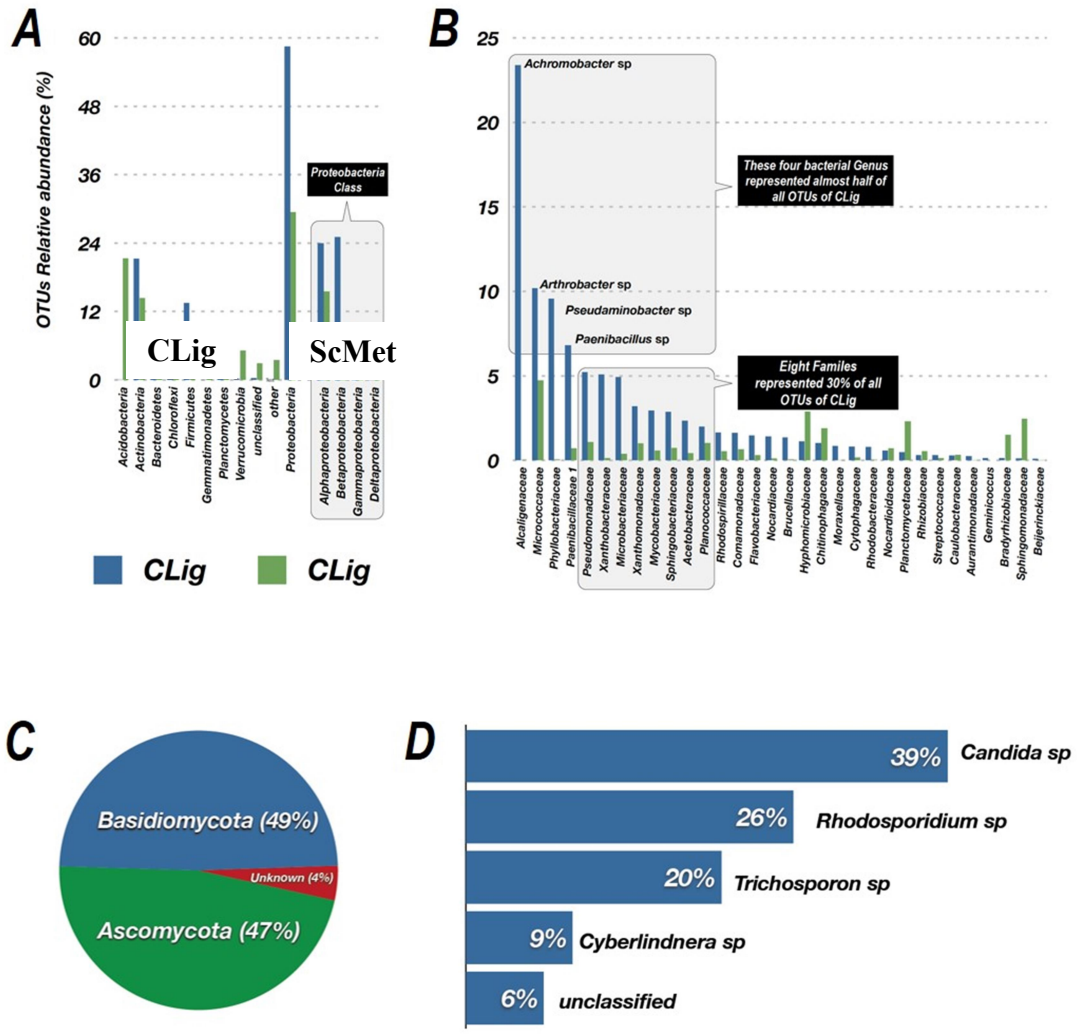


Fig. 2. Taxonomic composition of microbial communities from Clig and ScMet. **A,** Bacterial composition at the phylum level. **B**, The high abundance bacterial families from CLig and ScMet. **C and D**, Fungal communities at the phylum and species level from CLig based on ITS2 region sequencing, respectively.

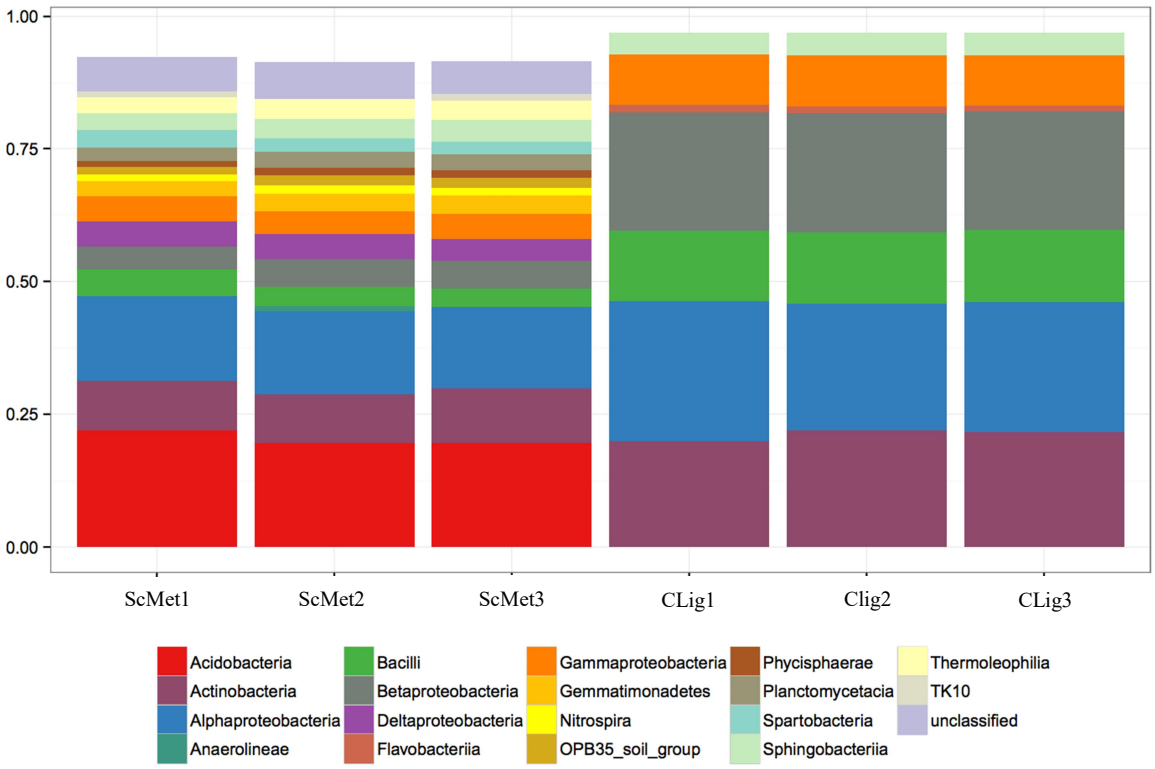


Fig. S6. The taxonomic profile from CLig and ScMet at class level based on 16S rRNA gene amplicons. The respective relative abundances of the replicate for each CLig and ScMet are shown.

CLig metagenome global metabolic profile

A total of 97.5 million high-quality paired-end reads (~ 18 gigabase of sequences) were generated from the CLig sample. Assembled reads resulted in approximately 260 megabases (Mb) of contiguous sequences ≥ 1 Kilobase (Kb). The N50 statistic from the assembled data was 35.5 Kb and the longest contig assembled reached up to 1.3 Mbp. Considering the complete assembly, 282,237 gene models were predicted (protein-coding regions), where 243,896 of these were larger than 300 bp.

The functional annotation was determined by different methods, where all predicted proteins obtained from the metagenome dataset were annotated using EggNOG, KEGG, Pfam and dbCAN databases. According to the results, 242,299 unigenes (85.8 % of the predicted proteins) were assigned based on the EggNOG database, 129,140 (45.7 %) based on KEGG, 208,248 (73.78 %) on Pfam and 8,800 (3.1 %) had at least one protein domain annotated in dbCAN. The functional annotation obtained based on EggNOG and

KEGG was quite similar (Fig. 3); the majority of proteins were classified as belonging to the metabolism category, which harbors a variety of pathways involved in degradation of aromatic compounds. According to the KEGG and Pfam annotation, several aromatic compound degradation pathways were found complete in the metagenome dataset, such as benzoate degradation (or methylbenzoate) to catechol (or methylcathecol), catechol ortho-cleavage, catechol meta-cleavage and phthalate degradation (data not shown).

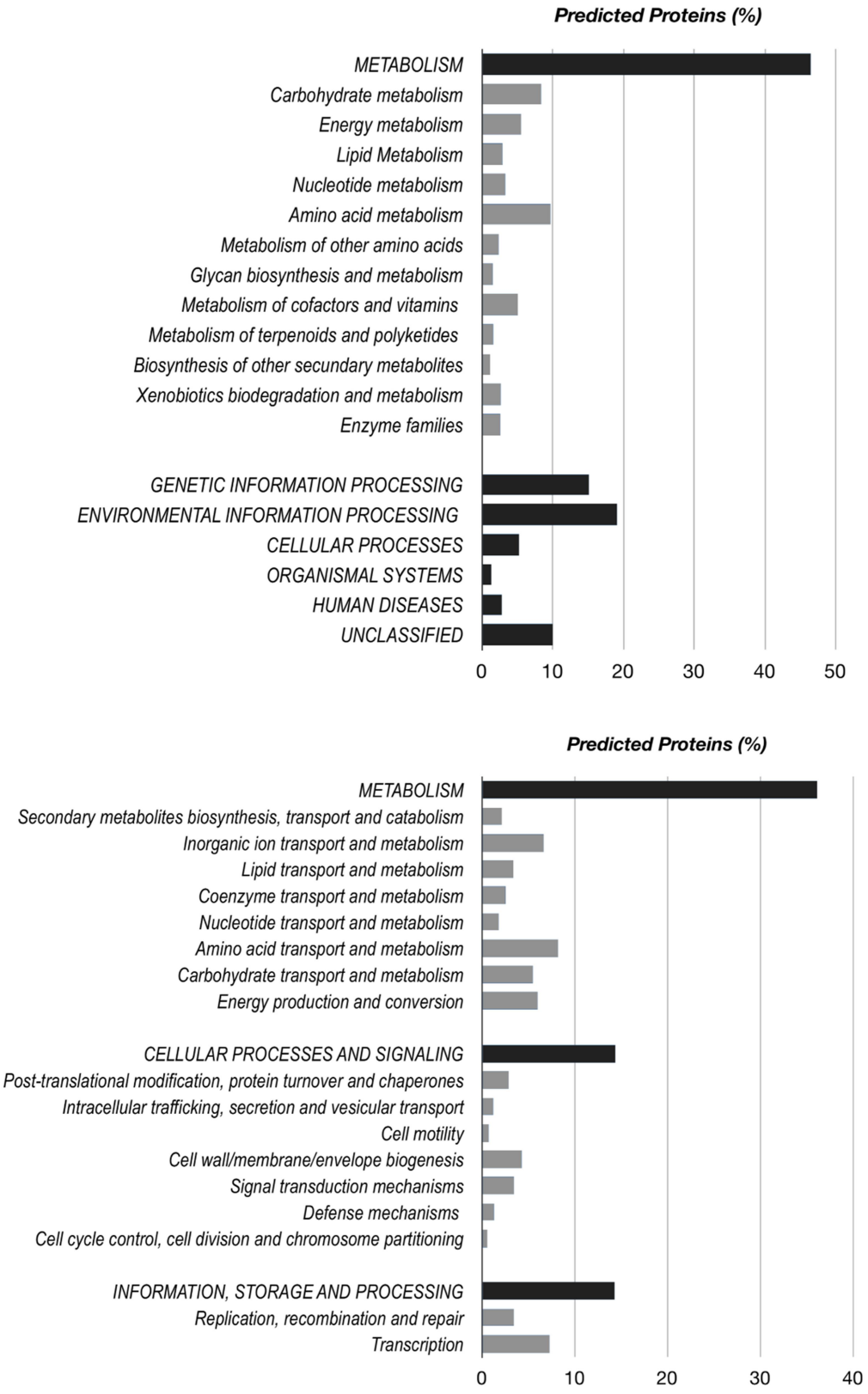


Fig. 3: Summary of the annotated predicted proteins from the CLig Metagenome according to KEEG (upper graphic) and eggNOG (lower graphic).

Reconstruction of draft genomes from metagenomic dataset

The assembly of metagenomic data resulted in the reconstruction of 65 draft bacterial genomes, of which 20 presented completeness greater than 75 % and contamination less than 10% (Table 1). The taxonomic assignments show that most of the genome bins were assigned to *Alphaproteobacteria* (9 bins), followed by *Actinobacteria* (6 bins), and *Betaproteobacteira* (2 bins). The remaining genome bins were assigned to *Gammaproteobacteria*, *Bacilli*, and *Sphingobacteriia*. These results are in accordance with previous studies that reported *Rhodococcus*, *Sphingobacterium*, *Ochrobactrum*, *Bacillus*, *Sphingobium* and *Arthrobacter* as microorganisms that harbor lignolytic enzymes and metabolize aromatic compounds [17,63,64]. Twenty genomes bins were further analyzed to depict their metabolic potential on lignin and aromatic compound degradation.

Lignin and aromatic compounds degradation pathways

Pfam-based analysis identified several conserved domains related to lignin and aromatic compound degradation in the CLig metagenome dataset and genome bins (Fig. 4). Among these, peroxidases (PF00141), dye-decolorizing peroxidases (PF04261) and laccases (PF00394, PF02578, PF07731, PF07732) domains, and enzymes that can generate radicals for cleavage of several lignin linkages, were found predominantly in genome bins belonging to the *Actinobacteria* class (bins 48,7,17, and 65) (Fig. 4). These findings corroborated previous studies reporting that *Actinobacteria* is able to produce extracellular lignolytic enzymes [17, 65, 66].

Glutathione-dependent beta-etherases of *Sphingobium* spp were recently described as able to cleave beta-aryl ether bonds of lignin from softwoods and hardwoods [67]. Corresponding domains for the beta-etherases LigE, LigP, LigG and LigF (considering the PF13417, PF00043 and PF02798) were predominantly found in bins 4, 20, 28, 35, and 37, which were assigned with *Alpha* and *Betaproteobacteria* origin (Fig. 4). Conversion of the beta-aryl ether lignin dimer involves LigEFG and LigD (PF00106) and results in vanillic acid [16], which can also be generated from vanillin by the action of LigV (PF00171). In this pathway, vanillic acid is converted into protocatechuic acid by LigM (PF01571, PF08669). All these protein domains (LigEFG, LigD, LigV and LigM) were identified in several bins,

indicating their potential to metabolize vanillin and vanillic acid. Nonetheless, they were found predominantly in bins assigned to *Mycobacterium* (bin 7).

The biphenyl linkage is also a common component of lignin structure. Its cleavage is carried out by the cascade pathway LigX (PF0035), LigZ (PF02900), and LigW, LigW2 and LigY (PF04909) [16], also resulting in the central intermediate vanillic acid. With the exception of bins 6 and 28, several *Actinobacteria* genome bins displayed the complete set of proteins needed for biphenyl and vanillic acid conversion (Fig. 4). However, as shown in Figure 4, this cleavage cascade pathway was identified in only a few bins of *Proteobacteria* origin (20, 37 and 3).

The depolymerization of lignin releases a mixture of aromatic monomers that can be used as carbon and energy sources by several microorganisms [68,69]. Phenolic aromatic monomers can be converted into metabolic intermediates, via catechol and protocatechuate, by the action of dioxygenases, which are classified according to the relative positions of hydroxyl groups (ortho- and meta-cleavage) [16]. The routes can be divided in three blocks: (i) the branch of catechol intermediate (ortho-cleavage), which involves the following enzymes: CatA (PF0775 and PF04444), CatB (PF02746 and PF13378) and CatC (PF02426), (ii) the branch of protocatechuate (meta-cleavage), which involves PcaG (PF00775), PcaH (PF00775 and PF12391), PcaB (PF10397 and PF00206) and PcaC (PF02627), and finally (iii) the reactions common for both branches, catalyzed by PcaD (PF00561), PcaI (PF01144), PcaJ (PF01144) and PcaF (PF02803 and PF00108). Three *Actinomycetes* genome bins (17, 48 and 65) and two of *Proteobacteria* origin (4 and 20) presented all protein domains related to routes i, ii and iii. Regarding route (ii), all *Actinobacteria* genome bins and two of *Proteobacteria* origin (bins 20 and 35) were clearly enriched with coding genes involved in protocatechuate degradation. With exception of the *Sphingobacterium* genome bin (24) and the *Bacillus* bin (30), the protein domains corresponding to enzymes related to route iii were mapped in all bins.

The present work disclosed several novel aromatic-degrading enzymes of a high and low degree of homology to previously identified variants. For instance, the BLASTp comparison with beta-etherase (LigE) from *S. paucimobilis* SYK-6 identified 18 orthologous in CLig, with sequence identity varying from 31 to 64% (e value from 4.0 E^{-30} to 0.0). Accordingly, CLig orthologous to beta-etherase LigF (26 hits), Glutathione S-transferase

homolog LigG (3 hits) and beta-etherase LigP (18 hits) from *S. paucimobilis* were recovered as well, with sequence identity varying from 26 to 93% (e value from 9 E^{-11} to 4 E^{-178}), 29 to 84% (e value from 3.0 E $^{-18}$ to 4.0 $E^{-164)}$ and 31 to 88% (e value from 4.0 E^{-30} to 0.0), respectively.

Moreover, to obtain deeper insight on potential of the microbial community to carry out redox and hydrolytic mechanisms related to lignocellulose degradation, the auxiliary activity (AA) and carbohydrate esterase (CE) families were predicted based on dbCAN analysis (Additional file 1: Fig. S8). The carbohydrate esterases from family 1 (CE1), including feruloyl esterases and p-coumaroyl esterases, can break down ester cross links of lignin and hemicelluloses [70, 71]; and glucuronyl esterases (CE15) were reported to catalyze ester linkage hydrolysis between glucuronoxylan and lignin [72]. Furthermore, acetyl esterases (described in families CE01, 02, 03, 04, 05, 06, 07, 12 and 16) play a role in degradation of xylose units esterified with acetic acid [71]. Nonetheless, only the families CE1 and CE4 were predominant in the CLig metagenome dataset and widespread in all genome bins (Additional file 1: Fig. S7 and S8). Similar to CE families, AA were abundant in the CLig metagenome, notably the AA3 family members which may be related to peroxidase activity, thus corroborating with the previous PFAM-based functional analysis.

Table 1: Assembly Statistics from draft genomes binned from CLig Metagenome (20 best results)

Bin Id	Phyla Amphora Taxonomy	Completeness (%)	Contamination (%)	Heterogeneity (%)	Total length (bp)	# contigs	# Predicted genes	GC (%)
3	*Sphingomonadaceae*	97.85	2.1	0	2,600,224	14	2487	63.85
4	*Alcaligenaceae*	96.64	3.15	8.33	4,071,016	139	3831	63.9
6	*Microbacteriaceae*	93.03	2.3	25	2,775,244	23	2744	70.21
7	*Mycobacterium* sp.	98.15	4.7	7.41	8,113,983	152	7824	67.03
14	*Agrobacterium* sp.	97	1.88	0	3,723,529	25	3637	62.37
17	*Rhodococcus* sp.	96.42	1.35	12.5	5,940,093	80	5417	67.93
19	*Rhodobacter* sp.	97.47	3.53	0	4,829,944	102	4636	66.36
20	*Comamonadaceae*	76.05	1.77	0	5,317,169	797	5668	69.45
24	*Sphingobacterium* sp.	92.2	2.22	0	5,342,826	905	5155	40.87
28	*Ochrobactrum* sp.	99.09	10.5	0	8,098,959	173	7885	56.49
30	*Bacillus* sp.	92.62	3.65	0	4,962,376	118	4977	35.39
35	*Acetobacteraceae*	96.72	1.99	0	5,217,664	48	5050	70.22
37	*Sphingobium* sp.	98.08	9.27	5.66	6,313,527	428	6265	59.7
47	*Pseudoxanthomonas* sp.	97.33	1.46	0	3,192,941	22	2886	71.25
48	*Arthrobacter* sp.	96.96	4.12	0	4,259,515	57	3930	68.16
55	*Arthrobacter* sp.	97.31	1.59	0	4,537,449	80	4185	63.52
56	*Acetobacteraceae*	92.74	1.66	25	5,069,773	252	4671	70.35
60	*Hyphomicrobium* sp.	97.93	0.86	0	3,563,628	12	3278	64.65
64	*Methylobacterium* sp.	94.8	1.62	25	3,854,043	176	3682	70.11
65	*Rhodococcus*	96.8	1.75	0	5,770,250	74	5292	70.8

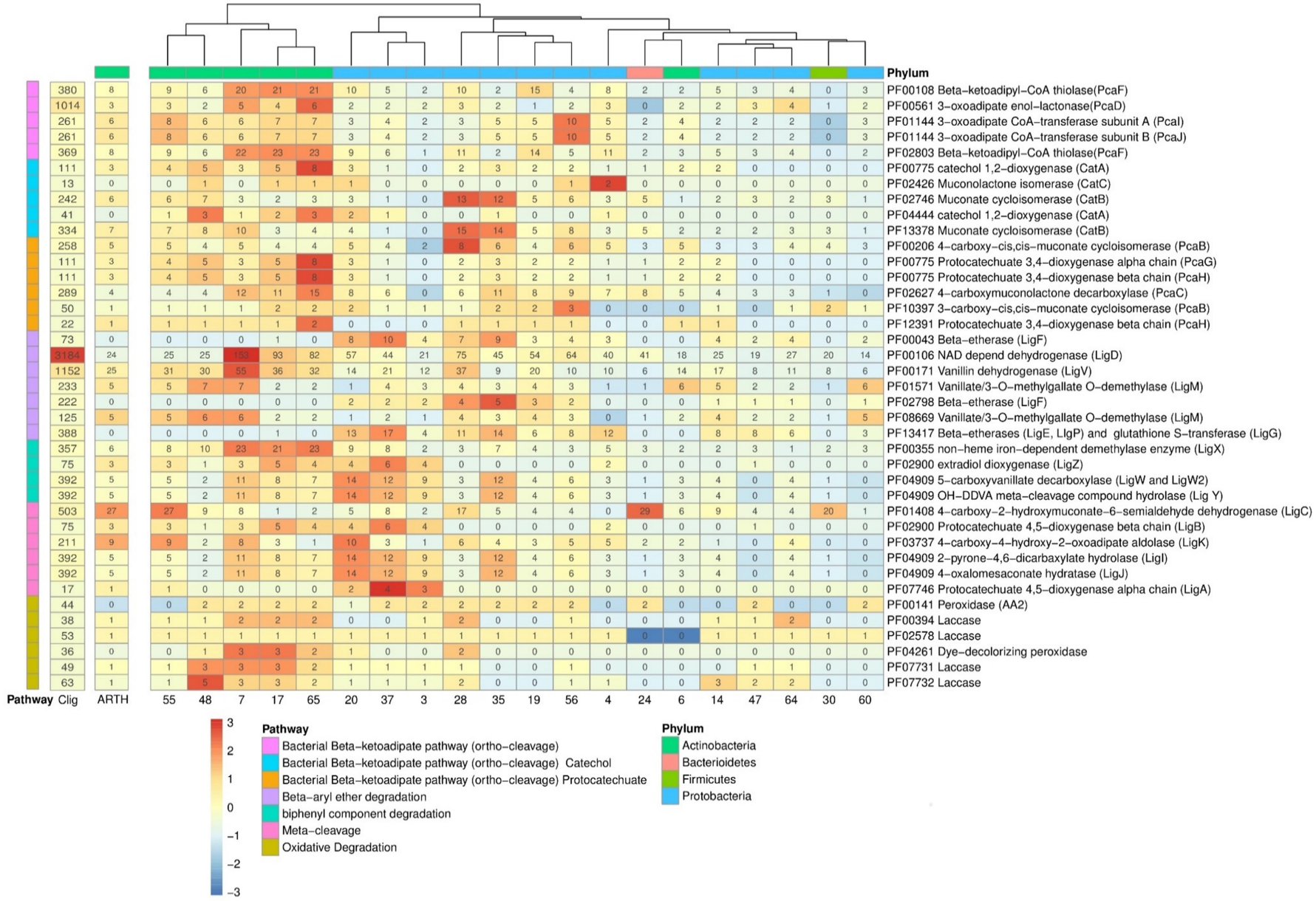


Figure 4. Pfam domains related to aromatic compounds degradation pathways identified in genome bins and CLig. Panel shows (from left to right): the aromatic compounds degradation pathways are indicated in different colors; total numbers of domains identified in the CLig according to the Pfam domain family; Pfam domains identified in strain HW13 (ARTH); numbers of Pfam domains identified each genome bins; name of Pfam family. Above and below main (or large) panel: multilocus phylogenetic analysis tree based on single copy conserved genes found across genome bins; identification number of each genome bins (only with completeness > 75 %).

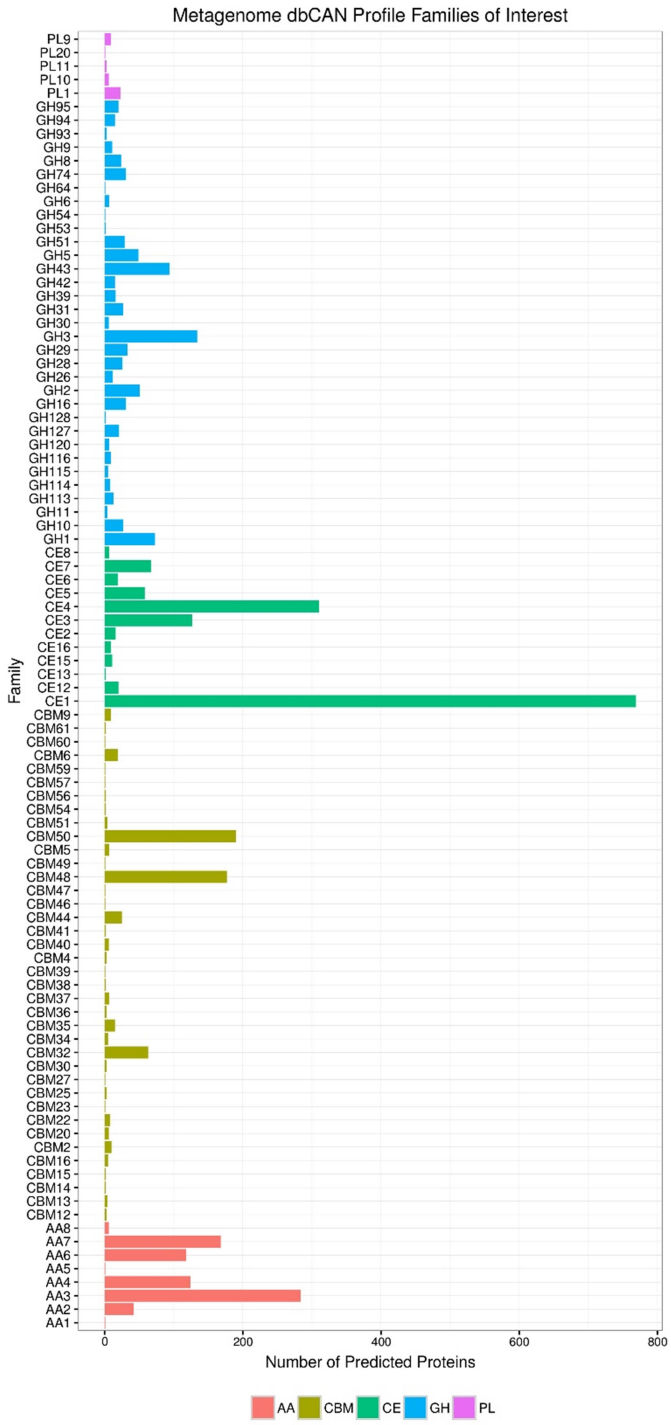


Fig. S7. Classification of the predicted proteins from CLig metagenome according to dbCAN database.

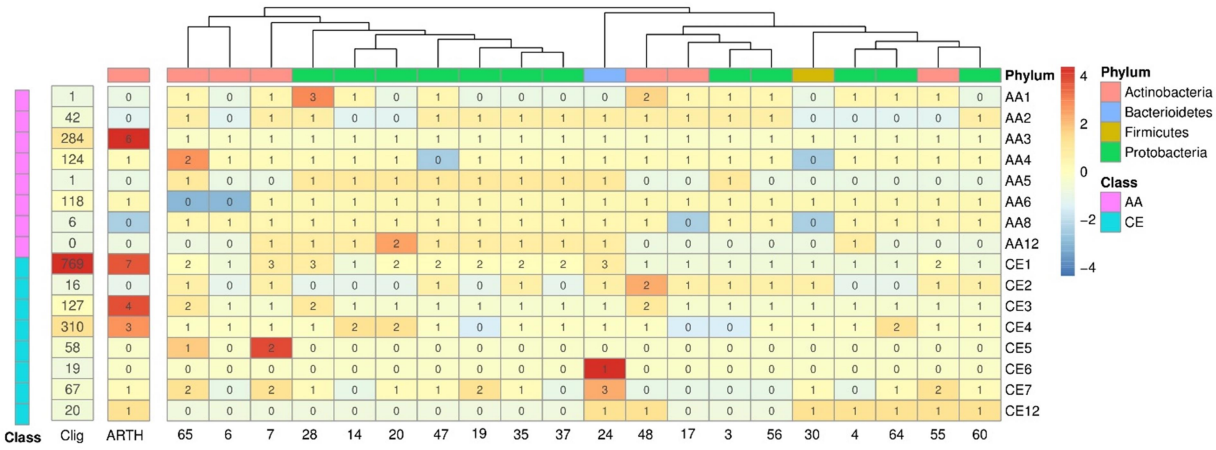


Fig. S8. Classification of the predicted proteins from draft genomes binned from CLig Metagenome based on dbCAN database. ARTH: strain HW13.

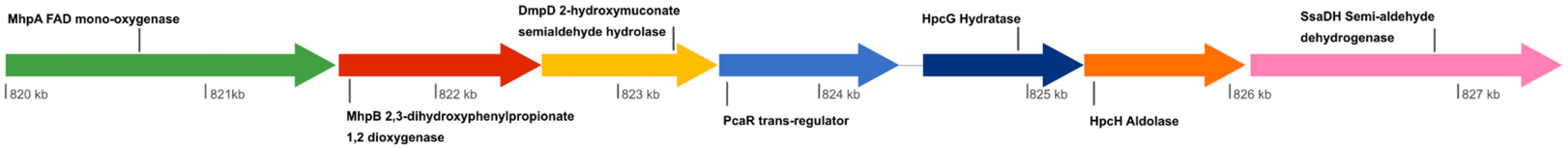
Paenarthrobacter sp. str. HW13 genome

To complement the metagenomic discovery approach, the isolation of novel lignin-degrading microorganisms was performed using HW as the sole carbon source in media. A bacterial strain showing capacity to grow on HW lignin was obtained and designated as strain HW13. Based on the 16S rRNA gene comparison, the analysis revealed that strain HW13 shares 96.8 % 16S rRNA gene sequence similarity with *P. ureafaciens* DSM 20126^{T} as the closest related species with a valid nomenclature (Additional file 1: Fig. S9). Therefore, the strain HW13 may represent a new species within the *Paenarthrobacter* genus.

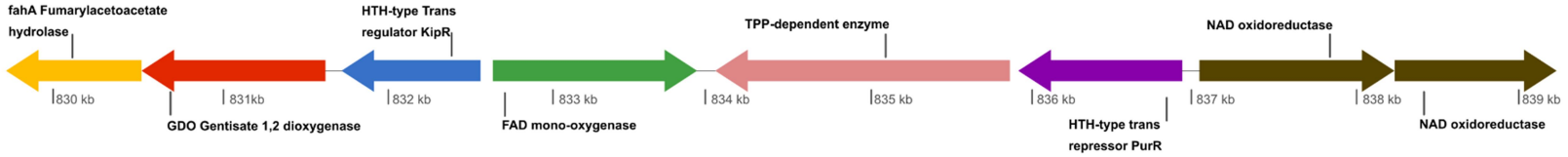
The genome sequence of strain HW13 depicted molecular features that explain the capacity of the strain to metabolize lignin fragments and phenolic compounds (Fig. 5). The paired-end MiSeq sequencing run yielded 4,091,031 bases of total sequence information. After data processing, the assembly resulted in three contigs. The strain HW13 draft genome size was estimated as 4,091 Mb, featuring a GC content of 63 %. Additional details of the genome features can be found in Additional file 1: Table S3. To determine whether strain HW13 correspond to bin 48 or 55 (both assigned to the *Paenarthrobacter* genera), the Microbial Species Identified (MiSi) method [73] was conducted, as implemented in IMG/ER. The P probability of bins 48 and 55 being assigned to the same species as HW13 was 0.0 and 0.99, respectively. Strain HW13 showed 100 % 16S rRNA gene sequence identity to the 16S

rRNA gene found in genome bin 55. The genomic analysis confirmed the prediction of several previously identified genetic determinants involved in aromatic compound metabolism, including the protocatechuate, catechol, phenylacetate, gentisate, and phenylpropionate degradation pathways (Fig. 5).

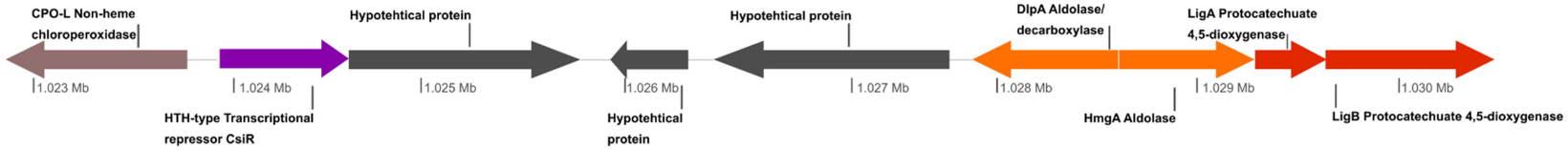
Phenylacetate or phenylpropionate



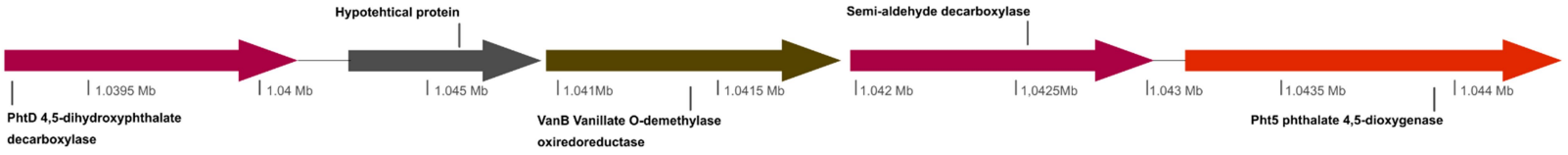
Gentisate/salicylic acid



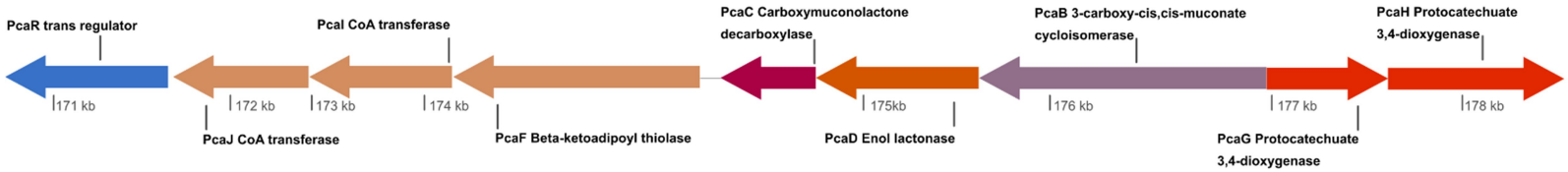
Meta-cleavage



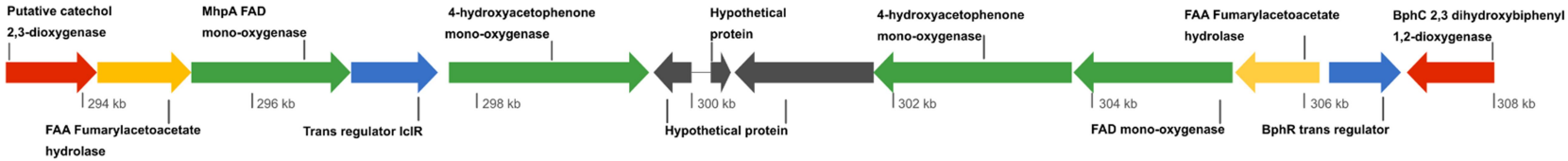
Phthalate



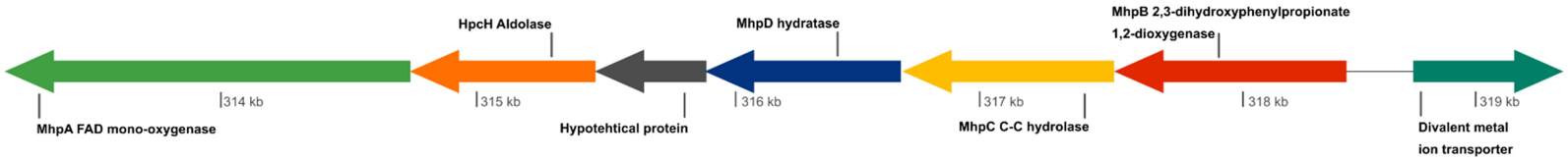
Beta-ketoadipate



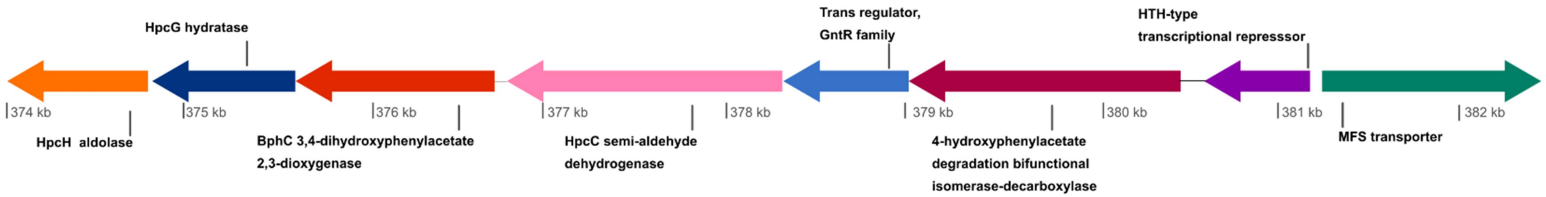
Possible biphenyl



Phenylpropionate



Phenylacetate



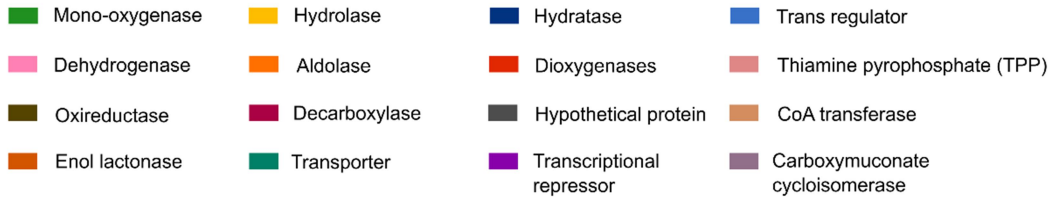


Fig. 5. Schematic representation of gene clusters involved in aromatic degradation in *Paenarthrobacter* sp. HW13. The predicted gene products are indicated.

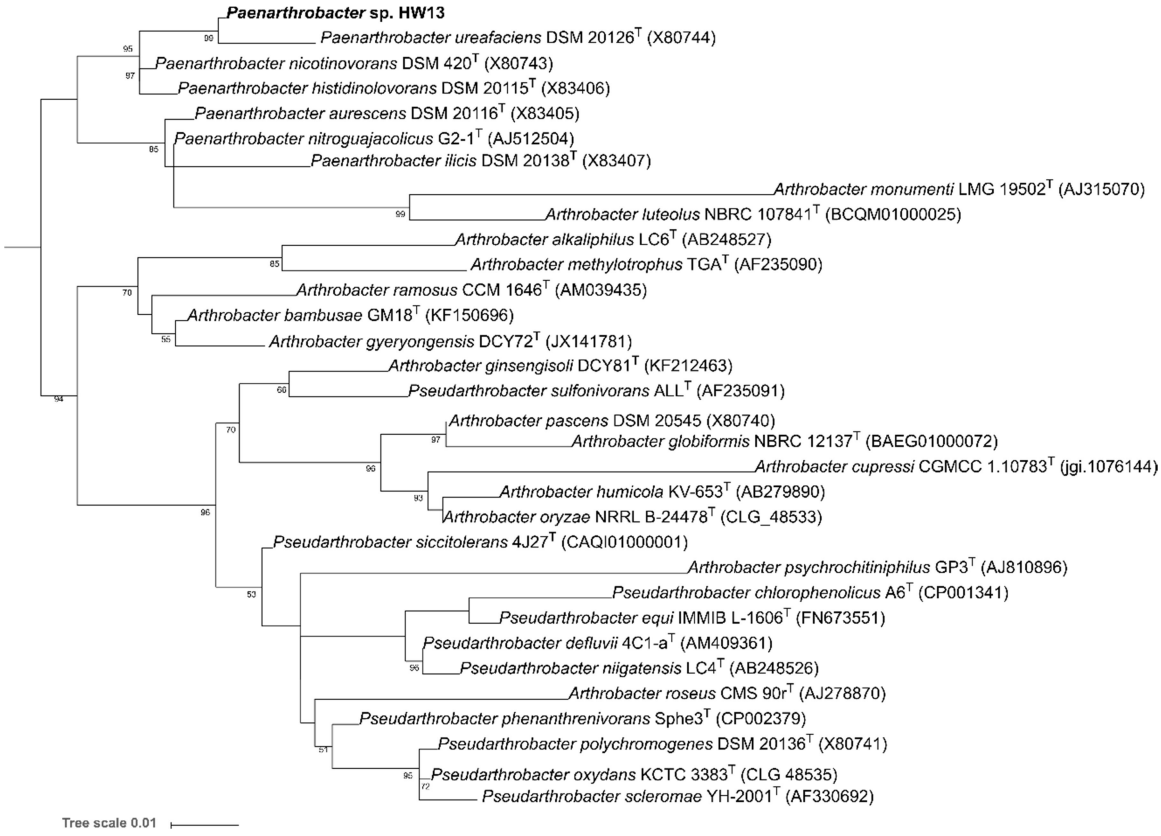


Fig S9. Phylogenetic position of strain HW13 relative to the most closely related strains of the genus *Paenarthrobacter*. EzBioCloud webserver was used to perform a similarity-based search of HW13 16S rRNA to retrieve the most closely related sequences. The resulting 16S rRNA sequences were aligned using Mafft v7.299b software. A phylogenetic tree was inferred using maximum likelihood method implemented in RAxML v8.2.0, evolutionary distances were based on the GRTGAMMAI model, inferred as the best model by jModelTest2. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. The type strains are marked with a superscript 'T'. Accession numbers are shown in parentheses.

Table S3. Genome statistics for *Paenarthrobacter* sp. str. HW13

Feature	Chromosome
Total length (bp)	4,091,031
Scaffold	3
Largest contig (bp)	2,064,645
GC (%)	63.43
Total number of genes	3,797
Protein coding genes	3,731
Protein coding genes with function prediction	3,033
rRNA operons	6
tRNAs	52
Gene space completeness	99.71%

Biosynthesis of vanillin from ferulic acid

The occurrence of aromatic conversion pathways in CLig to produce high value chemicals was validated through biochemical vanillin production. Ferulic acid is found in lignocellulosic biomass, and it is a precursor for biovanillin production [74,75]. The conversion of ferulic acid into vanillin has been reported in several microorganisms via coenzyme A-dependent, a non-beta-oxidative pathway [40, 76-79], including feruloyl-CoA synthetase (*fcs*/FerA) and enoyl-CoA hydratase/aldolase, (*ech*/FerB) [80], which are regulated by FerC, a MarR-type transcriptional regulator [81]. In this study, the protein domains related to FerA (PF13380) and FerB (PF00378) were found in several bins described in Fig. 4. The clusters of genes related to ferulic acid conversion into vanillin were manually identified in the CLig, two presenting similar genetic structures previously described in *Pseudomonas fluosrescens* BF13 [81] and *Sphingobium paucimobilis* SYK-6 [77] (Fig. 6). Additionally, another gene cluster were found in CLig of similar organization to the gluconate operon of *Bacillus subtilis* [83], which is regulated by a GntR family protein Fig. 6.

Candidates genes for FerA (derived from bin 3) and FerB (derived from bin 11) were selected for further analyses. The FerA_B3 (bin 3) when compared to other bacterial feruloyl-CoA synthetases, displays 60% amino acid identity to *Sphingomonas paucimobilis* (A0A031JK39), followed by *Sphingobium sp* (60%; BAK67177), *Delftia acidovorans* (55%; CAC83622) and *Burkholderia glumae* (52%; ACR31088), and. The FerB_B11 (bin 11) compared to other bacterial enoyl-CoA hydratases, shows 63% amino acid identity to *Pseudomonas nitroreducens* (C3VA24), followed by *Amycolatopsis methanolica* (59%;

A0A076MUC3), *Streptomyces sp* (55%; S5LPF1), and *Delftia acidovorans* (55%; Q8VNW7). Amino acid sequences from NCBI database and Uniprot with similarity higher than 20 % to FerA_B3 and FerB_B11 were considered to construct phylogenetic trees (Fig. S10). It evidenced the relationship FerA_B3 and FerB_B11 with other bacterial feruloyl-CoA synthetases and enoyl-CoA hydratases, respectively, but the low bootstrap value at most of the clades denoted the high divergence of FerA_B3 and FerB_B11.

For the bioproduction of vanillin, candidates genes for *FerA_B3* and FerB_B11 were then synthesized and successfully cloned for expression in *E. coli,* followed by purification by liquid chromatography. The conversion of ferulic acid into vanillin, after combining the two purified enzymes, was confirmed by GC/MS. The vanillin production from ferulic acid was detected after 6 h and 24 h incubation using the purified recombinant proteins (Fig. 7). The biochemical and structural characterization of the two proteins, FerA_B3 and FerB_B11, will be the focus for future studies.

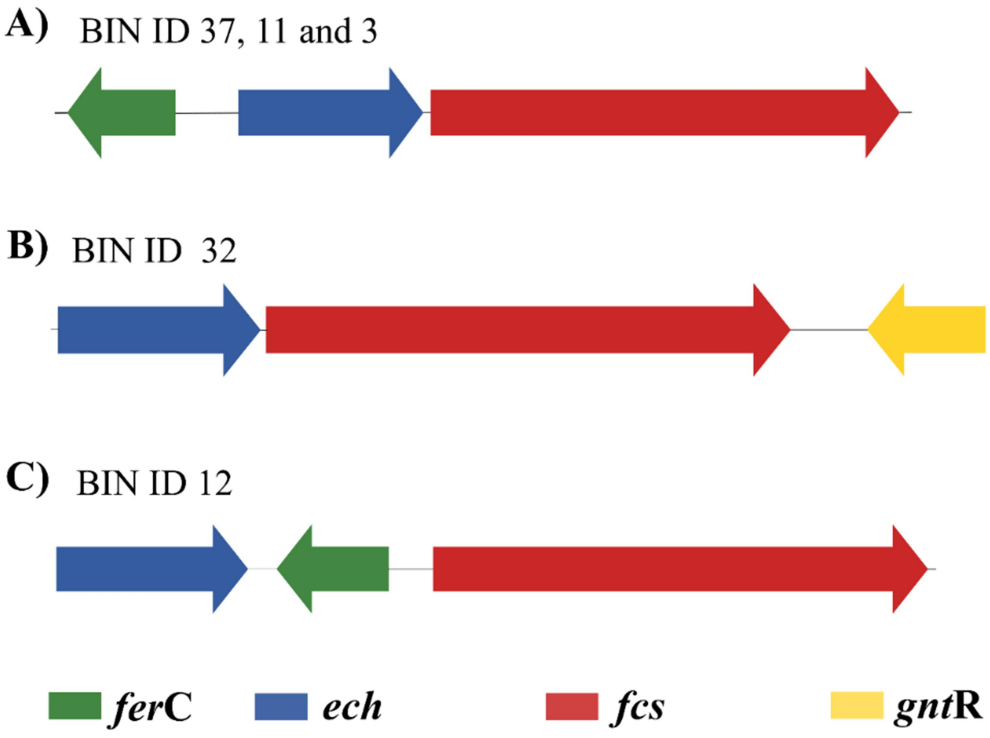


Fig. 6. The structure of *FerA* and FerB gene clusters identified in the CLig dataset. The structure of *FerA* and FerB gene clusters described in (A) *Sphingobium sp*, (B) *Bacillus subtilis*, which regulator gene belongs to the GntR family, and (C) *Pseudomonas sp* with different arrangement of the regulator (FerC). All these three gene clusters configuration were identified in the CLig [17-19]. The CLig genome bins presenting similar gene cluster configurations compared to **A**, **B** and **C** were depicted. The bins 37, 11, 3, 32, and 12 were assigned to *Sphingobium* sp., class *Alphaproteobacteria*, family *Sphingomonodaceae*, order *Actinomycetales* and order *Rhizobiales*, respectively. Homologous genes are shown in the same color, and the predicted gene names are indicated at the bottom.

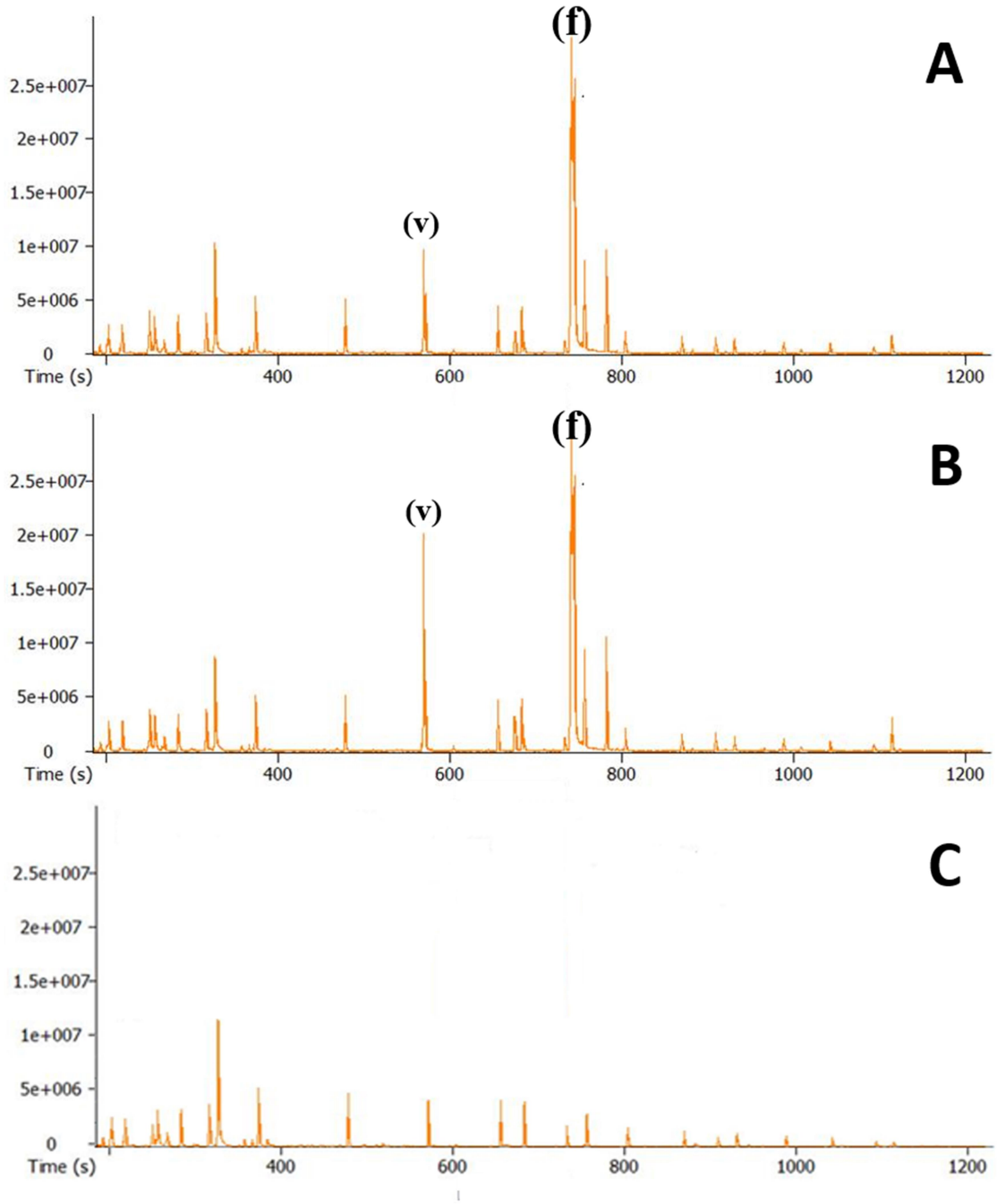


Fig. 7. GC-MS chromatograms of products from the reaction of ferulic acid into vanillin by the enzymes FERA and FERB. The Y-axis represents relative abundance (ion count) and the X-axis represents retention time (in seconds). Vanillin and ferulic acid peaks are depicted in chromatograms A and B by the letters (v) and (f), respectively. A, 6 hours reaction; B, 24 hours reaction. C, Blank.

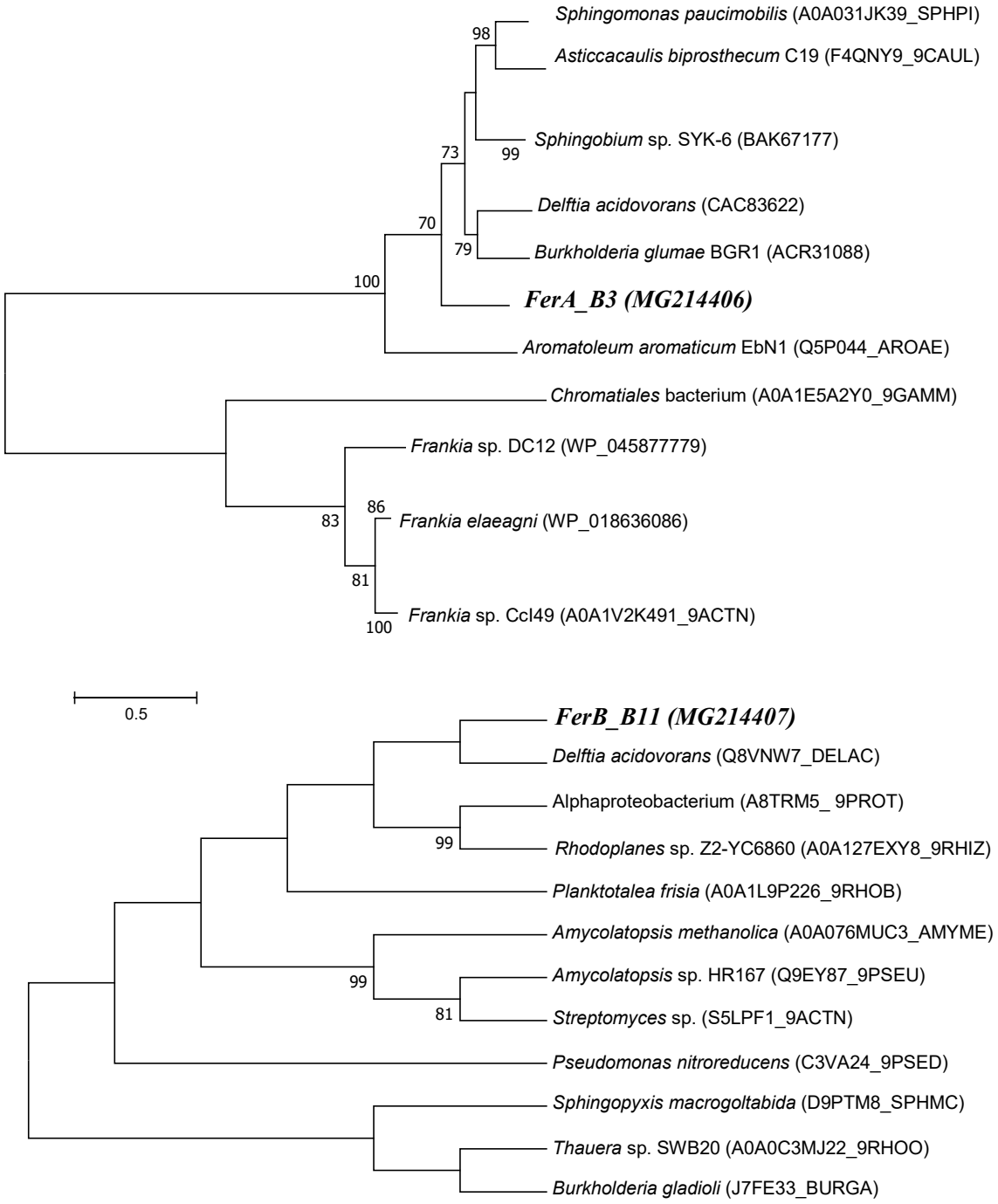


Fig S10. Phylogenetic relationships among feruloyl-CoA synthetase (upper) and Enoyl-CoA hydratase/aldolase. The phylogenetic tree was generated using amino acid sequences retrieved from NCBI and Uniprot database. The sequences were aligned using Mafft v7.299b software. The phylogenetic tree was reconstructed using maximum likelihood method implemented in RAxML v8.2.0, evolutionary distances were based on the GTRGAMMA model, inferred as the best model by jModelTest2. The bootstrap values (1,000 replicate runs, shown as %) higher than 70 % are represented. Accession numbers are listed in parentheses. The FerA_B3 and FerB_B11 amino acid sequence retrieved from CLig dataset is printed in bold.

Conclusions

The metagenomic discovery platform described in the present study, based on targeted community enrichment protocols, succeeded in depicting the major bacterial pathways involved in lignin degradation and aromatic compounds conversion, and also led to the identification of new lignin-oxidizing enzymes that could be expressed in recombinant form to produce vanillin. Utilization of complex microbial environments, such as soil as the starting material, are of special interest for engineering a specialized microbial consortium, because soil samples display countless versatility regarding phenotypic traits [82].

To validate applicability of the metagenomic strategy different approaches were conducted. From the analysis of massive DNA sequencing data, several microorganisms and enzymes linked to lignin degradation metabolic pathways were identified in CLig, thus constituting a promising resource of biological information for lignin conversion and valorization. A lignolytic bacteria of potential biotechnological interest was isolated and high-resolution genome sequencing was performed. Furthermore, gene clusters involved in a non-beta-oxidative pathway for vanillin production were depicted, followed by production of recombinant FerA and FerB in functional form for successful production of vanillin.

The targeted metagenomic discovery platform described in the present study is of potential interest for diverse fields, and can reveal optimized gene sets and microorganisms for initiatives based on synthetic biology principles. The CLig metagenome represents a vast reservoir of genes in relation to their participation in lignin depolymerization and assimilation. The straightforward metagenomic strategy described in the present study, targeting the discovery of genes and pathways involved in lignin degradation and valorization, could also applied to other fields such as the development of antibiotic producing microorganisms or the recycling of plastic polymers. We expect that our findings could provide a basis for the development of novel technologies based on synthetic biology to solve lignin recalcitrance and develop sustainable strategies to produce high value chemicals from lignocellulose.

Abbreviations

CLig*:* Lignin-degrading consortium*;* ScMet: sugarcane soil; PCR: Polymerase chain reaction; OTU: operational taxonomic unit; RDP database: Ribosomal database project database; ITS: nuclear ribosomal internal transcribed spacer; KEGG: Kyoto encyclopedia of genes and genomes; CAZy: carbohydrate-active enzyme; GH: glycoside hydrolase; AA: auxiliary activities; CE: carbohydrate esterase; HMM: hidden Markov model; GC-MS: gas chromatography mass spectrometry.

Author Contributions

Eduardo C. Moraes, Gabriela F. Persinotti, Douglas A. Paixão, Livia B. Brenelli, Juliana A. Aricetti and Gabriela Ematsu performed the experiments. Timotht D. H. Bugg, Camila Caldana and Neil Dixon contributed with reagents, materials and analyses. Eduardo M. Moraes, Thabata M. Alvarez, Gabriela F. Persinotti, Geizecler Tomazetto and Fabio M. Squina analyzed the results and wrote the manuscript. Neil Dixon and Fabio M. Squina revised the manuscript. All authors read and agreed with the submitted version of the paper.

Acknowledgements

E.C.M.,G.T, LBB. were supported by grants from the São Paulo Research Foundation (FAPESP, proc. numbers 2014/26152-4, 2015/23279-6, 2013/03061-0). ND is supported by the BBSRC David Phillips Fellowship (BB/K014773/1) and BBSRC-FAPESP (FAPPA) grant (BB/L026244/1) and BBSRC-FAPESP (BB/P01738X/1). FMS is supported by the FAPESP grant (2015/50590-4).

Conflict of Interest

The authors declare no conflict of interest.

References

1. Alonso DM, Bond JQ, Dumesic J a. Catalytic conversion of biomass to biofuels. Green Chem. 2010;12:1493–513.

2. Caspeta L, Buijs NAA, Nielsen J. Environmental Science. Energy Environ. Sci. 2013;6:1077–82.

3. Chemical OFTHE, Industry E. Lignocellulosic feedstock biorefinery—the future of the chemical and energy industry. 2009;4:456–7.

4. Zakzeski J, Bruijnincx PCA, Jongerius AL, Weckhuysen BM. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. Chem. Rev. 2010;110:3552–99.

5. Bruijnincx PC a., Weckhuysen BM. Biomass conversion: Lignin up for break-down. Nat. Chem. Nature Publishing Group; 2014;6:1035–6.

6. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, et al. Lignin valorization: improving lignin processing in the biorefinery. Science. 2014;344:1246843. 7. Boerjan W, Ralph J, Baucher M. Lignin Biosynthesis. Annu. Rev. Plant Biol. 2003;54:519–46.

8. Freudenberg K. Lignin: Its Constitution and Formation from p-Hydroxycinnamyl Alcohols: Lignin is duplicated by dehydrogenation of these alcohols; intermediates explain formation and structure. Science 1965;148:595–600.

9. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin biosynthesis and structure. Plant Physiol. 2010;153:895–905.

10. Beckham GT, Johnson CW, Karp EM, Salvachua D, Vardon DR. Opportunities and challenges in biological lignin valorization. Curr. Opin. Biotechnol. 2016;42:40–53.

11. Dias MOS, Cunha MP, Jesus CDF, Rocha GJM, Pradella JGC, Rossell CE V, et al. Second generation ethanol in Brazil: Can it compete with electricity production? Bioresour. Technol. Elsevier Ltd; 2011;102:8964–71.

12. Brenelli LB, Mandelli F, Mercadante AZ, Rocha GJ de M, Rocco SA, Craievich AF, et al. Acidification treatment of lignin from sugarcane bagasse results in fractions of reduced polydispersity and high free-radical scavenging capacity. Ind. Crops Prod. 2016;83:94–103.

13. Rahimi A, Azarpira A, Kim H, Ralph J, Stahl SS. Chemoselective metal-free aerobic alcohol oxidation in lignin. J. Am. Chem. Soc. 2013;135:6415–8.

14. Rahimi A, Ulbrich A, Coon JJ, Stahl SS. Formic-acid-induced depolymerization of oxidized lignin to aromatics. Nature. Nature Publishing Group; 2014;515:249–52.

15. Christopher LP, Yao B, Ji Y. Lignin Biodegradation with Laccase-Mediator Systems. Front. Energy Res. 2014;2:1–13.

16. Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. Pathways for degradation of lignin in bacteria and fungi. Nat. Prod. Rep. 2011;28:1883–96.

17. Taylor CR, Hardiman EM, Ahmad M, Sainsbury PD, Norris PR, Bugg TDH. Isolation of bacterial strains able to metabolize lignin from screening of environmental samples. J. Appl. Microbiol. 2012;113:521–30.

18. Kato S, Chino K, Kamimura N, Masai E, Yumoto I, Kamagata Y. Methanogenic degradation of lignin-derived monoaromatic compounds by microbial enrichments from rice paddy field soil. Sci. Rep. Nature Publishing Group; 2015;5:14295.

19. Scully ED, Geib SM, Hoover K, Tien M, Tringe SG, Barry KW, et al. Metagenomic Profiling Reveals Lignocellulose Degrading System in a Microbial Community Associated with a Wood-Feeding Beetle. PLoS One. 2013;8:1–22.

20. Strachan CR, Singh R, VanInsberghe D, Ievdokymenko K, Budwill K, Mohn WW, et al. Metagenomic scaffolds enable combinatorial lignin transformation. Proc. Natl. Acad. Sci. U. S. A. 2014;111:10143–8.

21. Rocha GJM, Martín C, da Silva VFN, Gómez EO, Gonçalves AR. Mass balance of pilot-scale pretreatment of sugarcane bagasse by steam explosion followed by alkaline delignification. Bioresour. Technol. 2012;111:447–52.

22. Miller GL. Use of DinitrosaIicyIic Acid Reagent for Determination of Reducing Sugar. Anal. Chem. 1959;31:426–8.

23. Caporaso JG, Lauber CL, Walters W a, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6:1621–4.

24. White, T. J., T. Bruns, S. Lee, J. W. Taylor. PCR Protocols: A Guide to Methods and Applications. Innis, M. A., D. H. Gelfand, J. J. Sninsky and TJW, editor. Academic Press, Inc., New York._©W; 1990.

25. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods. 2013;10:996–8.

26. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 2007;73:5261–7.

27. Nilsson RH, Tedersoo L, Ryberg M, Kristiansson E, Hartmann M, Unterseher M, et al. A Comprehensive, Automatically Updated Fungal ITS Sequence Dataset for Reference-Based Chimera Control in Environmental Sequencing Efforts. Microbes Environ. 2015;30:145–50.

28. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28:1420–8.

29. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38:1–15.

30. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, et al. The Pfam protein families database. Nucleic Acids Res. 2010;38:D211–22.

31. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40:W445-51.

32. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2015;44:D286-293.

33. Wang Z, Wu M. A phylum-level bacterial phylogenetic marker database. Mol. Biol. Evol. 2013;30:1258–62.

34. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.

35. Boetzer M, Henkel C V, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27:578–9.

36. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9.

37. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: A system for microbial genome annotation expert review and curation. Bioinformatics. 2009;25:2271–8.

38. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.

39. Aslanidis C, Jong PJ De. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 1990;18:6069–74.

40. Yang W, Tang H, Ni J, Wu Q, Hua D, Tao F, et al. Characterization of two Streptomyces enzymes that convert ferulic acid to vanillin. PLoS One. 2013;8.

41. Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, et al. Metabolic Profiling Allows Comprehensive Phenotyping of Genetically or Environmentally Modified Plant Systems. 2001;13:11–29.

42. Weckwerth W, Wenzel K, Fiehn O. Process for the integrated extraction , identification and quantification of metabolites , proteins and RNA to reveal their co-regulation in biochemical networks. 2004;78–83.

43. Cuadros-inostroza Á, Caldana C, Redestig H, Kusano M, Lisec J, Peña-cortés H, et al. TargetSearch - a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. 2009;12:1–12.

44. Giavalisco P, Li Y, Matthes A, Eckhardt A, Hubberten H, Hesse H, et al. Elemental formula annotation of polar and lipophilic metab- olites using 13 C , 15 N and 34 S isotope labelling , in combination with high-resolution mass spectrometry. 2011;364–76.

45. Huege J, Krall L, Steinhauser M, Giavalisco P, Rippka R, Marsac NT De, et al. Sample amount alternatives for data adjustment in comparative cyanobacterial metabolomics. 2011;3503–17.

46. Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. BIOINFORMATICS APPLICATIONS NOTE pcaMethods — a bioconductor package providing PCA methods for incomplete data. 2007;23:1164–7.

47. Dunbar J, Barns SM, Ticknor LO, Kuske CR. Empirical and Theoretical Bacterial Diversity in Four Arizona Soils. Appl. Environ. Microbiol. 2002;68:3035–45.

48. Lee SH, Ka JO, Cho JC. Members of the phylum Acidobacteria are dominant and metabolically active in rhizosphere soil. FEMS Microbiol. Lett. 2008;285:263–9.

49. White III RA, Bottos EM, Chowdhury TR, Zucker JD, Brislawn CJ, Nicora CD, et al. Moleculo Long-Read Sequencing Facilitates Assembly and Genomic Binning from Complex Soil Metagenomes. mSystems. 2016;1:1–15.

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

51. Chung SY, Maeda M, Song E, Horikoshi K, Kudo T. A Gram-positive Polychlorinated Biphenyl-degrading Bacterium, Rhodococcus erythropolis Strain TA421, Isolated from a Termite Ecosystem. Biosci. Biotechnol. Biochem. 1994;58.

52. Diamantidis G, Effosse A, Potier P, Bally R. Purification and characterization of the first bacterial laccase in the rhizospheric bacterium Azospirillum lipoferum. Soil Biol. Biochem. 2000;32:919–27.

53. Kellner H, Luis P, Zimdars B, Kiesel B, Buscot F. Diversity of bacterial laccase-like multicopper oxidase genes in forest and grassland Cambisol soil samples. Soil Biol. Biochem. 2008;40:638–48.

54. Le Roes-Hill M, Khan N, Burton SG. Actinobacterial peroxidases: An unexplored resource for biocatalysis. Appl. Biochem. Biotechnol. 2011;164:681–713.

55. Sampaio JP. Utilization of Low Molecular Weight Lignin-Related Aromatic Compounds for the Selective Isolation of Yeasts: Rhodotorula vanillica, a New Basidiomycetous Yeast Species. Syst. Appl. Microbiol. 1995;17:613–9.

56. Weng XY, Sun JY. Biodegradation of free gossypol by a new strain of Candida tropicalis under solid state fermentation: Effects of fermentation parameters. Process Biochem. 2006;41:1663–8.

57. Martorell MM, Pajot HF, de Figueroa LIC. Dye-decolourizing yeasts isolated from Las Yungas rainforest. Dye assimilation and removal used as selection criteria. Int. Biodeterior. Biodegrad. 2012;66:25–32.

58. Ettayebi K, Errachidi F, Jamai L, Tahri-Jouti MA, Sendide K, Ettayebi M. Biodegradation of polyphenols with immobilized Candida tropicalis under metabolic induction. FEMS Microbiol. Lett. 2003;223:215–9.

59. Yan J, Jianping W, Hongmei L, Suliang Y, Zongding H. The biodegradation of phenol at high initial concentration by the yeast Candida tropicalis. Biochem. Eng. J. 2005;24:243–7.

60. Glenn JK, Gold MH. Decolorization of several pymeric dies by the lignin degrading basidiomycete Phanerochete chrysosporium. Appl. Environ. Microbiol. 1983;6:1741–7.

61. Šimek K, Vrba J, Pernthaler J, Posch T, Hartman P, Nedoma J, et al. Morphological and compositional shifts in an experimental bacterial community influenced by protists with contrasting feeding modes. Appl. Environ. Microbiol. 1997;63:587–95.

62. Jürgens K, Pernthaler J, Schalla S, Amann R. Morphological and Compositional Changes in a Planktonic Bacterial Community in Response to Enhanced Protozoan Grazing Morphological and Compositional Changes in a Planktonic Bacterial Community in Response to Enhanced Protozoan Grazing. Appl. Environ. Microbiol. 1999;65:1241–50.

63. Campillo T, Renoud S, Kerzaon I, Vial L, Baude J, Gaillard V, et al. Analysis of hydroxycinnamic acid degradation in Agrobacterium fabrum reveals a coenzyme A-dependent, beta-oxidative deacetylation pathway. Appl. Environ. Microbiol. 2014;80:3341–9.

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

65. Větrovský T, Steffen KT, Baldrian P. Potential of cometabolic transformation of polysaccharides and lignin in lignocellulose by soil Actinobacteria. PLoS One. 2014;9.

66. Wang L, Nie Y, Tang Y-Q, Song X-M, Cao K, Sun L-Z, et al. Diverse Bacteria with Lignin Degrading Potentials Isolated from Two Ranks of Coal. Front. Microbiol. 2016;7:1–14.

67. Gall DL, Ralph J, Donohue TJ, Noguera DR. A Group of Sequence-Related

Sphingomonad Enzymes Catalyzes Cleavage of β-Aryl Ether Linkages in Lignin β-Guaiacyl and β-Syringyl Ether Dimers. Environ. Sci. Technol. 2014;48:12454–63.

68. Fuchs G, Boll M, Heider J. Microbial degradation of aromatic compounds — from one strategy to four. Nat. Rev. Microbiol. 2011;9:803–16.

69. Salvachúa D, Karp EM, Nimlos CT, Vardon DR, Beckham GT. Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria. Green Chem. Royal Society of Chemistry; 2015;17:4951–67.

70. Jeffries TW, Gifford O, Drive P. Biodegradation of lignin-carbohydrate complexes. 1990;163–76.

71. Subramaniyan S, Prema P. Biotechnology of Microbial Xylanases : Enzymology , Molecular Biology and Application. 2002;22:33–46.

72. Errico C, Jørgensen JO, Krogh KBRM, Spodsberg N. Enzymatic Degradation of Lignin-Carbohydrate Complexes (LCCs): Model Studies Using a Fungal Glucuronoyl Esterase from Cerrena Unicolor. 2015;112:914–22.

73. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis T, Mavrommatis K, Kyrpides NC, et al. Microbial species delineation using whole genome sequences. 2015;1–11.

74. Dalvi S, Youssef NH, Fathepure BZ. Microbial community structure analysis of a benzoate-degrading halophilic archaeal enrichment. Extremophiles. Springer Japan; 2016;20:311–21.

75. Naoto S. Phenolic acidsvand their cabohydrate esters in rice endosperm cell walls. Phytochemistry. 1984;23:2233–7.

76. Overhage J, Priefert H, Steinbuchel A. Biochemical and genetic analyses of ferulic acid catabolism in Pseudomonas sp. strain HR199. Appl Env. Microbiol. 1999;65:4837–47

77. Masai E, Harada K, Peng X, Kitayama H, Katayama Y. Cloning and Characterization of the Ferulic Acid Catabolic Genes of Sphingomonas paucimobilis SYK-6. 2002;68:4416–24.

78. Plaggenborg R, Overhage J, Loos A, Archer JAC, Lessard P, Sinskey AJ, et al. Potential of Rhodococcus strains for biotechnological vanillin production from ferulic acid and eugenol. 2006;745–55.

79. Achterholt S, Priefert H, Steinbüchel A. Identification of Amycolatopsis sp . strain HR167 genes , involved in the bioconversion of ferulic acid to vanillin. Appl Microbiol Biotecnhol. 2000;103:799–807.

80. Gasson MJ, Kitamura Y, Mclauchlan WR, Narbad A, Parr AJ, Parsons ELH, et al. Metabolism of Ferulic Acid to Vanillin. A bacterial gene of the enoyl-SCoA hydratase/isomerase superfamily encodes an enzyme for the hydration and cleavage of a hydroxycinnamic acid SCoA thioester. J Biol Chem. 1998;273:4163–70.

81. Calisti C, Ficca AG, Barghini P, Ruzzi M. Regulation of ferulic catabolic genes in Pseudomonas fluorescens BF13 : involvement of a MarR family regulator. 2008;475–83.

82. Loreau M, Loreau M, Naeem S, Naeem S, Inchausti P, Inchausti P, et al. Biodiversity and ecosystem functioning: current knowledge and future challenges. Science. 2001;294:804–8.

83. Reizer A, Deutscher J, Saier MH Jr, Reizer J. Analysis of the gluconate (gnt) operon of Bacillus subtilis. Mol Microbiol 1991; 5(5):1081-9.

Appendix: Additional file 1

Methods

Amplicons sequencing and analysis

The hypervariable V4 region of the 16S rRNA gene and ITS2 region (internal transcribed spacer) were amplified as previously described [1,2]. Amplicon libraries were performed in three technical replicates for each sample. PCR reactions were carried out using the specific primers sets to cover V4 region of the 16S rRNA gene and ITS2 region as previously described. Second PCR were performed to incorporate multiplex identified (MID) tags between Illumina adapter sequences and primers for enabling sample multiplexing during sequences. After the purification with AMPure XP SPRI, the PCR products were quantified using the fluorescent dye method by KAPA qPCR Library Quantification. The biological triplicate was pooled together in equimolar amount for sequencing on the Illumina MiSeq system according to standard procedures. The marker genes libraries were sequencing on Illumina MiSeq system applying the paired-end protocol and using the v2 kit chemistry with 500 cycles.

The raw datasets were preprocessed by removing adapter sequencing, quality assessment and sequences trimmed using Trimmomatic [3]. Quality-filtered reads were merged using Pear [4]. Obtained sequences were analyzed using Usearch pipepline [5]. The sequences were clustered into operational taxonomic unit (OTU) based on a 97 % identity using the USEARCH method. Chimeric sequences were identified and removed using UCHIME method. The representative sequence from each OTUs was assigned to taxonomic taxa using Ribosomal Database Project (RDP) Classifier [6] and UNITE Fungal ITS training sets [7]. The alpha diversity and diversity indices were calculated employing Phyloseq package using R statistical software version 3.2.

Metagenome sequencing, assembly and binning

For sequencing purposes, 50 ng of the total DNA were used to prepare an Illumina Nextera library according to the manufacturer's instructions. The library was

validated on an Agilent Bioanalyzer 2100 with the 12000 DNA assay kit (Agilent) and quantifies applying the KAPA Biosystem's next-generation sequencing library qPCR kit (KAPA). The libraries were sequenced on Illumina HiSeq 2500 applying the protocol paired-end. The metagenomic reads were first quality-filtered using Trimmomatic, to remove low quality reads, and assembled using IDBA-UD [8]. The assembled contigs were subject to MetaQuast [9] and MetaGeneMark [10] to quality assessment of the assembly and gene calling. Predicted coding genes were compared to Swiss Prot and UniRef90 databases, using Malt (https://ab.inf.uni-tuebingen.de/software/malt), and predicted proteins where further compared to Pfam (Finn et al., 2010), dbCAN [12] and EggNOG databases [13], using HMMER3 (http:// hmmer.janelia.org/) package. Pathway analysis was performed using KEGG web application GhostKoala [14].

Metagenomic contigs were binned by coverage and sequence composition using CONCOCT [15]. Assembled contigs were split in chunks of 10Kb prior to read mapping using Bowtie2 [16]. PCR duplicated reads were removed prior to genome coverage calculation. Genome bins assessed using CheckM [17] to determine completeness and contamination. Bins presenting completeness greater than 75%, contamination smaller than 10% and strain heterogeneity smaller than 25% were selected for further analysis using Phyla-AMPHORA [18], that incorporates bacterial phylum-level phylogenetic markers for taxonomy assignment.

Cloning, expression and purification of genes involved in bioconversion of ferulic acid into vanillin production

Two genes, *FerA* and *FerB* encoding feruloyl coenzyme A synthetase and enoyl-CoA hydratase, respectively, were employed in the biotransformation of ferulic acid into vanillin. CLig metagenomic dataset was screened based on Pfam families (PF13380 and PF00378) to find the candidate sequences coding *FerA* and *FerB* genes. The sequences were synthetized by Biomatik (Biomatik Corporation, Canada), inserted into the pET28a-vector and transformed into *Escherichia coli* BL21(DE). The *FerA* gene sequence was sub-cloned in the pETTRXA-1a/LIC by the ligase independent cloning (LIC) method [19]. The resulting plasmids were then transformed into *E. coli* BL 21 (DE) for expression. The recombinant *E. coli* were grown at 37 ºC in Luria Bertani medium containing 100 mg/L kanamycin. When the

culture reached an optical density (OD_{600} nm) of 0.6~0.8, IPTG was added to the culture at a final concentration of 0.5 mM. Thus, the culture was incubated for 4 hours at 30 º C for over-expression. Cells were harvested by centrifugation and suspended in bidding buffer (20 mM sodium phosphate pH7.4, 100 mM sodium chloride and 5mM imidazole) supplemented with phenylmethanesulfonyl fluoride (PMSF) at final concentration of 1 mM and lysozyme at final concentration of 0.5 mg/mL. Cells were incubated for 30 minutes on ice and then lysed by sonication. Soluble fractions were obtained by centrifugation at 14,000 x g at 4 ºC for 30 minutes and then loaded in His-Trap-Ni-NTA columns (GE Healthcare), pre-equilibrated with binding buffer. Recombinant proteins were eluted with an elution buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 400 mM imidazole). Eluted fractions were evaluated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and protein concentrations were determined according to the Bradford method. Purified proteins were stored at -80 ºC until further use.

Additional file references

1. Caporaso JG, Lauber CL, Walters W a, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. [Internet]. Nature Publishing Group; 2012;6:1621–4. Available from: http://dx.doi.org/10.1038/ismej.2012.8

2. White, T. J., T. Bruns, S. Lee, J. W. Taylor. PCR Protocols: A Guide to Methods and Applications. Innis, M. A., D. H. Gelfand, J. J. Sninsky and TJW, editor. Academic Press, Inc., New York._©W; 1990.

3. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

4. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30:614–20.

5. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods [Internet]. 2013 [cited 2014 Jul 10];10:996–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23955772

6. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 2007;73:5261–7.

7. Nilsson RH, Tedersoo L, Ryberg M, Kristiansson E, Hartmann M, Unterseher M, et al. A Comprehensive, Automatically Updated Fungal ITS Sequence Dataset for Reference-Based Chimera Control in Environmental Sequencing Efforts. Microbes Environ. [Internet]. 2015;30:145–50. Available from: https://www.jstage.jst.go.jp/article/jsme2/30/2/30_ME14121/_article

8. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28:1420–8.

9. Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: Evaluation of metagenome assemblies. Bioinformatics. 2016;32:1088–90.

10. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38:1–15.

11. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, et al. The Pfam protein families database. Nucleic Acids Res. 2010;38:D211–22.

12. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. [Internet]. 2012;40:W445-51. Available from: http://nar.oxfordjournals.org/content/40/W1/W445.short

13. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44:D286–93.

14. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J. Mol. Biol. [Internet]. The Authors; 2016;428:726–31. Available from: http://dx.doi.org/10.1016/j.jmb.2015.11.006

15. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat. Methods [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014 [cited 2014 Sep 14];11:1144–6. Available from: http://dx.doi.org/10.1038/nmeth.3103

16. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

17. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. [Internet]. 2015;25:1043–55. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4484387&tool=pmcentrez&rendertype=abstract

18. Wang Z, Wu M. A phylum-level bacterial phylogenetic marker database. Mol. Biol. Evol. 2013;30:1258–62.

19. Aslanidis C, Jong PJ De. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 1990;18:6069–74.

20. Rocha GJM, Martín C, da Silva VFN, Gómez EO, Gonçalves AR. Mass balance of pilot-scale pretreatment of sugarcane bagasse by steam explosion followed by alkaline delignification. Bioresour. Technol. [Internet]. Elsevier Ltd; 2012;111:447–52. Available from: http://dx.doi.org/10.1016/j.biortech.2012.02.005

21. Suguiyama VF, Silva E a, Meirelles ST, Centeno DC, Braga MR. Leaf metabolite profile of the Brazilian resurrection plant Barbacenia purpurea Hook. (Velloziaceae) shows two time-dependent responses during desiccation and recovering. Front. Plant Sci. [Internet]. 2014;5:96. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3953666&tool=pmcentrez&rendertype=abstract

CAPÍTULO 3 – CONSIDERAÇÕES FINAIS E PERSPECTIVAS

No presente trabalho, foi demonstrado que a estratégia de enriquecimento de uma comunidade de micro-organismos de solo se mostrou eficiente no estabelecimento de um consórcio especializado na degradação de compostos aromáticos derivados de lignina. A identificação de grupos taxonômicos relacionados à degradação de lignina, compostos fenólicos e corantes, bem como a descrição de rotas metabólicas envolvidas na degradação de lignina, reforçam a especialização do consórcio e sua capacidade de degradar compostos aromáticos. A prospecção de enzimas envolvidas na conversão de ácido ferúlico em vanilina e a validação da funcionalidade das mesmas em ensaios enzimáticos demonstram que o consórcio estabelecido é uma plataforma eficiente para a obtenção de enzimas que podem ser empregadas na conversão de compostos derivados de lignina em compostos de elevado valor comercial.

Como forma de complementar a estratégia de consórcio, realizou-se também o isolamento e caracterização de cepas a partir da comunidade microbiana estabelecida. Além da cepa descrita de *Paenarthrobacter* sp., realizamos também o isolamento das leveduras *Rhodosporidium fluviale* e *Candida utilis (dados não apresentados)*, as quais possuem grande potencial biotecnológico para a produção de biomassa a partir de resíduos agroindustriais (Lee e Kim, 2001; Watanabe et al., 2013; Polburee et al., 2015; Sitepu et al., 2013). Além disso, estão sendo conduzidos ensaios de caracterização funcional, cristalização, SAXS e dicroísmo circular das enzimas Fch e Ech triadas a partir do metagenoma do consórcio.

Este trabalho terá continuidade dento de uma parceria com grupos de pesquisa internacionais da University of Manchester, University of Warwick e University College London, no contexto do projeto temático “Lignin valorization in cellulosic ethanol plants: biocatalytic conversion via ferulic acid to high value chemicals" concedido pelo convênio entre a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e o Biotechnology and Biological Sciences Research Council (BBSRC) ao prof. Dr. Fabio Squina, serão desenvolvidas estratégias de triagem de enzimas do metagenoma a partir de biossensores (Machado e Dixon, 2016) para a conversão de compostos de alto valor agregado, tais como vanilina, L-DOPA e álcool coniferílico.

Espera-se que esta tese de doutorado contribua no desenvolvimento de tecnologias de valorização da lignina do bagaço de cana-de-açúcar, integradas, por exemplo, em processos de produção de etanol.

REFERÊNCIAS

Ahmad M, Roberts JN, Hardiman EM, Singh R, Eltis LD, Bugg TD. Identification of DypB from *Rhodococcus jostii* RHA1 as a lignin peroxidase. Biochemistry (2011) 50:5096-107.

Alves EF, Bose SK, Francis RC, El Moussaouiti M. Trapping of p-coumaryl and coniferyl alcohol during soda-anthraquinone treatment: a means of estimating uncondensed ß-O-4 structures in native lignin. J Agric Food Chem (2012) 60:9202-10.

Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individiual microbial cells without cultivation. Microbiol Rev (1995) 59:143-169.

Angelo PM, Jorge N. Compostos fenólicos em alimentos: Uma breve revisão. Revista Instituto Adolf Lutz (2007) 66:232-240.

Bugg TD, Ahmad M, Hardiman EM, Singh R. The emerging role for bacteria in lignin degradation and bio-product formation. Curr Opin Biotechnol (2011) 22:394-400.

Clark GS. Vanillin. Perfurmer Flavorist (1990) 15:45-54.

Coughlan LM, Cotter PD, Hill C and Alvarez-Ordóñez A. Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. Front Microbiol (2015) 6:672.

Cowan DA. Microbial genomes - the untapped resource. Trends Biotechnol (2000) 18:14-16.

Daniel R. The soil metagenome - a rich resource for the discovery of novel natural products. Curr Opin Biotechnol (2004) 15:199-204.

Das N, Chandran P. Microbial degradation of petroleum hydrocarbon contaminants: an overview. Biotechnology Research International (2011) Article ID 941810, 13 pages. doi:10.4061/2011/941810.

de Lima Brossi MJ, Jiménez DJ, Cortes-Tolalpa L, van Elsas JD. Soil-Derived Microbial Consortia Enriched with Different Plant Biomass Reveal Distinct Players Acting in Lignocellulose Degradation. Microb Ecol (2016) 71(3):616-627. doi: 10.1007/s00248-015-0683-7.

Esposito LJ et al. Vanillin, in Kirk-Othmer Encyclopedia of Chemical Technology, John Wiley & Sons, Inc., 2000.

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

Fitzgerald DJ, Stratford M, Gasson MJ, Narbad A. Structure-Function Analysis of the Vanillin Molecule and Its Antifungal Properties. J Agric Food Chem (2005) 53:1769-775.

Fon FN, Nsahlai IV, Scogings PF, Basha NAD. In Vitro Cellulase Production from Five Herbivore Microbial Ecosystems and Consortia. Annals of Animal Science (2014) 14(2)

Freundenberg K. Lignin: Its Constitution and Formation from p-Hydroxycinnamyl Alcohols: Lignin is Duplicated by Dehydrogenation of these Alcohols; Intermediates Explain Formation and Structure. Science (1965) 148:595-600.

Furukawa H, Wieser M, Morita H, Sugio T, Nagasawa T. Purification and characterization of eugenol dehydro¬genase from *Pseudomonas fluorescens* E118. Archives Microbiology (1998) 171: 37-43.

Gao ZM, Xu X, Ruan LW. Enrichment and characterization of an anaerobic cellulolytic microbial consortium SQD-1.1 from mangrove soil. Appl Microbiol Biotechnol (2014) 98(1):465-474. doi: 10.1007/s00253-013-4857-2.

Gargouri B, Karray F, Mhiri N, Aloui F, Sayadi S. Bioremediation of petroleum hydrocarbons-contaminated soil by bacterial consortium isolated from an industrial wastewater treatment plant. J Chem Tech Biotech (2014) 89(7):978-987.

Goldemberg, J. The Brazilian Biofuel Industry. Biotechnology for Biofuels (2008) 1:6.

Gülçin I, Beydemir S. Phenolic compounds as antioxidants: carbonic anhydrase isoenzymes inhibitors. Mini Rev Med Chem (2013) 13(3):408-430.

Hamid B, Kaushik G, Chawla J, Ahmad Baba Z. Isolation and Development of Efficient Bacterial Consortia for Bioremediation of Textile Dye Effluent. J Pollut Eff (2015) 3:142. doi:10.4172/2375-4397.1000142.

Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews (2004) 68:669.

Hansen EH, Møller BL, Kock GR, Bünner CM, Kristensen C, Jensen OR, Okkels FT, Olsen CE, Motawia MS, Hansen J. De novo biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). Appl Environ Microbiol (2009) 75:2765-74.

Hocking MB. Vanillin: synthetic fla¬voring from spent sulfite liquor. Journal of Chemical Education (1997) 74:1055-59.

Ichor T, Okerentugba PO, Okpokwasili, GC. Biodegradation of Total Petroleum Hydrocarbon by a Consortium of Cyanobacteria Isolated from Crude Oil Polluted Brackish Waters of Bodo Creeks in Ogoniland, Rivers State. Research Journal of Environmental Toxicology (2016) 10:16-27. doi: 10.3923/rjet.2016.16.27.

Jiménez DJ, de Lima Brossi MJ, Schückel J, Kracun SK, Willats WG, van Elsas JD. Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. Appl Microbiol Biotechnol (2016) 100(24):10463-77.

Joshi PA, Jaybhaye S, Mhatre K. Biodegradation of dyes using consortium of bacterial strains isolated from
textile effluent. European Journal of Experimental Biology (2015) 5(7):36-40.

Kato A., Azuma J, Koshijima T. Isolation and identification of a new feruloylated tetrasaccharide from bagasse lignin-carbohydrate complex containing phenolic acid. Agric Biol Chem (1987) 51:1691-93.

Kerbauy GB. Fisiologia Vegetal. 2ª edição. Rio de Janeiro: Guanabara Kogan, 2008. 431 p.

Korenblum E, Jiménez DJ, van Elsas JD. Succession of lignocellulolytic bacterial consortia bred anaerobically from lake sediment. Microb Biotechnol (2016) 9(2):224–234. doi: 10.1111/1751-7915.12338.

Kortekaas S, Vidal G, Yan-Ling H, Letting G, Field JA. Anaerobic-aerobic treatment of toxic pulping black liquor with upfront effluent recirculation. Journal Fermentation Bioengineering (1998) 86:97-110.

Kull, K. (2010). Ecosystems are Made of Semiosic Bonds: Consortia, Umwelten, Biophony and Ecological Codes. Biosemiotics 3, Issue 3, pp 347–357

Lade H, Kadam A, Paul D, Govindwar S. Biodegradation and detoxification of textile azo dyes by bacterial consortium under sequential microaerophilic/aerobic processes. EXCLI J (2015) 14:158–174.

Lee B, Kim JK. Production of Candida utilis biomass on molasses in different culture types. Aquacultural engineering 2001; 25(2):111-124.

Lee EG, Yoon SH, Das A, Lee SH, Li C, Kim JY, Choi MS, Oh DK, Kim SW. Directing vanillin production from ferulic acid by increased acetyl-CoA consumption in recombinant Escherichia coli. Biotechnol Bioeng (2009) 102:200-208.

Lora JH, Glasser WGJ. Recent industrial applications of lignin - a sustainable alternative to non-renewable materials. J. Polym. Environ (2002) 10:39-48.

Machado LFM, Dixon N. Development and substrate specificity screening of an in vivo biosensor for the detection of biomass derived aromatic chemical building blocks. Chem Commun 2016; 52:11402. doi: 10.1039/c6cc04559f.

Mamo G, Faryar R, Karlsson EN. Microbial glycoside hydrolases for biomass utilization in biofuels applications. In V. K. Gupta & M. G. Tuhoy (Eds.), Biofuel technologies: Recent developments (2013) 171–188. Heidelberg: Springer.

Masarin F, Gurpilhares DB, Milagres AMF. Chemical composition and enzymatic digestibility of sugarcane clones selected for varied lignin content. Biotechnology for Biofuels (2011) 4:55.

Mukhopadhyay AK. Industrial chemical cresols and downstream derivateds. New York: CRC, 2005.

Ogeda TL, Petri DFS. Hidrólise Enzimática de Biomassa. Química Nova (2010) 33:1549-58.

Pan X, Kadla JF, Ehara K, Gilkes N, Saddler JN. Organosolv Ethanol Lignin from Hybrid Poplar as a Radical Scavenger: Relationship between Lignin Structure, Extraction Conditions, and Antioxidant Activity. J Agric Food Chem (2006), 54:5806-5813.

Patowary K, Patowary R, Kalita MC and Deka S. Development of an Efficient Bacterial Consortium for the Potential Remediation of Hydrocarbons from Contaminated Sites. Front Microbiol (2016) 7:1092.

Phugare SS, Kalyani DC, Patil AV, Jadhav JP. Textile dye degradation by bacterial consortium and subsequent toxicological analysis of dye and dye metabolites using cytotoxicity, genotoxicity and oxidative stress studies. J Hazard Mater (2011) 186(1):713-723. doi: 10.1016/j.jhazmat.2010.11.049.

Polburee P, Yongmanitchai W, Lertwattanasakul N, Ohashi T, Fujiyama K, Limtong S. Characterization of oleaginous yeasts accumulating high levels of lipid when cultivated in glycerol and their potential for lipid production from biodiesel-derived crude glycerol. Fungal Biol 2015; 119(12):1194-204. doi: 10.1016/j.funbio.2015.09.002.

Research Council (U.S.). Committee on metagenomics: Challenges and functional applications, N. The new science of metagenomics: revealing the secrets of our microbial planet. 2007. 158 p.

Rezende CA, de Lima MA, Maziero P, de Azevedo ER, Garcia W, Polikarpov I. Chemical and morphological characterization of sugarcane bagasse submitted to a delignification process for enhanced enzymatic digestibility. Biotechnol Biofuels (2011) 4:54. doi: 10.1186/1754-6834-4-54.

Ribeiro C. Cosmetologia aplicada a dermocosmética. São Paulo: Pharmaboks; 270p, 2006.

Rocha GJD, Martin C, Soares IB, Maior AMS, Baudel HM, Abreu CAM. Dilute mixed-acid pretreatment of sugarcane bagasse for ethanol production. Biomass and Bioenergy (2011) 35:663-670.

Rubin EM. Genomics of Cellulosic Biofuels. Nature Reviews (2008) 454:841-845.

Sainsbury PD, Hardiman EM, Ahmad M, Otani H, Seghezzi N, Eltis LD, Bugg TD. Breaking down lignin to high-value chemicals: the conversion of lignocellulose to vanillin in a gene deletion mutant of Rhodococcus jostii RHA1. ACS Chem Biol (2013) 8:2151-2156.

Senan RC, Abraham TE. Bioremediation of Textile Azo Dyes by Aerobic Bacterial Consortium Aerobic Degradation of Selected Azo Dyes by Bacterial Consortium. Biodegradation (2004) 15:275. doi:10.1023/B:BIOD.0000043000.18427.0a.

Shah AA, Hasan F, Akhter JI. Degradation of polyurethane by novel bacterial consortium isolated from soil. Ann Microbiol (2008) 58:381. doi:10.1007/BF03175532.

Sitepu IR, Sestric R, Ignatia L, Levin D, German JB, Gillies LA, et al. Manipulation of culture conditions alters lipid content and fatty acid profiles of a wide variety of known and new oleaginous yeasts species. Bioresour Technol 2013; 144:360-9. doi: 10.1016/j.biortech.2013.06.047.

Skariyachan S, Manjunatha V, Sultana S, Jois C, Bai V, Vasist KS. Novel bacterial consortia isolated from plastic garbage processing areas demonstrated enhanced degradation for low density polyethylene. Environ Sci Pollut Res Int (2016) 23(18):18307-19. doi: 10.1007/s11356-016-7000-y.

Stewart D. Lignin as a base material for materials applications: Chemistry, application and economics. Ind Crops Prod (2008) 27:202-207.

Taylor CR, Hardiman EM, Ahmad M, Sainsbury PD, Norris PR, Bugg TD. Isolation of bacterial strains able to metabolize lignin from screening of environmental samples. J Appl Microbiol (2012) 113:521-530.

Tony BD, Goyal D, Khanna S. Decolorization of textile azo dyes by aerobic bacterial consortium. International Biodeterioration & Biodegradation (2009) 63(4):462-469.

Walton NJ, Mayer MJ, Narbad A. Vanillin. Phytochemestry (2003) 63:505-15.

Watanabe T, Iefuji H, Kitamoto HK. Treatment of, and Candida utilis biomass production from shochu wastewater; the effects of maintaining a low pH on DOC removal and feeding cultivation on biomass production. Springerplus 2013; 2: 514. doi: 10.1186/2193-1801-2-514.

Wong DW. Structure and action mechanism of ligninolytic enzymes. Appl Biochem Biotechnol (2009) 157:174-209.

Yamada M, Okada Y, Yoshida T, Nagasawa T. Vanillin production using Escherichia coli cells over-expressing isoeugenol monooxygenase of Pseudomonas putida. Biotechnol Lett (2008) 30:665-670.

Yang W, Tang H, Ni J, Wu Q, Hua D, Tao F, Xu P. Characterization of two Streptomyces enzymes that convert ferulic acid to vanillin. PLoS ONE (2013) 8:e67339.

Yun J, Ryu S. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. Microbial Cell Factories (2005) 4(1):8.

Zakzeski J, Bruijnincx PC, Jongerius AL, Weckhuysen BM. The catalytic valorization of lignin for the production of renewable chemicals. Chemical Reviews (2010) 110:3552-99.

ANEXO 1. MATERIAL SUPLEMENTAR

1. Esquematização do processo de obtenção da solução de lignina de baixo peso molecular (LW) / lignina solúvel

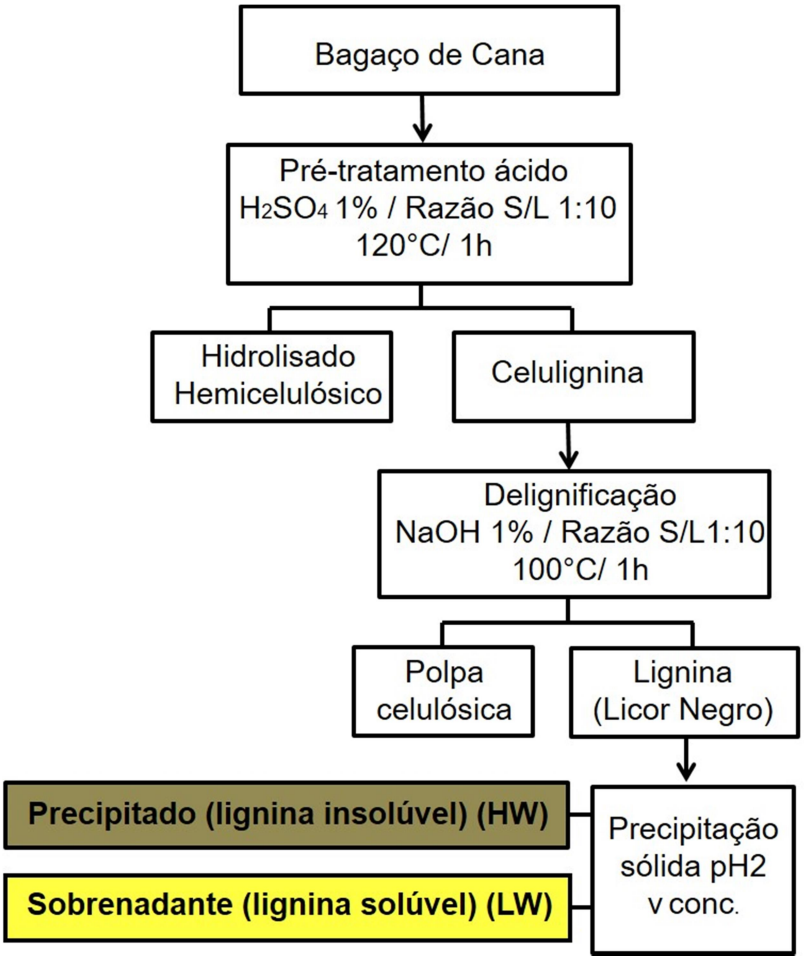


Figura 1. O licor negro é caracterizado como um líquido de coloração escura, contendo frações de diferentes pesos moleculares de lignina solúvel e em suspensão. As frações solúvel e insolúvel são separadas através da acidificação do licor negro, obtendo assim a fração de alto peso molecular no precipitado (lignina insolúvel) e a fração de baixo peso molecular em solução no sobrenadante (lignina solúvel). Denominamos as frações de alto e baixo peso molecular de HW (***h**igh **w**eight*) e LW (***l**ow **w**eight*), respectivamente.

2. Géis SDS Page e condições de expressão das proteínas utilizadas no projeto

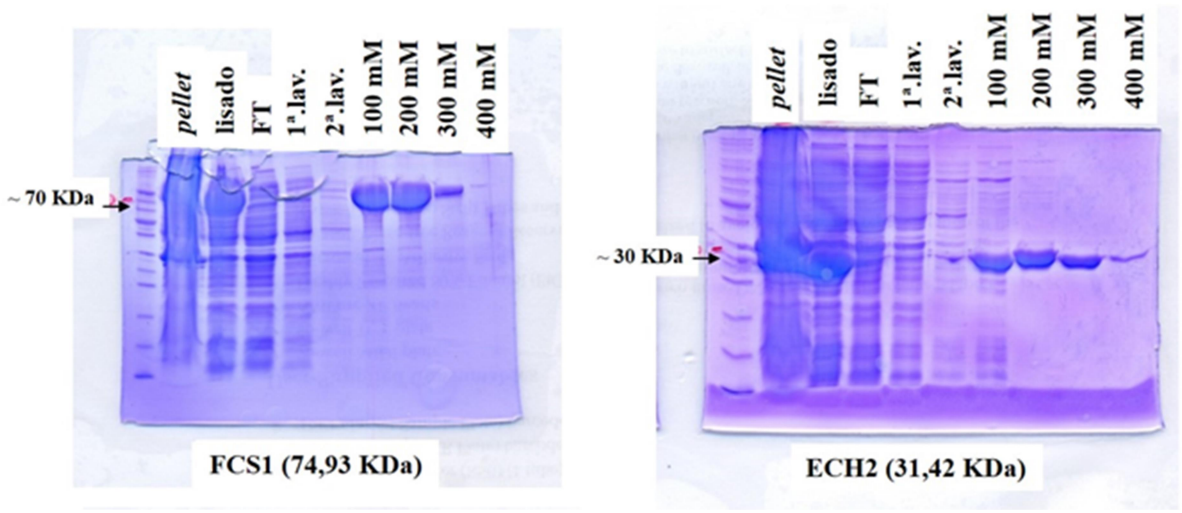


Figura 2. Testes de expressão das sequências clonadas de FCS1 (FerA_B3) e ECH2 (FerB_B11). As concentrações de imidazol (em mM) utilizadas na eluição das proteínas estão indicadas na figura. A concentração obtida, após a purificação em coluna de afinidade, foi de aproximadamente 0,79 mg/ml para ambas as proteínas. FT: flow throught.

ANEXO 2. FORMULÁRIO DE APROVAÇÃO DA COMISSÃO INTERNA DE BIOSSEGURANÇA (CIBIO)

Uso exclusivo da CIBio:

Número de projeto / processo: CTBE-2015-3

Formulário de encaminhamento de projetos de pesquisa para análise pela CIBio - Comissão Interna de Biossegurança do CNPEM – Centro Nacional de Pesquisa em Energia e Materiais

Título do projeto: Prospecção de Enzimas e Otimização de Vias Metabólicas Para Produção de Vanilina a partir de Lignina do Bagaço de Cana de Açúcar

Pesquisador responsável: Fábio Márcio Squina

Experimentador: Eduardo Cruz Moraes

Classe de risco do OGM: [x] Risco I [ ] Risco II

Nível do treinamento do experimentador: [ ]-Iniciação científica, [ ]-mestrado, [x]-doutorado, [ ]-doutorado direto, [ ]-pós-doutorado, [ ]-nível técnico, [ ]-outro, especifique:_____

Resumo do projeto: A lignina, um dos materiais mais recalcitrantes da natureza, consiste em um polímero tridimensional composto de vários anéis aromáticos estavelmente ligados, estando presente em combinação com celulose e hemicelulose na parede celular vegetal. A lignina se encontra fortemente associada a polissacarídeos da parede celular, o que dificulta seu aproveitamento e praticamente evita seu isolamento da forma não alterada. Portanto, a maioria dos dados sobre a estrutura da lignina é baseada na análise de frações isoladas de monômeros e dímeros derivados de sua degradação química. Os produtos solúveis em água resultantes da oxidação da lignina são predominantemente ácido p-hidroxibenzóico, vanilina e siringaldeído.

Vanilina, um dos produtos da oxidação de lignina, é obtida na forma natural da vagem de orquídeas tropicais do gênero *Vanilla*. Devido à sua baixa produtividade e à dificuldade de cultivo fora de seu habitat, a produção total de vanilina natural é de apenas 40 Ton/ano, suprindo apenas ~0,25% da demanda mundial total. Historicamente, esta demanda vem sendo atendida através de síntese química, primeiramente a partir de eugenol e da lignina presente no licor negro, e posteriormente substituída pela síntese a partir de recursos não renováveis como guaicol e ácido glioxílico. Recentemente, rotas fermentativas naturais tem sido desenvolvidas para a produção de vanilina a partir de fontes renováveis. Uma rota biossintética *de novo* foi desenvolvida em *S. pombe*, produzindo em torno de 45mg/L de vanilina a partir de glicose como matéria prima. A empresa Rhodia desenvolveu um método de produção de vanilina através de bioconversão microbial e vende seu produto com o nome comercial de Rhovanil Natural à $700/Kg. Por ser oriunda de fontes renováveis, este produto carrega o selo "natural" nos mercados dos EUA e da União Européia.

Propomos neste projeto a utilização da fração solúvel da lignina de bagaço de cana, a qual é um resíduo da extração de celulose e hemicelulose do bagaço e é atualmente descartado. Este líquido possui quantidade considerável de compostos fenólicos derivados de lignina que podem ser utilizados como precursores para outras moléculas, entre elas a vanilina. Nosso plano de trabalho para tal consiste em estabelecer um consórcio de microrganismos degradadores a partir do enriquecimento de uma amostra de solo de canavial cultivada em meio contendo lignina solúvel. Este consórcio será caracterizado quanto a sua capacidade de degradar lignina, onde então passaremos para abordagens de metagenômica, metatranscriptômica e tratamentos de dados por bioinformática para prospecção de novas enzimas. Os genes destas enzimas serão sintetizados e caracterizados quanto à eficiência catalítica e posteriormente montados em uma via de síntese, onde utilizaremos uma técnica desenvolvida por nossos parceiros que permite atenuar os níveis de co-expressão utilizando "botões moleculares" denominados *riboswitches*.

A CIBio analisou este projeto em reunião realizada no dia: 09/02/2015 .

Parecer final: [X]-projeto aprovado, [ ]-projeto recusado, [ ]-projeto com deficiências, comentários anexo.

Presidente da CIBio – CNPEM-LNBio
Marcio Chaim Bajgelman

Membro da CIBio da CNPEM -CTBE
Fabio Márcio Squina

Membro da CIBio – CNPEM -LNBio
Celso Eduardo Benedetti

Membro da CIBio da CNPEM -CTBE
Sindélia Freitas Azzoni

Membro da CIBio – CNPEM -LNBio
Carolina Borsoi Mortes H. Freitas

Membro da CIBio da CNPEM -CTBE
Roberto Ruller

ANEXO 3. DECLARAÇÃO SOBRE DIREITOS AUTORAIS

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Enzyme and metabolic pathways prospection for the biosynthesis of vanilin and valorization of lignin from sugarcane bagasse,** não infringem os dispositivos da Lei n.º 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 3 de Abril de 2018

Assinatura : Eduardo Moraes

Nome do(a) autor(a): **Eduardo Cruz Moraes**

RG: 341969035

Assinatura : _____

Nome do(a) autor(a): **Fabio Marcio Squina**

RG: 20981777