From genes to microbial consortia: multi-level bioprospecting of natural and artificial environments

Cristina Vilanova Serrador

VNIVERSITAT DE VALÈNCIA

FACULTAT DE CIÈNCIES BIOLÒGIQUES

DEPARTAMENT DE BIOQUÍMICA I BIOLOGIA MOLECULAR

PROGRAMA DE DOCTORAT EN BIOTECNOLOGIA

From genes to microbial consortia: multi-level bioprospecting of natural and artificial environments

MEMÒRIA PRESENTADA PER CRISTINA VILANOVA SERRADOR, CANDIDATA AL GRAU DE DOCTORA PER LA UNIVERSITAT DE VALÈNCIA

CO-DIRECTORS:
DR. MANUEL PORCAR MIRALLES I DRA. AMPARO LATORRE CASTILLO

VALÈNCIA, NOVEMBRE DE 2016

El Dr. MANUEL PORCAR MIRALLES, Investigador Contratat Doctor de la Universitat de València, i la Dra. AMPARO LATORRE CASTILLO, Catedràtica del Departament de Genètica de la Facultat de Ciències Biològiques (Universitat de València):

AUTORITZEN la presentació de la memòria titulada "From genes to microbial consortia: multi-level bioprospecting of natural and artificial environments" i CERTIFIQUEN que els resultats que inclou van ser obtinguts sota la seua co-direcció a l'Institut Cavanilles de Biodiversitat i Biologia Evolutiva per CRISTINA VILANOVA SERRADOR.

I perquè així conste, signen el següent certificat.

Dr. Manuel Porcar Miralles *Dra. Amparo Latorre Castillo*

València, a 4 de Novembre de 2016

"All we have yet discovered is but a trifle in comparison with what lies hid in the great treasury of nature."

Antoni van Leewenhoek (1680)

Agraïments

De vegades he sentit dir que la sort no es troba, que s'ha de buscar. Crec que és veritat, però mirant enrere tinc cada vegada més clar que la sort també es troba per casualitat. Ho sé perquè aquesta tesi és el resultat, en gran part, de la sort que he tingut en coincidir amb moltes persones:

En primer lloc, he sigut molt afortunada per conéixer els meus directors de tesi: Manel i Amparo. Poder treballar amb Amparo ha sigut, simplement, un luxe, per poder aprendre una miqueta de la seua dedicació i entusiasme per la ciència, i per estar sempre disponible i ajudar-me en tot el que he necessitat. I què dir de Manel, a qui vaig conéixer per pura casualitat, sense imaginar que aquella col.laboració puntual acabaria en quatre anys de treball colze a colze, una tesi i, més important encara, una amistat que espere mantindre sempre. Moltes gràcies als dos pel vostre suport incondicional durant aquestos anys. No puc oblidar-me de donar les gràcies també a Juli i Ximo, per la seua disposició constant a escoltar-me, i pels seus consells sempre sincers i útils.

La sort m'ha permés també coincidir amb gent com Sergio, Manzano, Diego, Quèlo, Rafa... una generació irrepetible de joves científics brillants, amb qui he pogut compartir experiments, converses, i unes quantes cervesses. Igualment, gràcies als meus companys actuals, Christian, Dani, José Manuel, Mariano, Mariana i Carlos per la seua ajuda (i paciència) en diferents parts del treball. Vull agraïr especialment a dos persones més brillants encara, Laura i Kristie, el seu suport durant l'ultima etapa, en forma de consells, ànims, carinyo, i molt de xocolate.

Però, res d'açò seria possible si no haguera tingut la sort de "caure" en una família com la meua, que m'ha suportat sempre, sense condicions. Gràcies especialment als meus pares i al meu germà, i també a Pedro, altra "casualitat" que ha acabat sent un element indispensable per a ser feliç.

A tots, moltíssimes gràcies per participar d'una manera o altra en aquesta tesi.

Contents

Introduction

1. A planet of useful microbes

The term "bioprospecting" is the systematic search for genes, molecules, whole organisms, and other biological elements that can be found within the global diversity of plants, animals, and microorganisms, with a potential for product development (Mateo *et al*., 2011). It is evident that this is not a new discipline, since the first human beings were already selecting biological resources in order to survive. The perception that early humans had on nature and biodiversity is clearly different from that of modern scientists involved in bioprospecting projects, but their goal is essentially similar: using nature for practical applications.

Millions of years of evolution have shaped trillions of life forms on Earth, which harbor a huge variety of structures and mechanisms at the molecular, cellular, and supracellular level. Natural selection has driven the adaptation of living organisms in virtually every habitat existing on our planet, and has "equipped" them with complex repertoires of tools in order to survive. Bioprospecting aims at identifying, isolating and exploiting such tools for particular applications. The most promising bioprospecting efforts, though, focus on microscopic life. Microorganisms are the most abundant and diverse living beings on Earth, and have been able to colonize even the harshest environments of the planet. One of the key aspects that makes microbes useful tools for humans is their extraordinary versatility in terms of metabolic capabilities: they can use plenty of organic and inorganic molecules as sources of energy and nutrients, and have evolved to produce a myriad of primary and secondary metabolites.

1.1. From the microscope to microbiome studies: history of a revolution in microbiology

In 1665, Robert Hooke was the first human to observe a microorganism: he described the fruiting bodies of molds through a microscope and was the first to define the term "cell" to describe what he was observing. Soon after that, Antoni van Leewenhoek examined the first bacteria after scraping his own teeth in 1683 (Brock, 1961). From the first sightings of microbial cells to the isolation of the first bacterial species in pure culture with artificial media, by Robert Koch, two centuries elapsed (Carter, 1987). Microbial culturing initiated the Golden Era of microbiology, allowing the exhaustive study of microorganisms, including the discovery of pathogenic bacteria causing epidemic outbreaks by Koch, Pasteur, and their disciples (Brock, 1999), or the isolation of antibiotic-producing fungi by Alexander Fleming in 1928. The identification and study of microbial species relied for another century in their cultivation and phenotypic characterization, until Carl Woese proposed in 1977 the comparison of ribosomal RNA (rRNA) sequences to identify microbial isolates and find out the relationships among them in the first phylogenetic trees of life (Figure 1A). However, it was soon discovered that only a small fraction (around 1%) of the microorganisms present in environmental samples were cultivable (Staley and Konopka, 1985), so the vast majority of the microbial diversity was, and still is, unexplored.

It was not until 1985 when Norman Pace and collaborators proposed molecular methods to identify microorganisms with no need of cultivation (Olsen *et al*., 1986). The methodology consisted of isolating the total DNA from an environmental sample, PCR-amplifying the 16S rRNA gene with universal oligonucleotides, cloning individual 16S rRNA sequences, and finally sequencing them. With this approach, the uncultivated pool of microorganisms

started to be explored, cultivation-dependent techniques were quickly put aside, and microbial phylogenetic trees dramatically changed to include many more species -most of them uncultivated- belonging to new phylogenetic branches (Figure 1B). The pool of rRNA sequences from the environment were found to diverge deeply in phylogenetic trees, suggesting that the environmental, uncultured organisms identified by rRNA sequencing are potentially more different from known organisms than known organisms are from each other (Pace *et al.*, 1997).

The true revolution in the study of microbial diversity began in 2002 with the birth of the so-called metagenomic sequencing. The seminal works by Tyson and Venter reported the application of random, whole-genome sequencing to the study of microbial populations in environmental samples (Tyson *et al.*, 2004; Venter *et al.*, 2004). Despite the computational effort needed to analyze metagenomic data, this new technology led to the explosion of microbiome studies, since the entire microbial community present in a sample could from that moment be massively identified in a single

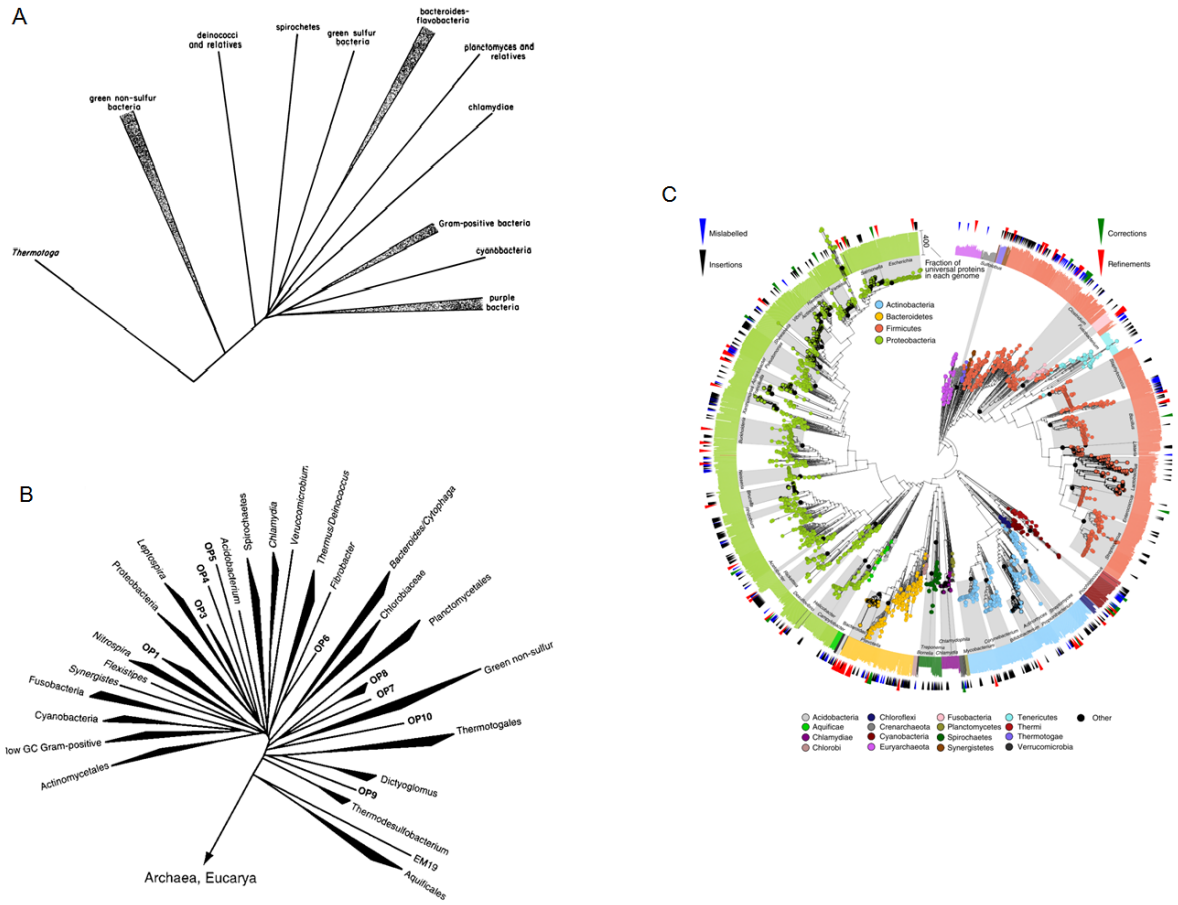


FIGURE 1: Comparison of the molecular trees of life throughout the evolution of microbiology and sequencing techniques. A) Tree of bacterial diversity proposed by Carl Woese (Woese, 1987) containing 11 eubacterial phyla based on 16S rRNA sequences; B) Tree of bacterial diversity published after the development of 16S rRNA cloning and sequencing techniques (Pace *et al.*, 1997): the number of eubacterial phyla increased up to 33, and branches labeled as "OP" indicate phyla without cultured representatives; C) Tree of bacterial diversity published after the explosion of shotgun metagenomic sequencing (Segata *et al.*, 2013) based on the nucleotide sequence of 400 broadly conserved genes.

experiment (up to 1800 species and several millions of genes in the first report by Venter and collaborators). Today, thousands of previously uncharacterized genomes have been sequenced, thus expanding the microbial biodiversity databases in an exponential trend (Figure 1C).

More recently, other -omic technologies complementing the information of metagenomics have arose (metatranscriptomics, metaproteomics, meta-metabolomics, etc.), focusing on the dynamic pool of biomolecules present in a community (Franzosa *et al*., 2015) and thus increasing the landscape of tools for the high-throughput ecological analysis of complex biomes. The increasing power of biomolecule extraction, identification, and analysis has contributed to today's notion of "multi-omics" as the most complete and thus suitable approach to tackle the complexity of microbial communities (Franzosa *et al*., 2015; Jansson and Baker, 2016).

1.2. Multi-level bioprospecting: approaches to the mining of microbial communities

From the most simple genetic elements, to the entangled network of interactions established in any ecosystem, the biological elements found within a microbial community are distributed in different levels of complexity (Figure 2). Cultivation in artificial media, metagenomic sequencing, functional metagenomics, and variations of these techniques are, to date, the main approaches for bioprospecting environmental samples. Since the biological elements targeted are different, their combined use in bioprospecting efforts results in an exhaustive mining of the microbial communities under study (Figure 3). In this thesis, "multi-level bioprospecting" refers to the holistic approach we have used to tackle with the whole diversity of biological elements found within microbial communities. The following sections provide a brief description of some of the techniques that might be combined in multi-level bioprospecting efforts.

1.2.1. Sequence-driven analyses

Metagenomics Metagenomics has accelerated the process of discovery of novel microorganims or enzymes by enabling the detection of the entire diversity held within natural microbial populations, with no need of cultivation. The basis of metagenomic sequencing is conceptually simple: total DNA is directly isolated from the samples, randomly fragmented and sequenced, and finally assembled and annotated by means of computational techniques. The current power of sequencing technologies and their dropping costs enable the identification of impressive catalogues of species, genes, and other genetic elements from environmental samples with a moderate economic effort (Carlson, 2010). Case examples of metagenomic studies resulting in the identification of genes with direct applications in industry are the metagenomic study performed on a biphenyl-contaminated river sediment, which led to the cloning of biphenyl dioxygenase genes useful for the bioremediation of byphenil-related compounds (Sul *et al*., 2009); the massive identification of biomass-degrading enzymes from cow rumen (Hess *et al*., 2011) or glycosyl hydrolases from a range of environments (Li *et al*., 2009) with applications for the transformation of plant feedstocks into biofuels.

16S/18S rRNA profiling An alternative approach, aiming at the exploration of community taxonomic composition rather than gene content, consists of targeting short fragments of ribosomal RNA (rRNA) genes and noncoding regions, such as the 16S rRNA gene in eubacteria and archaea, and the internal transcribed spacer (ITS) region in fungi. In this approach, the target genes are PCR-amplified, then sequenced, and finally taxonomically assigned (Sogin *et al*., 2006). One of the main advantages of this approach is that the oligonucleotides used in the PCR can be tailored to contain short nucleotide sequences (barcodes) that are specific for each of the metagenomic DNAs under study, thus allowing the combination of multiple samples in the same sequencing experiment (Caporaso *et al*., 2011).

The development of this technology brought out the challenge of confidently assigning sequences to taxonomy and obtaining estimations of species richness for the communities under study. Even though there are many approaches to tackle those problems, most of the algorithms currently available for the analysis of massive 16S/18S rRNA profiles group the sequences into the so-called operational taxonomic units (OTUs), which are groups of sequences that are clustered according

to a given threshold of similarity (Schloss and Handelsman, 2005; Caron *et al*., 2009). But, how does identity information translate into classical taxonomic classification? Sequences with greater than 97% identity are typically assigned to the same species, those with >95% identity are typically assigned to the same genus, and those with >80% identity are typically assigned to the same phylum, although these distinctions are still very controversial (Schloss and Westcott, 2011).

Single-cell genomics While metagenomics focuses on entire microbial communities, single-cell genomic approaches separate individual cells from the microbial community as a previous step to genome sequencing. This approach is more difficult from the experimental point of view, since it requires numerous steps, from the isolation of single cells from the environmental sample, to cell lysis, whole genome amplification, and genome sequencing (Woyke *et al*., 2009). However, performing single-cell genomics on environmental samples has the advantage of targeting the so-called "microbial dark matter", that is, the microorganisms belonging to the rare biosphere, and virtually impossible to sequence through traditional metagenomic techniques.

Some flagship studies on the exploration of such rare microorganisms are the works describing the isolation and sequencing of TM7 cells (belonging to a recently discovered phyla with no cultivated representatives to date) that were present at a frequency varying between 0.7 and 1.9% in the ecosystems studied (Marcy *et al*., 2007; Podar *et al*., 2007); or the very recent report by Woyke and collaborators (2016) on microbial dark matter lineages yielding the genomes of more than twenty bacterial and archaeal representatives from phyla composed exclusively of uncultivated members.

Other "-omics" A range of techniques have been developed for the massive analysis of other biomolecules different than DNA. For instance, metatranscriptomics involves the collection of the entire pool of RNA molecules present in a community, which is used to construct cDNA libraries that are sequenced. RNA-based analyses give a snapshot of the current state of the community by revealing which genes are active, this is, transcribed into RNA. Metaproteomics consists of the isolation of the whole protein content of a community and its analysis through mass spectrometry techniques, thus providing an image into the entire protein complement of the microbial communities and insights into the genes expressed and the key metabolic activities characterizing the community. Finally, meta-metabolomics provides information on the secreted or modulated metabolite composition of the microbial communities by massively identifying such molecules through gas chromatography and mass spectrometry techniques. Together, these approaches can complement metagenomics data and improve our understanding of the ecology of complex microbial communities (Franzosa *et al*., 2015).

This expanded set of "-omic" tools has allowed the exploration of microbial communities from a different perspective, especially in the studies of adaptation to specific ecological niches (Ram *et al*., 2005; Lo *et al*., 2007), and led to the identification of novel enzymes or proteins previously undetected with metagenomic approaches (Valenzuela *et al*., 2006; Frias-López *et al*., 2008; Gilbert *et al*., 2008).

1.2.2. Function-driven analyses

On the other hand, functional metagenomics has proved one of the most efficient tools for the isolation of new biocatalysts. This strategy consists of the isolation of genes -randomly cloned from a metagenome- into libraries that are expressed in culturable microorganisms, often the well-known model bacterium *Escherichia coli*. Through an appropriate screening assay, with a particular selection pressure, a single gene or a set of genes expressing a particular enzymatic function can be identified and their products further analyzed (Simon and Daniel, 2009). Functional metagenomics has the advantage of not requiring having the sample sequenced, nor the cultivation of microorganisms or previous sequence information for the targeted genes, thus representing a valuable approach for mining enzymes with new features (Mirete *et al*., 2016). This is the case of the cold-adapted enzymes (lipases and estereases) isolated from Antarctic soils, which display important applications in detergents to be used at low temperatures, as well as in food processing, or molecular biology applications (Berlemont *et al*., 2011). Another example is the thermostable

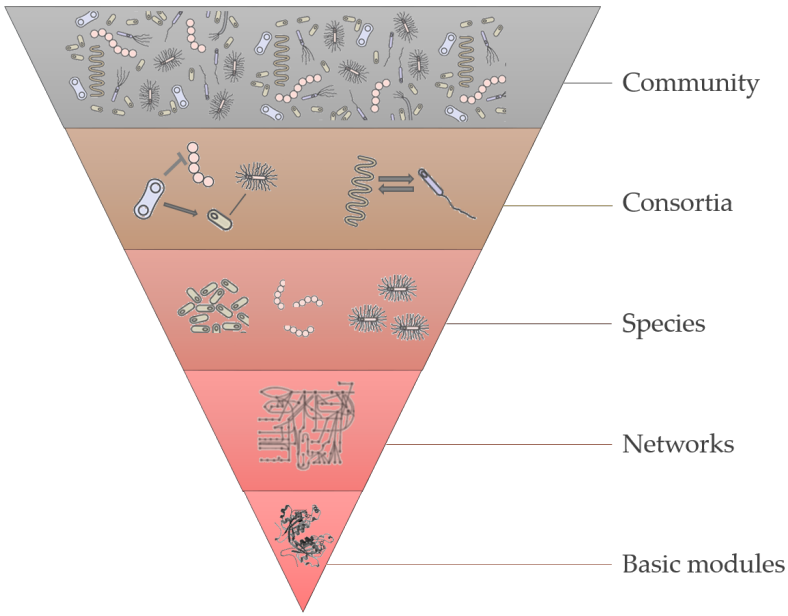


FIGURE 2: Top-down representation of complexity in the biological elements of a microbial community.

DNA polymerase with reverse transcriptase (RT) activity, now patented for its use in reverse transcription PCR (RT-PCR) reactions, isolated from a hot spring in Yellowstone National Park (Moser *et al.*, 2012). A more exhaustive list of enzymes isolated through functional metagenomics and currently commercialized is available in the report by Ekkers *et al.*, 2012. Today, new functional metagenomics approaches are being developed in order to achieve high-throughput functional screening. These approaches are focused, mainly, on the improvement of transformation (Colin *et al.*, 2015) and expression efficiency (Lam *et al.*, 2015) and the use of alternative hosts for the expression of the libraries (Cheng *et al.*, 2014; Liebl *et al.*, 2014).

1.2.3. Culturing

Finally, microbial culturing in artificial media is still the basis for most microbial bioprospecting efforts on environmental samples, as this both gives access to the totality of genomic information in the (culturable) isolates, and makes possible the study of their phenotype in the laboratory (de Pascale *et al.*, 2012). The number of microbial strains isolated by means of standard culturing for biotechnological purposes is actually impressive, and their applications vary from industrial processes of the food, chemical, and pharmaceutical sectors, to growing areas such as environmental bioremediation, biopesticide production and bioenergy. Nowadays, cultivation techniques are being revisited and new, innovative approaches are increasing our ability to isolate, domesticate, and characterize new microbial strains in the laboratory. In order to achieve this goal, three main approaches are developed:

- "Mimicking" the natural, physico-chemical conditions of a particular habitat in order to isolate hard-to-culture species from the environmental samples. This is the principle of the so-called *in situ* cultivation, which has been developed under different set ups (Nichols *et al.*, 2010; Jung *et al.*, 2014) and has proved successful for the culturing of novel species with outstanding applications (Ling *et al.*, 2015).

- Increasing the number of microbial cells that are intended to be artificially grown by using high-throughput microfluidic devices. As in the previous strategy, this experimental configuration encloses individual cells in micro-chambers, but do also allow the high-throughput exploration of the isolates, which can be one-by-one subjected to characterization through

molecular techniques (Liu *et al.*, 2009; Ma *et al.*, 2014).

- Targeting natural microbial consortia rather than individual species. This approach aims at the cultivation of particular strains, whose growth strongly depends on the presence of other species, by enriching the samples in the microbial consortia (Luo *et al.*, 2015; Liang *et al.*, 2015).

2. A particularly relevant level of complexity: microbial consortia

It is known that microbial consortia are ubiquitous in nature. Microbial species are often naturally assembled in the shape of consortia in order to perform complicated functions that individual species cannot carry out, and because consortia can be more robust to environmental fluctuations (Brenner *et al.*, 2008).

2.1. Types of interactions

Different types of interspecies interactions operate in the biosphere. Following the classical categories established for macroorganisms, these symbiotic interactions can be either positive -mutualism or commensalism- or negative -parasitism or competition- (Figure 4). Most of the microbial interactions studied to date are based on mutualism, in which two or more different species living in close proximity rely on each other for nutrients, protection, and/or other vital functions (Odum and Barrett, 2005; Moya *et al.*, 2008). A particular case of mutualistic relationships is syntrophy, which is based on providing trophic benefits for the partners of the consortia. An example of this is the consortium formed between fatty-acid oxidizing bacteria, which produce hydrogen as an end-product, and methanogenic archea, which use hydrogen to oxidize CO_2 to CH_4 (McInerney *et al.*, 2008). Another case of syntrophy is the division of labor in metabolic pathways. This is the case of the synthesis of tryptophan by *Buchnera aphidicola* and *Serratia symbiotica* in the bacteriocytes of the aphid *Cinara cedri*: *B. aphidicola* is able to convert chorismic acid (the precursor metabolite) to anthranilic acid, which is then used by *S. symbiotica* to produce tryptophan (Perez-Brocal *et al.*, 2006; Gosalbes *et al.*, 2008). But mutualism is not always restricted to the exchange of nutrients. In fact, an astonishing example of mutualism is established in *Chlorochromatium aggregatum*,

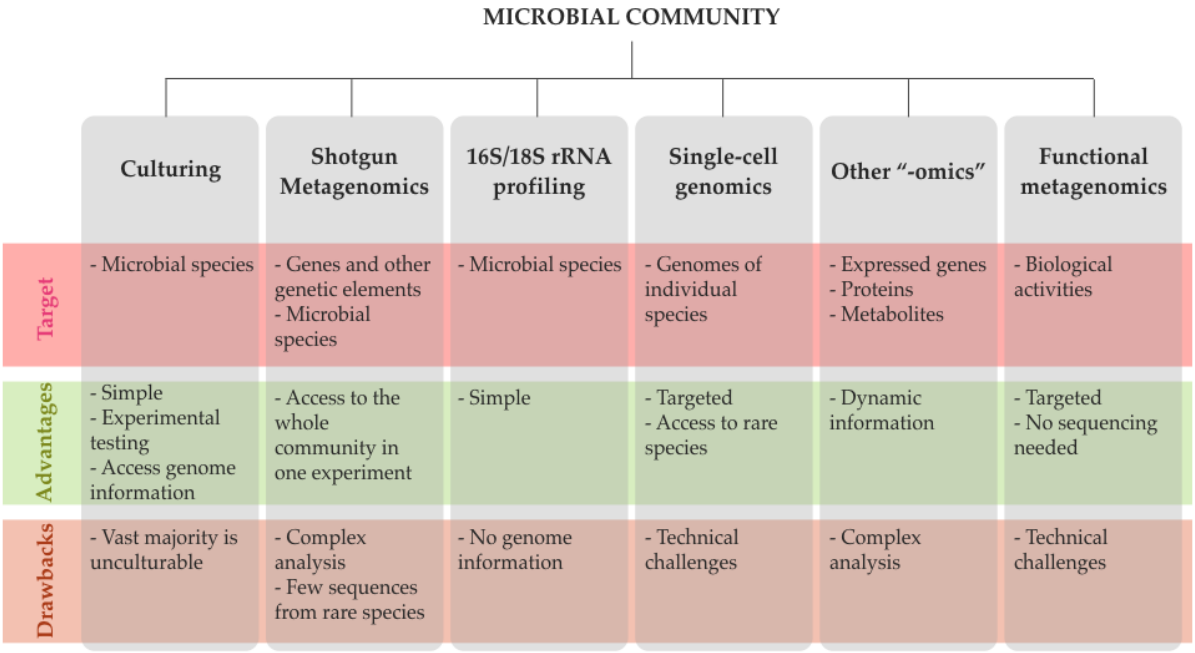


FIGURE 3: Approaches to microbial bioprospecting.

a motile photosynthetic community composed of photosynthetic, green sulfur bacteria, and a β-proteobacterium. These bacteria are naturally attached to each other, in such a way that the photosynthetic partner takes advantage of the β-proteobacterium, which has motility and chemotaxic and scotophobotaxic (movement away from the dark) ability. Reciprocally, the β-proteobacterium gains access to energy from its partner (Wanner *et al*., 2008). Commensalism describes a relationship between two living organisms where one benefits and the other is not significantly benefited or harmed. An example from the microbial world is the scavenging of metabolites released from a producer organism as waste products by a receiver organism (Jagmann and Philipp, 2014).

On the other hand, competition for nutrients is the most widespread negative interaction in microbial communities. In fact, microorganisms have developed efficient strategies for competition such as antibiotic production (Chater *et al*., 2010). Finally, parasitic or predator-prey interactions are also found in the microbial world. This is the case of the parasitic growth of *Pseudomonas aeruginosa* in co-culture with the chitinolytic bacterium *Aeromonas hydrophila* (Jagmann *et al*., 2010). *P. aeruginosa* uses secondary metabolites to manipulate the metabolism of *A. hydrophila*, in such a way that chitin is incompletely oxidized to acetate, which serves as a growth substrate for *P. aeruginosa*.

2.2. Types of microbial consortia and applications in industry

A range of microbial consortia are known to play key roles in many industrial bioprocesses, especially in food and beverage production and in wastewater treatment. Also, they are central in human health or bioremediation. Some of these applications are cited below and grouped depending on the origin of the consortia.

Natural consortia A microbial consortium is termed "natural" when its members have not been altered by human selection, this is, the composition of the consortium is the one found in nature. Some examples of this type of consortia are the communities naturally able to degrade tetrachloroethene (Kotik *et al*., 2013), crude oil (Darvishi *et al*., 2011; Bao *et al*., 2012) and herbicides (Marrón-Montiel *et al*., 2006), or the complex consortia inhabiting the human gut (Gilbert *et al*., 2016), which have a direct influence on health. Also, enrichment cultures of natural consortia have been used by humans without altering their composition in order to apply them in wastewater treatment (Daims *et al*., 2006), composting (He *et al*., 2012), or the production of traditional beverages and food (Paramithiotis *et al*., 2006; De Vuyst *et al*., 2014).

Artificial consortia Artificial consortia are the associations of microorganisms that have been shaped by humans, and they are often called "co-cultures". For instance, several microbial species have been artificially combined to achieve the production of biofuel from plant feedstocks (Ward *et al*., 1995; Abate *et al*., 1996); or combinations of selected bacterial and yeast strains have been used as starter cultures for the production of yoghurt, cheese, sourdough, kefir, and wines with particular organoleptic properties (Bader *et al*., 2009).

Synthetic consortia Finally, natural and artificial consortia can be converted in synthetic consortia if any of the members is genetically modified. The design of synthetic consortia is considered one of the ultimate goals of Synthetic Biology (Brenner *et al*., 2008), and some synthetic associations between genetically-engineered *E. coli* and *Saccharomyces cerevisiae* strains have been reported as efficient producers of biofuels from different substrates (Qian *et al*., 2006; Eiteman *et al*., 2009; Shin *et al*., 2010; Goyal *et al*., 2011).

2.3. Current techniques to detect and isolate microbial consortia from complex communities

To date, one of the most common methodologies to isolate natural microbial consortia has been a "top down" approach: a mixture of microorganisms present in an environmental sample is used to inoculate the "test material" (a substrate containing the physical or chemical elements for which the consortium is going to be selected), and the assumption is made that a working combination of these original microorganisms will emerge after several sub-culturing steps as a stable and effective consortium (Rawlings and Johnson, 2007). This approach has been used for the selection

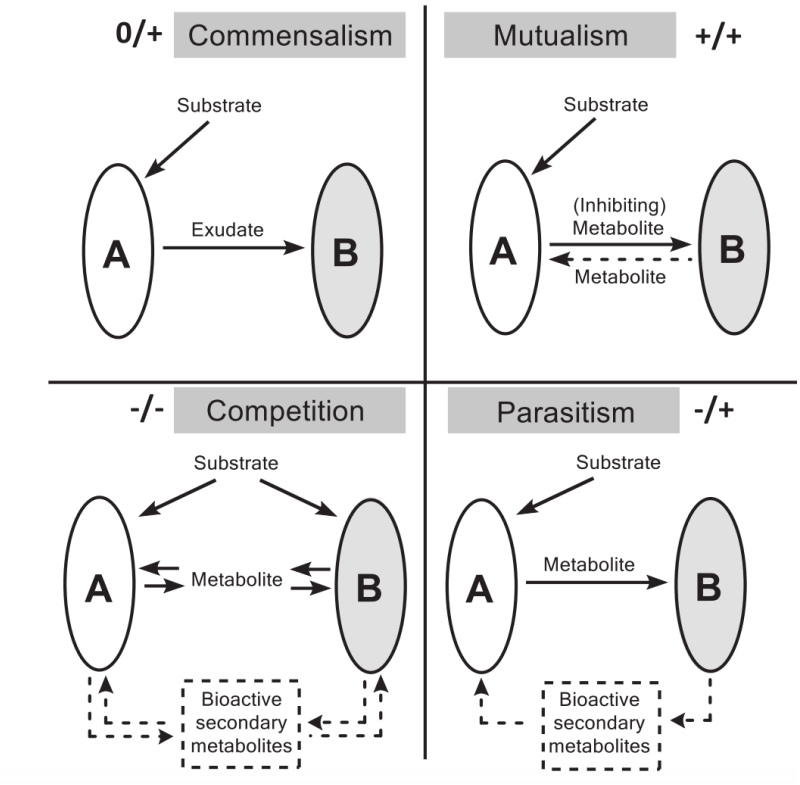


FIGURE 4: Schematic representation of possible interactions established between two bacterial species. Image modified from Jagmman and Philipp (2014).

of natural consortia with particular electron transfer abilities in microbial fuel cells (Rabaey *et al*., 2004), mineral-oxidizing activity (Rawlings and Johnson, 2007), or TNT and 2,4-D degrading ability (Marrón-Montiel *et al*., 2006; Muter *et al*., 2012).

Besides this "blind" approach, metagenomic sequencing offers the opportunity to access all the genomic information of the microbial communities -and the consortia therein- present in the sample under study; but at the same time, poses a challenge for the identification of the members of each consortium within all the members of the microbial community. In this context, a search of co-occurrence and correlation patterns among microbial species within metagenomic datasets has been carried out in order to identify putative microbial consortia (Faust and Raes, 2012). The rationale is simple: species A and B are positively correlated if, when the frequency of A increases, so does the frequency of B. Contrary, A and B are negatively correlated if, when the frequency of A increases, the frequency of B decreases. Finally, A and B are not correlated when their frequencies are independent. Obviously, these patterns can only be found when comparing multiple datasets, so two alternative approaches exist: the cross-sectional analysis of data from different samples/individuals (Faust and Raes, 2012) and the time-series analysis of data from a single sample/individual (Friedman and Alm, 2012; Stein *et al*., 2013). Due to the multiple number of samples that need to be sequenced with both approaches, the relatively unexpensive 16S rRNA sequencing (or ITS for fungal communities; Toju *et al*., 2016) is normally used for these experiments.

However, the detection of correlation patterns among bacterial species from 16S rRNA metagenomic profiles raises a major problem: 16S rRNA data provide relative abundances, which are strictly interdependent in a given sample, since they always sum 1, and this effect is of particular importance when the sample is dominated by few or even a single species (this problem is known as "compositional effect"). In this sense, the SparCC method reported by Friedman and Alm (2012) may overcome this issue by log-transforming the ratios of

relative abundances. It has to be emphasized, though, that statistical correlation does not imply biological interaction: if species A interacts with species B, and at the same time, B interacts with C; A and C are surely correlated, even if they are not establishing any direct biological interaction. Fisher and Mehta (2014) proposed a method to detect interactions among bacterial species fitting the discrete Lotka-Volterra model within time series analysis of 16S rRNA datasets, which has been used to detect keystone species in the human gut of different individuals. In the present thesis, a method to visualize both correlations and interactions among bacterial species within complex communities has been developed (Chapter 2).

3. Natural environments as sources of new biotechnological tools: the particular case of insects

In the early years of metagenomic sequencing, natural environments were the unique target of all the microbiome studies. The two first studies by Venter and collaborators on the Sargasso sea (Venter *et al*., 2004) and Tyson and collaborators on a natural acidophilic biofilm (Tyson *et al*., 2004), were the paradigm of the next ones to follow: on one hand, the focus was on habitats with an extremely high diversity of microorganisms -including soil (Daniel, 2005; Delmont *et al*., 2011), ocean (Sogin *et al*., 2006), cow rumen (Ross *et al*., 2012), and human gut (Turnbaugh *et al*., 2007; Qin *et al*., 2010)-; and on the other hand, many research groups focused on the exploration of the so-called "extreme" environments such as thermal vents (Xie *et al*., 2011), hot springs (Menzel *et al*., 2015), and permafrost (Mackelprang *et al*., 2011), which were supposed to host microbial communities of limited diversity but highly specialized in terms of their physical or chemical adaptation to the harsh environment.

The study of the microbial communities associated to insects has traditionally been limited to endosymbiotic bacteria. From the very first works at the beginning of the XX century by Pierantoni and Buchner (Buchner, 1965), to the most detailed deep-sequencing analysis of the current research, the very intimate relationships established between some insect species and their symbiotic bacteria have been unveiled even at the molecular level. One of the paradigmatic examples is the microbial consortia -mentioned previously- formed by *B. aphidicola* and *S. symbiotica* in the cedar aphid *C. cedri*, in which a metabolic complementation is established among the members of the consortia to achieve a global supply of tryptophan (Lamelas *et al*., 2011). Similarly, other types of complementation have been reported for psyllids (Sloan and Moran, 2012), sharpshooters (Wu *et al*., 2006), whiteflies (Rao *et al*., 2015), cicadas (McCutcheon *et al*., 2009), spittlebugs (McCutcheon and Moran, 2010), and leafhoppers (Nishino *et al*., 2016). Besides the reduced and highly specialized endosymbiotic bacteria found in many insect species, it is known that more diverse microbial communities do also colonize insect's gut, often playing an important role in regulating the host metabolism, promoting an efficient digestion from ingested foods (Kaufman and Klug, 1991) and protecting the host from other, potentially harmful microbes (Dillon and Charney, 2002). In this sense, the microbial communities associated to a range of insect taxa have been already reported: termites (Kohler *et al*., 2012; Boucias *et al*., 2013), ants (Russell *et al*., 2009; Funaro *et al*., 2011; Poulsen and Sapountzis, 2012), fire bugs (Sudakaran *et al*., 2012; Salem *et al*., 2013), fruit flies (Chandler *et al*., 2011; Wong *et al*., 2011), beetles (Reid *et al*., 2011; Arias-Cordero *et al*., 2012; Hulcr *et al*., 2012), bees (Mohr and Tebbe, 2006; Martinson *et al*., 2011; Engel *et al*., 2012), mosquitos (Wang *et al*., 2011; Bossière *et al*., 2012), and cockroaches (Carrasco *et al*., 2014; Schauer *et al*., 2014; Berlanga *et al*., 2016). In the particular case of Lepidoptera, reports have been limited to species considered as major pests, such as the gypsy moth *Lymantria dispar* (Broderick *et al*., 2004; Mason and Raffa, 2014); the diamondback moth *Plutella xylostella* (Lin *et al*., 2015); or the cotton bollworm *Helicoverpa armigera* (Xiang *et al*., 2006). Also, some focus has been put on how diet changes influence the gut microbiota of polyphagous insects such as *Bombyx mori* (Liang *et al*., 2014), *Spodoptera littoralis* (Tang *et al*., 2012), or *Ostrinia nubilalis* (Belda *et al*., 2011).

The gut microbiota of insects has also been a source of microbial species with biotechnological applications. The most famous example of this, is the *Bacillus thuringiensis* strain Berliner, first described by Ishiwata in 1905, and later isolated from the gut of a flour

moth in 1915 by Berliner. *B. thuringiensis* Cry proteins are the basis of many bioinsecticides, mostly used against lepidopteran, dipteran and coleopteran larvae; and their coding genes are actually the heterologous sequences expressed in the now widely used transgenic crops resistant to the attack of different insect pests (Roh *et al*., 2007). Also, microbial isolates from phytophagous insects (Zhou *et al*., 2009; Arnand *et al*., 2010; Vilanova *et al*., 2012) and wood-feeding insects, especially termites (Kuhnigk and König, 1997; Kato *et al*., 1998), have been reported to host a biological arsenal of enzymes for the degradation of cellulose and ligninic compounds, and might be of great interest for the production of biofuels from plant feedstocks.

In this thesis, the microbial communities associated to three lepidopteran species -*Retinia resinella* (Tortricidae), *Brithys crini* (Noctuidae), and *Hyles euphorbiae* (Sphingidae)- with very particular, highly toxic-rich diets, have been investigated through both metagenomic sequencing and culturing techniques (Chapter 1A).

4. The virtually unexplored world of "artificial" environments

Although most of the metagenomic studies of microbial communities have been traditionally focused on the so-called "natural" ecosystems, recently, some researchers have turned the focus to "artificial" environments. Of particular importance is the field of the built environment, this is, the human-made surroundings that provide the setting for human activity. To date, the microbiome of public restrooms (Gibbons *et al*., 2015), hospitals (Kanden *et al*., 2014; Pereira *et al*., 2016), residencial kitchens (Flores *et al*., 2012), university classrooms (Bright *et al*., 2010; Hayleeyesus *et al*., 2014; Ross and Neufeld, 2015), office buildings, and metropolitan subways (Leung *et al*., 2014), among many others, has been studied through metagenomic sequencing or related techniques. Results from these studies suggest that outdoor air and human-associated microorganisms are robust sources of the indoor microbiota, and that the microbial communities of the built environment are strongly influenced by factors such us building type, occupancy, or geography (Kembel *et al*., 2014; Adams *et al*., 2015).

But even more scarce are the reports describing the microbiome of artificial, human-manufactured devices. Most of this pioneer studies are focused on hand-manipulated devices such as computer touch screens (Gerba *et al*., 2016), mobile phones (Meadow *et al*., 2014), and money (Vriesekoop *et al*., 2010); while others describe the microbiome of artificial devices with implications for human health: drinking water distribution systems (Revetta *et al*., 2016) and air conditioning units (Diekmann *et al*., 2012). Table 1 summarizes the main findings of these works. This thesis contributes to the exploration of artificial devices as sources of useful biological tools by analyzing the microbial communities associated to human-manufactured environments subjected to extreme physical and chemical conditions: solar photovoltaic panels (Chapter 1A) and coffee machines (Chapter 1B).

5. Beyond microbiome studies: bioprospecting meets Synthetic Biology

Beyond its applications in environmental microbiome studies, bioprospecting is an indispensable step for Bioengineering and Synthetic Biology to be developed. These disciplines focus on the architecture of circuits composed of biological bricks, and bioprospecting is, indeed, the tool for searching those useful bricks. In more detail, Synthetic Biology understands biological systems as biological machines that can be subjected to engineering principles (i.e.: standardization, modularity, orthogonality) as any other machine (Endy, 2005). In the same way that nuts and bolts are the basis of the machines surrounding us (Porcar *et al*., 2015), biological parts are the basis of biological machines. For instance, a synthetic microbial consortium is composed of different biological parts belonging to increasing complexity levels: genes, gene networks, individual organisms, and the ultimate consortium. To date, most of the systems designed through Synthetic Biology have been centered on individual genes, which are "simple" biological parts (Cameron *et al*., 2014). According to the engineering principles of Synthetic Biology, those genes should behave in a similar way in different

TABLE 1: Microbiome studies performed to date on artificial devices.

Device	Experimental approach	Main findings	Reference
Computer touch screens	Cultivation	Detection of opportunistic pathogenic bacteria on touch screens in hospitals and enteric bacteria on grocery store touch screens.	Gerba *et al.*, 2016
Mobile phones	16S rRNA massive sequencing	The microbial communities on smartphone touchscreens significantly overlap with the skin microbiome from the owners.	Meadow *et al.*, 2014
Money	Cultivation	The bacterial communities are influenced by the material of the notes, the economic prosperity of the country, and the age of the notes.	Vriesekoop *et al.*, 2010
Drinking water distribution system	16S rRNA cloning and sequencing	A biofilm community is formed after some months, dominated by *Mycobacterium* species. The distribution positively correlates with the drinking water distribution system's temperature.	Revetta *et al.*, 2016
Air conditioning units	16S rRNA SSCP and sequencing	Bacterial communities are dominated by Sphingomonadales, Burkholderiales, Bacillales, *Alcanivorax* sp. and *Stenotrophomonas* sp.	Diekmann *et al.*, 2012
Dental unit waterlines	16S rRNA massive sequencing	High bacterial diversity shifting towards Gamma- and Alpha-Proteobacteria as water circulates in the dental unit.	Costa *et al.*, 2015

organisms (standard behavior), and under changing conditions (stability), and should interact with other genetic elements only in a predictable way, without unexpected, emergent properties (orthogonality). However, different studies carried out during this thesis proved that those engineering principles are not always met by biological parts. First, we observed that the biological parts collected in repositories of the Synthetic Biology community (particularly, the Registry of Standard Biological Parts of the iGEM competition) are not properly characterized and rarely re-used by the community (Vilanova and Porcar, 2014; available in Appendix D). Second, we experimentally tested a set of biological parts from the same repository, and confirmed the lack of standard, stable, and orthogonal behavior (Vilanova *et al.*, 2015; available in Appendix D).

To overcome this issue, one possibility is to engineer the sequence of each biological part in order to optimize its standard behavior (Wang *et al.*, 2011; Rhodius *et al.*, 2013). However, bioprospecting biological parts (genes) meeting Synthetic Biology principles is actually a powerful approach, since such biological parts do naturally occur in nature. A clear example of this are the so-called genomic islands (Rodríguez-Valera *et al.*, 2016), genomic regions encoding basic biological functions –parts- that act in combination conferring a defined function to the receiving organism. One of the most characterized paradigms of this phenomenon are the pathogenicity islands of uropathogenic strains of *E. coli* (Kaper *et al.*, 2004), where up to four protein-coding genes (*hlyC*, *hlyA*, *hlyB*, and *hlyD*) are the parts that determine the synthesis, activation and transport of alpha-hemolysin outside the bacterial cell, constituting a virulence-conferring device. Other examples of natural parts able to work similarly in different organisms are the metabolic islands found in *Xanthomonas campestris* (allowing the production of xanthum gum; Lima *et al.*, 2005), or the resistance islands of several

Salmonella species (Michael and Schwarz, 2016). Therefore, bioprospecting techniques aiming at the selection of naturally standard parts, or combinations of parts, are a promising tool for the robust development of Synthetic Biology. This techniques may include a screening of biological parts with a given activity (i.e.: heat-activated promoter of gene expression), followed by their characterization in different hosts in order to select the ones exhibiting the highest standard behavior. Similarly, bioprospecting microbial consortia can be a key step for the design of synthetic microbial consortia. For instance, the natural assemblages observed for microbial consortia in insects, cow rumen, and sludge have inspired the metabolic configuration and also the spatial distribution of synthetic consortia (Agapakis *et al*., 2012, and references therein) or synthetic enzymatic assemblages (Goyal *et al*., 2011; Alcalde, 2015). Reciprocally, Synthetic Biology circuits can be applied to new bioprospecting techniques based on the detection of specific activities in environmental isolates through a synthetic reporter system (Kim *et al*., 2016).

On the other hand, bioprospecting can be directly linked to another goal of Synthetic Biology: the creation of an artificial, minimal cell. Even though the first artificial cell with a synthetic chromosome containing a minimal set of genes for life has been recently reported (Hutchinson *et al*., 2016), there are still many shadows regarding the gene content of minimal cells (Moya *et al*., 2009; Hutchinson *et al*., 2016) and their fitness for future biotechnological or field applications. In this sense, the metagenomic exploration of natural minimal cells, such us insect endosymbionts (Moya *et al*., 2008) has been essential for the definition of the minimal number of genes able to sustain life (Gil *et al*., 2004; Gabaldon *et al*., 2007). More recently, innovative bioprospecting approaches, such as analyzing the fraction of cells present in an environmental sample passing through a 0.2 and 0.1 μm filter, have led to the discovery of new, ultra-small bacteria with very reduced genomes and intriguing forms of metabolism (Luef *et al*., 2015). This offers new insights into alternative forms of minimal life still able to thrive in natural environments.

Objectives

This thesis aims at the bioprospecting of yet-unexplored, natural and artificial environments in search of biological elements with potential applications in biotechnology and Synthetic Biology. In our view, the most powerful approach for this is multi-level bioprospecting, consisting of a compendium of approaches targeting different levels of biological complexity: from the most simple one, genes, to gene networks, proteins, metabolites, individual microbial species, and finally the most complex microbial consortia.

This is an ambitious goal that requires a range of microbiology, molecular biology, and bioinformatics techniques, and therefore exceeds the time limitations of a PhD. In the context of this broad framework, the objectives of the present thesis have been defined as:

- Unveiling the microbial communities associated to previously unexplored, natural and artificial environments (Chapter 1A and 1B, respectively) by means of metagenomic sequencing in order to identify microbial species and genes with potential applications in biotechnology.

- Isolating and characterizing individual strains from those environments with potential applications in biotechnology by means of culturing in artificial media, selection, and/or determination of the biological activity (Chapters 1A and 1B).

- Developing a new approach for the identification of microbial consortia from metagenomic data, and unveiling the microbial associations present in a selection of the natural and artificial environments studied (Chapter 2).

Chapter 1A. Microbial communities associated to natural environments: the case of toxic-feeding insects

Publication I:
Vilanova C, Marín M, Baixeras J, Latorre A, Porcar M. (2014). Selecting microbial strains from pine tree resin: biotechnological applications from a terpene world. *PLoS One* **9**:e100740.

Publication II:
Vilanova C, Baixeras J, Latorre A, Porcar M. (2016). The generalist inside the specialist: gut bacterial communities of two insect species feeding on toxic plants are dominated by *Enterococcus* sp. *Front Microbiol* **7**:1005.

Publication I

Selecting microbial strains from pine tree resin: biotechnological applications from a terpene world

Cristina Vilanova[1], Maria Marín[1], Joaquín Baixeras[1], Amparo Latorre[1,2] and Manuel Porcar[1,3]

[1] Cavanilles Institute of Biodiversity and Evolutionary Biology (Universitat de València), Spain.
[2] Unidad Mixta de Investigación en Genómica y Salud, CSISP, Valencia, Spain.
[3] Fundació General de la Universitat de València, Spain.

Abstract

Resin is a chemical and physical defensive barrier secreted by many plants, especially coniferous trees, with insecticidal and antimicrobial effects. The degradation of terpenes, the main components accounting for the toxicity of resin, is highly relevant for a vast range of biotechnological processes, including bioremediation. In the present work, we used a resin-based selective medium in order to study the resin-tolerant microbial communities associated with the galls formed by the moth *Retinia resinella*; as well as resin from *Pinus sylvestris* forests, one of the largest ecosystems on Earth and a yet-unexplored source of terpene-degrading microorganisms. The taxonomic and functional diversity of the cultivated, resin-tolerant fraction of the whole microbiota were unveiled by high-throughput sequencing, which resulted in the detection of more than 40 bacterial genera among the terpene-degrading microorganisms, and a range of genes involved in the degradation of different terpene families. We further characterized through culture-based approaches and transcriptome sequencing selected microbial strains, including *Pseudomonas* sp., the most abundant species in both environmental resin and *R. resinella* resin-rich galls, and three fungal species, and experimentally confirmed their ability to degrade resin and also other terpene-based compounds and, thus, their potential use in biotechnological applications involving terpene catabolism.

Introduction

Coniferous resin is a complex mixture of secondary metabolites. Resin protects injured tree tissues from phytophagous insects (Hanover, 1975) and plant pathogens (Buijtenen and Santamour, 1972; Marpeau *et al.*, 1989; Himejima *et al.*, 1992; Pearce, 1996). Terpenoids, flavonoids, and fatty acids are the main components of resin (Trapp and Croteau, 2001; Tomlin *et al.*, 2000).

Among these, terpenes (containing a variable number of complete repetitions of isoprene) are the best characterized metabolites because they can be easily identified with techniques such as gas chromatography. Monoterpenes, sesquiterpenes, and diterpenes (containing two, three, and four repetitions of isoprene, respectively) are the most abundant families of terpenes in pine tree resin (Trapp and Croteau, 2001). Since most of these *de novo* synthesized compounds display antibacterial

and antifungal properties, they are considered to be phytoalexins accounting for the toxicity of resin (Grayer and Harborne, 1994; Gershenzon and Dudareva, 2007; Wang *et al*., 2012). Many terpenes display interesting features for the chemical industry, specifically in the production of fragrances, essential oils, and food additives (Yermakova *et al*., 1997). They are also valuable molecules in medicine because of their cytotoxic, cardiotonic, and anti- inflammatory properties (de Araújo *et al*., 2011; Xu *et al*., 2011). However, terpenes are one of the main pollutants in effluents of pulp mill industries (Leuenberger *et al*., 1985; Suntio *et al*., 1988), and terpene-based materials such as tire rubber or latex account for tons of solid waste per year. Thus, these molecules are among the main chemical targets for bioremediation (Dagley, 1975). Given all their applications in biotechnology, bioprospection aiming to identify single genes, gene networks, and microorganisms able to transform or catabolize these molecules is a key starting point in the development of a range of biotechnological applications. The toxicity of terpenes and complexity of their chemical structure hinder their degradation by microorganisms, and few studies describe the ability of microorganisms to use individual components of resin as a sole carbon source (Martin, 1999) or to biotransform particular terpene molecules (de Carvalho and da Fonseca, 2006). The ability of some insects to overcome tree's defensive compounds (Douglas, 2013) and terpenes in particular (Adams *et al*., 2013) has been previously attributed to their association with microorganisms.

A yet unexplored source of potential terpene-degrading microorganisms is related to the insect *Retinia resinella* Linnaeus. The larvae of *R. resinella* feed on young twigs of the Scotch pine *Pinus sylvestris* (Nieukerken *et al*., 2011; Zhang, 1994). They cause small wounds that induce the secretion of resin, which is manipulated by the larva to construct a nodule-like resin capsule –commonly known as "resin gall"- as a hard protective cocoon. A single larva develops inside this resin blister for nearly two years, completely isolated from the external environment by its terpene-rich shelter (Heliövaara and Väisänen, 1988). *R. resinella*'s gut microbiota has not been studied to date, and represents a potential reservoir of resin-tolerant microorganisms. This work, however, aimed at selecting and identifying the cultivable microbial communities associated to *P. sylvestris* resin and *R. resinella* resin-rich galls with a potential ability to degrade terpenes. In order to do so, we used a holistic approach starting from strain selection on a resin-containing medium; genomic and transcriptomic analyses and microbial confrontation assays. The combination of high-throughput sequencing, with the selection of bacterial and fungal strains with outstanding terpene degradation ability, enabled us to characterize natural isolates with promising biotechnological applications.

Materials and Methods

Sample collection. Samples of Scot pine (*Pinus sylvestris*)-associated environmental resin -taken from lopped trees-, and *R. resinella* galls were collected from different trees covering an area of about 1 km^2 along the forest trail "Fuente del Tajo" (N 40º 17', W 0º 36'), Mora de Rubielos, Teruel, Spain) from September 2011 to March 2012. Oficial permissions for collecting were provided by INAGA (Instituto Aragonés de Gestión Ambiental). Larvae were removed from the galls, and both galls and resin coming from different trees were pooled and separately ground with a porcelain pestle to obtain small particles of 4-5 mm. The associated microbiota was harvested by washing 10 g of these particles in sterile PBS buffer (NaCl 8 g/L, KCl 0.2 g/L, Na_2HPO_4 1.44 g/L, KH_2PO_4 0.24 g/L, pH adjusted to 7.4), obtaining a suspension that was subsequently cultured in selective media.

Culture media and growth conditions. A selective minimal medium containing pine resin was employed. As the selection factor, a 10% (w/v) stock solution of resin was obtained by dissolving resin samples in absolute ethanol. Plant debris and other insoluble particles in suspension were eliminated by centrifugation (2,000xg, 5 min) and the supernatant was filter-sterilized through a 0.2 μm pore diameter filter (Corning Inc., NY, USA). A minimal medium was prepared (2 g/L $NaNO_3$, 1 g/L K_2HPO_4, 0.5 g/L $MgSO_4$, 0.5 g/L KCl, 0.2 g/L bacteriological peptone; and 15 g/L agar for solid medium). Once sterilized by autoclave, this minimal medium was kept at 80 ºC and mixed with the resin:ethanol stock solution (ethanol evaporated under these conditions), yielding a selective medium, hereafter called

RM (Resin-based Medium), with increasing resin concentrations (0.05%, 0.1%, 0.2%, 0.4%, and 0.8% w/v) as the main carbon source.

Pine tree resin (from healthy pine wounds) and galls suspensions, including small environmental resin and galls resin particles, were spread on the selective medium plates (in 5 replica for each resin concentration) and incubated at 30 ºC for 14 days.

DNA extraction. Microbial colonies observed on the plates of increasing resin concentrations were harvested by washing the plates with sterile PBS. Each plate was washed using the same volume of PBS, and the resulting suspensions were pooled in two independent tubes in order to isolate the total DNA of the communities cultivated from galls and environmental resin, respectively. The Power Soil DNA Isolation kit (MO BIO Laboratories) was used following the manufacturer's instructions with an additional pretreatment with DNA-free lysozyme at 37 ºC for 10 min. The quantity and quality of the DNA was determined on a 1.5% agarose gel and with a Nanodrop-1000 Spectophotometer (Thermo Scientific, Wilmington, DE).

DNA sequencing, assembly and ORF prediction. Two shotgun libraries were created from 1 μg of the total DNA of the communities cultivated from galls and environmental resin, respectively, according to manufacturer instructions (Roche, Rapid Library Preparation Method Manual GS FLX+ Series XL+, May 2011). Average insert size was 1,800 bp. Each library was sequenced at the CSISP (Centro Superior de Investigación en Salud Pública, Valencia, Spain) using a half pyrosequencing plate in a Roche 454 FLS GS Titanium sequencer. The sequences obtained were assembled using the NEWBLER software (454 LifeSciences Roche) with the default parameters, and the resulting assembly was manually revised and curated with the software Gap4 of the Staden Package (Bonfield *et al*., 1995). Finally, a prediction of Open Reading Frames (ORFs) was carried out on the assemblies using the MetaGeneAnnotator program (Noguchi *et al*., 2008), based on Hidden Markov Models (HMMs). Predicted ORFs smaller than 100 bp were not considered for further analysis.

All sequences were deposited and made publicly available in the MG-RAST server with the following accession numbers: 4455198.3 and 4454707.3.

Sequence analysis. Taxonomic assignations were performed by combining different methods based on sequence similarity. In a first approach, sequences from the 16S and 18S ribosomal RNA genes were used as phylogenetic markers, but only a few sequences of this type were found among both the unassembled and the assembled reads due to the high number of sequences corresponding to *Pseudomonas* sp., the majoritary species in both metagenomes. In order to improve the detection of other taxa, an alternative protein-based taxonomic binning was performed. First, unassembled reads from each cultivated community were analyzed with the MG-RAST server, based on the SEED framework (Meyer *et al*., 2008), and taxonomic assignations were obtained based on sequence similarity searches. Only assignments with an e-value less than or equal to 10^{-5} and a similarity percentage greater than or equal to 80% were accepted. Second, an *ad hoc* sequence analysis pipeline was used. The analysis consisted of BLASTX searches against the non-redundant protein division of GenBank with the ORFs predicted from the assemblies. Again, an e-value of 10^{-5} was used as threshold. BLASTX results were processed with the MEGAN software (Huson *et al*., 2011), assuming a direct correlation between the number of reads corresponding to a particular taxon and the number of ORFs found for it. The more abundant a taxon, the more reads it generates, resulting in longer contigs (Supplementary Figure I.1) and a higher number of ORFs identified. Finally, the putative taxonomic assignments obtained for the most abundant species according to the above-described methods were revised and confirmed by analyzing the coding sequences of housekeeping genes extracted from the assemblies. To do this, 16S, rpoD and gyrB sequences of the NCBI nucleotide database belonging to different species of the most abundant genera were retrieved and aligned using software MEGA. Then, a phylogenetic reconstruction of the sequences was made, and taxonomic assignations were performed accordingly.

In order to obtain functional information of the predicted ORFs from each sample, BLASTP searches were performed against

the COG (Cluster of Orthologous Groups of proteins) database (Tatusov and Fedorova, 2003). Metabolic reconstructions based on gene content were obtained with the program KAAS through BLASTP searches against the KEGG Genes database. In all cases, only hits with an associated e-value less than or equal to 10^{-5} were kept. In addition, the functional assignments provided by the MG-RAST server, based on searches against the non-redundant protein subdivision of GenBank, were also considered for the functional annotation of the ORFs.

Isolation of microbial strains in pure cultures. *Pseudomonas* strain PS was isolated from a Petri plate of selective medium (containing a concentration 0.1% w/v of resin) where a galls suspension had been spread and that had been incubated for 14 days at 30 ^{o}C. Several fungal strains were isolated from other plates (containing different amounts of resin -0.05%, 0.1%, and 0.2% w/v-) where environmental resin samples had been cultured under the same conditions. In all cases, individual colonies were picked and two consecutive re-isolations were performed in resin selective medium (0.1% w/v). Then, pure cultures were set up in resin selective liquid medium. For the cryopreservation of bacteria, a stock solution of glycerol 50% was prepared by mixing equal volumes of glycerol (Panreac Química S.L.U., Barcelona, Spain) and sterile water. Aliquots from each liquid culture were stored in 500 mL of this solution at -20 ^{o}C until required.

Fungal inhibition assays. The antifungal properties of the *Pseudomonas* sp. isolate were tested with a confrontation assay following the procedure previously described (Mela *et al.*, 2011). As a control, all the isolates were grown alone, under the same conditions but without *Pseudomonas* confrontation. All confrontation assays were performed in both RM (0.1% resin w/v) and LB solid media. The fungal growth inhibition was assessed by comparing the diameter of the confronted vs the isolated fungal colonies at different times and in three independent replica.

Total RNA isolation, amplification and cDNA synthesis. To perform the transcriptomic analysis, confronted and isolated fungal colonies from three independent replica of the confrontation assays were separately pooled, and total RNA was extracted with TRI Reagent solution (Ambion). The quantity and the integrity of the total RNA was determined on a 0.8% agarose gel and with a Nanodrop-1000 Spectophotometer (Thermo Scientific, Wilmington, DE). Then, mRNAs were amplified with MessageAmp^{TM} II aRNA Amplification Kit (Ambion). The resulting RNA was converted to double-stranded cDNA

TABLE I. 1: Summary of the sequencing, assembly, and ORF prediction statistics for the microbial communities cultivated from environmental resin and galls.

	Resin	**Galls**
Megabases generated	283.79	279.56
Number of reads	721,575	660,246
Average read length	393.3	423.42
Number of contigs	14,102	2,613
Max. contig length (bp)	87,262	793,584
Average contig length (bp)	1,504 $\pm$ 2,682	2,586 $\pm$ 20,534
Mean coverage	12.91	34.72
N50	4,956	152,572
N90	773	1,040
Number of unique proteins	19,136	5,868
Number of unique non-coding RNAs	1,345	246

using random hexamers. Again, the quantity and quality of the cDNA was assessed on a 0.8% agarose gel and with a standard PicoGreen Assay for dsDNA (Thermo Scientific, Wilmington, DE).

cDNA sequencing and assembly. The cDNA of both confronted and isolated fungal samples was sequenced in a Roche 454 FLS GS Titanium sequencer, using 1/8 of a pyrosequencing plate for each sample. The resulting reads were trimmed and assembled with the NEWBLER software (454 LifeSciences Roche), using the default parameters for cDNA sequences.

Functional analysis of putative mRNA. The contigs obtained from the assemblies of cDNA reads were aligned to the NCBI nr protein database using BLASTX searches in order to obtain functional annotations. Only hits with an e-value less than or equal to 10^{-5} were considered in the assignments. In order to study mRNA distribution in standardized categories, GO (Gene Ontology) terms for each transcript were retrieved with the BLAST2GO software (Götz *et al*., 2008). The number of reads associated to each contig was considered as an indicator of the corresponding transcript expression level. In order to detect significantly overrepresented or underrepresented GO categories in a mRNA subset, a Fisher two-tailed enrichment test was performed, using an e-value of 0.05 as threshold.

Results

Taxonomic diversity of resin-tolerant microbial communities

The main goal of this work was to study the diversity of microorganisms able to degrade terpenes, the main components of pine tree resin. Galls and environmental resin suspensions were cultured in a selective medium with resin as the main carbon source, and the microbial communities able to grow in such medium were then identified by high-throughput sequencing. After 14 days of incubation, a lawn of microbial colonies was observed in those Petri plates with low (0.05% w/v) to slightly high (0.2% w/v) concentration of resin, whereas a significantly lower number of colonies was observed in media with high resin concentration (0.4% and 0.8% w/v). Some of the plates were clearly dominated by a fluorescent bacterial species. This was particularly striking in plates where galls suspensions had been inoculated and, interestingly, no fungal colonies were observed in these plates (data not shown).

The cultivable, resin-tolerant microbial communities associated with environmental resin and galls were pyrosequenced, yielding 721,575 and 660,246 reads, respectively, representing approximately 280 Mb of DNA sequence for each community. Prior to taxonomic and functional analysis of the data, all reads were *de novo* assembled, and ORFs predicted and extracted from the assemblies. The results of these procedures are summarized in Table I.1. The assembly of gall sequences exhibited a lower number of contigs (2,613) compared to the 14,102 obtained for the environmental resin, and also longer contigs (Table I.1).

The taxonomic binning of the samples performed by the MG-RAST server (using the pyrosequencing reads) and the taxonomic composition inferred by an alternative *ad hoc* method (using the ORFs predicted from the assemblies) yielded similar results. *Pseudomonas* proved the most abundant genus in both samples (Figure I.1). Taxonomic assignments on the basis of BLASTX searches suggested that *Pseudomonas fluorescens* was the predominant species in both cases. In parallel, sequences from the house-keeping genes 16S, rpoD, and gyrB were retrieved from the sequencing data, and a phylogenetic reconstruction was carried out in order to confirm BLASTX-based taxonomic assignations. Only a single copy of these genes belonging to the genus *Pseudomonas* was found in the galls metagenome, whereas two copies were detected in the environmental resin metagenome. One of the copies of each gene, the most represented in terms of number of reads, was more than a 99% identical to the single copy found in galls, suggesting that they might belong to the same species. The phylogenetic analysis based on these house-keeping genes indicated *Pseudomonas abietaniphila* rather than *Pseudomonas fluorescens* as the closest match for the most abundant –and unique- *Pseudomonas* species in galls samples, and the presence of at least two different *Pseudomonas* species in the microbial communities associated with environmental resin (Supplementary Figure I.2). The incongruence between BLASTX searches

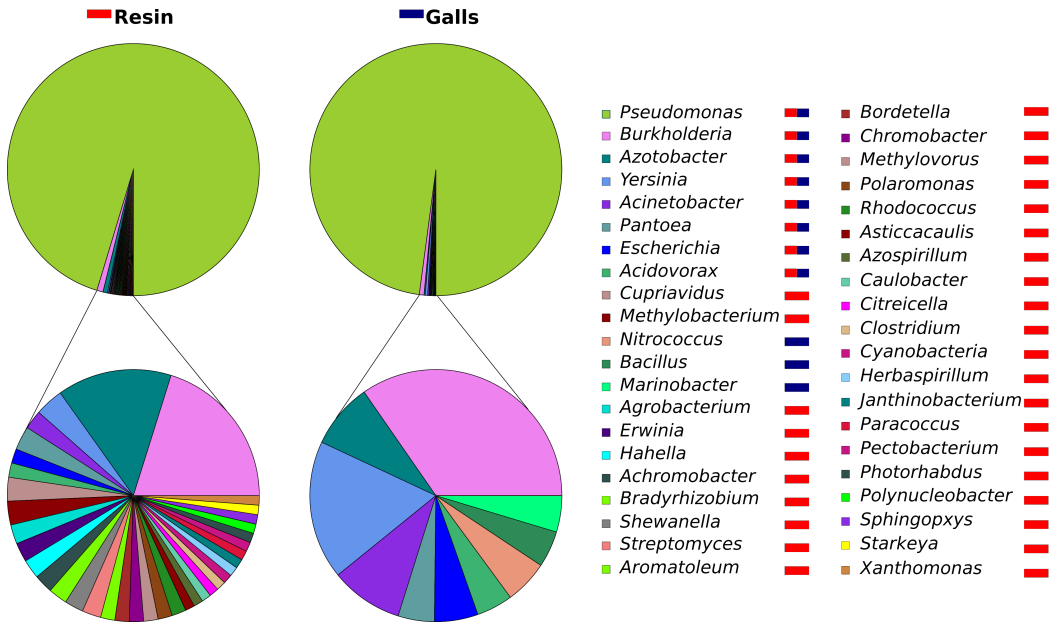


FIGURE I. 1: Relative abundance of bacterial genera in the communities cultivated from environmental resin and galls. Horizontal bars indicate the presence of a particular genus in environmental resin (red) or galls (blue).

and the phylogenetic analysis performed with house-keeping genes is likely a consequence of the lack of complete *P. abietaniphila* genomes in the databases used in the analysis.

Besides *Pseudomonas*, ten different genera able to grow on resin medium were identified from gall samples: *Burkholderia*, accounting for 35% of the remaining sequences; *Yersinia*, with 18%; *Acinetobacter*, 9.3%; *Azotobacter*, 8.4%; and *Pantoea*, *Escherichia*, *Acidovorax*, *Nitrococcus*, *Bacillus*, and *Marinobacter*, which were represented with about 5% of sequences each (Figure I.1). The diversity of potential terpene-degrading bacteria found in galls was lower than that associated with environmental (pine tree-associated) resin, where a total of 38 different genera were detected. *Burkholderia* and *Azotobacter* were, besides *Pseudomonas*, the two most abundant taxa (21% and 15% of the ORFs not belonging to *Pseudomonas*). Among the remaining genera, only *Yersinia*, *Acinetobacter*, *Pantoea*, *Escherichia*, and *Acidovorax* were also found within the community isolated from galls (Figure I.1), with the other 28 genera being exclusive of environmental resin. The rarefaction curves obtained for each community suggested that the difference between environmental resin and galls in terms of taxonomic diversity is rather high, since an increase in the number of sequences analyzed did not result in saturation of the number of taxa identified in the case of environmental resin (Supplementary Figure I.3). It has to be noted that the taxonomic diversity of environmental resin and, in particular, the diversity of the microbial communities associated with galls might be underrepresented due to the overwhelming abundance of a single *Pseudomonas* species, which accounts for 85% and 95% of the total ORFs found in the cultivable communities of environmental resin and gall samples, respectively.

Under our experimental conditions, cultivable fungal species were rare, particularly in the gall samples, where only 195 out of the 99,509 BLASTX hits corresponded to fungi. This was in accordance with the inhibition of fungal growth observed in those Petri dishes dominated by a fluorescent species (putatively, *Pseudomonas* sp.). In the case of environmental resin, where a similar number of total BLASTX hits were obtained, fungal sequences were three-fold more abundant (Supplementary Figure I.4). All the putative taxonomic assignations in the culturable pools from both gall and environmental resin samples corresponded to ascomycetes, with the only exception being that of the class Tremellomycetes, detected in environmental resin, belonging to

Basidiomycetes. As shown in Supplementary Figure I.4, the diversity of fungal classes found in environmental resin was higher compared to resin-rich galls, where only sequences from Dothideomycetes, Eurotiomycetes, Leotiomycetes, and Sordariomycetes were detected.

Functional diversity and metabolic reconstruction

The functional study of the sequencing data was carried out by assigning each one of the ORFs found to a COG category and a KO (KEGG Orthology) identifier. This yielded a very similar distribution of COG categories in both environmental resin and gall-associated microbial communities (Supplementary Figure I.5), with amino acid metabolism, energy production, and translation being the most represented functional groups in the metagenomes. Most metabolic pathways were also common to both environmental resin and galls (Supplementary Figure I.5). This might be a consequence of the high proportion of sequences belonging to the same species (*Pseudomonas* sp.) in both samples.

The complete sequence of a diterpene-degradation cluster, first reported by Martin and Mohn (2000) in *P. abietaniphila* BKME-9, a natural isolate from pulp mill effluents, was found in the gall pool, whereas several partial copies of all the genes were detected in the case of environmental resin (Supplementary Table I.1). The analysis of the 11 kb sequence of the cluster found in galls revealed that it belonged to *Pseudomonas* sp. (hereafter *Pseudomonas abietaniphila* strain PS, taken after *Pinus sylvestris*), the most abundant species in both metagenomes, and displayed the same gene synteny as *P. abietaniphila* BKME-9. On the other hand, several copies of most of the genes involved in pinene degradation (KEGG Pathway 00903) were found in both samples. Sequences encoding particular genes of the acyclic terpene degradation pathway were also detected (Supplementary Table I.1).

Characterization of the terpene-degrading ability of PS and several resin-associated fungal strains

As expected from the abundance of PS found in the bioinformatic analysis, this strain, which we isolated and grew on resin-containing medium, exhibited good performance in terms of resin degradation. The resin content (estimated as described in Supporting Methods) during the exponential phase of a PS culture in RM decreased from 5.4 g/L to 4.1 g/L after 4 days, indicating a degradation of the resin originally present in the medium by nearly 25% (Figure I.2A). After 5 days, when the culture reached the stationary phase, the amount of resin could not be estimated properly, since the number of viable cells significantly differed from the total number of cells. The growth of PS in RM was accompanied by the secretion of a fluorescent compound, probably one of the siderophores usually produced by different *Pseudomonas* species under conditions of iron starvation (Cox and Adams, 1985; Ochsner *et al.*, 2002).

Several fungi able to grow on RM were isolated from environmental resin samples, and pure cultures were inoculated in liquid medium. Three strains, identified as *Aspergillus terreus* (isolate F1), *Aspergillus flavus* (isolate F8), and *Penicillium decumbens* (isolate F9) through 18S rDNA sequencing, were subjected to further characterization in terms of terpene degradation. We managed to follow the changes in resin content in RM cultures of these fungi using a simple method based on measuring optical density of the medium, as described in Supporting Methods. The number of resin colloids dramatically decreased as fungi grew in the medium, resulting in decreasing optical densities reaching nearly zero after 5 days of cultivation, in the case of isolate F8; and 7 days, in the case of isolates F1 and F9 (Figure I.2B). As shown in Figure I.2C, RM broth became virtually transparent as a consequence of the degradation of resin colloids and setting of fungal hyphae. The formation of mycelium spheres was observed in all the cultures (Supplementary Figure I.6).

The ability of the four selected microbial strains to degrade other terpene-based materials was tested in minimal media with latex and rubber as the sole carbon sources. All the strains proved able to grow on latex. F1, F8 and F9 fungal hyphae grew in close association with the latex particles, as observed by scanning electron microscopy (SEM) (Figure I.3D). *Pseudomonas* strain PS also grew in the latex-containing medium, producing a fluorescent substance (Cox and Adams, 1985; Ochsner *et al.*, 2002).

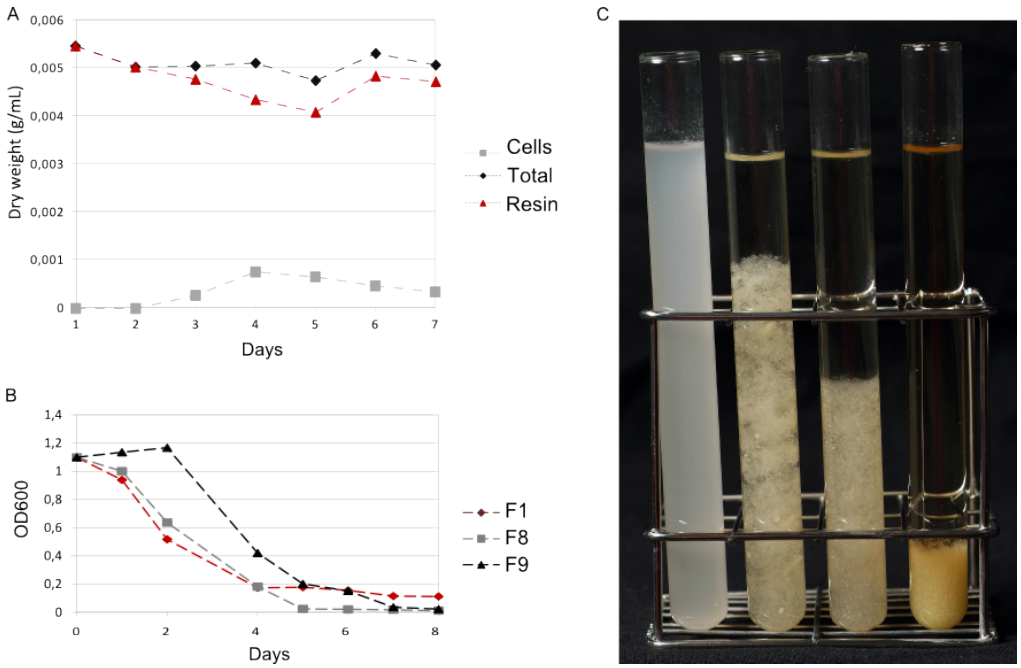


FIGURE I. 2: A) Bacterial and total dry weight of RM cultures of strain PS and evolution of the estimated amount of resin in the medium. B) Variation in resin content in RM cultures of different fungi isolated from resin, estimated from optical density of the medium (OD600), as described in Supporting Methods. C) Resin colloid removal by fungal isolates grown for 7 days compared to a non inoculated RM control (from left to right, control, F9, F8, and F1 cultures).

The growth of this strain correlated with the appearance of cracks on the surface of many of the latex particles (Figure I.4B). Interestingly, after one month of growth under relatively strong agitation (250 rpm), PS cells exhibited a clear attachment pattern to the latex particles, being virtually embed into the latex structure as a consequence of in situ degradation around cells, which resulted in cell-shaped cavities or niches (Figure I.4C and 4D). All isolates, with the only exception being PS, also grew in the rubber-containing medium displaying, again, an association with rubber particles. F1 formed dense biofilms covering the rubber and fungal hyphae (Figure I.3E and 3F).

Fungal growth inhibition by strain PS

PS proved able to inhibit the growth of isolates F1 and F8 in confrontation assays regardless of the medium used (RM or LB). As shown in Figure I.5, a reduction of almost 40% was observed in the diameter of F1 colonies on RM, and they stopped growing after 5 days of confrontation. Similar results were obtained on LB, with a reduction of 46% in colony diameter and a halt in fungal growth after 3 days. The inhibition of F8 displayed the same pattern, with a reduction of 37% and 46% in colony diameter on RM and LB, respectively. In the case of F9, no significant differences in terms of colony size were found on RM, whereas a slight inhibition was observed in the confrontations performed on LB. Control experiments carried out with *E. coli* and with a filter-sterilized supernatant of PS cultures, showed no fungal growth inhibition in any case (data not shown).

The transcriptomes of F1 grown both isolated and confronted with strain PS on LB (showing the highest fungal inhibitions, as shown in Figure I.5) were pyrosequenced in order to check whether transcriptional changes occurred in the fungi as a response to PS. The sequencing statistics of these transcriptomes are shown in Supplementary Table I.2. The analysis of nearly 300 and 200 protein-coding transcripts corresponding to the fungus grown isolated and confronted with PS, respectively, revealed that both transcriptomes had a similar global distribution of GO terms in their mRNAs.

However, further analysis revealed that 44 mRNAs of the confronted fungus were not detected in the transcriptome of the isolated fungus (Supporting Data). A two-tailed Fisher enrichment test performed with this subset of 44 genes against the whole transcriptome of the confronted fungal strain revealed that GO terms involved with transmembrane transportation were significantly overrepresented in the subset (p-value < 0.05) (Figure I.6).

Discussion

From the Spanish mountains to the Far East of Russia, and from the Arctic Circle to the Mediterranean, we find *Pinus sylvestris*, one of the most widespread tree species on Earth. The economic and ecological relevance of *P. sylvestris* forests is undeniable, representing major sources of wood and pulpwood, the processing of which generates terpenes as main contaminants of industrial wastewater (Leuenberger *et al.*, 1985; Suntio *et al.*, 1988). The diversity of microorganisms associated with *P. sylvestris* resin and selected on resin-containing medium as reported for the first time here can be likened to a biological arsenal with terpene-degrading ability. Albeit influenced by the composition of the resin-containing medium or the antimicrobial compounds secreted by the isolates (in particular, by *Pseudomonas abietaniphila* strain PS), we report here the highest diversity of terpene-degrading microorganisms described to date. Previously, surveys of microorganisms able to degrade individual terpenes have been carried out in environments such as pulp mill industry effluents (Bicho *et al.*, 1995), sequencing batch bioreactors (Wilson *et al.*, 1996), hydrocarbon-contaminated soils (Yu *et al.*, 1999), or forest soils (Mohn *et al.*, 1999). Those works were based on culture-dependent techniques and PCR-based identifications, and reported the ability of several bacterial species belonging to the genera *Pseudomonas*, *Burkholderia*, and *Cupriavidus* to degrade specific resin acids such as dehydroabietic or isopimaric acid (diterpenes). All these genera were detected in our resin-selected samples, and proved to be moderately to highly abundant.

We further characterized particular microbial strains to assess their potential applications for bioremediation of terpene-contaminated environments. Strain PS proved able to both tolerate and degrade significant amounts of resin. Our results indicate that PS degraded

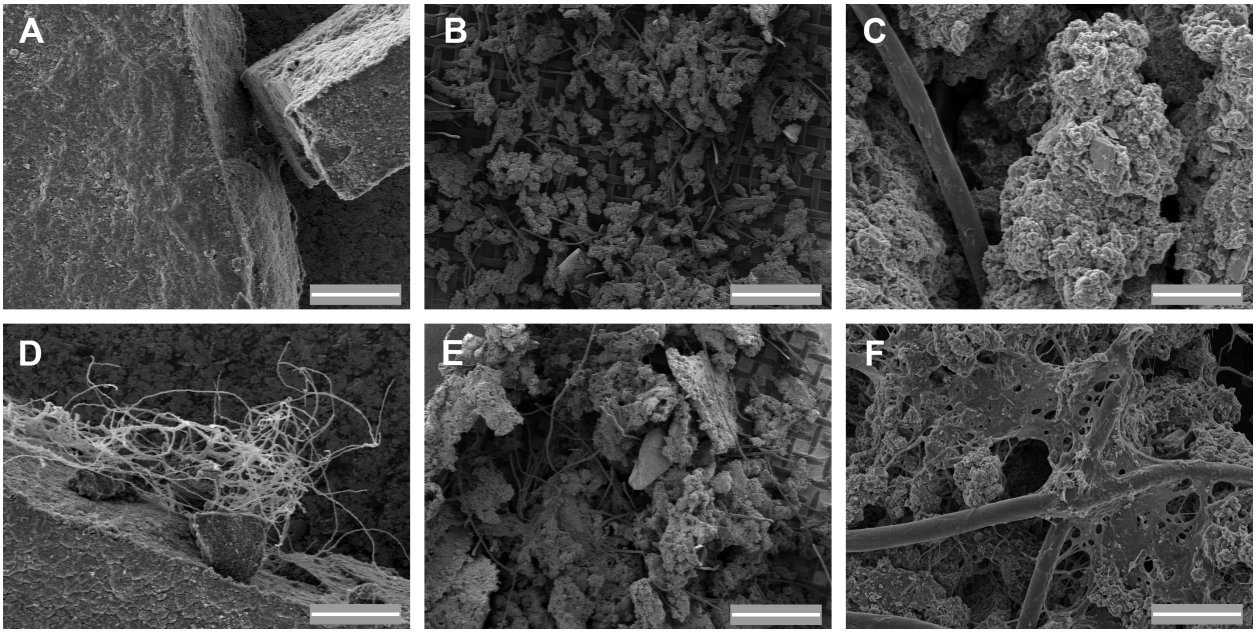


FIGURE I. 3: SEM images of latex (A and D) and rubber (B, C, E, and F) used as the sole carbon source in the selective media. A, B and C show non-inoculated control media, whereas 15-day-old cultures of isolates F9 and F1 are shown in subfigures D and E-F, respectively. A, C, D and F scale bars= 50 μm; B, E scale bars= 500 μm.

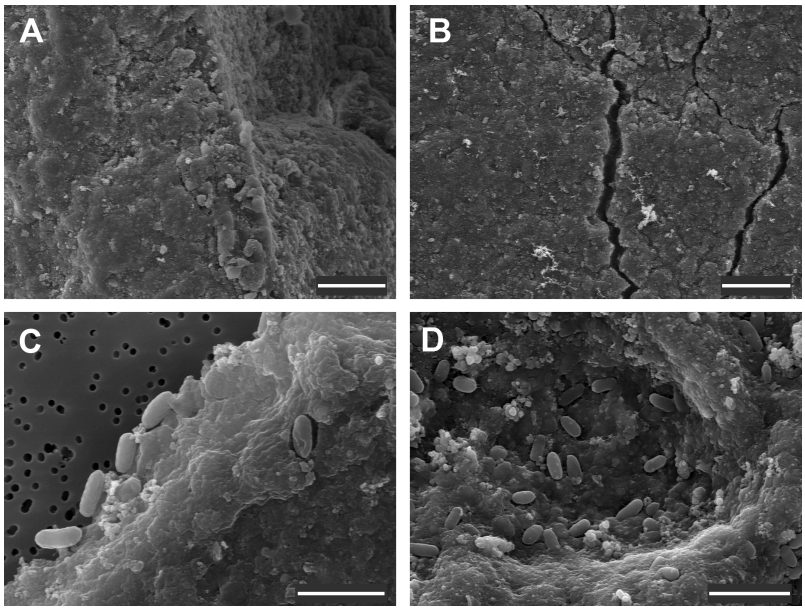


FIGURE I. 4: SEM images of latex particles from (A) a non-inoculated latex-containing medium; and a 15-days (B) and a one-month culture (C and D) of strain PS in the same medium. Arrows indicate the cell-shaped niches formed on the latex surface. A and B scale bars= 10 μm, C and D scale bars= 2 μm.

nearly 1.5 g of resin per liter after 4 days of exponential growth. This is probably an underestimation of the real value, since the protocol we used to determine the resin content only took into account the amount of resin that is converted into biomass and hence contributes to the increase in dry weight of viable bacteria. The production of any carbonated compound as a result of terpene processing, may also contribute to a decrease in resin content, which suggests that the performance of PS in terms of resin degradation is actually higher than we calculated. On the other hand, different fungal strains isolated from environmental resin samples were also cultivated in resin-containing medium and found to display a dramatic ability to degrade not only resin (Figure I.2) but also terpene-based materials such as latex gloves or non-vulcanized rubber. Some of the fungi were able to form a biofilm covered by a dense EPS matrix (Figure I.3). These strains may be of interest not only for the bioremediation industry, particularly in the treatment of terpene-contaminated effluents, but they are also potential candidates for the waste management of large-scale disposal materials, such as tire rubber.

The gene content of strain PS was analyzed based on the vast number of sequences obtained, which might cover almost 35% of its complete genome, and a cluster of genes responsible for degradation of diterpenes was detected. This cluster was previously reported in *Pseudomonas abietaniphila* BKME-9 (Martin and Mohn, 2000) and *Burkholderia xenovorans* LB400 (Smith *et al*., 2007), and has recently been detected in other species, whose ability to degrade diterpenes remains unstudied. Case examples of these species are *P. fluorescens* F113 (Redondo-Nieto *et al*., 2012) and *P. aeruginosa* 2192 (Mathee *et al*., 2008). The cluster found in the selected gall-associated community corresponded to PS, and displayed the same gene synteny as *P. abietaniphila* BKME-9, pointing out, again, that both strains are closely related.

The compatibility of strain PS with other terpene-degrading isolates is also relevant for its application in bioremediation processes. PS proved able to inhibit the growth of different fungi naturally present in resin (Figure I.5). Sequencing of the transcriptomes of the fungal isolate F1 (identified as *A. terreus*), grown in isolation as well as under confrontation with PS, revealed changes in the populations of mRNAs involved in transmembrane transportation or coding for proteins with unknown function. Our results are in concordance with a previous transcriptomic analysis of confrontations

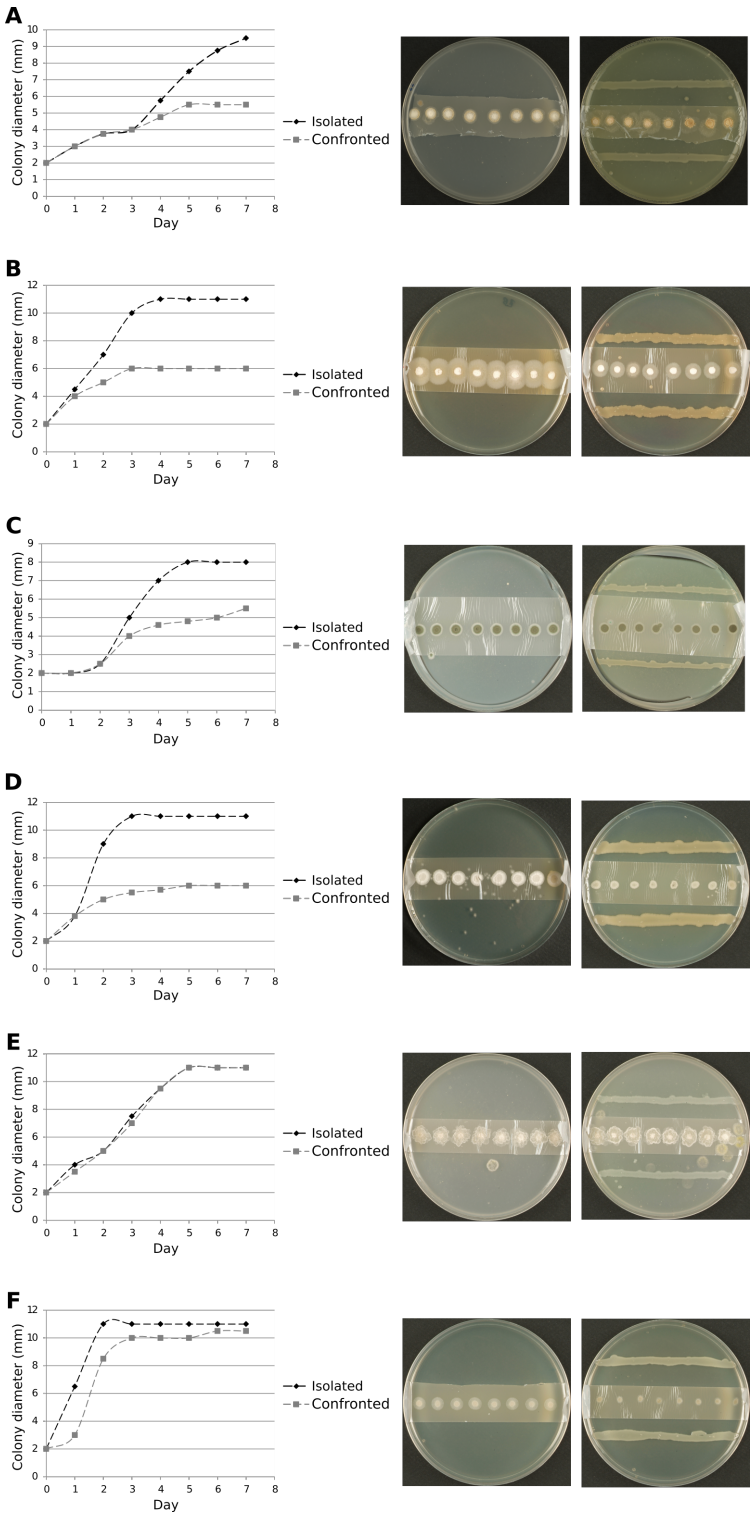


FIGURE I. 5: Colony size (mm) of three fungal isolates during confrontation assays with strain PS. A and B, F1; C and D, F8; and E and F, F9. The experiments were carried out on RM (A, C, and E) and LB (B, D, and F) medium. Pictures of particular experiments were taken after 3 and 5 days in the case of LB and RM media, respectively.

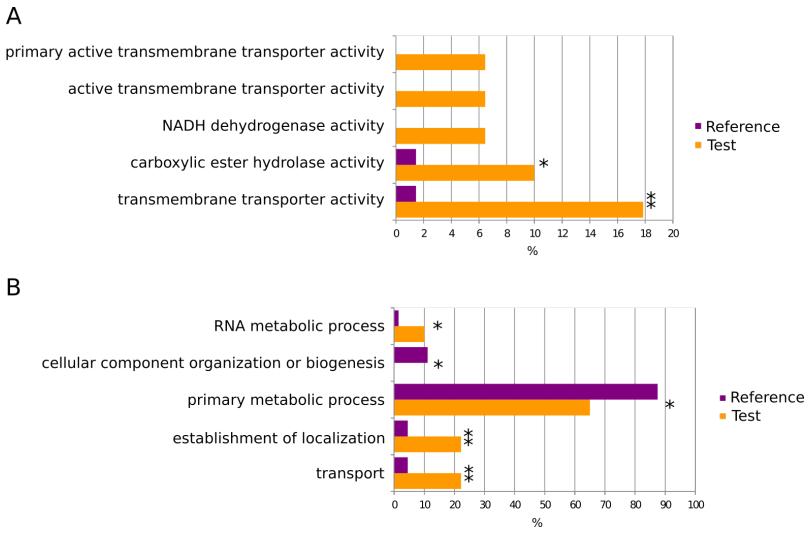


FIGURE I. 6: Relative abundance of selected GO terms corresponding to molecular functions (A) and cellular processes (B) in the transcriptome of the fungal isolate F1 confronted with PS (reference, purple bars), and in the subset of genes not detected in the transcriptome of F1 grown isolated (test, orange bars). Asterisks indicate statistically significant differences for p-value<0.05 (∗∗) and p-value<0.1 (∗).

between *Collimonas fungivorans* and *Aspergillus niger* (Mela *et al*., 2011), where moderate transcriptional changes were observed (affecting 0.4% of the transcriptome) in early stages of confrontation, and several genes linked to the fungal cell membrane or coding membrane transporters were up- or down-regulated in response to *C. fungivorans*. As proposed in other works (Mela *et al*., 2011), these changes might be linked to nutrient shortage (mainly nitrogen) experienced by the fungi during confrontation, leading to the overexpression of genes involved in nutrient intake.

This is the first holistic and bioprospection-oriented study of microbial communities associated with *P. sylvestris* resin and *R. resinella*-induced galls. The genes, species, and interactions described in this work represent potential tools for several biotechnological processes involving terpenes, and, particularly, for the bioremediation of environments contaminated with these recalcitrant metabolites.

Acknowledgements

Financial support was provided by grant BFU2012-39816-C02-01 (Ministerio de Ciencia e Innovación, Spain) to A. Latorre and CGL2008-00605 (Ministerio de Ciencia e Innovación) to J. Baixeras. Cristina Vilanova is a recipient of a FPU fellowship from the Ministerio de Educación (Spain). Oficial permissions for collecting were provided by INAGA (Instituto Aragonés de Gestión Ambiental). The authors want to express their gratitude to Fabiola Barraclough for the English correction of the manuscript and to the technical staff of the Forestry Laboratory of Mora de Rubielos in Teruel for field collaboration. Strain PS, deposited in the Spanish Type Culture Collection (CECT) under reference number CECT 8327, has been found by us to hold not only for scientific publication, but also for patenting (Application number P201300612 at Spanish Office of Patents and Trademarks, OEPM). The authors have prepared the patent and the registration in collaboration with the Research Transfer Office (OTRI) of the University of Valencia (contact person, Silvia Bort: silvia.bort@uv.es). SEM images were obtained at the Electron Microscopy Services of the University of Valencia.

References

Adams AS, Aylward FO, Adams SM, Erbilgin N, Aukema BH, *et al*. 2013. Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community

highly enriched in genes contributing to terpene metabolism. *Appl Environ Microb* **79**: 3468–3475.

Bicho P, Martin V, Saddler JN. 1995. Growth, induction, and substrate specificity of dehydroabietic acid-degrading bacteria isolated from a kraft mill effluent enrichment. *Appl Environ Microb* **61**: 3245–3250.

Bonfield JK, Smith K f, Staden R. 1995. A new DNA sequence assembly program. *Nucleic Acids Res* **23**: 4992–4999.

Buijtenen JP van, Santamour FS. 1972. Resin crystallization related to weevil resistance in white pine Pinus strobus. *Can Entomol* **104**: 215–219.

Cox CD, Adams P. 1985. Siderophore activity of pyoverdin for Pseudomonas aeruginosa. *Infect Immun* **48**: 130–138.

Dagley S. 1975. Microbial Degradation of Organic Compounds in the Biosphere. *Am Sci* **63**: 681–689.

de Araújo DAM, Freitas C, Cruz JS. 2011. Essential oils components as a new path to understand ion channel molecular pharmacology. *Life Sci* **89**: 540–544.

de Carvalho CCCR, da Fonseca MMR. 2006. Biotransformation of terpenes. Biotechnol Adv 24: 134–142. doi:10.1016/j.biotechadv.2005.08.004.

Douglas AE. 2013. Microbial brokers of insect-plant interactions revisited. *J Chem Ecol* **39**: 952–961.

Gershenzon J, Dudareva N. 2007. The function of terpene natural products in the natural world. *Nat Chem Biol* **3**: 408–414.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, *et al*. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**: 3420–3435.

Grayer RJ, Harborne JB. 1994. A survey of antifungal compounds from higher plants, 1982–1993. *Phytochemistry* **37**: 19–42.

Hanover JW. 1975. Physiology of Tree Resistance to Insects. *Annu Rev Entomo*l **20**: 75–95.

Heliövaara K, Väisänen R. 1988. Periodicity of Retinia resinella in Northern Europe Lepidoptera: Tortricidae. *Entomologia Gen* **14**: 37–45.

Himejima M, Hobson KR, Otsuka T, Wood DL, Kubo I. 1992. Antimicrobial terpenes from oleoresin of ponderosa pine tree Pinus ponderosa: A defense mechanism against microbial invasion. *J Chem Ecol* **18**: 1809–1818.

Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21**: 1552–1560.

Leuenberger C, Giger W, Coney R, Graydon JW, Molnar-Kubica E. 1985. Persistent chemicals in pulp mill effluents. *Water Res* **19**: 885–894.

Marpeau A, Walter J, Launay J, Charon J, Baradat P, *et al*. 1989. Effects of wounds on the terpene content of twigs of maritime pine Pinus pinaster Ait. *Trees* **3**: 220-6.

Martin VJ, Mohn WW. 2000. Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by Pseudomonas abietaniphila BKME-9. *J Bacteriol* **182**: 3784–3793.

Martin VJ, Yu Z, Mohn WW. 1999. Recent advances in understanding resin acid biodegradation: microbial diversity and metabolism. *Arch Microbiol* **172**: 131–138.

Mathee K, Narasimhan G, Valdes C, Qiu X, Matewish JM, *et al*. 2008. Dynamics of Pseudomonas aeruginosa genome evolution. *PNAS* **105**: 3100–3105.

Mela F, Fritsche K, de Boer W, van Veen J, de Graaff LH, *et al*. 2011. Dual transcriptional profiling of a bacterial/fungal confrontation: Collimonas fungivorans versus Aspergillus niger. *ISME J* **5**: 1494–1504.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, *et al*. 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Mohn WW, Wilson AE, Bicho P, Moore ERB 1999. Physiological and phylogenetic diversity of bacteria growing on resin acids. *Syst Appl Microbiol* **22**: 68–78.

Nieukerken EJ, Van Kaila L, Kitching IJ, Kristensen NP, Lees DD, *et al*. 2011. Order Lepidoptera Linnaeus, 1758. In: Animal biodiversity: An outline of higher level classification and survey of taxonomic richness. *Zootaxa* **3148**: 212-221.

Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* **15**: 387–396.

Ochsner UA, Wilderman PJ, Vasil AI, Vasil ML. 2002. GeneChip® expression analysis of the iron starvation response in Pseudomonas aeruginosa: identification of novel pyoverdine biosynthesis genes. *Mol Microbiol* **45**: 1277–1287.

Pearce R. 1996. Antimicrobial defences in the wood of living trees. *New Phytol* **132**: 203–233.

Redondo-Nieto M, Barret M, Morrisey JP, Germaine K, Martínez-Granero F, *et al*. 2012. Genome sequence of the biocontrol strain Pseudomonas fluorescens F113. *J Bacteriol* **194**: 1273–1274.

Savluchinske-Feio S, Curto MJM, Gigante B, Roseiro JC. 2006. Antimicrobial activity of resin acid derivatives. *Appl Microbiol Biot* **72**: 430–436.

Smith DJ, Park J, Tiedje JM, Mohn WW. 2007. A large gene cluster in Burkholderia xenovorans encoding abietane diterpenoid catabolism. *J Bacteriol* **189**: 6195–6204.

Suntio LR, Shiu WY, Mackay D. 1988. A review of the nature and properties of chemicals present in pulp mill effluents. *Chemosphere* **17**: 1249–1290.

Tatusov R, Fedorova N. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **14**: 1–14.

Tomlin ES, Antonejevic E, Alfaro RI, Borden JH. 2000. Changes in volatile terpene and diterpene resin acid composition of resistant and susceptible white spruce leaders exposed to simulated white pine weevil damage. *Tree Physiol* **20**: 1087–1095.

Trapp S, Croteau R. 2001. Defensive resin biosynthesis in conifers. *Annu Rev Plant Phys* **52**: 689–724.

Wang J, Chen YP, Yao K, Wilbon PA, Zhang W, *et al*. 2012. Robust antimicrobial compounds and polymers derived from natural resin acids. *Chem Commun* **48**: 916–918.

Wilson AE, Moore ER, Mohn WW. 1996. Isolation and characterization of isopimaric acid-degrading bacteria from a sequencing batch reactor. *Appl Environ Microb* **62**: 3146–3151.

Xu M-F, Shen L-Q, Wang K-W, Du Q-Z, Wang N. 2011. A new rearranged abietane diterpenoid from Clerodendrum kaichianum Hsu. *J Asian Nat Prod Res* **13**: 260–264.

Yermakova TG, Kurochkin VN, Suslov SN, Myachina GF, Salaurov VN, *et al*. 1997. Complex utilization of wood chemistry products. *Chem Sust Dev* **5**: 149–153.

Yu Z, Martin VJJ, Mohn WW. 1999. Occurrence of two resin acid-degrading bacteria and a gene encoding resin acid biodegradation in pulp and paper mill effluent biotreatment systems assayed by PCR. *Microb Ecol* **38**: 114–125.

Zhang BC. 1994. Index of economically important Lepidoptera. CAB International University Press: Cambridge.

Publication II

The generalist inside the specialist: Gut bacterial communities of two insect species feeding on toxic plants are dominated by *Enterococcus* sp.

Cristina Vilanova[1,2], Joaquín Baixeras[1], Amparo Latorre[1,2,3] and Manuel Porcar[1,2]

[1] Cavanilles Institute of Biodiversity and Evolutionary Biology (Universitat de València), Spain.
[2] Institute for Integrative Systems Biology (I^{2}SysBio, University of Valencia-CSIC), Valencia, Spain.
[3] Unidad Mixta de Investigación en Genómica y Salud, CSISP, Valencia, Spain.

Abstract

Some specialist insects feed on plants rich in secondary compounds, which pose a major selective pressure on both the phytophagous and the gut microbiota. However, microbial communities of toxic plant feeders are still poorly characterized. Here, we show the bacterial communities of the gut of two specialized Lepidoptera, *Hyles euphorbiae* and *Brithys crini*, which exclusively feed on latex-rich *Euphorbia* sp. and alkaloid-rich *Pancratium maritimu*m, respectively. A metagenomic analysis based on high-throughput sequencing of the 16S rRNA gene revealed that the gut microbiota of both insects is dominated by the phylum Firmicutes, and especially by the common gut inhabitant *Enterococcus* sp. *Staphylococcus* sp. is also found in *H. euphorbiae* though to a lesser extent. By SEM, we found a dense ring-shaped bacterial biofilm in the hindgut of *H. euphorbiae*, and identified the most prominent bacterium in the biofilm as *Enterococcus casseliflavus* through molecular techniques. Interestingly, this species has previously been reported to contribute to the immobilization of latex-like molecules in the larvae of *Spodoptera litura*, a highly polyphagous lepidopteran. The *E. casseliflavus* strain was isolated from the gut and its ability to tolerate natural latex was tested under laboratory conditions. This fact, along with the identification of less frequent bacterial species able to degrade alkaloids and/or latex, suggest a putative role of bacterial communities in the tolerance of specialized insects to their toxic diet.

Introduction

Plants have biochemical and molecular mechanisms to defend themselves from insects attack. Among those, plants produce a vast range of secondary metabolites with anti-herbivore effects, which are produced either constitutively or in response to tissue damage (War *et al*., 2012). Some plant biochemicals are toxic, repellent, or antinutritive for herbivores. Among these compounds, alkaloids, terpenoids and complex mixtures of macromolecules such as latex are among the most frequent plant biochemical defense barriers.

Plant alkaloids, are toxic to a wide range of insects (Nuringtyas *et al*., 2014). However, a few species of insects are unaffected by even high concentrations of alkaloids. The Amaryllidaceae, with more than 300 alkaloids isolated to date (Bastida *et al*., 2011), are among the most deterrent plants. Alkaloids present in *Pancratium maritimum* have demonstrated both cytotoxic and antimicrobial activity (Hetta and Shafei, 2013). The aposematic larvae of the noctuid moth *Brithys crini* –the "lily borer"- are well known to feed monophagously on the sand lily *Pancratium maritimum*. Sequestration of alkaloids by this species has never been tested

but the larvae of the closely related species complex *Xanthopastis timais* – the "Spanish moth", another Amaryllidaceae specialist- was early included by Rothschild (1973) in her seminal work on insect chemical defense and is a typical example of sequestration of phenanthridine alkaloids (Nishida, 2002). Terpenes are chemical compounds that are present in large amounts in a large variety of plants: in conifers, for example, they are the main components of resin. Plant terpenes are involved in defense against herbivory, even at the belowground level (Vaughan *et al*., 2013). Terpenes, along with alkaloids, natural gum and many other compounds, are also present in Euphorbiaceae and other plants exuding latex. Unsurprisingly, latex-producing plants are particularly resistant to many insects and other pests (Hagel *et al*., 2008). The larvae of the sphingid moth *Hyles euphorbiae* -the "spurge hawk moth"- feed on a broad variety of *Euphorbia* plants from which they sequester the cytotoxic ingenane diterpene esters (Marsh and Rothschild, 1984). Beyond biochemicals, plant defenses against herbivory also involve mutualistic microorganisms. Endophytic fungi, plant symbionts living asymptomatically within the host tissues, produce alkaloid-based herbivore deterrents that contribute to the defense of the plant (Koh and Hik, 2007). Reciprocally, bacteria associated with insects can play a role in disturbing plant defensive barriers (Hansen and Moran, 2014; Hammer and Bowers, 2015). For example, Colorado potato beetle (*Leptinotarsa decemlineata*) larvae have been reported to bear bacteria in their oral secretions that suppress antiherbivore defenses in tomato (*Solanum lycopersicum*), the plant the beetle feeds on (Chung *et al*., 2013). In summary, plant-insect interactions are complex ecological processes mediated by secondary metabolites, alkaloids and terpenes among them, but also by microorganisms, which play key roles as both defense and attack allies for the plants and phytophagous insects, respectively. Surprisingly enough, though, there are few reports on that topic. The few studies available on the microbial communities associated to the gut of insects, and particularly to Lepidoptera, are focused on species considered as agricultural or forest pests worldwide. This is the case of the gypsy moth *Lymantria dispar* (Broderick *et al*., 2004; Mason and Raffa, 2014), the diamondback moth *Plutella xylostella* (Lin *et al*., 2015), or the cotton bollworm *Helicoverpa armigera* (Xiang *et al*., 2006). Also, some focus has been put on how diet changes influence the gut microbiota of polyphagous insects such as *Bombyx mori* (Liang *et al*., 2014), *Spodoptera littoralis* (Tang *et a*l., 2012), or *Ostrinia nubilalis* (Belda *et al*., 2011).

Latex- and alkaloid-rich plants constitute a particularly strong selection pressure not only for phytophagous insects (Kirk *et al*., 2012; Ramos *et al*., 2015), but also for their gut microbiota, which is subjected to a constant flow of toxic compounds. The gut of insects feeding on toxic plants is thus a unique and extreme habitat. We present here a complete characterization of the bacterial gut symbionts of two monophagous Lepidoptera feeding on plants rich in, among other toxic compounds, a cocktail of alkaloids or terpenes (Figure II.1). This is the first report of the microbial larval gut communities associated to such toxic diets.

Materials and Methods

Sampling. Larvae of *B. crini* and *H. euphorbiae* (Figure II.1) were obtained in the field (coastal dunes from Pinedo and El Saler, Valencia, Spain) by direct inspection of the food plants *P. maritimum* and *Euphorbia* sp., respectively, in the adequate moment of the year: spring for most larvae of *B. crini* and autumn for *H. euphorbiae*.

Gut dissection and DNA extraction. Last instar larvae from *B. crini* and *H. euphorbiae* were kept in starvation for one day to promote the elimination of plant material from the gut. Larvae were immobilized by placing them on ice and both the midgut and hindgut were dissected under sterile conditions. Guts from three different individuals of each species were independently disaggregated and manually homogenized in PBS buffer (NaCl 8 g/L, KCl 0.2 g/L, Na_2HPO_4 1.44 g/L, KH_2PO_4 0.24 g/L, pH adjusted to 7.4) with an Eppendorf-adapted pestle. Total DNA was obtained from the homogenate with a standard purification protocol consisting of alkaline lysis followed by precipitation with potassium acetate and isopropanol (Latorre *et al*., 1986). An initial incubation step with 2 μg/mL lysozyme at 37 ^{o}C for 30 min was performed to ensure the lysis of Gram-positive bacteria. The quality of the DNA was finally checked on a 0.8% (w/v) agarose gel and quantified with

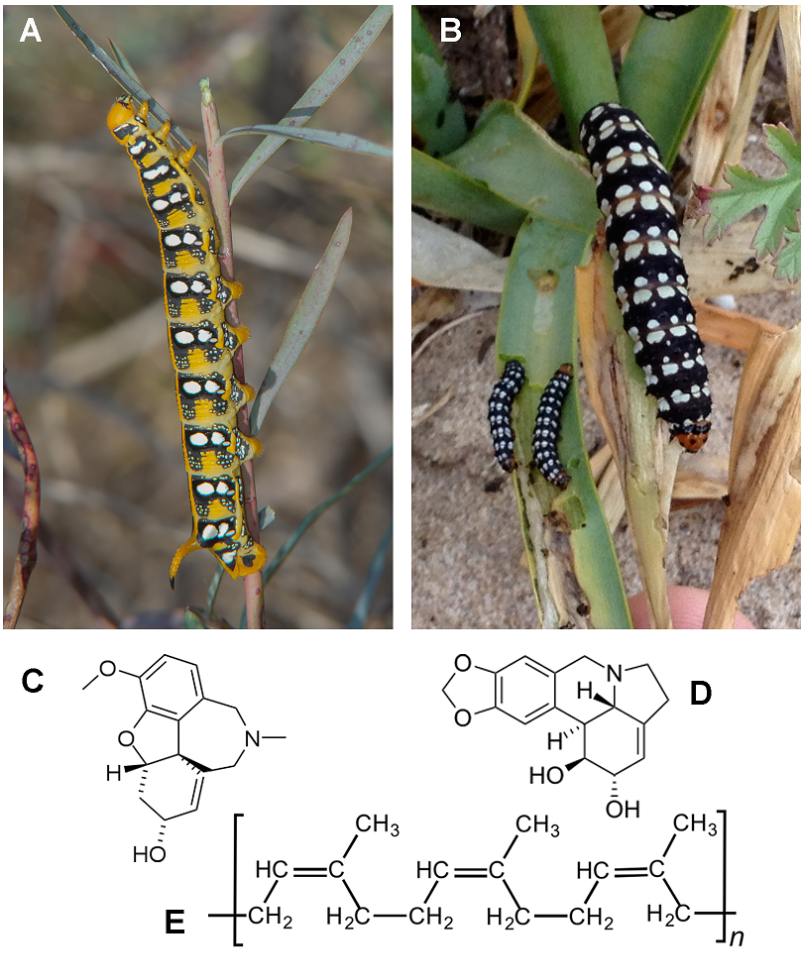


FIGURE II. 1: Larvae and representative secondary compounds found in the diet. A: *Hyles euphorbiae*; B: *Brithys crini*; C: galanthamine (*Pancratium maritimum*); D: narciclasine (*P. maritimum*); E: latex (*Euphorbia* sp.).

Nanodrop-1000 Spectophotometer (Thermo Scientific, Wilmington, DE).

PCR amplification and 16S rRNA profiling. A 700 bp fragment of the V1-V3 hypervariable region of the 16S rRNA genes was PCR-amplified from all the samples with universal primers 28F (5'-GAG TTT GAT CNT GGC TCA G-3') and 519R (5'-GTN TTA CNG CGG CKG CTG-3'). A short (9-11 nucleotides) barcode sequence followed by a four-nucleotide spacer (CGAT) was included at the 5' end of the oligonucleotides used as forward primers to enable assignment of sequences to samples after high-throughput sequencing. All the amplifications were performed under the following thermal cycling conditions: initial denaturing at 95 ºC for 5 min, followed by 35 cycles of denaturing at 95 ºC for 30 s, annealing at 55 ºC for 30 s, and extension at 72 ºC for 1 min, finalized by a 10-min elongation at 72 ºC. Amplicons were checked on a 0.8% (w/v) agarose gel and purified by precipitation with 3M potassium acetate (pH 5) and isopropanol. Pure amplicons were quantified with the Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and an equimolar pool of amplicons was prepared from all the samples.

Amplicons of the 16S rRNA gene for all the samples were pyrosequenced with a Roche GS FLX sequencer and Titanium chemistry in the Center for Public Health Research (FISABIO-Salud Pública, Valencia, Spain). All the sequences obtained were split into groups (a maximum of 2 mismatches were allowed for primer search, whereas no mismatches were

allowed for barcode search), trimmed with a minimum quality score of 20, and filtered to remove short reads (<150 nt). Then, sequences were clustered and taxonomically assigned with the open-reference OTU picking pipeline implemented in the QIIME software (Caporaso *et al.*, 2010). Clustering was performed at a similarity threshold of 97% (species-level OTUs) and the 16S rRNA Greengenes database (version 13.8) was used as reference. Finally, the resulting OTU table was processed and analyzed with software MEGAN (Huson *et al.*, 2011). A summary of statistics is available as Supplementary Table II.1. Sequences were deposited in the MG-RAST public repository under accession numbers 4639004.3 - 46390015.3.

Scanning electron microscopy. Small fragments (2-5 mm) of *B. crini* and *H. euphorbiae* hindguts were dissected. Fragments were fixed by immersion into paraformaldehyde 2% - glutaraldehyde 2.5 % for more than two hours, washed with water and refixed by osmium tetroxide for 20 minutes, washed and dehydrated in absolute ethanol. These pieces were placed inside microporous capsules (30 μm pore size, available from Ted Pella Inc. product number 4619) immersed in absolute ethanol, following critical point drying in an Autosamdri 814 (Tousimis). Dry samples were then arranged on SEM stubs with silver conducting paint TAAB S269. Pieces were manipulated under a stereomicroscope Leica MZ9.5 with Dumont forceps number 5. Stubs were examined under a scanning electron microscope Hitachi S-4100. Images were edited with Photoshop CS3 (Adobe).

Metagenomic sequencing of the bacterial biofilm. The biofilm of a particular *H. euphorbiae* specimen was dissected. Total DNA was isolated with the same protocol described above and then subjected to shotgun metagenomic sequencing in the Center for Public Health Research (FISABIO-Salud Pública, Valencia, Spain). A Nextera Illumina library was built from 100 ng of total DNA following the protocol indications by Illumina. The library was sequenced in a MiSeq sequencer (Illumina) in a combination of 500 cycles, in order to obtain 250 bp paired-end sequences. The MG-RAST platform (Meyer *et al.*, 2008) was used to filter out sequences matching the insect's genome and to taxonomically classify the 16S rRNA sequences belonging to bacteria. To do that, sequence similarity searches were performed against the non-redundant m5RNA database (Supplementary Figure II.1). Sequences were deposited in the MG-RAST public repository under accession number 4633589.3.

Culture media and growth conditions. The *E. casseliflavu*s strain isolated from *H. euphorbia*e hindgut was maintained on LB medium (10 g/L NaCl, 10 g/L bacteriological peptone, 5 g/L yeast extract; and 15 g/L agar for solid medium) at room temperature. The ability of the strain to tolerate or degrade latex was tested in both LB and artificial minimal synthetic medium (2 g/L $NaNO_3$, 1 g/L K_2HPO_4, 0.5 g/L $MgSO_4$, 0.5 g/L KCl, 0.2 g/L bacteriological peptone; and 15 g/L agar for solid medium) supplemented with 10-20% (v/v) *Euphorbia* sp. plant extract or 1-3% (v/v) natural liquid latex (Chemionics Corp., Tallmadge, OH) as the sole carbon sources, respectively, at room temperature. Plant extracts enriched in latex were obtained by grinding 300 g of *Euphorbia* sp. with a domestic blender. The mixture was infused overnight with 150 ml of pure ethanol and then filtered through Whatmann paper. The resulting raw plant extracts were added to the sterilized media at 65-70ºC. The ability of the resulting media to inhibit bacterial growth was tested in *E. coli* XL1-Blue strain, which proved unable to grow in the presence of either the plant extract or the natural latex.

Identification of *E. casseliflavus* strain He. A colony PCR was performed to identify the taxonomy of the *Enterococcus* strain isolated from *H. euphorbiae* gut. A fragment of the 16S rDNA gene was amplified with universal primers 28F (5'-GAG TTT GAT CNT GGC TCA G-3') and 519R (5'-GTN TTA CNG CGG CKG CTG-3'). The PCR program was as follows: an initial denaturing step at 95 ºC for 300 s, followed by 35 cycles of denaturing, annealing and extension (95 ºC, 30 s; 48 ºC, 30 s; and 72 ºC, 60 s) and a final extension step at 72 ºC for 480 s. PCR amplicons were purified by the High Pure PCR Product Purification Kit (Roche Diagnostics GmbH, Mannheim, Germany) and sequencing was carried out with the ABI PRISM BigDye Terminator v3.1 system (Applied Biosystems) on an ABI 3730 automated sequencer. PCR products were sequenced in both senses with the 28F and 519R primers. Sequences were

verified and both strands assembled using the STADEN package. Sequence taxonomy was attributed with BLASTN searches against the RefSeq database of the NCBI. The closest match corresponded to *E. casseliflavus* strain RTCLI14 (sequence similarity=99%; e-value=5e-115).

Results

The bacterial composition of gut extracts from triplicates of larvae of *B. crini* and *H. euphorbiae* was investigated by high throughput sequencing of the 16S rRNA amplicons. A total of 182 species-level OTUs representing 87 different genera were detected in total. As Figure II.2A shows, Firmicutes (OTUs 1, 3, and 4) were, by far, the most abundant bacterial taxa in all cases. Nevertheless, the overall taxonomic profiles from the two species exhibited a clear difference: whereas *B. crini* was characterized by the overwhelming presence of *Enterococcus* sp., (OTU 1, accounting for 94-99% of reads), *H. euphorbiae* harbored a more heterogeneous community. *Enterococcus* sp. was found at high frequencies (10-50%), and another species of the Enterococcaceae family (OTU 3) was also detected at similar frequencies (10-60%). A species belonging to the Enterobacteriaceae family (OTU 2) and *Staphylococcus* sp. (OTU 4) were very common in *H. euphorbiae* (8-70% of sequences depending on the specimen) but were rare in *B. crini*. In that species, the bacterial composition of both midgut and hindgut sections of the insect gut were similar, although some differences were detected (see below). In the case of *H. euphorbiae*, midgut samples exhibited higher amounts of Enterococcaceae (OTU 3) in comparison to hindgut samples, which were richer in *Staphylococcus* sp. and enterobacteria (Figure II.2A).

The remaining bacterial taxa were detected at very low levels in both insects and in both the medium and final sections of the gut. Of these, 58 were exclusive of *H. euphorbiae* (of which 27 were exclusive of the midgut and 15 were exclusive of the hindgut); 29 were exclusive of *B. crini* (20 of which were exclusive of the hindgut); and 95 were found in both species in at least one sample. Figure II.2B shows a Venn diagram with overlapping and exclusive genera occurring in *B. crini* and *H. euphorbiae* samples. A systematic bibliographic search was made with bacterial taxa (those which could be identified to the level of genus or species) in order to identify alkaloid- or latex-degrading abilities. As Figure II.2C shows, we identified minoritary taxa (accounting for near 1% of the total number of reads) reported to have latex and/or alkaloid degradation abilities in both insect species. Three latex degraders (*Nocardioides* sp., *Gordonia* sp., and *Curtobacterium* sp.) were exclusively present in *H. euphorbiae*, whereas two alkaloid degraders (*Klebsiella* sp. and *Corynebacterium* sp.) were detected only in *B. crini*.

Electron micrographs of the gut surface of five individuals of both species showed a virtual absence of bacteria in the midgut. However, some *B. crini* individuals and all the *H. euphorbiae* specimens analyzed showed a detectable ring-like layer of bacteria at the level of the pyloric valve (Figure II.3) of the hindgut. Bacteria concentrated in both species on more sclerotized areas and were particularly common around acanthae. An *H. euphorbiae* hindgut sample with a particularly dense bacterial layer was subjected to total DNA isolation and metagenomic sequencing, which allowed the identification of the most frequent bacterium of the biofilm ring as *Enterococcus casseliflavus* (Supplementary Figure II.1). This *E. casseliflavus* strain (hereafter called *E. casseliflavus* He) was isolated in pure culture. In order to check that this isolate corresponded to the dominant bacterium of the biofilm, the 16S rRNA gene was sequenced and confirmed to be more than 99% identical to the most abundant 16S rRNA gene detected through the metagenomic sequencing of the ring. Also, the sequence was found to be more than 99% identical to the representative 16S rRNA sequence of OTU 3 (the Enterococcaceae species detected at moderate abundance in *H. euphorbiae*). *E. casseliflavus* He exhibited strong growth on media supplemented either with natural latex or *Euphorbia* sp. plant extract, but failed to use natural latex as the sole carbon source (data not shown), suggesting that this strain is able to tolerate –rather than grow on- latex molecules.

Discussion

The gut microbial communities of two lepidopteran species feeding on toxic plants rich in latex and alkaloids are examined for the first time. Although *B. crini* and *H. euphorbiae* belong to different Lepidopteran families

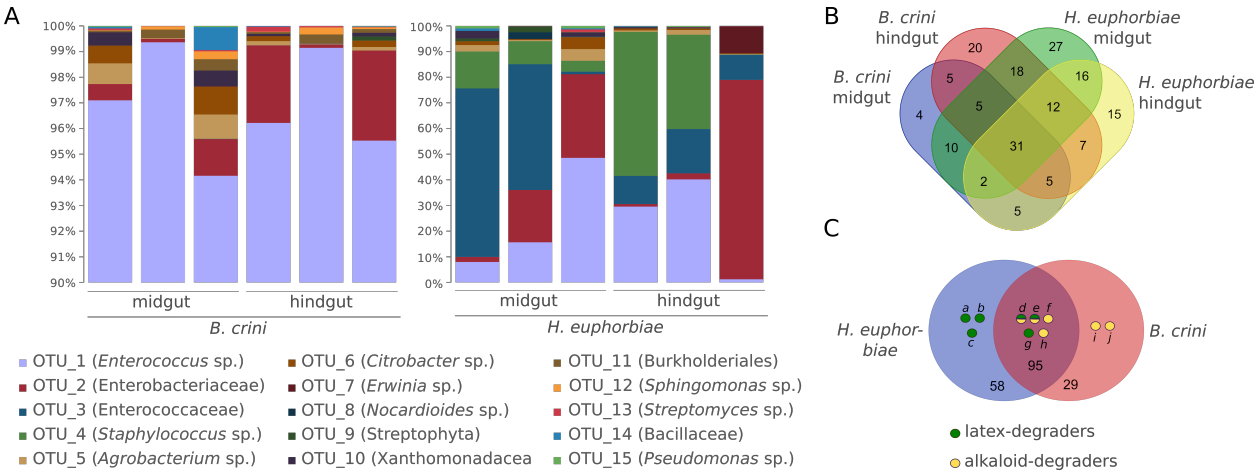


FIGURE II. 2: Bacterial composition of *B. crini* and *H. euphorbiae* guts as deduced by massive 16S rRNA sequencing. A) Relative abundance of bacterial OTUs in the midgut and hindgut sections of three different individuals of each species. Note that the scale of Y-axis starts at 90% for *B. crini* samples. B) Venn diagram showing overlapping and exclusive OTUs occurring in *B. crini* and *H. euphorbiae* guts. C) Occurrence of potential latex-degrading and alkaloid-degrading bacteria in the guts of *B. crini* and *H. euphorbiae* in correlation with their diet (a: *Nocardioides* sp.; b: *Gordonia* sp.; c: *Curtobacterium* sp.: d: *Pseudomonas* sp.; e: *Bacillus* sp.; f: *Sphingomonas* sp.; g: *Streptomyces* sp.; h: *Propionibacterium* sp.; i: *Klebsiella* sp.; j: *Corynebacterium* sp.).

(Noctuidae and Sphingidae, respectively), they proved to harbor similar bacterial communities, surprisingly dominated by the bacterium *Enterococcus* sp., and to a lesser extent in the case of *H. euphorbiae*, by other Enterococcaceae species, an enterobacterium, and *Staphylococcus* sp. The taxonomic profile of midgut and hindgut samples did not significantly differ in *B. crini*, but showed some differences in *H. euphorbiae*. This, as well as the scarce presence of bacteria detected through SEM in midgut samples of both species, might be a consequence of the strong alkaline pH of the midgut of Lepidoptera (Dow, 1992), even though the particular pH conditions of the insects analyzed in this study have not been determined to date. Additionally, a range of genera containing species which are known to degrade latex and/or alkaloids were detected in *H. euphorbiae* and *B. crini*, respectively, which is in correspondence with their diets. Among these, genus *Pseudomonas* is known to harbor several species able to degrade alkaloids (*P. putida* and other unidentified isolates) and natural latex or rubber (*P. aeruginosa* and *P. citronellolis*); and genus *Streptomyces* is considered especially rich in latex- and rubber-degrading species (*S. coelicolor*, *S. griseus*, *S. lividans*, etc.) (Jendrossek *et al.*, 1997; Bode *et al.*, 2001; Rathbone and Bruce, 2002 and references therein). Bacteria belonging to the genera *Enterococcus* and *Staphylococcus* are prevalent in Lepidoptera of the families Sphingidae and Noctuidae (Visôtto *et al.*, 2009), and have been traditionally considered generalist bacteria, since they are widely present in insects (Martin and Mundt, 1972; Tholen *et al.*, 1997; Geiger *et al.*, 2009; Tang *et al.*, 2012). For instance, *E. casseliflavus* has recently been isolated from the lepidopteran *Spodoptera litura*, a highly polyphagous major pest on many crops (Thakur *et al.*, 2015), and also from *Manduca sexta* (Lepidoptera, Sphingidae), a specialist species that feeds on toxic Solanaceae, rich in phenolic derivatives of caffeic acid (Brinkmann *et al.*, 2008). The formation of biofilms dominated by a single bacterial species, such as the one we report in this work, in insect's gut has been traditionally related to entomopathogenic bacteria (Vodovar *et al.*, 2006; Vallet-Gely *et al.*, 2008), although some studies have demonstrated that biofilm formation is essential for the establishment of symbiotic relationships between bacteria and the host insect (Maltz *et al.*, 2012; Vásquez *et al.*, 2012). In particular, *E. casseliflavus* has been found associated with larvae of *Spodoptera litura* (Lepidoptera, Noctuidae) feeding on lima beans, which are especially rich in toxic terpenes

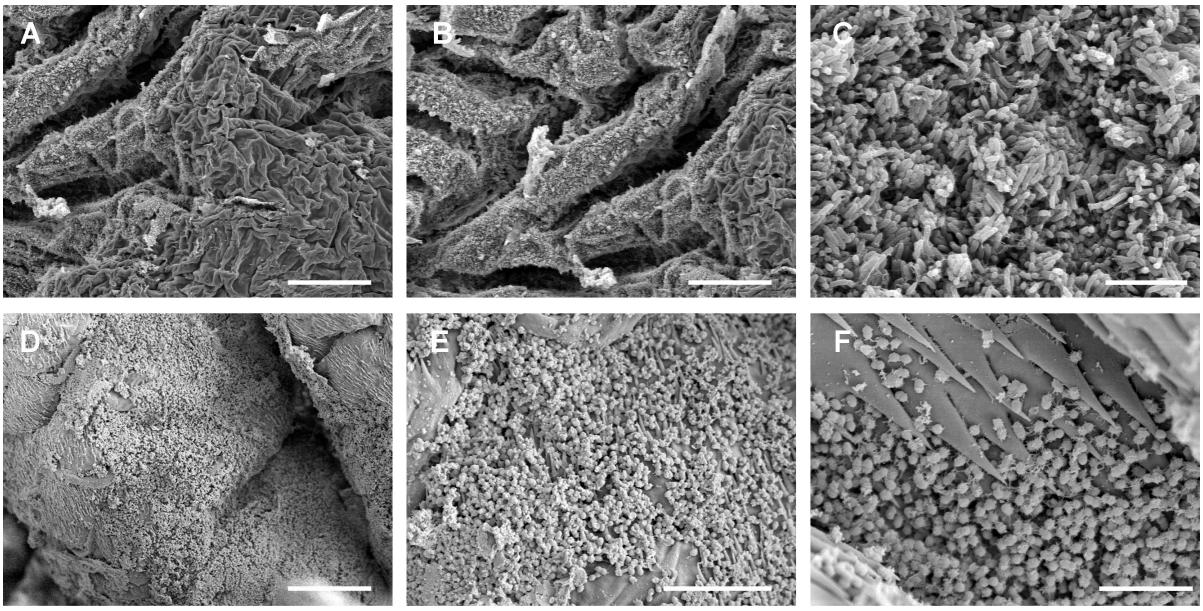


FIGURE II. 3: Scanning electron microscope images of the luminal surface of the hindgut of *Brithys crini* (A, B, C) and *Hyles euphorbiae* (D, E, F). A biofilm of bacilli was observed in some *B. crini* specimens, whereas a biofilm of cocci was detected in all the *H. euphorbiae* specimens analyzed. Notice the presence of acanthae around the biofilm area in *H. euphorbiae*. Scale bars: A, B = 50μm; C = 7 μm; D = 25 μm; E = 10μm; F= 5μm.

such as carotenes. In this case, *E. casseliflavus* forms a monospecific biofilm in which toxic alpha- and beta-carotenoids are crystallized, and larvae failing to develop the biofilm exhibit increased mortality (Shao *et al*., 2011). Given the similarity between carotenes and latex in terms of chemical structure, and the ability of *E. casseliflavus* He to tolerate latex, it is tempting to hypothesize that this strain might be involved in latex immobilization in *H. euphorbiae* hindgut.

This work describes and sheds light in a putatively new case of close relationship between an –apparently- generalist bacterium and a specialist insect. The microbiota described in this work, and especially the *E. casseliflavus* He strain isolated from *H. euphorbiae* hindgut, may be of interest not only for understanding the ecology of such specialist insects, but also for the biotechnological industry, where microorganisms and/or enzymes able to transform alkaloids or latex-like molecules may have biotechnological applications such as bioremediation.

Acknowledgements

We are indebted to the Electron Microscopy Service of the Servei Central de Suport a la Investigació i l'Experimentació (SCSIE, University of Valencia) and Nuria Jiménez (FISABIO) for their technical support. Financial support was provided by grant BFU2012-39816-C02-01 (Ministerio de Ciencia e Innovación, Spain) co-financed by FEDER funds to Amparo Latorre and CGL2008-00605 (Ministerio de Ciencia e Innovación) to Joaquín Baixeras. Cristina Vilanova is a recipient of a FPU fellowship from the Ministerio de Educación (Spain). We are grateful to Sergio Montagud (University of Valencia) for photography in Figure II.1. Local authorities provided permission for collecting the specimens in the field.

References

Bastida, J., Berkov, S., Torras, L., Pigni, N.B., Andrade, J. P. de, Martínez, V., Codina, C. and Viladomat, F. (2011). "Chemical and biological aspects of Amaryllidaceae alkaloids", In Recent Advances in Pharmaceutical Sciences, ed. D. Muñoz-Torrero, D., (Ed.), *Recent Advances in*

Pharmaceutical Sciences, 2011: 65-100.

Belda, E., Pedrola, L., Peretó, J., Martínez-Blanch, J. F., Montagud, A., Navarro, E., *et al*. (2011). Microbial diversity in the midguts of field and lab-reared populations of the European corn borer Ostrinia nubilalis. *PLoS One* **6**, e21751.

Bode, H. B., Kerkhoff, K., and Jendrossek, D. (2001). Bacterial Degradation of Natural and Synthetic Rubber. *Biomacromolecules* **2**, 295–303.

Brinkmann, N., Martens, R., and Tebbe, C. C. (2008). Origin and diversity of metabolically active gut bacteria from laboratory-bred larvae of Manduca sexta (Sphingidae, Lepidoptera, Insecta). *Appl. Environ. Microbiol*. **74**, 7189–96.

Broderick, N. A., Raffa, K. F., Goodman, R. M., and Handelsman, J. (2004). Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. *Appl. Environ. Microbiol.* **70**, 293–300.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., *et al*. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–6.

Chung, S.H., Rosa, C., Scully, E.D., Peiffer, M., Tooker, J.F., Hoover, K., Luthe, D.S., Felton, G.W. (2013) Herbivore exploits orally secreted bacteria to suppress plant defenses. *Proc Natl Acad Sci* **110**, 15728–33.

Dow, J. (1992). pH gradients in lepidopteran midgut. *J. Exp. Biol.* **172**, 355–375.

Geiger, A., Fardeau, M.-L., Grebaut, P., Vatunga, G., Josénando, T., Herder, S., *et al*. (2009). First isolation of Enterobacter, Enterococcus, and Acinetobacter spp. as inhabitants of the tsetse fly (Glossina palpalis palpalis) midgut. *Infect. Genet. Evol.* **9**, 1364–70.

Hagel, J., Yeung, E., and Facchini, P. (2008). Got milk? The secret life of laticifers. *Trends Plant Sci.* **13**, 631–639.

Hammer, T. J., and Bowers, M. D. (2015). Gut microbes may facilitate insect herbivory of chemically defended plants. *Oecologia* **179**, 1–14.

Hansen, A. K., and Moran, N. A. (2014). The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol. Ecol.* **23**, 1473–96.

Hetta, M. H. and Shafei A. A. (2013) Comparative cytotoxic and antimicrobial activities of the alkaloid content of Egyptian Pancratium maritimum L. fruits and flowers. *Journal of American Science* **9**: 104-109.

Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–60.

Jendrossek, D., Tomasi, G., and Kroppenstedt, R. M. (2006). Bacterial degradation of natural rubber: a privilege of actinomycetes? *FEMS Microbiol. Lett.* **150**, 179–188.

Kirk, H., Vrieling, K., Pelser, P. B., and Schaffner, U. (2012). Can plant resistance to specialist herbivores be explained by plant chemistry or resource use strategy? *Oecologia* **168**, 1043–55.

Koh, S., and Hik, D. S. (2007). Herbivory mediates grass-endophyte relationships. *Ecology* **88**, 2752–7.

Latorre, A., Moya, A., and Ayala, F. J. (1986). Evolution of mitochondrial DNA in Drosophila subobscura. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 8649–53.

Liang, X., Fu, Y., Tong, L., and Liu, H. (2014). Microbial shifts of the silkworm larval gut in response to lettuce leaf feeding. *Appl. Microbiol. Biotechnol.* **98**, 3769–76.

Lin, X.-L., Kang, Z.-W., Pan, Q.-J., and Liu, T.-X. (2015). Evaluation of five antibiotics on larval gut bacterial diversity of Plutella xylostella (Lepidoptera: Plutellidae). *Insect Sci.* **22**, 619–28.

Maltz, M. A., Weiss, B. L., O'Neill, M., Wu, Y., and Aksoy, S. (2012). OmpA-mediated biofilm formation is essential for the commensal bacterium Sodalis glossinidius to colonize the tsetse fly gut. *Appl. Environ. Microbiol.* **7**8, 7760–8.

Marsh, N. and Rothschild M. (1984) "A New Look at Lepidoptera Toxins"., In The Biology of Butterflies, Symposium of the Royal Entomological Society of London, ed. R.I. Vane-Wright and P.R. Ackery (Eds.), The Biology of Butterflies, *Symposium of the Royal Entomological Society of London* **11**: 135-139.

Martin, J. D., and Mundt, J. O. (1972). Enterococci in insects. *Appl. Microbiol.* **24**, 575–80.

Mason, C. J., and Raffa, K. F. (2014). Acquisition and structuring of midgut bacterial communities in gypsy moth (Lepidoptera: Erebidae) larvae. *Environ. Entomol.* **4**3, 595–604.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., *et al*. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386.

Nishida, R. (2002). Sequestration of defensive substances from plants by Lepidoptera. *Annu. Rev. Entomol.* **47**, 57–92.

Nuringtyas, T. R., Verpoorte, R., Klinkhamer, P. G. L., van Oers, M. M., and Leiss, K. A. (2014). Toxicity of Pyrrolizidine Alkaloids to Spodoptera exigua Using Insect Cell Lines and Injection Bioassays. *J. Chem. Ecol.* **40**, 609–616.

Ramos, M. V, Pereira, D. A., Souza, D. P., Silva, M.-L. S., Alencar, L. M. R., Sousa, J. S., *et al*. (2015). Peptidases and peptidase inhibitors in gut of caterpillars and in the latex of their host plants. *Planta* **241**, 167–78.

Rathbone, D. A., and Bruce, N. C. (2002). Microbial transformation of alkaloids. *Curr. Opin. Microbiol.* **5**, 274–281.

Rothschild, M. (1973) "Secondary plant substances and warning colouration in insects"., In Insect/Plant Relationships, Symposia of the Royal Entomological Society of London, ed. H. F. van Emden (Ed.), Insect/Plant Relationships, *Symposia of the Royal Entomological Society of London*, **6**: 59-83.

Shao, Y., Spiteller, D., Tang, X., Ping, L., Colesie, C., Münchberg, U., *et al*. (2011). Crystallization of α- and β-carotene in the foregut of Spodoptera larvae feeding on a toxic food plant. *Insect Biochem. Mol. Biol.* **41**, 273–81. doi:10.1016/j.ibmb.2011.01.004.

Tang, X., Freitak, D., Vogel, H., Ping, L., Shao, Y., Cordero, E. A., *et al*. (2012). Complexity and variability of gut commensal microbiota in polyphagous lepidopteran larvae. *PLoS One* **7**, e36978.

Tang, X., Freitak, D., Vogel, H., Ping, L., Shao, Y., Cordero, E. A., *et al*. (2012). Complexity and variability of gut commensal microbiota in polyphagous lepidopteran larvae. *PLoS One* **7**, e36978.

Thakur, A., Dhammi, P., Saini, H. S., and Kaur, S. (2015). Pathogenicity of bacteria isolated from gut of Spodoptera litura (Lepidoptera: Noctuidae) and fitness costs of insect associated with consumption of bacteria. *J. Invertebr. Pathol.* **127**, 38–46.

Tholen, A., Schink, B., and Brune, A. (2006). The gut microflora of Reticulitermes flavipes, its relation to oxygen, and evidence for oxygen-dependent acetogenesis by the most abundant Enterococcus sp. *FEMS Microbiol. Ecol.* **24**, 137–149.

Vallet-Gely, I., Lemaitre, B., and Boccard, F. (2008). Bacterial strategies to overcome insect defences. *Nat. Rev. Microbiol.* **6**, 302–13.

Vásquez, A., Forsgren, E., Fries, I., Paxton, R. J., Flaberg, E., Szekely, L., *et al*. (2012). Symbionts as major modulators of insect health: lactic acid bacteria and honeybees. *PLoS One* **7**, e33188.

Vaughan, M. M., Wang, Q., Webster, F. X., Kiemle, D., Hong, Y. J., Tantillo, D. J., *et al*. (2013). Formation of the unusual semivolatile diterpene rhizathalene by the Arabidopsis class I terpene synthase TPS08 in the root stele is involved in defense against belowground herbivory. *Plant Cell* **25**, 1108–25.

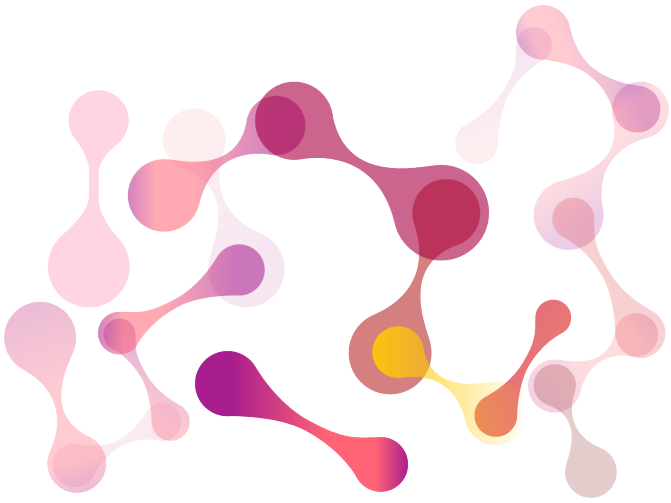
Visôtto, L. E., Oliveira, M. G. A., Ribon, A. O. B., Mares-Guia, T. R., and Guedes, R. N. C. (2009). Characterization and identification of proteolytic bacteria from the gut of the velvetbean caterpillar (Lepidoptera: Noctuidae). *Environ. Entomol.* **38**, 1078–85.

Vodovar, N., Vallenet, D., Cruveiller, S., Rouy, Z., Barbe, V., Acosta, C., *et al*. (2006). Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium Pseudomonas entomophila. *Nat. Biotechnol.* **24**, 673–9.

War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., *et al*. (2012). Mechanisms of plant defense against insect herbivores. *Plant Signal. Behav.* **7**, 1306–20.

Xiang, H., Wei, G.-F., Jia, S., Huang, J., Miao, X.-X., Zhou, Z., *et al*. (2011). Microbial communities in the larval midgut of laboratory and field populations of cotton bollworm (Helicoverpa armigera). *Can. J. Microbiol.* **52**, 1085-1092.

Chapter 1B. Microbial communities associated to artificial environments: the case of solar panels



Publication III:
Dorado-Morales P, Vilanova C, Peretó J, Codoñer FM, Ramón D, Porcar M. (2016). A highly diverse, desert-like microbial biocenosis on solar panels in a Mediterranean city. *Sci Rep* **6**:29235.

Publication III

A highly diverse, desert-like microbial biocenosis on solar panels in a Mediterranean city

Pedro Dorado-Morales[1,+], Cristina Vilanova[1,2,+], Juli Peretó[1,2,3], Francisco M. Codoñer[4], Daniel Ramón[5] and Manuel Porcar[1,2,*]

[1] Cavanilles Institute of Biodiversity and Evolutionary Biology (Universitat de València), Spain.
[2] Darwin Bioprospecting Excellence, SL., Parc Cientific Universitat de València, Paterna, Valencia, Spain.
[3] Departament de Bioquímica i Biologia Molecular, Universitat de València, Burjassot 46100, Spain.
[4] Lifesequencing S.L., Parc Científic Universitat de València, Paterna, Spain.
[5] Biópolis S.L., Parc Cientific Universitat de València, Paterna, Valencia, Spain.

[+] These authors contributed equally to this work.

Abstract

Microorganisms colonize a wide range of natural and artificial environments although there are hardly any data on the microbial ecology of one the most widespread man-made extreme structures: solar panels. Here we show that solar panels in a Mediterranean city (Valencia, Spain) harbor a highly diverse microbial community with more than 500 different species per panel, most of which belong to drought-, heat- and radiation-adapted bacterial genera, and sun-irradiation adapted epiphytic fungi. The taxonomic and functional profiles of this microbial community and the characterization of selected culturable bacteria reveal the existence of a diverse mesophilic microbial community on the panels' surface. This biocenosis proved to be more similar to the ones inhabiting deserts than to any human or urban microbial ecosystem. This unique microbial community shows different day/night proteomic profiles; it is dominated by reddish pigment- and sphingolipid-producers, and is adapted to withstand circadian cycles of high temperatures, desiccation and solar radiation.

Introduction

Today, photovoltaic panels cover around 4000 square kilometers, and are forecasted to be the world's main electricity source by 2050 (`http://www.epia.org`). Solar panels are unique biotopes characterized by a smooth flat glass or glass-like surface, minimum water retention capacity and maximum sunlight exposure, all of which determine circadian and annual peaks of irradiation, desiccation and heat. Extreme natural habitats such as thermal vents, mountain plateaus or hyper arid deserts are known to host microbial biocenoses adapted to those particular selection pressures (Corliss *et al.*, 1979; Singh and Lal, 2009; Neilson *et al.*, 2012); and artificial or humanized environments, such as industrial reactors (Rosche *et al.*, 2009), radioactive waste (Williamson *et al.*, 2014) or oil spills (Pham and Anonye, 2014) are also colonizable by specialized microorganisms. Among artificial environments, indoor biomes constitute around 0.5% of the ice-free land area, a surface comparable to the subtropical coniferous forest (Martin *et al.*, 2015). Geography and building type, among other factors, structure indoor microbes, which are inoculated by human skin and outdoor air (Adams *et al.*, 2015) and even by indoor plants (Mahnert *et al.*, 2015). One of the recently studied indoor habitats is the subway. A recent

study of the aerosol microbial communities of the New York City subway platforms (Robertson *et al*., 2013) revealed that such communities were a mixture of soil, environmental water, and human skin commensal bacteria. In a similar work in the Honk Kong subway networks (Leung *et al*., 2014), researchers found that each subway line harbored a different phylogenetic community, depending on architectural characteristics, nearby (outdoor) microbiomes, as well as connectedness with other lines.

Most of the so-called built environment studies have so far focused in the indoor biomes but less attention has been paid to artificial, outdoor environments. A recent example of those is an ambitious city-scale metagenomics screening on the bacterial diversity of New York (including both indoor and outdoor biomes), which found a very high diversity of microorganisms -roughly 1700 bacterial taxa-, half of which did not match any known organism (Afshinnekoo *et al*., 2015). However, many artificial outdoor environments remains unexplored to date. In the present work, we aimed at studying solar panels, an outdoor artificial environment whose role as biotope had not previously been reported. The goal of our study was to identify possible microbial communities thriving in the harsh conditions of the panels' surfaces, and to determine to which extent this highly irradiated environment was home of ecologically extremophile bacteria (thermophilic and radio-resistant taxa, for example), as well as to identify microbial strains with biotechnological applications, such as carotenoid producers.

Our report documents a complete bioprospection and characterization of the microbial community on photovoltaic panels of a Mediterranean city, using high throughput 16S/18S rRNA analysis, metagenomic sequencing, metaproteomics, and culture-based characterization of selected isolates.

Materials and Methods

Sampling. Sampling was performed during the summer solstice of 2013 and 2014. Sampling consisted of a simple harvesting procedure, by pouring sterile PBS (sodium phosphate buffer) on the panel and immediately harvesting the liquid by strongly and repeatedly scraping the surface with a modified window cleaner with an autoclaved silicone tube measuring 5 mm in diameter. The resulting suspension was collected by using a sterile plastic pipette and transferred to sterile Falcon tubes, placed on ice and immediately transported to the lab. Solar panels had a surface of 1.28 m^2, had been in operation for at least 5 years, and the different panels sampled were separated at least 10 meters to avoid statistical biases in microbial composition. In 2013, nine samples from solar panels on the three campuses of the University of Valencia (Valencia, Spain) were collected at noon (2 PM) and pooled together for further analysis through 16S/18S rRNA sequencing and cultivation methods. Air temperature was 33 ºC and relative humidity was 60%. In 2014, we sampled three solar panels from a single location (Faculty of Economics, University of Valencia, Valencia, Spain) under day (2 PM) and night conditions (4 AM). Day-collected samples were independently analyzed through 16S/18S rRNA sequencing, Illumina shotgun sequencing (only two of the panels sampled), and metaproteomics; whereas night-collected samples were only used for the metaproteomic analysis. Air temperature and relative humidity were 32 ºC and 56% (2 PM), and 23 ºC and 83% (4 AM).

The average temperature of the panels' surface during the sampling process (at 2 PM) was 51 ºC. The average solar irradiance in Valencia at 2 PM is 461.3 W/m^2, whereas the accumulated solar irradiance during an average day is 19.6 kJ/m^2. According to the reports of AEMET (Spanish State Meteorological Agency), the weather conditions in Valencia were similar between May-June 2013 and May-June 2014 (the periods prior to summer solstice when sampling took place). In 2013, the temperature of the one-month period prior sampling was 0.8 ºC lower than the average, and the precipitation levels dropped a 20% with respect to the last decade. In contrast, in 2014, the temperature of the one-month period prior sampling was 0.5 ºC higher than the average, and normal precipitation levels were recorded. The microbial profiles displayed no major changes between 2013 and 2014.

Microbiological media and growth conditions. Three culture media (LB medium, R2A medium, and marine agar) were used in this work according to our previous experience in other highly irradiated

environments, where those media proved optimal to detect pigment-producing strains. Aliquots of 100 μL from the solar-panel samples were spread on Petri dishes containing LB medium (composition in g/L: peptone 10.0, NaCl 10.0, yeast extract 5.0), R2A medium (composition in g/L: peptone 0.5, casaminoacids 0.5, yeast extract 0.5, dextrose 0.5, soluble starch 0.5, K_2HPO_4 0.3, $MgSO_4$ 0.05, sodium pyruvate 0.3) or marine medium (composition in g/L: peptone 5.0, yeast extract 1.0, ferric citrate 0.1, NaCl 19.45, $MgCl_2$ 5.9, Na_2SO_4 3.24, $CaCl_2$ 1.8, KCl 0.55, $NaHCO_3$ 0.16, KBr 0.08, $SrCl_2$ 0.034, H_3BO_3 0.022, Na_4O_4Si 0.004, NaF 0.024, NH_4NO_3 0.0016, Na_2HPO_4 0.008), and incubated at room temperature (previous tests were performed at different temperatures, but the largest number of colonies was observed at room temperature) for 7 days. Individual colonies were independently re-streaked on new media and pure cultures were finally identified through 16S rRNA sequencing and criopreserved in 20% glycerol (v/v) until required. A total of 53 bacterial strains were characterized under stress conditions and serially confronted with each other in solid medium to detect interactions in terms of resistance to harsh conditions.

Strains stress tests. Each strain was subjected to a range of stress assays to test tolerance to salinity, heat, low pH, and UV radiation. Overnight liquid cultures were adjusted to an OD600 value of 0.03. Then, stress tests were carried out by plating several 20 μL droplets of the diluted culture on LB, R2A, or marine agar with the following modifications. In the case of salinity stress, increasing amounts of NaCl (from 1 to 9% w/v) were added to the media (final concentration ranging from 1 to 26% w/v). To test pH resistance, culture media was adjusted to pH 5, 6, and 8. In the case of heat, plates were incubated overnight at 60 ºC; whereas resistance to radiation was tested by applying UV pulses of different length (30 s, 2 min, and 8 min) with a VL-4C lamp (254 nm, 340 μW cm^{-2}; Labolan, S.L., Spain). The XL1-Blue *E. coli* strain was used as control. For each experiment, two independent replicates were performed. In order to detect inhibition or synergistic effects between strains, experiments were performed as described above and strain suspensions (20 μl) were closely (3 mm) placed on the same dish.

DNA isolation. Selected DNA purification methods were used to process the solar panels samples and DNA yields were compared (data not shown). Metagenomic DNA was isolated using the Power Soil DNA Isolation kit (MO BIO Laboratories) following the manufacturer's instructions with an additional pretreatment with DNA-free lysozyme at 37 ºC for 10 min. The quantity and quality of the DNA was determined on a 1.5% agarose gel and with a Nanodrop-1000 Spectophotometer (Thermo Scientific, Wilmington, DE).

PCR amplification and 16S/18S rRNA massive sequencing. A set of primers adapted to massive sequencing for the Ion Torrent platform (Lifetechnologies) were used to capture 16S (modified from a previous study; Sim *et al*., 2012) and 18S (modified from a previous study; La Duc *et al*., 2012) rRNA from the solar-panel DNA extraction in a PCR reaction. PCR reactions were performed with 30 ng of metagenomic DNA, 200 μM of each of the four deoxynucleoside triphosphates, 400 nM of each primer, 2.5 U of FastStart HiFi Polymerase, and the appropriate buffer with $MgCl^2$ supplied by the manufacturer (Roche, Mannheim, Germany), 4% of 20 g/mL BSA (Sigma, Dorset, United Kingdom), and 0.5 M Betaine (Sigma). Thermal cycling consisted of initial denaturation at 94ºC for 2 minutes followed by 35 cycles of denaturation at 94ºC for 20 seconds, annealing at 50ºC for 30 seconds, and extension at 72ºC for 5 minutes.

Amplicons were combined in a single tube in equimolar concentrations. The pooled amplicon mixture was purified twice (AMPure XP kit, Agencourt, Takeley, United Kingdom) and the cleaned pool requantified using the PicoGreen assay (Quant-iT, PicoGreen DNA assay, Invitrogen). Subsequently, sequencing on the Ion Torrent platform was performed at LifeSequencing S.L. (Valencia, Spain).

Shotgun metagenomic sequencing. The metagenomic DNA of two of the solar panels sampled in 2014 (solar panels 1 and 3, from which enough DNA was available) was shotgun sequenced. A Nextera Illumina library was built from 100 ng total DNA following the protocol indications marked by Illumina. Those libraries were sequenced in a MiSeq sequencer (Illumina) at Lifesequencing SL, in a combination of 500 cycles, in order to obtain 250 bp paired-end

sequences.

16S/18S rRNA profiling. The resulting sequences from the taxonomical identification, based on PCR capturing of the 16S and 18S rRNA, were split taking into account the barcode introduced during the PCR reaction, providing a single FASTQ file for each of the samples. We performed quality filtering (Q20) using fastx tool kit version 0.013, primer (16S and 18S rRNA primers) trimming using cutadapt version 1.4.1 and length (minimum 300 bp read length) trimming using in-house perl scripting over those FASTQ files to obtain a FASTQ file with clean data. Those clean FASTQ files were converted to FASTA files and UCHIME (Edgar *et al.*, 2011) program version 7.0.1001 was used to remove chimeras arising during the amplification and sequencing step. Those clean FASTA files were BLAST against NCBI 16S rRNA and fungi database using blastn version 2.2.29+. The resulting XML file were processed using a pipeline developed by Lifesequencing S.L. (Paterna, Valencia, Spain) in order to annotate each sequence at different phylogenetic levels (Phylum, Family, Genera and Species). Statistical analysis was performed using R version 3.1.1. A summary of sequencing statistics and results is available in Supplementary Table III.2.

Taxonomic and functional analysis of metagenomic sequences. Two FASTQ files per sample were obtained during the sequencing step, each coming from each of the directions on the paired-end sequencing. Those files were trimmed for adapters and low quality reads using *cutadapt* version 1.4.1 with the paired-end option. Trimmed sequences were used for taxonomical identification using a local alignment tool against nt database from NCBI as described before (Ames *et al.*, 2015). The trimmed sequences from each solar panel were also assembled using different combinations of k-mers in Abyss version 1.5.2 (Simpson *et al.*, 2009) and Velvet version 1.2.1 (Zerbino *et al.*, 2008) in order to find the best combination. The best assembly in each solar panel was used to perform a prediction of ORFs by using MetaGeneMark (Besemer *et al.*, 2011). BLASTP against nr NCBI database was used for annotation and webMGA (Wu *et al.*, 2011) for COG assignation. All our data have been deposited in the MG-RAST server, and is publicly available under accession numbers 4629146.3 and 4629747.3.

In order to compare the taxonomic profile of solar panels with other environments, the taxonomic information of 25 metagenomes belonging to different habitats was obtained from the MG-RAST server (IDs 4455835.3, 4455836.3, 4477803.3, 4477872.3, 4477873.3, 4441205.3, 4445129.3, 4445126.3, 4477903.3, 4477904.3, 4477901.3, 4514299.3, 4543019.3, 4543020.3, 4441347.3, 4441363.3, 4441215.3, 4441214.3, 4441679.3, 4441682.3, 4447192.3, 4447102.3, 4497390.3, 4497389.3, 4497397.3, 4516651.3, and 4516403.3). The different profiles were processed with MEGAN. Data were normalized, and the distances between pairs of profiles calculated with the Bray-Curtis method. Finally, the calculated distances were used to build a Principal Coordinates Analysis. We employed the Statistical Analysis of Metagenomic Profiles (STAMP) (version 1.08; Faculty of Computer Science, Dalhousie University) software to compare the functional profile (according to subsystems categories) of our samples with those of the metagenomes previously cited. The functional data of metagenomes 4455835.3, 4455836.3, 4477803.3, 4477872.3, 4477873.3, 4477903.3, 4477904.3, 4477901.3, 4514299.3 was poor or absent, and was thus eliminated from the analysis. This comparison was represented in a heatmap, where the different metagenomes are clustered according to their similarity. The functional contents of solar panels 1 and 3 were compared to each other with a Fisher's exact test combined with the Newcombe-Wilson method for calculating confidence intervals (nominal coverage of 95%). As a multiple-hypothesis test correction, a false-discovery-rate (FDR) method was applied.

Pangenome reconstruction. Trimmed sequences from both solar panels for the total DNA experiment were blasted against a database containing all sequences for the genera Thermo/Deinococcus. Only sequences with a positive hit in against this database were used for assembly them using different k-mers with the Abyss assembler. We performed a genome annotation in different steps i) ORFs were predicted with GeneMark version 3.25 (Besemer *et al.*, 2001), ii) rRNA version 1.2 (Lagesen *et al.*, 2007) for rRNA prediction, and iii) tRNA-Scan (Lowe and Eddy, 1997) for tRNA prediction. The functional annotation using COG classification

was performed using webMGA. DNAPlotter (Carver *et al*., 2009) from the Artemis Package was used to represent a circular map of the pangenome.

Proteomics. Protein samples were precipitated with TCA (trichloroacetic acid) and pellets were dissolved with 75 μL of 50 mM ABC (ammonium bicarbonate). The protein concentration in the samples was determined by fluorometric analysis. Then, 10 μg of each sample were digested as described in the following protocol. Cysteine residues were reduced by 2 mM DTT (DL-Dithiothreitol) in 50 mM ABC at 60ºC for 20 min. Sulfhydryl groups were alkylated with 5 mM IAM (iodoacetamide) in 50 mM ABC in the dark at room temperature for 30 min. IAM excess was neutralized with 10 mM DTT in 50 mM ABC, 30 min at room temperatura. Each sample was subjected to trypsin digestion with 250 ng (100 ng/μl) of sequencing grade modified trypsin (Promega) in 50 mM ABC at 37ºC overnight. The reaction was stopped with TFA (trifluoroacetic acid) at a final concentration of 0.1%. Final peptide mixture was concentrated in a speed vacuum and resuspended in 30 μL of 2% ACN, 0.1% TFA. Finally, 5 μl of each sample were loaded onto a trap column (NanoLC Column, 3μ C18-CL, 75μm x 15cm; Eksigen) and desalted with 0.1% TFA at 2 μl/min during 10 min.

The peptides were then loaded onto an analytical column (LC Column, 3 μC18-CL, 75μm x 25cm, Eksigen) equilibrated in 5% acetonitrile 0.1% FA (formic acid). Elution was carried out with a linear gradient of 5:35% B in A for 40 min (A: 0.1% FA; B: ACN, 0.1% FA) at a flow rate of 300 nl/min in a label free mode. Peptides were analyzed in a mass spectrometer nanoESI qQTOF (5600 TripleTOF, ABSCIEX). The tripleTOF was operated in information-dependent acquisition mode, in which a 0.25s TOF MS scan from 350–1250 m/z, was performed, followed by 0.05s product ion scans from 100–1500 m/z on the 25 most intense 2-5 charged ions.

ProteinPilot default parameters were used to generate a peak list directly from 5600 TripleTof wiff files. The Paragon algorithm of ProteinPilot was used to search NCBI protein database with the following parameters: trypsin specificity, cys-alkylation, no taxonomy restriction, and the search effort set to through. To avoid using the same spectral evidence in more than one protein, the identified proteins were grouped based on MS/MS spectra by the ProteinPilot Progroup algorithm. The PeakView v 1.1 (ABsciex) software was used to generate the peptides areas from Protein Pilot result files and to perform a principal component (PCA) and a t-test analysis.

Results

Culturing of solar panel samples from the 2013 solstice on LB, R2A, and marine media yielded a relatively high number of colony forming microorganisms, mostly bacteria, displaying a wide range of color and shapes. Many of the isolates displayed red, orange or pink pigmentation (Supplementary Figure III.1), particularly those incubated on marine agar. A total of 53 pigmented isolates were selected and subjected to taxonomic (16S rRNA) characterization and tested for resistance to heat (incubation at 60ºC), UV exposure (2 to 30 s pulses of a 340 μW cm^{-2} UV light), high NaCl contents (1 to 26%) and different pH values (5 to 9). Most of the solar panel isolates displayed resistance to very high salt concentrations (20-26% w/v NaCl) and short exposures to UV light, whereas only a few number of isolates proved resistant to a low pH or extreme heat (Supplementary Figure III.2). Interestingly, during these characterization assays we were able to identify isolates able to restore the growth of nearby isolates under conditions of extreme salt or pH values, in the latter case because of local buffering of the pH of the plate (Supplementary Figure III.1). Full characterization of the 53 isolates is provided in Supplementary Table III.1 and Supplementary Figure III.2.

The taxonomic composition of bacterial and eukaryotic taxa was first studied through 16S and 18S rRNA genes massive sequencing; the results are shown in Figure III.1. As many as 800 different bacterial species were identified in the 2013 pool (nine solar panels from different locations within the University of Valencia); and around 500 different species were found in each of the individual panels sampled from a single building in 2014 (Supplementary Table III.2). Two orders, Sphingobacteriales (families Flexibacteriaceae and Sphingomonadaceae) and Deinococcales comprised the highest number of species. *Deinococcus*, *Sphingomonas*,

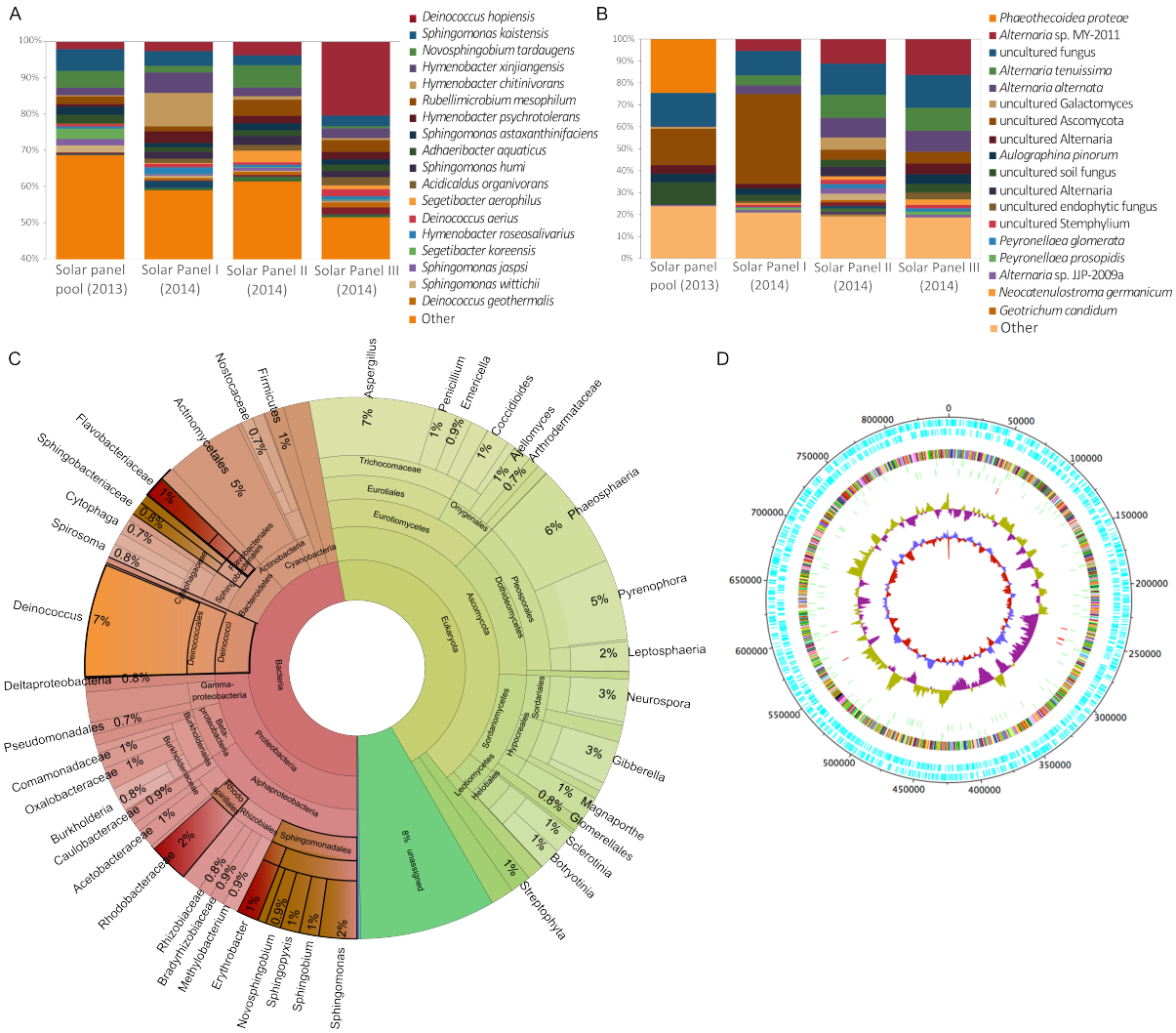


FIGURE III. 1: Taxonomic diversity of solar-panel biocenosis, deduced by culture-independent techniques. Diversity of bacteria (A) and fungi (B) analyzed by 16S and 18S rRNA gene sequencing, respectively, of solar panels sampled during the Summer solstice of 2013 (pool of samples) and 2014 (three solar panels independently analyzed). Histograms show the relative abundance (%) of the species identified, as described in Methods. Species representing less than 1% of total reads were clustered and labeled as "Other". Taxonomic diversity of one of the panels (panel 3) sampled in the summer solstice of 2014 as deduced from shotgun metagenomic sequencing (C). Carotenoid and sphingolipid-producing bacteria are highlighted in red and brown, respectively. Radiation-resistant phyla are highlighted in orange. The taxonomic diversity of solar panel 1 is represented in Supplementary Figure III.3. Circular representation of the *Deinococcus* solar panel pangenome (D) obtained from the metagenomic sequences of panels 1 and 3. The map includes (from the outer to the inner circle) the ORFs in forward and reverse sense, a colour-coded COG functional annotation, the predicted tRNAs and rRNAs, the GC count, and the GC skew.

Novosphingobium or *Hymenobacter* were the dominant genera, with 9.2% or 28% of the assigned sequences (*H. chitinivorans* in panel 1 and *D. hopiensis* in panel 3, respectively). The remaining sequences were distributed among 17 phyla and 146 families. Other well-represented genera, in order of abundance, were *Rubellimicrobium*, *Adhaeribacter*, *Acidicaldus*, *Segetibacter*, or *Modestobacter*. In the case of fungi, lower biodiversity was found (Figure III.1B). Taxonomic eukaryotic profiles were dominated by the phylum Ascomycota and the families Pleosporaceae and Teratosphaeriaceae, with genera *Phaeothecoidea* and *Alternaria* representing the majority in the 2013 and 2014 samples, respectively.

In 2014, the metagenomic DNA of two independent solar panels was shotgun sequenced. As shown in Figure III.1C, the ratio of bacteria:fungi reads was close to 50%, and species distribution was similar to that found for 16S and 18S rRNA sequencing. Genus *Deinococcus*, one of the clearest taxonomic markers of extremophily, again proved highly abundant in all our samples. A summary of sequencing statistics and diversity indexes can be accessed in Supplementary Table III.3. The core microbiome, understood as the number of species detected in all the samples analyzed, comprised around 120 species (26% of the total number of species detected), and was composed, mainly, by members of the *Deinococcus*, *Sphingomonas*, *Novosphingobium*, and *Hymenobacter* genera. We analyzed *Deinococcus* sequences from our metagenomic analysis and a draft *Deinococcus* solar panel pangenome of 2098 contigs was obtained, covering more than 0,8 Mb (25%) of standard *Deinococcus* genomes (3.3Mb with 2 chromosomes and 2 plasmids), with 2166 and 149 ORFs and tRNAs, respectively (Figure III.1D). The low identity level of the solar panel pangenome with previously sequenced *Deinococcus* species strongly suggests that at least one previously undescribed *Deinococcus* species is present in the sampled panels.

Regarding the functional profile, that of two independent solar panels (1 and 3) was deduced from the metagenomic data, and statistically analyzed with the STAMP software. When compared to a range of metagenomes from diverse habitats, solar panel functional profiles clustered together with those described for polar microbial mat and saline desert datasets, and distant to those of other environments such as air or sediments (Figure III.2A). Both solar panels proved very similar to each other in terms of functions, as shown in Figure III.2B. The bioactivity of the biocenosis was studied through a metaproteomic analysis conducted on solar panels sampled at noon (solar time) and at night (4 AM). Protein composition differed between the day and night samples (Figure III.2C). Significantly, a protein involved in modulating bacterial growth on surfaces and biofilm formation -diguanylate cyclase- (Römling *et al*., 2013) was particularly abundant. Also among the more expressed proteins, we identified fungal and bacterial enzymes involved in respiration and ATP synthesis or ribosomal proteins (bacterial L7/L12 and archaeal L7, the latter being a moonlighting protein involved in rRNA processing (Diaconu *et al*., 2005). Other abundant proteins have been reported to confer resistance/tolerance to the extreme conditions found in solar panels, namely, salt stress and drought (membrane-bound proton-translocating pyrophosphatase mPP; Baykov *et al*., 2013), nutrient starvation (mPP and cold-shock protein; Etchegaray *et al*., 1999), heat-shock (molecular chaperone GroEL), as well as proteins involved in the preservation of membrane integrity under harsh conditions (S-layer protein and lipoprotein 1; Cascales *et al*., 2002; Rothfuss, 2006).

Discussion

Despite the harsh conditions to which microorganisms deposited -or permanently inhabiting- the solar panels of a Mediterranean city during summer, standard culturing resulting in important microbial growth, with a diversity in shape, color and textures of colony forming microorganisms from the 2013 solstice suggesting high biodiversity of the environment. Although culturable isolates are typically only a small fraction of the global biocenosis, we were able to identify several strain-to-strain effects that proved able to restore the sensitivity of neighboring isolates to stress factors (salinity and low pH). These results suggest that microbial interactions and the particular physical location of microorganisms on the solar panels, rather than individual cell properties, might play a major role in bacterial survival on solar panels.

High throughput sequencing allowed confirming the high diversity of the habitat

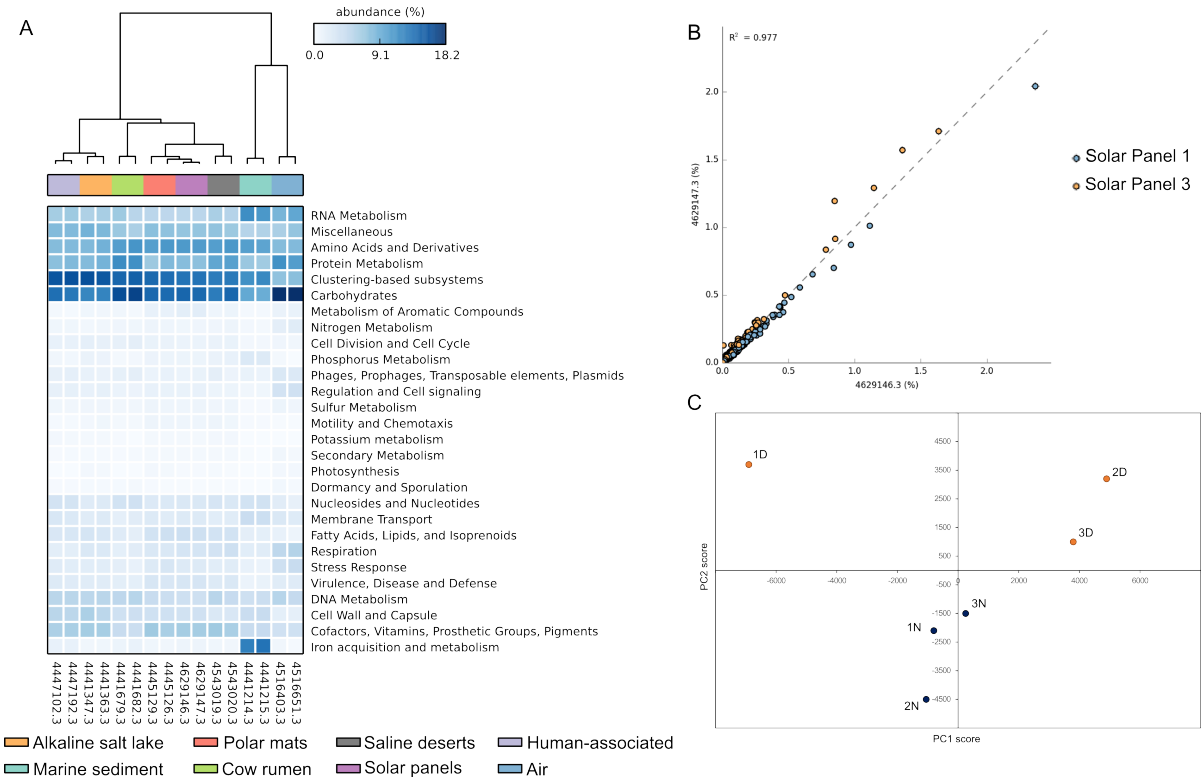


FIGURE III. 2: Functional analysis of the solar-panel metagenomes. Heatmap representation (A) based on the functional profiles of solar panels 1 and 3 compared with a range of metagenomes from different environments. Comparison of the set of functions found in the metagenomes of solar panels 1 and 3 (B). Each dot corresponds to one function of the subsystems classification. Principal Component Analysis (C) performed with the proteomic profile of solar panels sampled at noon (yellow dots) and night (dark blue).

in the form of a sun-adapted taxonomic profile. Indeed, and in accordance with the majoritary phenotype observed within culturable isolates, most of the species identified by high throughput sequencing, and particularly the most frequent ones, are known to produce pink (*H. xingiangensis*, Zhang *et al.*, 2007; *H. psychrotolerans*, Zhang *et al.*, 2008), orange (*Sphingomonas humi*, Yi *et al.*, 2010), orange-red (*S. kaistensis*, Kim *et al.*, 2007) or reddish pigments (*Hymenobacter chitinivorans*, Buczolits *et al.*, 2006; *Rubellimicrobium mesophilum*, Dastager *et al.*, 2008), in most cases carotenoids; as well as sphingolipids (*Sphingomonas* spp., Yi *et al.*, 2010, Kim *et al.*, 2007; *Novosphingobium* spp., Fujii *et al.*, 2003). Carotenoids have been reported to play a major role in radiation tolerance in bacteria (Tian and Hua, 2010) and sphingolipids have recently been described to mediate bacteria-to-silica and –polyamide adhesion (Gutman *et al.*, 2014). Therefore, carotenoids and sphingolipids are candidates accounting for the sunlight resistance and fixation properties that microorganisms need to survive on a smooth, south-facing surface.

A review of the ecology of the main bacterial taxa we identified gives more insights of the extremophile character of the solar panel bacteriome. Indeed, several of the most frequent *Deinococcus* spp. and other solar-panel bacteria have been described as inhabitants of relatively mild desertic areas as well as polar environments. *D. hopiensis* was isolated from the Sonora desert (Rainey et al., 2005), while other *Deinococcus* species that we detected were first described in the Sahara desert (de Groot *et al.*, 2005). Regarding other very abundant species, such as *Hymenobacter xingiangensis* or *Sphingomonas kaistensis*, they have previously been reported on the high Tibet plateau (Zhang *et al.*, 2008), on dry Antarctic valleys (Hirsch *et al.*, 1998), or in the Chinese desert of Xingiang

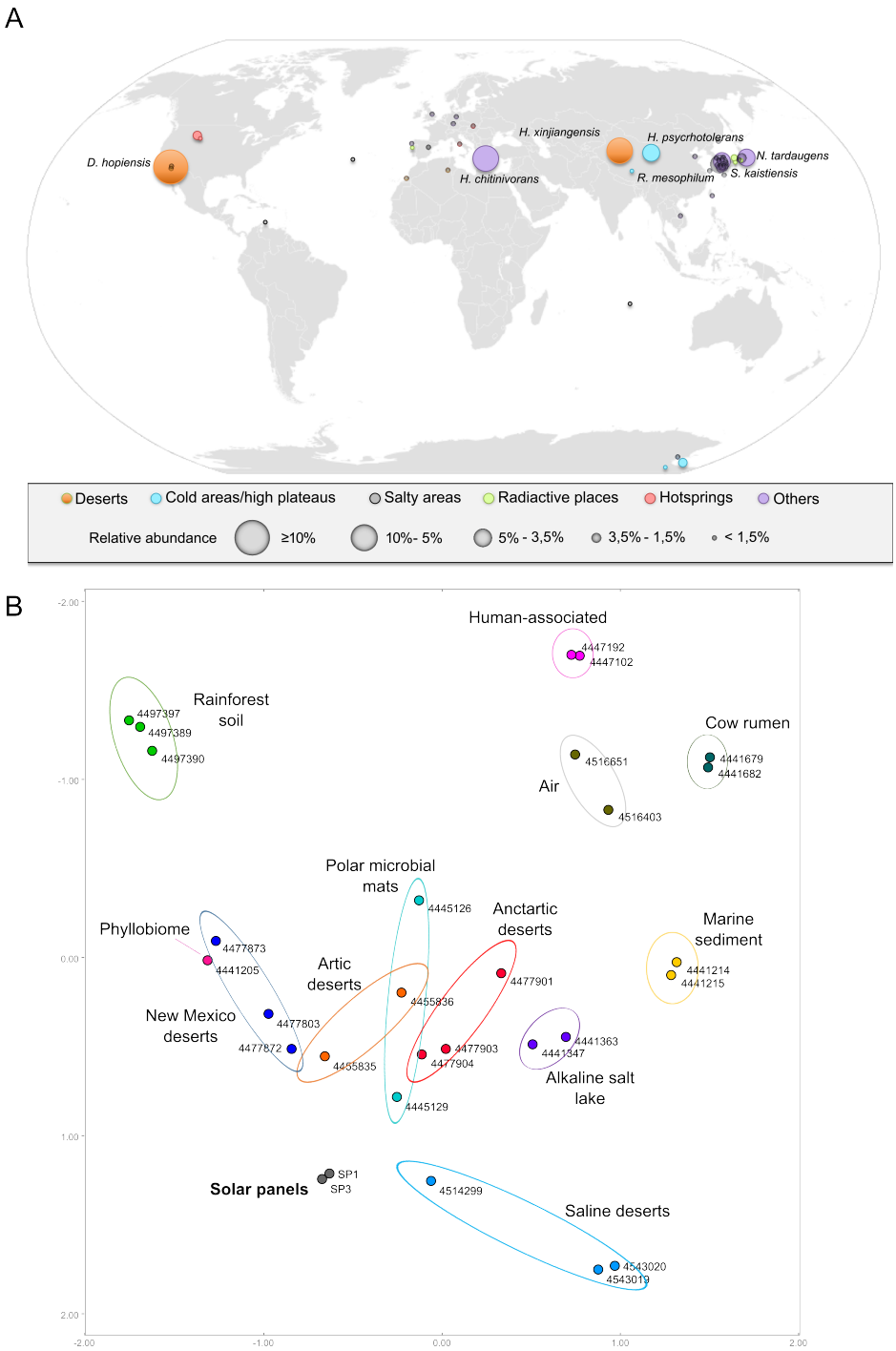


FIGURE III. 3: Biogeographical context of the solar-panel microbiomes as deduced from their taxonomic profile. Geographic distribution (A) of the 50 most abundant (more than 1% of the reads) bacterial species detected by high-throughput sequencing of 16S rRNA amplicons in solar panels sampled at the University of Valencia in 2014. Each circle corresponds to a different species and the size of the circles is proportional to the number of reads. Species found at a frequency higher than 3.5% are shown. Colors indicate type of environment. Principal Coordinates Analysis (B) performed with the taxonomic profile of a range of metagenomes from diverse ecosystems. The solar-panel metagenomes (panels 1 and 3 from the 2014 sampling, grey dots) map within desert and circumpolar metagenomes.

(Zhang *et al*., 2007). Others were first reported in high salinity areas (Nübel *et al*., 2000), thermal springs (Alarico *et al*., 2002) or during bioprospections of soil (Yoon *et al*., 2008; Pradhan *et al*., 2010), or air samples (Yang *et al*., 2009; Yoo *et al*., 2008).

A systematic review of the locations where the 50 most abundant solar panel bacteria were first isolated reveals their adaptation to extreme environments: most of them occur in drought, high radiation and/or high temperature habitats. Most of the species we found on solar panels were originally reported to inhabit a relatively narrow geographical band in the temperate zone of the Northern hemisphere. The distribution of others in the dry Antarctic valleys suggests a major role of radiation as a key selective factor (Figure III.3A), and the PCoA analysis of the solar-panel taxonomic profile compared with other metagenomes reveals a clear link with extremophile environments, such as temperate and cold deserts (Figure III.3B). Regarding fungi, the ecological niches of several of the most frequent genera (*Neocatenulostroma*, *Xenophacidiella* and *Metschnikowia*) are sunny habitats, such as the phylloplane (Crouss *et al*., 2008; Sláviková *et al*., 2007) or the surface of rocks (*Coniosporium* spp., particularly abundant in the 2013 samples; Zakharova *et al*., 2013). As in the case of bacteria, this taxonomic profile strongly suggests sunlight exerts a major selective pressure, shaping the fungal community on the panels.

The abundance and diversity of microorganisms in solar panels can be solely due to wind-deposition or correspond to an *in situ* active ecological community. Protein composition differed significantly between the day and night samples (Figure III.2C), implying that the microbial communities populating the solar panel surface are biologically active. The abundance of proteins involved in resistance to harsh conditions and biofilm formation on surfaces proves the presence of stress-response mechanisms in the microbial communities inhabiting solar panels. These results, along with the abundance of radiation-resistant taxa and the desert-like taxonomic profile of the solar samples, that strikingly plot within desert microbiomes (Figure III.3B), strongly suggest an *in situ* adaptation from (probably) wind-transported microorganisms that are immediately subjected to selection with radiation, heat and dessication as main shapers of this microbial ecosystem. Indeed, the solar panels microbiome proved taxonomically very distant from that associated to air samples from similar latitude (see metagenomes 4516651.3 and 4516403.3 in Figure III.3B). Furthermore, a recent analysis of air samples in Sardinia (Rosselli *et al*., 2015), a known Mediterranean crossroad of dust-conveying winds from Saharan Africa, revealed an extremely low abundance of extremophiles such as Deinococcus-Thermus species (< 1% of relative abundance, in contrast to up to 30% in solar panels). Taken together, all these results strongly suggest that the diverse biocenosis on solar panels we report here is not a mere consequence of physical accumulation of air-driven microorganisms, but a resident microbial community adapted to desert-like selection pressures.

This is the first report of a highly diverse microbial community on solar panels. A recently published study reported limited microbial diversity on solar panels in Brazil, including some fungal species which hindered the panel's photovoltaic efficiency (Shirakawa *et al*., 2015). Our data show for the first time that solar panels of a temperate Mediterranean city support a highly diverse and active ecological community, one of the richest extremophile biocenoses described to date. Moreover, this community is metabolically active and displays striking taxonomic and functional similarities with highly irradiated environments: temperate deserts and polar environments. The detailed analysis of the habitats where the solar panel microorganisms have previously been detected indicates their strong adaptation to sun exposure, which can only be partially reproduced by stress characterization on pure microbial cultures. Microbial interactions (including pH and salinity tolerance restoration), physical effects such as shading, DNA repair mechanisms and production of pigments and adhesion molecules might play a major role in the adaptation of a unique microbial ecosystem to the abrupt circadian cycles in desert-like conditions. This previously undescribed ecosystem is the first urban microdesert reported to date, and it may provide a valuable new source of compounds with biotechnological applications.

Acknowledgements

Data reported in this paper have been deposited in the MG-RAST server, and are publicly available under accession numbers 4629146.3 and 4629147.3. The authors are indebted to the Vicerrectorat d'Infraestructures (University of Valencia) for their support before and during sampling. Biópolis SL funded this work. C.V. is a recipient of a FPU (Formación de Personal Universitario) grant from the Spanish MECD (Ministerio de Educación, Cultura y Deporte). We also thank the proteomics service (SCSIE) of the University of Valencia for its technical support and Fabiola Barraclough for English revision. P.D. and C.V. contributed equally to this work.

References

Adams RI, Bateman AC, Bik HM, Meadow JF. 2015. Microbiota of the indoor environment: a meta-analysis. *Microbiome* **3**:49.

Afshinnekoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, *et al*. 2015. Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics. *Cell Syst* **1**:72–87.

Alarico S, Rainey FA, Empadinhas N, Schumann P, Nobre MF, da Costa MS. 2002. Rubritepida flocculans gen. nov., sp. nov., a new slightly thermophilic member of the alpha-1 subclass of the Proteobacteria. *Syst Appl Microbiol* **25**:198–206.

Ames SK, Gardner SN, Marti JM, Slezak TR, Gokhale MB, Allen JE. 2015. Using populations of human and microbial genomes for organism detection in metagenomes. *Genome Res* gr.184879.114–.

Baykov AA, Malinen AM, Luoto HH, Lahti R. 2013. Pyrophosphate-fueled Na+ and H+ transport in prokaryotes. *Microbiol Mol Biol Rev* **77**:267–76.

Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* **29**:2607–18.

Buczolits S, Denner EBM, Kämpfer P, Busse H-J. 2006. Proposal of Hymenobacter norwichensis sp. nov., classification of 'Taxeobacter ocellatus', 'Taxeobacter gelupurpurascens' and 'Taxeobacter chitinovorans' as Hymenobacter ocellatus sp. nov., Hymenobacter gelipurpurascens sp. nov. and Hymenobacter chitinivo. *Int J Syst Evol Microbiol* **56**:2071–8.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**:119–20.

Cascales E, Bernadac A, Gavioli M, Lazzaroni J-C, Lloubes R. 2002. Pal Lipoprotein of Escherichia coli Plays a Major Role in Outer Membrane Integrity. *J Bacteriol* **184**:754–759.

Corliss JB, Dymond J, Gordon LI, Edmond JM, von Herzen RP, Ballard RD, *et al*. 1979. Submarine thermal sprirngs on the galapagos rift. *Science* **203**:1073–83.

Crous PW, Summerell BA, Mostert L, Groenewald JZ. 2008. Host specificity and speciation of Mycosphaerella and Teratosphaeria species associated with leaf spots of Proteaceae. *Persoonia* **20**:59–86.

Dastager SG, Lee J-C, Ju Y-J, Park D-J, Kim C-J. 2008. Rubellimicrobium mesophilum sp. nov., a mesophilic, pigmented bacterium isolated from soil. *Int J Syst Evol Microbiol* **58**:1797–800.

de Groot A, Chapon V, Servant P, Christen R, Saux MF-L, Sommer S, *et al*. 2005. Deinococcus deserti sp. nov., a gamma-radiation-tolerant bacterium isolated from the Sahara Desert. *Int J Syst Evol Microbiol* **55**:2441–6.

Diaconu M, Kothe U, Schlünzen F, Fischer N, Harms JM, Tonevitsky AG, *et al*. 2005. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**:991–1004.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**:2194–200.

Etchegaray JP, Inouye M. 1999. CspA, CspB, and CspG, major cold shock proteins of Escherichia coli, are induced at low temperature under conditions that completely block protein synthesis. *J Bacteriol* **181**:1827–30.

Fujii K, Satomi M, Morita N, Motomura T, Tanaka T, Kikuchi S. 2003. Novosphingobium tardaugens sp. nov., an oestradiol-degrading bacterium isolated from activated sludge of a

sewage treatment plant in Tokyo. *Int J Syst Evol Microbiol* **53**:47–52.

Gutman J, Kaufman Y, Kawahara K, Walker SL, Freger V, Herzberg M. 2014. Interactions of glycosphingolipids and lipopolysaccharides with silica and polyamide surfaces: adsorption and viscoelastic properties. *Biomacromolecules* **15**:2128–37.

Hirsch P, Ludwig W, Hethke C, Sittig M, Hoffmann B, Gallikowski CA. 1998. Hymenobacter roseosalivarius gen. nov., sp. nov. from continental Antartica soils and sandstone: bacteria of the Cytophaga/Flavobacterium/Bacteroides line of phylogenetic descent. *Syst Appl Microbiol* **21**:374–383.

Kim MK, Schubert K, Im W-T, Kim K-H, Lee S-T, Overmann J. 2007. Sphingomonas kaistensis sp. nov., a novel alphaproteobacterium containing pufLM genes. *Int J Syst Evol Microbiol* **57**:1527–34.

La Duc MT, Vaishampayan P, Nilsson HR, Torok T, Venkateswaran K. 2012. Pyrosequencing-derived bacterial, archaeal, and fungal diversity of spacecraft hardware destined for Mars. *Appl Environ Microbiol* **78**:5912–22.

Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**:3100–8.

Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, *et al*. 2014. Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* **80**:1048–1052.

Leung MHY, Wilkins D, Li EKT, Kong FKF, Lee PKH. 2014. Indoor-air microbiome in an urban subway network: diversity and dynamics. *Appl Environ Microbiol* **80**:6760–70.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**:955–64.

Mahnert A, Moissl-Eichinger C, Berg G. 2015. Microbiome interplay: plants alter microbial abundance and diversity within the built environment. *Front Microbiol* **6**:887.

Martin LJ, Adams RI, Bateman A, Bik HM, Hawks J, Hird SM, *et al*. 2015. Evolution of the indoor biome. *Trends Ecol Evol* **30**:223–32.

Neilson JW, Quade J, Ortiz M, Nelson WM, Legatzki A, Tian F, *et al*. 2012. Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles* **16**:553–66.

Nübel U, Garcia-Pichel F, Muyzer G. 2000. The halotolerance and phylogeny of cyanobacteria with tightly coiled trichomes Spirulina Turpin and the description of Halospirulina tapeticola gen. nov., sp. nov. *Int J Syst Evol Microbiol* **50**:1265–77.

Ortiz M, Legatzki A, Neilson JW, Fryslie B, Nelson WM, Wing RA, *et al*. 2014. Making a living while starving in the dark: metagenomic insights into the energy dynamics of a carbonate cave. *ISME J* **8**:478–91.

Pham N TA, Anonye BO. 2014. Vying over spilt oil. Nat Rev Microbiol 12:156. Pradhan S, Srinivas TNR, Pindi PK, Kishore KH, Begum Z, Singh PK, *et al*. 2010. Bacterial biodiversity from Roopkund Glacier, Himalayan mountain ranges, India. *Extremophiles* **14**:377–95.

Rainey FA, Ray K, Ferreira M, Gatz BZ, Nobre MF, Bagaley D, *et al*. 2005. Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus Deinococcus obtained from a single soil sample. *Appl Environ Microbiol* **71**:5225–35.

Robertson CE, Baumgartner LK, Harris JK, Peterson KL, Stevens MJ, Frank DN, *et al*. 2013. Culture-independent analysis of aerosol microbiology in a metropolitan subway system. *Appl Environ Microbiol* **79**:3485–93.

Römling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev* **77**:1–52.

Rosche B, Li XZ, Hauer B, Schmid A, Buehler K. 2009. Microbial biofilms: a concept for industrial catalysis? *Trends Biotechnol* **27**:636–43.

Rosselli R, Fiamma M, Deligios M, Pintus G, Pellizzaro G, Canu A, *et al*. 2015. Microbial immigration across the Mediterranean via airborne dust. *Sci Rep* **5**:16306.

Rothfuss H. 2006. Involvement of the S-layer proteins Hpi and SlpA in the maintenance of cell envelope integrity in Deinococcus radiodurans R1. *Microbiology* **152**:2779–2787.

Shirakawa MA, Zilles R, Mocelin A, Gaylarde

CC, Gorbushina A, Heidrich G, *et al*. 2015. Microbial colonization affects the efficiency of photovoltaic panels in a tropical environment. *J Environ Manage* **157**:160–167.

Sim K, Cox MJ, Wopereis H, Martin R, Knol J, Li M-S, *et al*. 2012. Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing. *PLoS One* **7**:e32543.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res* **19**:1117–23.

Singh A, Lal R. 2009. Sphingobium ummariense sp. nov., a hexachlorocyclohexane HCH-degrading bacterium, isolated from HCH-contaminated soil. *Int J Syst Evol Microbiol* **59**:162–6.

Sláviková E, Vadkertiová R, Vránová D. 2007. Yeasts colonizing the leaf surfaces. *J Basic Microbiol* **47**:344–50.

Tian B, Hua Y. 2010. Carotenoid biosynthesis in extremophilic Deinococcus-Thermus bacteria. *Trends Microbiol* **18**:512–20.

Williamson AJ, Morris K, Law GTW, Rizoulis A, Charnock JM, Lloyd JR. 2014. Microbial reduction of UVI under alkaline conditions: implications for radioactive waste geodisposal. *Environ Sci Technol* **48**:13549–56.

Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* **12**:444.

Yang Y, Itoh T, Yokobori S, Itahashi S, Shimada H, Satoh K, *et al*. 2009. Deinococcus aerius sp. nov., isolated from the high atmosphere. *Int J Syst Evol Microbiol* **59**:1862–6.

Yi TH, Han C-K, Srinivasan S, Lee KJ, Kim MK. 2010. Sphingomonas humi sp. nov., isolated from soil. *J Microbiol* **48**:165–9.

Yoo S-H, Weon H-Y, Noh H-J, Hong S-B, Lee C-M, Kim B-Y, *et al*. 2008. Roseomonas aerilata sp. nov., isolated from an air sample. *Int J Syst Evol Microbiol* **58**:1482–5.

Yoon J-H, Kang S-J, Lee S-Y, Oh T-K. 2008. Sphingomonas insulae sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **58**:231–6.

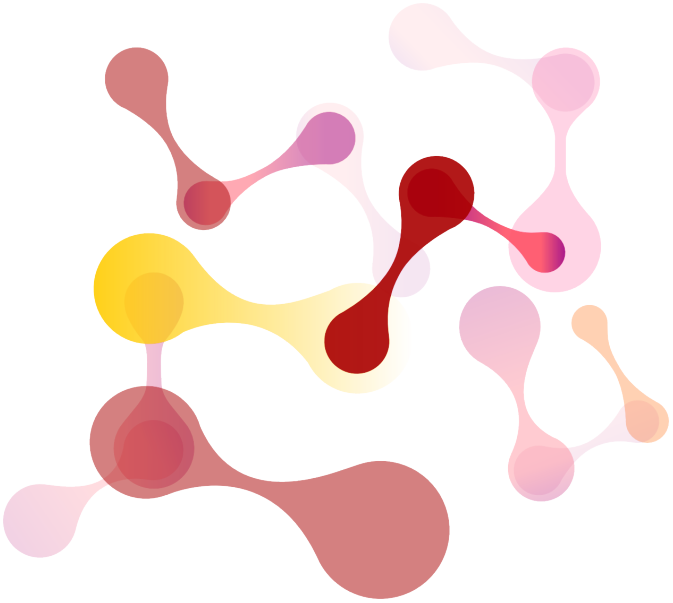
Zakharova K, Tesei D, Marzban G, Dijksterhuis J, Wyatt T, Sterflinger K. 2013. Microcolonial fungi on rocks: a life in constant drought? *Mycopathologia* **175**:537–47.

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

Zhang G, Niu F, Busse H-J, Ma X, Liu W, Dong M, *et al*. 2008. Hymenobacter psychrotolerans sp. nov., isolated from the Qinghai-Tibet Plateau permafrost region. *Int J Syst Evol Microbiol* **58**:1215–20.

Zhang Q, Liu C, Tang Y, Zhou G, Shen P, Fang C, *et al*. 2007. Hymenobacter xinjiangensis sp. nov., a radiation-resistant bacterium isolated from the desert of Xinjiang, China. *Int J Syst Evol Microbiol* **57**:1752–6.

Chapter 2. Beyond organisms: identification and isolation of microbial consortia from environmental samples



Publication IV:
Dorado-Morales P, Vilanova C, P Garay C, Martí JM, Porcar M. (2015). Unveiling bacterial interactions through multidimensional Scaling and dynamics modeling. *Sci Rep* **5**:18396.

Publication V:
Vilanova C, Iglesias A, Porcar M. (2015). The coffee-machine bacteriome: biodiversity and colonisation of the wasted coffee tray leach. *Sci Rep* **5**:17163.

Publication IV

Unveiling bacterial interactions through multi-dimensional scaling and dynamics modeling

Pedro Dorado-Morales[1], Cristina Vilanova[1], Carlos P. Garay[2], José Manuel Martí[2] and Manuel Porcar[1,3]

[1] Cavanilles Institute of Biodiversity and Evolutionary Biology (Universitat de València), Spain.
[2] Instituto de Física Corpuscular, CSIC-UVEG, Valencia, Spain.
[3] Fundació General de la Universitat de València, Spain.

Abstract

We propose a new strategy to identify and visualize bacterial consortia by conducting replicated culturing of environmental samples coupled with high-throughput sequencing and multidimensional scaling analysis, followed by identification of bacteria-bacteria correlations and interactions. We conducted a proof of concept assay with pine-tree resin-based media in ten replicates, which allowed detecting and visualizing dynamical bacterial associations in the form of statistically significant and yet biologically relevant bacterial consortia.

Introduction

There is a growing interest on disentangling the complexity of microbial interactions in order to both optimize reactions performed by natural consortia and to pave the way towards the development of synthetic consortia with improved biotechnological properties (Brenner *et al.*, 2008; Mee and Wang, 2012). Despite the enormous amount of metagenomic data on both natural and artificial microbial ecosystems, bacterial consortia are not necessarily deduced from those data. In fact, the flexibility of the bacterial interactions, the lack of replicated assays and/or biases associated with different DNA isolation technologies and taxonomic bioinformatics tools hamper the clear identification of bacterial consortia. We propose here a holistic approach aiming at identifying bacterial interactions in laboratory-selected microbial complex cultures. The method requires multi-replicated taxonomic data on independent subcultures, and high-throughput sequencing-based taxonomic data. From this data matrix, randomness of replicates can be verified, linear correlations can be visualized and interactions can emerge from statistical correlations. The whole procedure can be summarized as follows:

1. Taxonomic data from multi-replicated, independent assays is obtained.
2. Fluctuation scaling of replicates, i.e. slope of Taylor's law, is validated for the samples of the first time step against the expectation of a Poisson-distributed selection.
3. Linear correlation coefficients are determined, converted into distances and displayed by multidimensional scaling.
4. Interaction matrix is inferred from the correlation matrix using the discrete Lotka-Volterra model with relative abundances (Fisher and Mehta, 2014).

As a proof of concept, we chose to analyze independent laboratory cultures grown on a natural, recalcitrant compound. A range of recalcitrant substrates, from synthetic dyes to polycyclic aromatic hydrocarbons,

polychlorinated biphenyls, and other organic pollutants can be efficiently degraded by mixed microbial cultures combining catabolic enzyme activities of individual consortium members (Mikesková *et al.*, 2012). Therefore, a carbon source that requires complex pathways for degradation is expected to shape the structure of the microbial community and behave as a strong selection pressure towards the establishment of microbial consortia with biodegradation properties. In a previous work, we characterized the cultivable microbial communities present in coniferous resin, and detected a rather diverse microbial community, including several fungal and bacterial strains with potential use in bioremediation as deduced from their ability for the degradation of different terpenic compounds (Vilanova *et al.*, 2014). In the proof of concept we present here, we used pine-tree resin as the main carbon source of a resin-rich semi-synthetic medium (prepared as described in our previous report), which we inoculated with environmental resin in ten independent subcultures. We designed this multi-replicate experimental evolution assay to address three main issues: i) Time-course variation in biodiversity fate (does sub-culturing in a recalcitrant compound lead to taxonomic impoverishment?); ii) reproducibility of the selection process (do the replicates behave as such?); and iii) strain-to-strain bacterial interactions: can microbial interactions be deduced from a close analysis of the biodiversity dataset?

Materials and Methods

Culture medium and growth conditions. A pine-tree resin-based medium previously described (Vilanova *et al.*, 2014) containing 0.1% (w/v) of resin as the main carbon source was used for microbial culturing. A starter culture of environmental resin in the resin-rich semi-synthetic medium was set up, and ten 5 mL-aliquots were taken from it and incubated independently for 4 days (time series 1, t_1). Then, a 50 μL aliquot of each culture was inoculated in a new set of ten independent tubes containing fresh medium, and incubated for another 4 days (t_2). New 50 μL aliquots were taken from the second round of cultivation, and the whole process was repeated a total of 9 times. Since previous observations in the laboratory proved that few sub-culturing steps resulted in an accelerated microbial growth (resin colloids were consumed faster), the time between sub-cultures was shortened throughout the experiment from 4 to 2 days in order to promote the selection of efficient resin-degrading consortia. Tubes were grown for 4 days in t_1-t_2 and t_2-t_3; 3 days in t_3-t_4; and 2 days in t_4-t_5, t_5-t_6, t_6-t_7, and t_8-t_9. Three days of incubation in t_7-t_8 were applied to compensate an unexpected decrease (from 30 to 28 ºC) in the temperature of the incubator. The complete assay lasted 24 days, and aliquots were taken at every sub-culturing step to obtain a total of 90 (10 replicated cultures x 9 sub-culturing steps) culture samples.

DNA isolation and quantification. A two mL aliquot of each culture was sampled and resin colloids from the culture medium were pelleted by mild centrifugation (800 g, 5 min). The supernatant was transferred to a new tube, cells were harvested at 11,000 g for 3 min and washed twice with sterile PBS buffer (NaCl 8 g/L, KCl 0.2 g/L, Na_2HPO_4 1.44 g/L, KH_2PO_4 0.24 g/L, pH adjusted to 7.4). Then, DNA was isolated with a standard protocol consisting of alkaline lysis followed by precipitation with potassium acetate and isopropanol. The quality of the DNA was finally checked on a 0.8% (w/v) agarose gel and quantified with Nanodrop-1000 Spectophotometer (Thermo Scientific, Wilmington, DE).

PCR amplification. A 500 bp fragment of the V1-V3 hypervariable region of the 16S ribosomal RNA gene was PCR-amplified from all the samples with primers 28F (5'-GAG TTT GAT CNT GGC TCA G-3') and 519R (5'-GTN TTA CNG CGG CKG CTG-3'). A short (9-11 nucleotides) barcode sequence followed by a four-nucleotide spacer (CGAT) was included at the 5' end of the oligonucleotides used as forward primers to enable assignment of sequences to samples after high-throughput sequencing. All the amplifications were performed under the following thermal cycling conditions: initial denaturing at 95 ºC for 5 min, followed by 35 cycles of denaturing at 95 ºC for 30 s, annealing at 54 ºC for 30 s, and extension at 72 ºC for 1 min, finalized by a 10-min elongation at 72 ºC. The resulting amplicons were checked on a 0.8% (w/v) agarose gel and purified by precipitation with 3M potassium acetate (pH=5) and isopropanol. Pure

amplicons were quantified with the Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and an equimolar pool of amplicons was prepared from all the samples.

Sequencing. A sequencing library was constructed with 100 ng of the pool by amplicon fusion (Ion Plus Fragment Library Kit, MAN0006846, Life Technologies). The library was quantified with the Agilent 2100 Bioanalizer (Agilent Technologies Inc, Palo Alto, California) prior to clonal amplification. Emulsion PCRs were carried out applying the Ion PGM Template OT2 400 kit as described in the user-guide (MAN0007218, Revision 3.0 Lifetechnologies) provided by the manufacturer. Finally, the library was sequenced in an Ion 318 Chip v2 on a Personal Genome Machine (PGM IonTorrentTM, Lifetechnologies) at Lifesequencing S.L (Lifesequencing, Valencia, Spain), using the Ion PGM Sequencing 400 kit following the manufacturer's protocol (Publication Number MAN0007242, Revision 2.0, Lifetechnologies).

Bioinformatic analysis. Raw sequences were filtered to remove short (<200 bp) and low quality (<Q10) reads with the NextGENe® Software for Ion Torrent PGM^{TM} System, and the resulting sequences were processed with the QIIME package (Caporaso *et al.*, 2010) according to the following pipeline. First, a mapping file was generated by assigning reads to samples according to barcode sequences (allowing a maximum of 1 mismatch in primer search and 1 mismatch in barcode search). Second, an OTU table was generated with the uclust algorithm, using a similarity threshold of 0.97, corresponding to species-level OTUs. Then, representative sequences were picked from each OTU and classified with the BLAST algorithm (e-value<1e-05) against the 16S Greengenes database (version 13.8). Finally, an OTU table containing the taxonomical identifications and the absolute abundance of each OTU in every sample was built. In order to determine whether microbial composition significantly changed among replica and/or with time (variation between sub-culturing steps and/or variation of the time between these steps), a matrix of Bray-Curtis beta-diversity dissimilarities was calculated with QIIME, and an ANOSIM test with 999 permutations was carried out for each hypothesis.

Biostatistical analysis: multi-dimensional scaling. The OTU table containing the frequency of each OTU in every sample was used to compute the linear correlation matrix of the fluctuations observed in the 9 time series with the SparCC software (Friedman and Alm, 2012). Correlations were linearly converted into positive distances, so that positive full correlation was d = 0 and negative full correlation was d = 2. A matrix of distances, if all are precisely known, can be uniquely converted into positions up to dimension, translation and rotational symmetry. In our study, we chose non-metric multidimensional scaling in the MATLAB® function *mdscale*. We devised using three dimensions (3D) in the plot as a requirement to minimize 2D image ambiguity effects in the visualization of correlated, uncorrelated and anticorrelated taxa.

Biostatistical analysis: interactions. For a determined taxonomical level, the discrete time Lotka-Volterra model (dLV) relates the abundance of taxon i at an arbitrary future time $t + \delta t$ to the abundances of all the taxa at the present time t. The interactions (Hofbauer *et al.*, 1987) are described by interaction coefficients c_{ij} that describe the influence of taxon j on the abundance of taxon i. If the number of taxons is large, the dynamics of relative abundances can be described by a modified dLV model generalized to include stochasticity, like environmental and demographic stochastic effects (Equation 1):

$$x_i(t+\delta t) \approx \eta_i(t) x_i(t) exp(\delta t \Sigma c_{ij}(x_j(t) - \langle x_j \rangle))$$

where $\eta_i(t)$ is a log-normally distributed multiplicative noise, x_i is the relative abundance of taxon i and x_j is the equilibrium abundance of taxon j and is set by the carrying capacity of the environment. In this problem, the interaction coefficients are known up to an arbitrary multiplicative constant and the design matrix is singular. We inferred the interaction coefficients with the LIMITS algorithm (Fisher and Mehta, 2014).

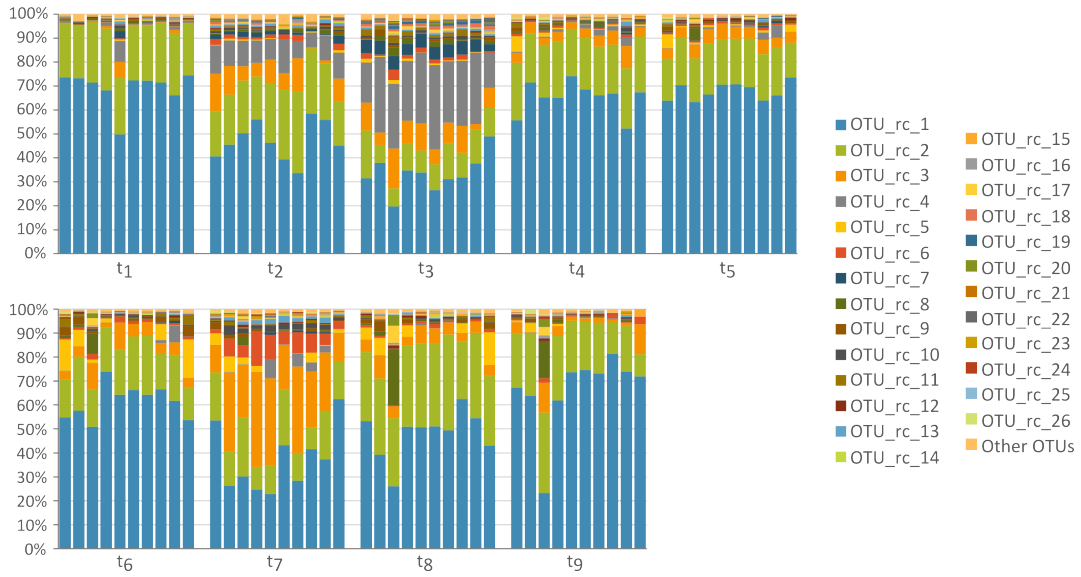


FIGURE IV. 1: Bacterial biodiversity as deduced from 16S amplicon high-throughput sequencing of ten replicated culture lines subcultured in parallel nine times (t_1 to t_9). A pool of pine-tree resin samples was used to inoculate ten independent minimum broth media supplemented with resin as the main carbon source. Sub-culturing was carried out after 4 (t_1-t_2, t_2-t_3), 3 (t_3-t_4, t_7-t_8), or 2 days (t_4-t_5, t_5-t_6, t_6-t_7, t_8-t_9) and the complete assay lasted 24 days. Each bar represents an independent experiment at a given time. The taxonomic assignation of each OTU is available in Table IV.1.

Results and Discussion

16S profiles-based monitoring of bacterial populations throughout the experiment

Ten replicated cultures of resin samples in resin-rich semisynthetic medium were set up and independently subcultured (nine times) during 24 days, as described in the Methods section. Rather than within replicates, the taxonomic profiles displayed statistically significant variations within time steps (p-value=0.01, see Methods). Indeed, significant alterations (p-value=0.02) were found at times 2, 3 and 7, in correlation with changes in the number of days between subculturing events (Figure IV.1).

For nearly all time points and replicates, *Pseudomonas* sp. (OTU_rc_1) followed by an enterobacterium (OTU_rc_2) were the most frequent species, being overwhelmingly present at subculturing steps 1, 4, 5, 6, 8, and 9. A species belonging to the Xanthomonadaceae family (OTU_rc_3) and *Comamonas terrigena* (OTU_rc_4) were also found at relatively high rates, particularly at subculturing steps 3 (*C. terrigena*), 7 (Xanthomonadaceae species) and 2 (both). Table IV.1 shows the taxonomical assignations for the 26 most abundant OTUs. Twenty-two other relatively frequent OTUs were detected at a comparatively lower, but still significant frequency (more than 100 counts), and as many as 87 different OTUs could be detected in total.

In order to find out whether sub-culturing was associated with a change in sample biodiversity, Shannon and Simpson indexes, as well as Species Richness were calculated for each sample. Richness index was virtually constant throughout the whole experiment; whereas Shannon and Simpson indexes fluctuated in time: they increased between subculturing steps 1 and 3, then dropped, and then increased again until subculturing step 7, showing a slightly decreasing trend at the end of the experiment (Supplementary Figure IV.1). These fluctuations coincided with changes in the number of days between subculturing steps, suggesting, again, the influence of this factor in community composition.

Identification of microbial consortia

Components of the consortia were identified through a three-phase analysis of a selected pool of data consisting of correlated time-series of relative abundances for the 26 more frequent OTUs in our samples. We first characterized the randomness of our ten subcultures at the first sampled time. The fluctuation scaling of the Taylor law (log of the standard deviation versus the mean frequency of each OTU) of our samples was 0.60 $\pm$ 0.04 (Supplementary Figure IV.2), which is consistent with Poisson-distributed replica (0.5). Data of each replicated time series were analyzed both independently and together, as a mixture of all ten replica. The linear correlation coefficients among OTUs were calculated and the correlation matrix was converted into a distance matrix by shifting the correlation coefficients by two units (Friedman and Alm, 2012; Huse *et al.*, 2012). Multidimensional scaling (MDS) was used to display the positions of the OTUs in a given dimension, where distances are well preserved. The stress value of our MDS solution was 0.092, which falls between "fair" and "good" goodness of fit according to Kruskal's criterion (Kruskal, 1964). We show the distances among OTUs by statistical correlation in Figure IV.2, where the size of the spheres scale logarithmically with the relative abundance. We determined the statistical significance of the inferred correlations by performing permutation tests and also by simulating 10,000 experiments with the assumption that the counts have lognormal distribution (both approaches yielded similar results, data not shown). Three main consortia (OTU groups positively correlated) were observed: OTUs 3, 6, 10, 13, 23, and 24; OTUs 4, 7, 18, 19, and 22; and OTUs 8, 11, 17, and 20. The first association involved only gamma-proteobacteria (mainly, three

TABLE IV. 1: Taxonomical classification obtained for the OTUs analyzed in this study.

ID	Taxonomical classification				
	Class	**Order**	**Family**	**Genus**	**Species**
OTU_rc_1	γ[a]	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	-
OTU_rc_2	γ	Enterobacteriales	Enterobacteriaceae	-	-
OTU_rc_3	γ	Xanthomonadales	Xanthomonadaceae	-	-
OTU_rc_4	β[b]	Burkholderiales	Comamonadaceae	*Comamonas*	*terrigena*
OTU_rc_5	β	Burkholderiales	Alcaligenaceae	*Achromobacter*	-
OTU_rc_6	γ	Xanthomonadales	Xanthomonadaceae	*Stenotrophomonas*	*geniculata*
OTU_rc_7	β	Burkholderiales	Comamonadaceae	*Comamonas*	-
OTU_rc_8	γ	Pseudomonadales	Moraxellaceae	-	-
OTU_rc_9	β	Burkholderiales	Comamonadaceae	*Delftia*	-
OTU_rc_10	γ	Xanthomonadales	Xanthomonadaceae	*Pseudoxanthomonas*	*mexicana*
OTU_rc_11	β	Burkholderiales	Comamonadaceae	-	-
OTU_rc_12	γ	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	*stutzeri*
OTU_rc_13	γ	Xanthomonadales	Xanthomonadaceae	*Stenotrophomonas*	-
OTU_rc_14	β	Burkholderiales	Alcaligenaceae	-	-
OTU_rc_15	γ	Pseudomonadales	Pseudomonadaceae	-	-
OTU_rc_16	γ	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	*viridiflava*
OTU_rc_17	γ	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	*alcaligenes*
OTU_rc_18	β	Neisseriales	Neisseriaceae	-	-
OTU_rc_19	β	Rhodocyclales	Rhodocyclaceae	*Zoogloea*	-
OTU_rc_20	γ	Pseudomonadales	Moraxellaceae	*Acinetobacter*	-
OTU_rc_21	γ	Aeromonadales	Aeromonadaceae	-	-
OTU_rc_22	β	Burkholderiales	-	-	-
OTU_rc_23	γ	Xanthomonadales	Xanthomonadaceae	*Luteimonas*	-
OTU_rc_24	γ	Xanthomonadales	Xanthomonadaceae	*Stenotrophomonas*	*acidaminiphila*
OTU_rc_25	γ	Alteromonadales	Shewanellaceae	*Shewanella*	-
OTU_rc_26	S[c]	Sphingobacteriales	Sphingobacteriaceae	*Sphingobacterium*	*multivorum*

[a] γ: γ-proteobacteria
[b] β: β-proteobacteria
[c] S: Sphingobacteria

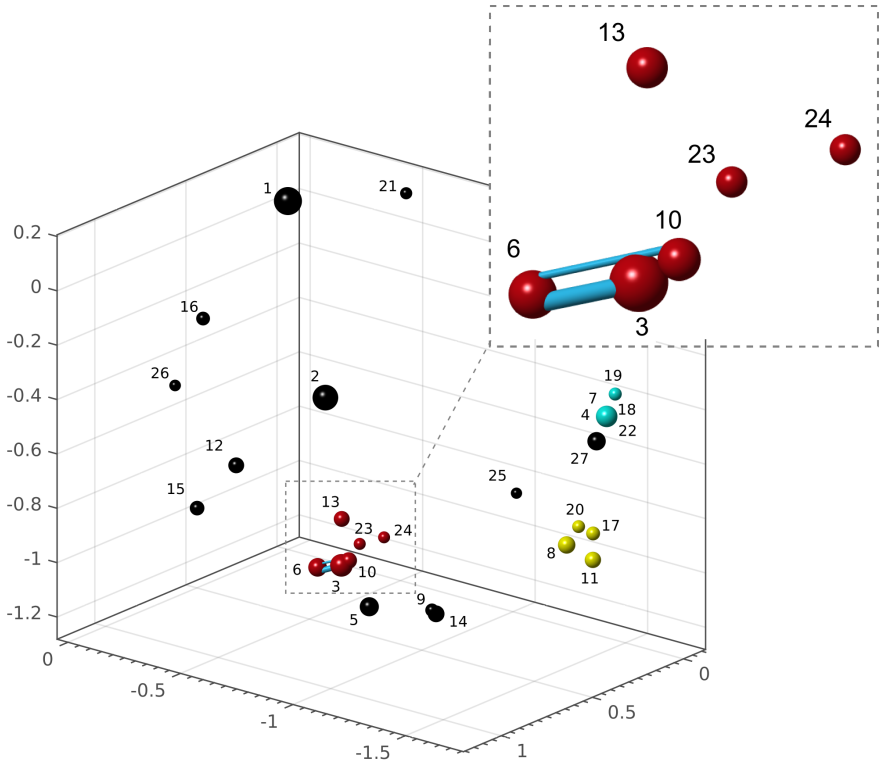


FIGURE IV. 2: Correlations and interactions among OTUs. Representation of the 3D positions of identified OTUs in our samples, where the distance corresponds to the amount of linear statistical correlation. Smaller distances mean stronger positive correlation and larger distances show stronger negative correlation. Sizes of the spheres are proportional to the relative abundances in logarithmic scale, while colors indicate putative bacterial consortia (OTU groups positively correlated). Blue bonds show the strength of significant positive interactions linking two OTUs, where the width of the bonds linearly scales with the relative strength of the interaction. Bonds linking OTUs 4-7 and 3-10 are hidden in this 3D perspective. The plotting code is available as Supplementary Data. The main consortia are colored as follows: yellow (*Pseudomonas alcaligenes*, *Acinetobacter* sp., *Moraxellaceae* OTU, *Comamonadaceae* OTU); blue (two species of the genus *Comamonas*, *Zooglea* sp., *Neisseriaceae* OTU, *Burkholderiales* OTU); and red (three species of the *Stenotrophomonas* genus, *Pseudoxanthomonas mexicana*, *Luteimonas* sp.). OTUs are numbered according to Table IV.1.

species of the *Stenotrophomonas* genus, and also *Pseudoxanthomonas mexicana*, and *Luteimonas* sp.); the second comprised only beta-proteobacteria (two species of the genus *Comamonas*, *Zooglea* sp., and two species belonging to the Neisseriaceae family and the Burkholderiales order, respectively); and, finally, the third one was composed of three gamma-proteobacteria (*Pseudomonas alcaligenes*, *Acinetobacter* sp., and a species of the Moraxellaceae family) and one beta-proteobacterium of the Comamonadaceae family. A significant negative correlation was found between OTU 12 (*Pseudomonas stutzeri*) and the consortia composed of OTUs 4, 7, 18, 19, and 22.

Statistical correlation does not imply interaction (Fisher and Mehta, 2014). We inferred the interaction matrix by using the discrete Lotka-Volterra model with relative abundances, a forward stepwise regression to include the strongest interactions and a bootstrap aggregation method to cure instabilities. Four

statistically significant, positive interactions (Figure IV.2) were found between OTUs 4 and 7 (*Comamonas terrigena* and *Comamonas* sp.); OTUs 3 and 10 (*Pseudoxanthomonas mexicana* and other Xanthomonadaceae species); OTUs 3 and 6 (the same Xanthomonadaceae species and *Stenotrophomonas geniculata*); and OTUs 6 and 10 (*S. geniculata* and *P. mexicana*). Interestingly, the three bacterial associations detected in this work include well known terpenoid degraders, particularly *Pseudomonas* (Molina *et al.*, 2013) and, to a lesser extent, *Comamonas* (Morgan and Wyndham, 2002) and *Stenotrophomonas* (Liao *et al.*, 2014). The resolution of the taxonomic assignations for the OTUs detected in this work poses difficulties in identifying putative metabolic complementation established among the members of each consortia. However, the associations unveiled with our analysis have been experimentally reported in previous studies. Particularly, *Pseudomonas*, *Comamonas*, and *Acinetobacter* species (the main taxa found in the third association) are known to form natural consortia able to degrade a range of recalcitrant compounds such as hydrocarbons (Pepi *et al.*, 2003), phenolic compounds (Prpich and Daugulis, 2005) and herbicides (Cordova-Rosa *et al.*, 2009). Interestingly, the degradation of these compounds involves similar steps (such as the cleavage of aromatic rings by dioxygenase enzymes; see Martin and Mohn, 2000) to those of the degradation pathway of the terpenes present in pinetree resin. *Pseudoxanthomonas* and *Stenotrophomas* species (detected in the first association and displaying strong, positive interaction) have also been reported as part of other consortia involved in the bioremediation of TNT-contaminated soils (Muter *et al.*, 2012). Further genomic-scale analyses are needed to better understand the relationships established at the molecular level among the members of the microbial consortia unveiled in this study. This work shows how microbial consortia can be visualized after a robust analysis of their relative frequencies in multireplicated cultures. The resulting plot can be very helpful to rapidly identify bacterial key-players and consortia in a wide range of bioprocesses. From the proof of concept of our method, we can conclude that unambiguous consortia can be detected within a pool of coexisting species on the basis not only of their relative abundance, but also by statistical correlation and verification of inferred biological interactions.

Acknowledgements

This work was supported by Generalitat Valencia Prometeo Grants II/2014/050, and by the Spanish Grants FPA2011-29678 of MINECO and by PITN- GA-2011-289442-INVISIBLES. CV was funded with a FPU (Formación de Personal Universitario) grant from the Spanish MECD (Ministerio de Educación, Cultura y Deporte), and JMM is a recipient of an FPI grant from MINECO (Ministerio de Economía y Competitividad. We are indebted to José Vicente Carrion (IFIC, CSIC) for continuous computing support. The sequences used in this study have been deposited in the SRA repository (BioProject ID: PRJNA296420).

References

Brenner K, You L, Arnold FH. 2008. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol* **26**:483–9.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**:335–6.

Cordova-Rosa SM, Dams RI, Cordova-Rosa E V, Radetski MR, Corrêa AXR, *et al.* 2009. Remediation of phenol-contaminated soil by a bacterial consortium and Acinetobacter calcoaceticus isolated from an industrial wastewater treatment plant. *J Hazard Mater* **164**:61–6.

Fisher CK, Mehta P. 2014. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* **9**:e102451.

Hofbauer J, Hutson V, Jansen W. 1987. Coexistence for systems governed by difference equations of Lotka-Volterra type. *J Math Biol* **25**:553–70.

Kruskal JB. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**:1–27.

Liao F, Wang Y, Huang Y, Mo Q, Tan H, *et al.* 2014. Isolation and identification of bacteria capable of degrading euptox A from Eupatorium adenophorum Spreng. *Toxicon* **77**:87–92.

Marrón-Montiel E, Ruiz-Ordaz N, Rubio-Granados C, Juárez-Ramírez C,

Galíndez-Mayer CJ. 2006. 2,4-D-degrading bacterial consortium. *Process Biochem* **41**:1521–1528.

Martin VJ, Mohn WW. 2000. Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by Pseudomonas abietaniphila BKME-9. *J Bacteriol* **182**:3784–93.

Mee MT, Wang HH. 2012. Engineering ecosystems and synthetic ecologies. *Mol Biosyst* **8**:2470–83.

Mikesková H, Novotný C, Svobodová K. 2012. Interspecific interactions in mixed microbial cultures in a biodegradation perspective. *Appl Microbiol Biotechnol* **95**:861–70.

Molina G, Pimentel MR, Pastore GM. 2013. Pseudomonas: a promising biocatalyst for the bioconversion of terpenes. *Appl Microbiol Biotechnol* **97**:1851–64.

Morgan CA, Wyndham RC. 2002. Characterization of tdt genes for the degradation of tricyclic diterpenes by Pseudomonas diterpeniphila A19-6a. *Can J Microbiol* **48**:49–59.

Muter O, Potapova K, Limane B, Sproge K, Jakobsone I, *et al.* 2012. The role of nutrients in the biodegradation of 2,4,6-trinitrotoluene in liquid and soil. *J Environ Manage* **98**:51–5.

Pepi M, Minacci A, Di Cello F, Baldi F, Fani R. 2003. Long-term analysis of diesel fuel consumption in a co-culture of Acinetobacter venetianus, Pseudomonas putida and Alcaligenes faecalis. *Antonie Van Leeuwenhoek* **83**:3–9.

Prpich GP, Daugulis AJ. 2005. Enhanced biodegradation of phenol by a microbial consortium in a solid-liquid two phase partitioning bioreactor. *Biodegradation* **16**:329–339.

Vilanova C, Marín M, Baixeras J, Latorre A, Porcar M. 2014. Selecting microbial strains from pine tree resin: biotechnological applications from a terpene world. *PLoS One* **9**:e100740.

Publication V

The coffee-machine bacteriome: biodiversity and colonisation of the wasted coffee tray leach

Cristina Vilanova[1], Alba Iglesias[1], and Manuel Porcar[1,2]

[1] Cavanilles Institute of Biodiversity and Evolutionary Biology (Universitat de València), Spain.
[2] Fundació General de la Universitat de València, Spain.

Abstract

Microbial communities are ubiquitous in both natural and artificial environments. However, microbial diversity is usually reduced under strong selection pressures, such as those present in habitats rich in recalcitrant or toxic compounds displaying antimicrobial properties. Caffeine is a natural alkaloid present in coffee, tea and soft drinks with well-known antibacterial properties. Here we present the first systematic analysis of coffee machine-associated bacteria. We sampled the coffee waste reservoir of ten different Nespresso machines and conducted a dynamic monitoring of the colonisation process in a new machine. Our results reveal the existence of a varied bacterial community in all the machines sampled, and a rapid colonisation process of the coffee leach. The community developed from a pioneering pool of enterobacteria and other opportunistic taxa to a mature but still highly variable microbiome rich in coffee-adapted bacteria. The bacterial communities described here, for the first time, are potential drivers of biotechnologically relevant processes including coffee decaffeination and bioremediation.

Introduction

Caffeine (1,3,7-trimethylxanthine) is a natural alkaloid with anti-herbivorous properties produced by *Coffea arabica* and *Coffea canephora*, which is present in a wide range of beverages including coffe, tea and soft –mainly cola- drinks. Caffeine is a well-known bioactive compound with stimulating effects on the central nervous system, along with a range of other potentially positive effects on human health. Such effects range from enhancing long-term memory (González de Mejía and Ramírez-Mares, 2014), improving sports performance (Spriet, 2014), inactivating breast cancer-associated myofibroblasts (Al-Ansari and Aboussekhra, 2014), reducing the risk of type 2 diabetes mellitus (Jiang *et al.*, 2014), or even reducing the risk of mortality among coffee consumers (Je and Giovannucci, 2014). However, it should also be noted that caffeine intake habits are often linked to living styles and thus it is difficult to draw epidemiological conclusions linking caffeine intake to health.

Caffeine may be an environmental pollutant (Mustard, 2014), and has also been proposed as an easily detectable marker for untreated wastewater (Buerge *et al.*, 2006). Indeed, the presence of caffeine in the natural environment is one of the best indicators of anthropogenic contamination. Caffeine bioactivity on human health and the environment has led to the development of processes to remove caffeine, either to yield decaffeinated products or to degrade environmental caffeine. Decaffeination, namely the caffeine removal process, is used industrially to produce low-caffeine beverages and can also be implemented for environmental remediation. One intriguing option is to use microorganisms to perform

decaffeination processes. Some microorganisms have been reported to degrade caffeine, such as *Aspergillus tamari*i (Gutiérrez-Sánchez *et al*., 2013), *Trichosporon asahii* (Lakshmi and Das, 2013), *Pseudomonas* sp. (Gokulakrishnan *et al*., 2007; Yu *et al*., 2015) or *P. putida* (Summers *et al*., 2011).

In this work, we report a diversity analysis aiming to characterise bacterial communities growing on coffee leach waste, using high througput sequencing, culturing, and electron microscocopy techniques. To achieve this goal, we have chosen one of the most widespread coffee preparation systems, Nespresso, due to its popularity and standard nature. In fact, Nespresso-compatible machines are highly standardized coffee making devices (same capsule type, same basic design, same pressure: 19 bars), and they represent a unique oportunity for a massive biological screening. Here we present the first attempt to do so. We have sampled the inner drip tray below the capsule container, in which coffee lyxiviate accumulates. We have analysed ten domestic and semi-domestic machines and studied the dynamic colonisation process in a brand new Inissia Krups machine operated in our laboratory. This is the first systematic analysis of the microbial diversity associated to coffee machines. Our results may shed light on the microbial arsenal of caffeine degraders with important implications for both medicine and biotechnology.

Materials and Methods

Sampling. Nine Nespresso machines, which had been operated for at least one year, used either at home (domestic) or in academic departments, institutes or biotechnology companies (communal) in the Valencia (Spain) area were aseptically sampled (Table V.1). The coffee lixiviate from wasted capsules present in the drip tray placed below the capsule container was sampled using a sterile Pasteur pipette. In one case (CityZ), the small space under the drip grid and cup support facilities was sampled and treated separately (Figure V.1A; Table V.1). In all cases, the average temperature of the room where the coffee machines were operated was close to 25 ºC. Additionally to these machines, a Krups Inissia machine was purchased for the present work and normally operated in our laboratory for five months with a daily use of around 1-5 capsules per day. Sampling was performed at increasing time lapses and consisted of the removal of most (except 5 mL, approximately) of the lixiviated liquid. When the volume of the lixiviate was smaller than 5 mL, no samples were taken. All the detachable pieces (drip grid and cup support, drip tray and capsule container were thoroughly rinsed once (day 28) with tap water. In all cases, 2 mL aliquots of the samples were immediately used for DNA isolation, whereas the remaining volume was stored at -80 ºC.

Scanning electron microscopy. Aliquots from each sample were filtered through a 0.2 micrometre filter (Corning Inc.), fixed with a 2% paraformalin - 2.5% glutaraldehyde solution, and lightly washed with filtered-sterile phosphate buffer (PBS). Then, small pieces of the filters were placed inside microporous capsules (30 μm pore size, available from Ted Pella Inc. product number 4619) and subjected to successive dehydration steps in growing ethanol solutions up to 100% ethanol, following critical point drying in an Autosamdri 814 (Tousimis). These fragments were then placed on SEM stubs with silver conducting paint TAAB S269 and examined under a scanning electron microscope Hitachi S-4100.

DNA isolation and PCR amplification. Two mL aliquots of each sample were centrifuged at 11,000 g for 3 min and the resulting pellets were washed twice with sterile PBS buffer (NaCl 8 g/L, KCl 0.2 g/L, Na_2HPO_4 1.44 g/L, KH_2PO_4 0.24 g/L, pH adjusted to 7.4). A widely-used DNA isolation protocol based on lysozyme treatment and alkaline lysis, suitable for the treatment of both Gram negative and Gram positive bacteria, was used to obtain metagenomic DNA. Briefly, each sample was treated with 2 mg/mL lysozyme (30 min, 37 ºC) to ensure the lysis of Gram positive bacteria, and DNA was isolated with an alkaline solution (Tris 300 mM, SDS 1.25%, sucrose 5%, EDTA 10 mM; adjusted to pH 8.0) followed by precipitation with 3M potassium acetate (pH 5.0) and isopropanol. The quality of the DNA was checked on a 0.8% (w/v) agarose gel and quantified with Nanodrop-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE).

Universal primer sequences 28F (5'-GAG TTT

TABLE V. 1: Nespresso machines sampled in this work.

Sample nº.	Machine model	Use*
1	*Essenza*	Communal (12 uses/day)
2	*Pixie*	Domestic (3 uses/day)
3	*Essenza*	Domestic (2 uses/day)
4	*CityZ (capsule container)*	Communal (10 uses/day)
5	*CityZ (drip base)*	
6	*Essenza*	Communal (20 uses/day)
7	*Inissia*	Domestic (4 uses/day)
8	*Essenza*	Communal (10 uses/day)
9	*Pixie*	Domestic (4 uses/day)
10	*Pixie*	Domestic (3 uses/day)
11-30	*Inissia Krups*	This work (3 uses/day)

* Frequency of use indicated in parenthesis as the average number of uses per day, as stated by the users of each machine

GAT CNT GGC TCA G-3') and 519R (5'-GTN TTA CNG CGG CKG CTG-3'), were chosen for the PCR-amplification of the 16S ribosomal RNA gene since they targeted the V1-V3 hypervariable region (commonly analysed in metagenomic studies) and produced an amplicon of a suitable length (500 bp) for the subsequent sequencing library construction. A short (9-11 nucleotides) barcode sequence followed by a four-nucleotide spacer (CGAT) was included at the 5' end of the oligonucleotides used as forward primers to enable sequence assignment to samples after high-throughput sequencing. All the amplifications were performed under the following conditions: initial denaturing at 95ºC for 5 min, followed by 35 cycles of denaturing at 95ºC for 30 s, annealing at 54ºC for 30 s, and extension at 72ºC for 1 min; finalized by a 10-min elongation at 72ºC.

The resulting amplicons were checked on a 0.8% (w/v) agarose gel and purified by precipitation with 3M potassium acetate (pH 5) and isopropanol. Pure amplicons were quantified with the Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and an equimolar pool of amplicons was prepared from all the samples.

16S rRNA gene sequencing and bioinformatic data analysis. A sequencing library was constructed with 100 ng of the equimolar pool by the amplicon fusion method (Ion Plus Fragment Library Kit, MAN0006846, Life Technologies). The library was quantified with the Agilent 2100 Bioanalizer (Agilent Technologies Inc, Palo Alto, California) prior to clonal amplification, and emulsion PCRs were carried out applying the Ion PGM Template OT2 400 kit as described in the user-guide (MAN0007218, Revision 3.0 Lifetechnologies) provided by the manufacturer. The library was sequenced in an Ion 318 Chip v2 on a Personal Genome Machine (PGM IonTorrent, Lifetechnologies) at Lifesequencing S.L (Lifesequencing, Valencia, Spain), using the Ion PGM Sequencing 400 kit following the manufacturer's protocol (Publication Number MAN0007242, Revision 2.0, Lifetechnologies). Raw sequences obtained from the sequencing centre were processed with the MOTHUR software (Schloss *et al.*, 2009). A summary of sequencing statistics is available in Supplementary Table V.1. Short (<100 bp) and low quality (<Q10, <90% accuracy) reads were removed in a first step, and sequences were then assigned to samples based on barcode matches (allowing a maximum of 3 mismatches in primer search and 1 mismatch barcode search). The resulting sequences were trimmed by removing primer, barcode, and spacer sequences. All sequences were aligned to the ribosomal 16S Greengenes database using the kmer method (8-mers) for finding template sequences and the Needleman method for sequence alignment. Penalties for mismatch, gap opening, and gap extension were set as default (-1, -2, and -1, respectively). Finally, sequences were classified using BLAST searches against the same database. The similarity percentage cut off was set at 70%.

Consortia identification. In order to identify microbial interactions, we used taxonomic data of samples from the ten machines. A method used a recently described multidimensional scaling and a biological significance filtering of the interactions (Dorado-Morales *et al.*, 2015). Basically, the procedure was as follows: fluctuation scaling of replicas was validated by a Poisson distribution selection. Then, linear correlation coefficients were converted into distances and displayed by multidimensional scaling. Finally, a discrete Lotka-Volterra model with relative abundances (Fisher and Mehta, 2014) was used to filter biologically significant interactions from the correlations identified in the last step.

Results and Discussion

The waste coffee in the capsule container of nine different Nespresso machines operated for at least one year was sampled (Figure V.1A). In one case (CityZ model), the cup tray was also sampled independently as it does not connect with the capsule container. The high throughput sequencing and analysis of the 16S rRNA gene amplicons from all the machines revealed a significant bacterial diversity, with the total number of identified genera ranging from 35 to 67. Although relatively similar microbial profiles were detected, there was an important variation in the frequency of particular taxa. *Enterococcus* sp. and *Pseudomonas* sp. proved to be the main taxa as they were moderately to highly abundant in nine out of the ten samples analysed. Other frequent genera were *Stenotrophomonas*, *Sphingobacterium*, *Acinetobacter* and, to a lesser extent, *Coprococcus*, *Paenibacillus* or *Agrobacterium*. *Dysgomonas* was very frequent in the Inissia machine, accounting for 15% of the sequences (Figure V.1B). No differences were detected between machine models (Table V.1) or use (domestic vs. communal). One of the two most frequent genera found in the coffee machines was *Pseudomonas*, which is also one of the few reported examples of a caffeine-degrading bacterium. Indeed, *Pseudomonas* sp. has been known to catabolise caffeine since the seventies (Woolfolk, 1975), and is reported to degrade up to 15 g/L of caffeine through an N- demethylation reaction, which along with C-8 oxydation represent the two potential catabolic pathways (Dash *et al.*, 2008). Species reported to display caffeine degradation abilities are *P. alcaligenes* (Babu *et al.*, 2007) and *P. putida* (strains C1, CBB1 or CBB5). In fact, *P. putida* N-demethylation genes have been used to genetically engineer a caffeine "addicted" version of *E. coli* (Quandt *et al.*, 2013), and caffeine removal from sewage by bioremediation with *P. putida* has also been proposed (Ogunseitan, 1996). The abundance of *Enterococcus* sp. in caffeine-rich leach might not necessarily involve unreported caffeine degradation abilities in *Enterococcus*, but it might simply be a consequence of tolerance to certain caffeine levels. The same applies to other frequent taxa. Interestingly, this genus has previously been associated with coffee (Wang *et al.*, 2008), along with several others detected in this work. For example, *Acinetobacter* sp. has been isolated during coffee fermentation

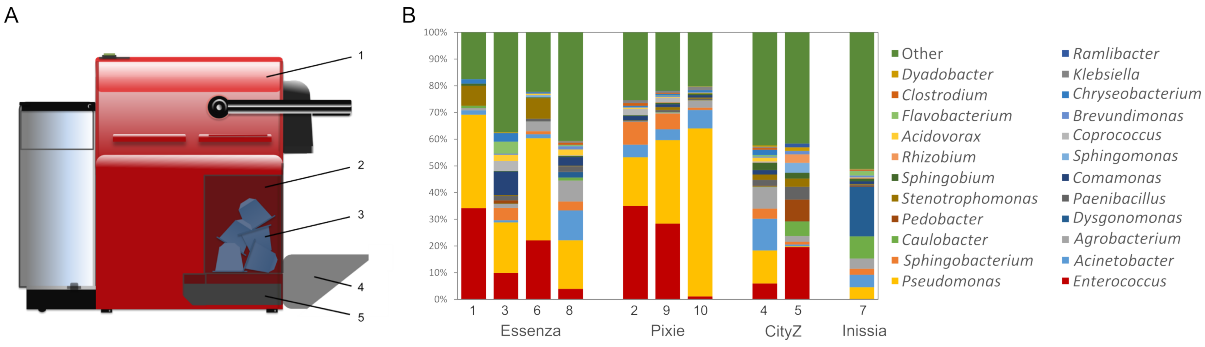


FIGURE V. 1: A) Schematic representation of a Nespresso machine 1) including a capsule 3) container 2), cup tray 4) a drip tray 5), which was sampled in this work. B) Bacterial profile of the nine Nespresso machines sampled according to 16S rRNA gene sequencing. Samples numbered in accordance to Table V.1.

(Silva *et al.*, 2008), while *Stenotrophomonas* sp., *Curtobacterium* sp., and *Pseudomonas* sp. are abundant in the coffee seed (Vega *et al.*, 2005).

The colonisation process of the wasted coffee leach was studied in an experiment using a brand new Krups Inissia machine (located in a separate room within our laboratory). The experiment lasted two months, during which leach samples were taken and bacterial diversity analysed, with a significant variation in the taxonomic profiles detected. The initially high species richness was substituted by a relatively simpler, but still highly variable, species composition (species richness significantly dropped 14 days after the beginning of the experiment; t-test p-value=0.039). During the first 11-13 days, *Pantoea* sp., *Cloacomonas* sp. and, to a lesser extent, *Brevundimonas* sp. were relatively abundant but amounts decreased to undetectable levels by the end of the experiment. All these taxa were largely substituted by *Pseudomonas* sp., *Acinetobacter* sp., and *Sphingobium* species, which reached a peak and then fluctuated (*Sphingobium* sp., *Bacillus* sp.) or reached the highest levels at the end of the experiment (*Pseudomonas* sp., *Acinetobacter* sp.) as shown in Figure V.2A. The first 30 days exhibited greater instability in the bacterial communities, as deduced by the consecutive peaks of very abundant taxa, which were substituted by a more balanced bacterial composition after one month. As in other studies on different environments (Kapplan and Kitts, 2004; Sulaiman *et al.*, 2014), these results strongly suggest a long ecological succession during the first month, in which generalist bacterial taxa, including enterobacteriaceae genera such as *Pantoea*, are the first colonizers but are then

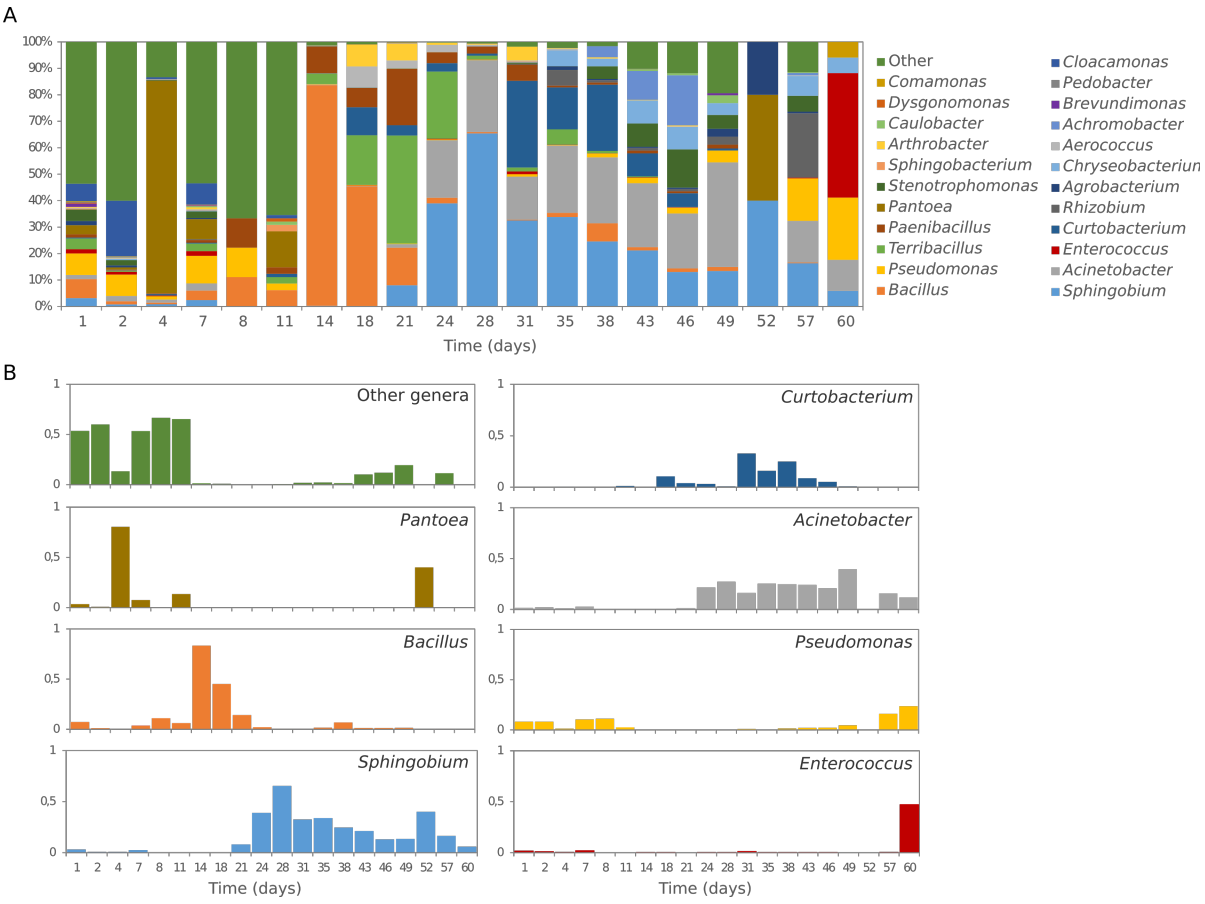


FIGURE V. 2: Bacterial colonisation in a brand new Nespresso Krups Inissia machine. A) Bacterial profile in the drip tray during the first month of operation according to 16S rRNA gene monitoring. B) Ecological succession of the main taxa during the experiment, represented as the variation of their relative frequencies.

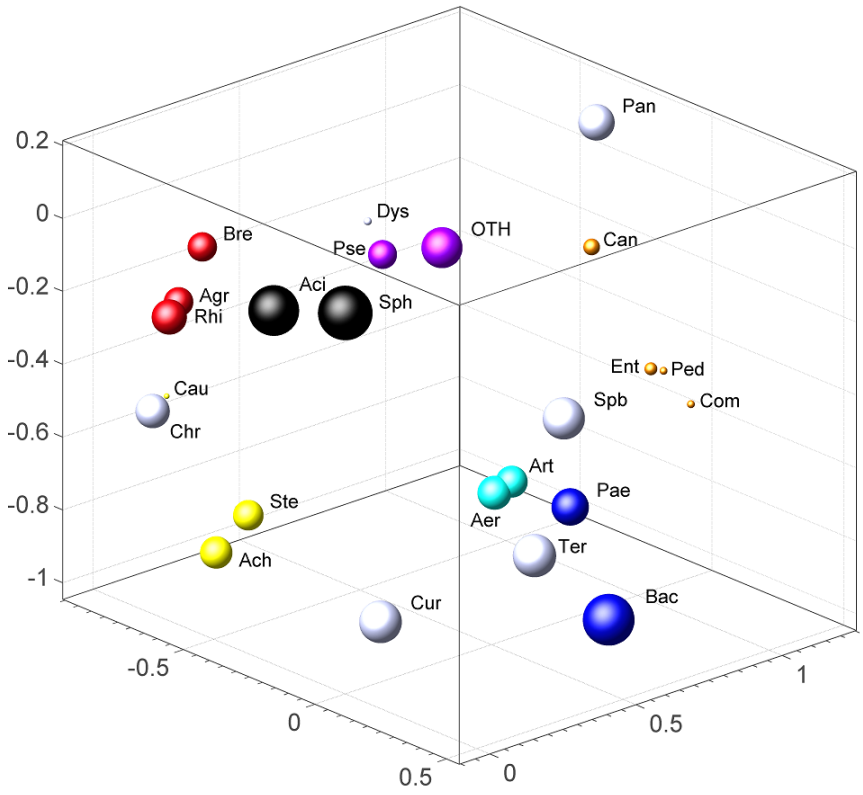


FIGURE V. 3: Correlations among the bacterial genera detected in this work. Distances correspond to the linear statistical correlation. Sizes of the spheres are proportional to the relative abundances in logarithmic scale. Highly correlated genera are shown in the same colour. Sph, *Sphingobium*; Bac, *Bacillus*; Aci, *Acinetobacter*; Ter, *Terribacillus*; Cur, *Curtobacterium*; Pae, *Paenibacillus*; Pan, *Pantoea*; Rhi, *Rhizobium*; Chr, *Chryseobacterium*; Aer, *Aerococcus*; Art, *Arthrobacter*; Ste, *Stenotrophomonas*; Ach, *Achromobacter*; Pse, *Pseudomonas*; Can, Candidatus *Cloacamonas*; Agr, *Agrobacterium*; Bre, *Brevundimonas*; Ent, *Enterococcus*; Cau, *Caulobacter*; Dys, *Dysgonomonas*; Spb, *Sphingobacterium*; Ped, *Pedobacter*; Com, *Comamonas*; Oth, Other genera.

displaced by successive waves of other taxa. The main keyplayers observed during this succession were, in order (Figure V.2B): enterobacteria (genus *Pantoea*; peaking 4-11 days after the beginning of the experiment), Firmicutes (three genera of the bacillaceae family: *Bacillus*, *Terribacillus*, *Paenibacillus*; peaking after 14-21 days); and, finally, the Sphingomonadales genus *Sphingobium* (proteobacteria), the Actinomycetales genus *Curtobacterium* (actinobacteria), and the Pseudomonadales genus *Acinetobacter* (proteobacteria), peaking after 28, 31 and 49 days, respectively. These taxa gave way to a different bacterial profile dominated by *Pseudomonas* sp. and *Enterococcus* sp. after two months of the experiment. This profile was very similar to that found in the nine other coffee machines sampled (Figure V.1B) which had been operated for a longer time, suggesting that the particular physico-chemical conditions (cycles of high temperature, constant caffeine accumulation, etc.) of coffee leach, rather than the influence of the user or the number of uses, are the main force shaping

the composition of the microbial community. A mathematical modelling performed on the dynamic series of 16S rRNA gene data revealed statistically significant correlations among the detected taxa, indicating that the distribution of bacterial genera in time is not random (Figure V.3).

Most of the taxa we identified during the colonisation process of the coffee machine operated in our laboratory have previously been found in natural coffee-related environments. Species belonging to the genera *Acinetobacter* and *Bacillus*, and also some enterobacteria, have been detected during the natural fermentation of coffee beans (Silva *et al*., 2008; Yu and Chen, 2013), whereas *Paenibacillus* and other Bacteroidetes and Firmicutes species have proved abundant in the composting process of coffee hulls (Silva *et al*., 2000; de Gannes *et al*., 2013). Despite some reports describing the ability of different *Sphingobium* species to degrade toxic molecules, such as bisphenols (Ogata *et al*., 2013) and hydroquinones (Machonkin and Doerner, 2011), this is the first report where *Sphingobium* sp. has been associated to a caffeine-rich environment.

In addition to the 16S rRNA gene monitoring, we followed up changes in the coffee leach microbial diversity through scanning electron microscopy (SEM). Figure V.4 shows a dynamic series of samples taken at different time points (4, 8, 14, and 21 days after the first day of operation) during the first month. Microbial biomass increased throughout the analysis, and variations in the composition and viscosity of the coffee leach were also evident. For example, a filamentous matrix was observed at days 8 and 21 (Figure V.4). At day 14, the sample was dominated by a single shape of bacterial cells, which interestingly coincided

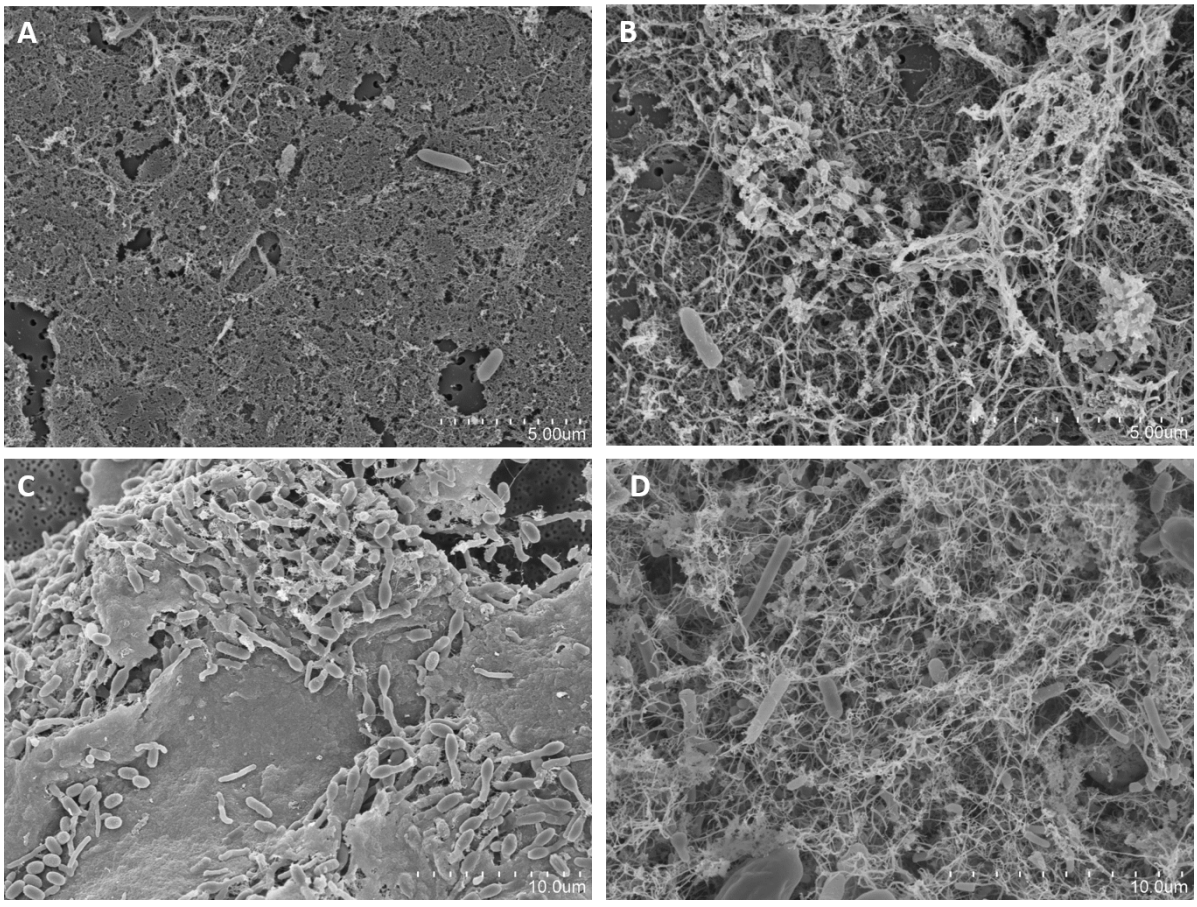


FIGURE V. 4: Chronological series of SEM images of the drip tray samples taken after 4 A), 8 B), 14 C), and 21 days D) of operation. Figure C corresponds to a sample highly abundant in *Bacillus* spp. Scale bars are indicated in each case.

with an overwhelming relative abundance of *Bacillus* spp. in that sample (Figure V.2A). Further experiments are needed to determine whether microbial community changes are the cause or the effect of the variations in the composition of coffee leach as shown by SEM. Our results show, for the first time, that coffee leach from standard capsule machines is a rich substrate for bacterial growth; that caffeine content does not prevent a rich bacterial biodiversity from rapidly colonising coffee leach; and that microbial succession from an initial pool of generalist bacteria gives way to an apparently coffee-adapted but still highly variable bacteriome. This bacteriome is rich in species previously reported to be associated with the coffee plant and/or the coffee fermentation processes. Colonising bacteria might be of environmental origin (no cultivable microorganisms nor bacterial DNA was detected in coffee capsules, data not shown), whereas heterogeneity of bacterial composition may relate to factors such as cleaning habits and, specially, the frequency of machine use (with higher frequencies presumably correlating with increased volume and temperature of the coffee leach). Further studies comprising more coffee machines, deep genome sequencing of the microbial communities therein, and even functional metagenomics, are required to contribute to shed light on the microbial ecology of coffee leach in capsule machines.

The presence of bacterial genera with pathogenic properties and the fast recovery of the communities after rinsing the capsule container, strongly suggest the need for frequent maintenance of the capsule container of these machines. Maintenance should employ bacteriostatic compounds, and avoid contact of the coffee leach with other parts of the machine to avoid unintended contamination of the beverage. On the other hand, the resistant microbial communities we describe here (microbial consortia, individual caffeine degrading/tolerant species or as a source of metabolic pathways and genes) may represent a promising tool for biological coffee decaffeination processes and for environmental caffeine decontamination.

Acknowledgements

C.V. was funded with a FPU Formación de Personal Universitario) grant from the Spanish MECD Ministerio de Educación, Cultura y Deporte). We are indebted to Dr. Carlos P. Garay IFIC, University of Valencia – CSIC) for his technical support in Figure V.3. SEM images were obtained at the Electron Microscopy Services of the University of Valencia.

References

Al-Ansari MM, Aboussekhra A. 2014. Caffeine mediates sustained inactivation of breast cancer-associated myofibroblasts via up-regulation of tumor suppressor genes. *PLoS One* **9**:e90907.

Babu VRS, Patra S, Karanth NG, Kumar MA, Thakur MS. 2007. Development of a biosensor for caffeine. *Anal Chim Acta* **582**:329–34.

Buerge IJ, Poiger T, Müller MD, Buser H-R. 2006. Combined sewer overflows to surface waters detected by the anthropogenic marker caffeine. *Environ Sci Technol* **40**:4096–102.

Dash SS, Sailaja NS, Gummadi SN. 2008. Chemotaxis of Pseudomonas sp. to caffeine and related methylxanthines. *J Basic Microbiol* **48**:130–4.

de Gannes V, Eudoxie G, Hickey WJ. 2013. Prokaryotic successions and diversity in composts as revealed by 454-pyrosequencing. *Bioresour Technol* **133**:573–80.

Dorado-Morales P, Vilanova C, P Garay C, Martí JM, Porcar M. 2015. Unveiling Bacterial Interactions through Multidimensional Scaling and Dynamics Modeling. *Sci Rep* **5**:18396.

Fisher CK, Mehta P. 2014. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* **9**:e102451.

Gokulakrishnan S, Chandraraj K, Gummadi SN. 2007. A preliminary study of caffeine degradation by Pseudomonas sp. GSC 1182. *Int J Food Microbiol* **113**:346–50.

Gonzalez de Mejia E, Ramirez-Mares MV. 2014. Impact of caffeine and coffee on our health. *Trends Endocrinol Metab* **25**:489–92.

Gutiérrez-Sánchez G, Roussos S, Augur C. 2013. Effect of caffeine concentration on

biomass production, caffeine degradation, and morphology of Aspergillus tamarii. *Folia Microbiol Praha* **58**:195–200.

Je Y, Giovannucci E. 2014. Coffee consumption and total mortality: a meta-analysis of twenty prospective cohort studies. *Br J Nutr* **111**:1162–73.

Jiang X, Zhang D, Jiang W. 2014. Coffee and caffeine intake and incidence of type 2 diabetes mellitus: a meta-analysis of prospective studies. *Eur J Nutr* **53**:25–38.

Kaplan CW, Kitts CL. 2004. Bacterial Succession in a Petroleum Land Treatment Unit. *Appl Environ Microbiol* **70**:1777–1786.

Lakshmi V, Das N. 2013. Removal of caffeine from industrial wastewater using Trichosporon asahii. *J Environ Biol* **34**:701–8.

Machonkin TE, Doerner AE. 2011. Substrate specificity of Sphingobium chlorophenolicum 2,6-dichlorohydroquinone 1,2-dioxygenase. *Biochemistry* **50**:8899–913.

Mustard JA. 2014. The buzz on caffeine in invertebrates: effects on behavior and molecular mechanisms. *Cell Mol Life Sci* **71**:1375–82.

Ogata Y, Goda S, Toyama T, Sei K, Ike M. 2013. The 4-tert-butylphenol-utilizing bacterium Sphingobium fuliginis OMI can degrade bisphenols via phenolic ring hydroxylation and meta-cleavage pathway. *Environ Sci Technol* **47**:1017–23.

Ogunseitan OA. 1996. Removal of caffeine in sewage by Pseudomonas putida: Implications for water pollution index. *World J Microbiol Biotechnol* **12**:251–6.

Quandt EM, Hammerling MJ, Summers RM, Otoupal PB, Slater B, *et al*. 2013. Decaffeination and measurement of caffeine content by addicted Escherichia coli with a refactored N-demethylation operon from Pseudomonas putida CBB5. *ACS Synth Biol* **2**:301–7.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, *et al*. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**:7537–41.

Silva CF, Schwan RF, Sousa Dias Ë, Wheals AE. 2000. Microbial diversity during maturation and natural processing of coffee cherries of Coffea arabica in Brazil. *Int J Food Microbiol* **60**:251–260.

Silva CF, Batista LR, Abreu LM, Dias ES, Schwan RF. 2008. Succession of bacterial and fungal communities during natural coffee Coffea arabica fermentation. *Food Microbiol* **25**:951–7.

Spriet LL. 2014. Exercise and sport performance with low doses of caffeine. *Sports Med* **44**:S175–84.

Sulaiman J, Gan HM, Yin W-F, Chan K-G. 2014. Microbial succession and the functional potential during the fermentation of Chinese soy sauce brine. *Front Microbiol* **5**:556.

Summers RM, Louie TM, Yu CL, Subramanian M. 2011. Characterization of a broad-specificity non-haem iron N-demethylase from Pseudomonas putida CBB5 capable of utilizing several purine alkaloids as sole carbon and nitrogen source. *Microbiology* **157**:583–92.

Vega FE, Pava-Ripoll M, Posada F, Buyer JS. 2005. Endophytic bacteria in Coffea arabica L. *J Basic Microbiol* **45**:371–380.

Wang X, Allen TD, May RJ, Lightfoot S, Houchen CW, Huycke MM. 2008. Enterococcus faecalis induces aneuploidy and tetraploidy in colonic epithelial cells through a bystander effect. *Cancer Res* **68**:9909–17.

Woolfolk CA. 1975. Metabolism of N-methylpurines by a Pseudomonas putida strain isolated by enrichment on caffeine as the sole source of carbon and nitrogen. *J Bacteriol*. **123**, 1088-1106.

Yu CL, Summers RM, Li Y, Mohanty SK, Subramanian M, Pope RM. 2015. Rapid Identification and Quantitative Validation of a Caffeine-Degrading Pathway in Pseudomonas sp. CES. *J Proteome Res* **14**:95–106.

Yu P, Chen Y. 2013. Purification and characterization of a novel neutral and heat-tolerant phytase from a newly isolated strain Bacillus nealsonii ZJ0702. *BMC Biotechnol* **13**:78.

General Results and Discussion

1. The analysis of microbial communities in the light of traditional bioprospecting techniques

1.1. Metagenomic sequencing

In the present thesis, a metagenomic approach was used as a first step towards the exploration of both natural (pinetree resin and the gut of three toxic-feeding Lepidoptera) and artificial (solar panels and the leach tray of coffee machines) environments. This technique allowed the characterization of different aspects of these microbial communities, namely the taxonomic identification of the species inhabiting each particular habitat, the estimation of microbial diversity, the exploration of the genetic pool, and the contextualization of the microbial communities through comparative metagenomics with other environments.

(i) Taxonomic identification of the species inhabiting each particular habitat, and estimation of microbial diversity. In the context of this thesis, the use of metagenomic sequencing has unveiled the taxonomic composition of the microbial communities associated to the environments cited above. These communities proved notably different in terms of specialization and diversity. On the one hand, extremely specialized communities were found in two particular environments: pinetree resin, *R. resinella* galls, and solar panels. The microbiota associated to resin and *R. resinella* was dominated by a single bacterial species, *Pseudomonas abietaniphila*, with known ability to degrade diterpenes, the most abundant terpenes in coniferous resin. Other species belonging to genera with known terpene-degrading representatives (such as *Burkholderia* and *Cupriavidus*) were also detected at moderate frequencies in these samples (Figure I.1). In the case of solar panels, the microbial communities were characterized by the abundance of microorganisms with one or more of the following features (Figure III.1): radiation-resistance (species from the *Deinococcus* and *Hymenobacter* genera); carotenoid production (*H. xingiangensis*, *S. kaistensis*, *R. mesophilum*, and others), which is known to confer protection against oxidative stress (Tian and Hua, 2010; Sandmann, 2015); and sphingolipid production (members of the Sphingomonadales and Sphingobacteriales orders), which has been linked to biofilm formation on glass-like surfaces (Gutman *et al*., 2014). Despite the similarity between solar panels and resin-associated samples in terms of specialization, the communities found on solar panels samples proved to be much more diverse, with up to 500 different microbial species detected per sample.

On the other hand, and somehow surprisingly, microbial communities were found to be apparently generalist –and also less diverse- in the case of the specialized toxic plant-feeders *B. crini* and *H. euphorbiae* guts, as well as in the leach tray of coffee machines. These communities were dominated by *Enterococcus* sp. and *Staphylococcus* sp., in the case of insects (Figure II.2); and *Pseudomonas* sp. and *Enterococcus* sp., in the case of coffee leach (Figure V.1). The only exceptions to this apparently generalist community were the detected representatives of genera *Klebsiella* and *Corynebacterium*, detected in *B. crini* and previously reported to degrade alkaloids; and species from *Nocardioides* and *Gordonia* genus, detected in *H. euphorbiae* and known to efficiently degrade natural latex. Also, specialist species with known ability to degrade alkaloids (*Pseudomonas* sp., *Sphingobium* sp., *Acinetobacter* sp.) were detected during the colonization of coffee leach (Figure V.2), in which an ecological succession was observed, and microbial profiles were strikingly similar to those detected in the gut of the coffee bean-borer beetle (Ceja-Navarro *et al*., 2015).

(ii) Exploration of the genetic pool harbored by the microbial communities. The analysis of some of the communities described above through shotgun metagenomic sequencing allowed the massive identification of genetic elements encoded in the metagenomes. As a result of this, a list of genes involved in the degradation of different types of terpenes were detected in resin and *R. resinella*-associated samples (Supplementary Table I.1). Of particular relevance was the complete cluster of genes necessary for diterpene degradation detected in *P. abietaniphila*, the most abundant species in resin samples. Similarly, a range of elements related to the resistance of microorganisms to harsh conditions such as UV-radiation, desiccation, and heat stress were identified in the solar panel communities not only through metagenomic sequencing but also through metaproteomics. By exhibiting such adaptation(s) to particular environments, those diverse microbial communities can have a broad range of applications in the field of bioremediation, protection against oxidative stress, and other applications involving super-resistance to extreme physical conditions (see section 1.2).

(iii) Ecological contextualization of the microbial communities. Comparative metagenomics is a powerful tool for the ecological characterization of a microbial community. An example of this is the comparison of the solar panel communities with those associated to a range of desertic and non-desertic environments in order to find similarities among microbial profiles from the taxonomic and the functional (gene content) point of view. This analysis unveiled that the solar panels communities were strikingly similar to those of other highly irradiated environments such as saline deserts, polar deserts, or the phylloplane, and notably distant to those associated to air, soil, or human-related samples. This similarity was found both at the taxonomic (Figure III.3) and the functional level (Figure III.2).

In concordance with our work, new studies report –almost daily- the metagenomic-based description of previously unexplored environments and the putative applications of the genes and microbial strains identified. This is the case of the microbiota associated to insect pests, which might be key for future pest management programmes (Malacrinò *et al*., 2016; Cheng *et al*., 2016), or the communities found in the gut of a toxic-feeding scorpion, which have been found to be rich in metabolic pathways for toxic compound degradation (Bolaños *et al*., 2016). Also, new species are being identified by increasing sequencing depth, like the novel methanogenic archaeal lineages identified in a sample from an anaerobic digester in a biogas plant (Vanwonterghem *et al*., 2016). Moreover, there is a growing interest in the microbiome of artificial habitats, due to its implications in human health (Triadó-Margarit *et al*., 2016) or in the conservation of the cultural heritage (Li *et al*., 2016). A very recent report analyzing the bacterial communities associated to washer machines, dishwashers, refrigerators, and other domestic devices operating at extreme conditions (high concentrations of detergents, low pH, high temperature) unveiled an unexpected diversity of bacteria able to grow in such harsh environments (Savage *et al*., 2016). These communities, similarly to those associated to solar panels or coffee machines, are a promising source of microorganisms with practical applications in biotechnological processes requiring the adaptation to extreme conditions.

Despite the gigabytes of information yielded by metagenomic analysis, this technique has several limitations, which have actually shown up in the context of this thesis. First, the taxonomic information obtained in community-scale analysis is not optimum, since the taxonomic assignations obtained from short reads are often restricted to the family, order, or even phylum level (Mizrahi-Man *et al*., 2013). Also, community composition can be biased due to several factors such as the DNA extraction method (Albertsen *et al*., 2015; de Bruin and Birnboim, 2016), the protocol for library construction (Danhorn *et al*., 2012; Klindworth *et al*., 2012), or the oligonucleotide sequences used as primers in the massive analysis of marker genes such as 16S or 18S/ITS rDNA (Albertsen *et al*., 2015). Second, it is difficult to obtain genomic information from species present at very low frequencies, especially when the community is dominated by one or a few microbial strains, as it was the case of resin and *R. resinella*-associated samples. And third, it is evident that metagenomic sequencing fails to identify microbial species or

genetic elements which have not been previously sequenced or annotated, and this could be particularly relevant when the microbiome of unexplored environments, which might harbor novel species or gene sequences, is analyzed. Last, particularly in the case of bioprospecting efforts, the mere identification of microbial strains with a potential use in biotechnology could be pointless without techniques allowing the isolation and experimental characterization of those strains.

1.2. Cultivation

Traditional culturing techniques have been used in this thesis as a complementary approach overcoming some of the limitations of metagenomic sequencing. Mainly, culturing allows the isolation of microbial strains identified in the metagenomic analysis in order to characterize them in the laboratory and experimentally assess their potential applications in biotechnology. Interestingly, one of the most relevant examples of this is the *P. abietaniphila* strain PS, the most abundant species detected in *R. resinella* samples. The cultivation of this strain allowed not only the experimental characterization of its resin-degrading ability (Figure I.2), but its proper taxonomic identification (it was misidentified as *P. fluorescens* due to the lack of sequenced *P. abietaniphila* genomes); and more importantly, the finding of other abilities that were not predictable from the genomic information obtained from the metagenomic sequences: the ability to degrade acyclic terpenes (Figure I.4) and its antifungal properties (Figure I.5). As a result of this experimental characterization, *P. abietaniphila* strain PS was patent-protected (Appendix C) and its industrial application is currently being tested in the frame of a collaboration with an agroalimentary company. In addition, three fungal strains with an outstanding ability to eliminate resin colloids from resin-rich media (Figure I.2) were also isolated and characterized in the laboratory.

The experimental characterization of the *Enterococcus casseliflavus* strain He, isolated from *H. euphorbiae* gut, was also essential to hypothesize its potential role in latex detoxification in this insect. This strain still needs to be subjected to further characterization, and its role might be, in accordance with a previous study, related to crystallization -rather than degradation- of latex molecules in a ring-shaped biofilm present in *H. euphorbiae* gut, as it has been previously reported (Shao *et al*., 2011) for *Spodoptera litura* larvae feeding on toxic-rich diets (Figure II.3).

Finally, the culture-dependent analysis of the solar panel samples revealed, surprisingly, a very diverse microbial community in which an extremely low number of isolates were thermophile, and that the survival of the community in such a harsh environment might be based on inter-species cooperation rather than in the multi-stress resistance ability of individual species (Supplementary Figure III.1). Cultivation allowed, again, the characterization of interesting species, in this case, carotenoid producers, which may be valuable for the biomedical and cosmetic industry due to the outstanding antioxidant properties of these biomolecules. However, cultivation techniques still have major limitations for the isolation of some microbial taxa, even when they are highly abundant in the samples. The most striking case of this are the *Deinococcus* and *Hymenobacter* species accounting for nearly 30% of the sequences in some of the solar panels samples (Figure III.1). These species might be of great interest as new and robust chassis for biotechnology, representing an alternative to traditional ones (i.e.: *E. coli*, *S. cerevisiae*, *P. putida*) with an improved performance in terms of intrinsic resistance to stressful conditions. However, its cultivation is a mandatory step for this kind of applications.

In summary, cultivation complements metagenomics in terms of experimental characterization of the isolates, determination of abilities that are difficult to predict from genomic data, and access to whole genome information, especially in the case of species present at low relative frequencies. Today, new repositories of microbial species are being developed in order to explore the culturable diversity of microorganisms with applications in health (Lagkouvardos *et al*., 2016). Nonetheless, current culturing techniques still fail to recover the vast majority of environmental microorganisms.

2. One step further: microbial interactions

Combining *in silico* analysis and cultivation-based experiments yields a more realistic picture of

the real applications of a particular microbial community. However, both approaches are traditionally centred on individual species, even though microorganisms are known to establish natural assemblies in the shape of microbial consortia in virtually any environment (Brenner, 2008). Therefore, new bioprospecting techniques targeting microbial associations rather than individual species are strongly needed.

In the present thesis, an innovative approach complementary to traditional metagenomic analysis of microbial communities has been developed. This approach consists of predicting the composition of putative microbial consortia through the detection of statistically significant correlations and biological interactions imprinted in the metagenomic sequencing data. Similarly to other methodologies previously described (Faust and Raes, 2012; Friedman and Alm, 2012), our analysis is based in the time-series analysis of the relative frequencies of the OTUs defined through the metagenomic sequencing of marker genes (in this case, the 16S rRNA gene). Microbial associations are predicted according to correlation coefficients, whereas biological interactions are detected on the basis of the Lotka-Volterra dynamic model. The key feature of our approach, which is actually its main advantage with respect to other methodologies, is that all the information obtained from the microbial community (identified OTUs, relative frequencies, positive and negative correlations, and positive and negative interactions) is directly visualized in a single, interactive 3D plot, which can be tuned by the user. This is achieved thanks to the use of MDS (Multi-Dimensional Scaling), which provides a unique and optimum representation of the data matrix (Figure IV.2 and Figure V.3).

When we used this method on resin-associated samples and coffee leach samples, the algorithm for the prediction of microbial consortia yielded a different view of the communities, in which different microbial associations were detected. Even though the predicted consortia could not be experimentally reproduced, bibliographic research suggests that the predictions might be reliable. Indeed, the resin-degrading consortia detected with our methodology have been reported in previous studies as natural associations able to degrade hydrocarbons, phenolic compounds, TNT, and herbicides. Interestingly, the degradation of these compounds involves similar biochemical transformations (such as the cleavage of aromatic rings by dioxygenase enzymes) to those of the degradation pathway of the terpenes present in pinetree resin. Regarding the degradation of caffeine or other alkaloids, very few microbial consortia have been reported so far. Only one microbial consortium formed by species of *Klebsiella* and *Rhodococcus* (which were not detected in our study) have been reported to degrade caffeine (Madyastha and Sridhar, 1998; Madyastha *et al*., 1999), and the only reports on natural consortia able to degrade other alkaloids are focused on anaerobic environments such as animal rumen (Lodge-Ivey *et al*., 2005; Rattray and Craig, 2007). Our studies fit within very recent works studying how microbes colonize particular substrates and achieve a defined population structure through ecological successions and microbial interactions (Wright *et al*., 2016; Korenblum *et al*., 2016).

Obviously, further experiments are needed in order to test the microbial associations predicted in our analyses. Today, our current efforts are focused in the isolation of the members of the associations in order to reconstruct the predicted consortia and experimentally test if they show an improved performance in resin/caffeine degradation with respect to the individual strains. This has notable limitations. For instance, the taxonomical assignations obtained are sometimes limited to order or family levels, and the actual composition of the consortia is therefore difficult to determine. Also, the isolation of candidate species can be difficult when they are present at very low frequencies in the samples, or when they establish obligate interdependencies among the members of the consortium. Therefore, bioinformatic analysis and cultivation techniques targeting individual species -or microbial consortia as an individual unit- are still to be improved.

From the microbial communities enclosed in the gut of toxic-feeding insects to the sun-exposed communities inhabiting the surface of solar panels, the microbial communities analyzed throughout this thesis represent a treasure of diversity hosting a wide range of useful biotechnological tools. In fact, mining those tools is our current goal in the context of new research projects. Also, reflecting on the "lights and shadows" of each of the bioprospecting techniques used in this thesis has been central to understand the limitations and challenges of current microbiome-scale analyses.

This raises a key question: which is the ultimate approach to tackle with the huge microbial diversity of our planet?

3. Epilogue: A new, holistic approach for microbiome studies

Even though the integration of multi-omic data into the recently called "trans-omic" data is able to generate unprecedentedly complete results (Yugi *et al.*, 2016), the analysis of such datasets often ignores the search of ecologically relevant conclusions, and focuses, instead, on getting increasingly exhaustive catalogs of species, expressed genes, or metabolite inventories. As a consequence, multi-omics has the risk of increasing the complexity it is supposed to address. As an example of "trees hidden by the forest", biological interactions among members of a microbial community (this is, microbial consortia) remain often buried beneath the massive multi-omic datasets. It is not because metagenomics is complemented with metatranscriptomics, metaproteomics and meta-metabolomics that relevant biological interactions will magically emerge. It is known that microbes are naturally assembled into interacting communities and that these community structures are directly linked to microbial processes (Bier *et al.*, 2015). Therefore, the identification of key-player species in a taxonomically complex sample is necessary to understand the ecology of a particular habitat. This is especially true when it comes to the study of biotechnologically relevant microbial consortia, such as those present in the biogas industry (Abendroth *et al.*, 2015). Interestingly, engineers tend to consider their fermenters as "black boxes" yielding biogas, and although omic approaches certainly contribute to shed light on the taxonomic or functional complexity of a fermenting biomass, such strategy might fail at identifying the ecological -and economical- core of the process. In plain words, the challenge is to reduce complexity to understandable informational "pills".

We argue that a core of ecological conclusions has to emerge beyond the combination of the complex information obtained through multi-omics from microbiome studies. Multi-omic analyses should yield

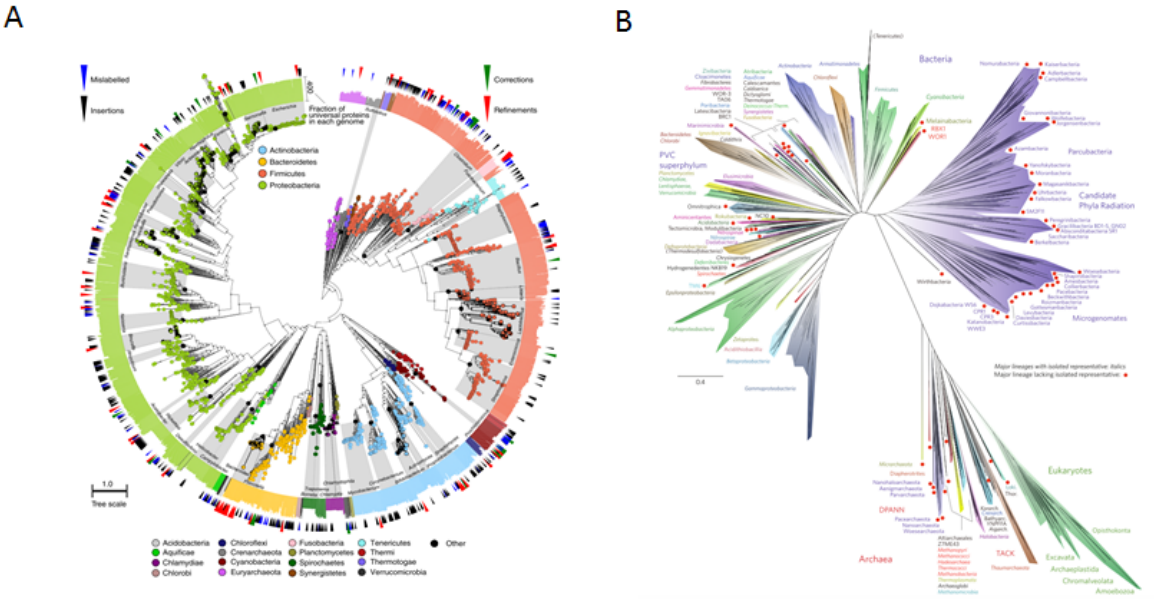


FIGURE 1: Representations of microbial diversity before (A) and after (B) the development of innovative bioprospecting techniques cited in section 3. The upper left branch in sub-figure B represents the Candidate Phyla Radiation (CPR). Sub-figure A has been modified from Segata *et al.*, 2013; sub-figure B taken from Hug *et al.*, 2016.

more than "the sum of their parts", as claimed by some ecologists (Morales and Holben, 2011). Even though there is still a long way to go, the relatively recent birth of the so-called reverse ecology might prove helpful for the prediction of interactions among species and for better understanding their metabolic networks in the context of their natural habitats (Levy and Borenstein, 2012). Moreover, alternative approaches to predict ecologically-relevant information have emerged in the last couple of years. For instance, artificial neural networks have been proposed for modelling microbial communities as functions of environmental parameters and intra-microbial interactions (Larssen *et al*., 2015); the dynamics and composition of microbial consortia have been unveiled by measuring temporal variations in interspecies metabolic interactions (Zomorrodi *et al*., 2014); and a new approach for visualizing microbial consortia through mathematical modelling and multi-dimensional scaling has been reported in the present thesis. Nevertheless, the integration of experimental multi-omic data with predictive mathematical models based on mechanistic understanding is still considered a "missing link" in microbial ecology (Widder *et al*., 2016). We envision a bright, relatively close future in microbial ecology in which multi-omic unbearable databases will be analyzed through approaches able to "milk" all the gigabytes of information into simpler, ecologically relevant, core conclusions.

Reached this point, does this mean that we will never need to culture microorganisms again? We strongly believe that culturing is more needed today than ever. In fact, the mere identification and characterization of bacterial species or consortia through multi-omics may not be enough when the interest of the study is focused on rare species, whose complete functional and metabolic repertoire is difficult to unveil even with high-coverage metagenomic sequencing (Lagier *et al*., 2012), or when one pursues the biotechnological application of a particular species or group of species. The development of single-cell genomic approaches (Rinke *et al*., 2013; Paez-Espino *et al*., 2016), improved analysis of metagenomic data (Albertsen *et al*., 2013), or innovative sampling methods, like analyzing the biomass passed through a 0.2-μm filter (Luef *et al*., 2015), have unveiled the extraordinary phylogenetic and functional diversity of many new microorganisms (Figure 1; Hug *et al*., 2016).

Therefore, an approach able to fully characterize rare species, and to experimentally test the *in silico* predictions for particular microorganisms of biotechnological interest is needed. Such approach does exist: culturing. In this sense, improved bioprospecting techniques aiming at the cultivation of hard-to-culture species (Liu *et al*., 2009; Nichols *et al*., 2010; Jung *et al*., 2014) have been developed in the last years. Among such techniques, in situ cultivation methods have proved successful to isolate bacterial species carrying novel gene sequences (Rasmussen *et al*., 2008) or producing new antibiotics (Ling *et al*., 2015).

These features cannot be predicted with multi-omic approaches, since they might only match with sequences of unknown function in the databases. In this sense, it is estimated that at least 7-60% of the sequences obtained through metagenomic sequencing cannot be properly classified due to the limiting number of reference annotated genomes in the public databases (Prakash and Taylor, 2012). Another obvious obstacle is when rare microbes are "ultra-small" bacteria, with incomplete metabolic networks, and with an intrinsic difficulty for *in vitro* culture. In this case, novel approaches for isolating natural microbial consortia (Luo *et al*., 2015) supported by efficient methods to predict microbial interactions such as the one developed during this thesis are essential. The isolation of rare, new-to-science species or microbial consortia allows not only their experimental characterization in the laboratory, but also the complete analysis of their genomes -which will be used as new reference data for ulterior microbiome studies-, and the understanding of organismal and community biology (Lazcano, 2011). Interestingly, cultivation-dependent and cultivation-independent approaches do not only complement each other, but strictly need each other. Multi-omics needs more reference genomes to better analyze new, complex microbiomes; and microbial ecologists need multi-omics to know what else is alive out there, and thus what they can attempt to culture. Another example of the required interphase between strategies is that multi-omic data analyzed through metabolic modelling can be used to predict the essential nutrients required for the cultivation of hard-to-culture species on the basis of its metabolic network (Yus *et al*.,

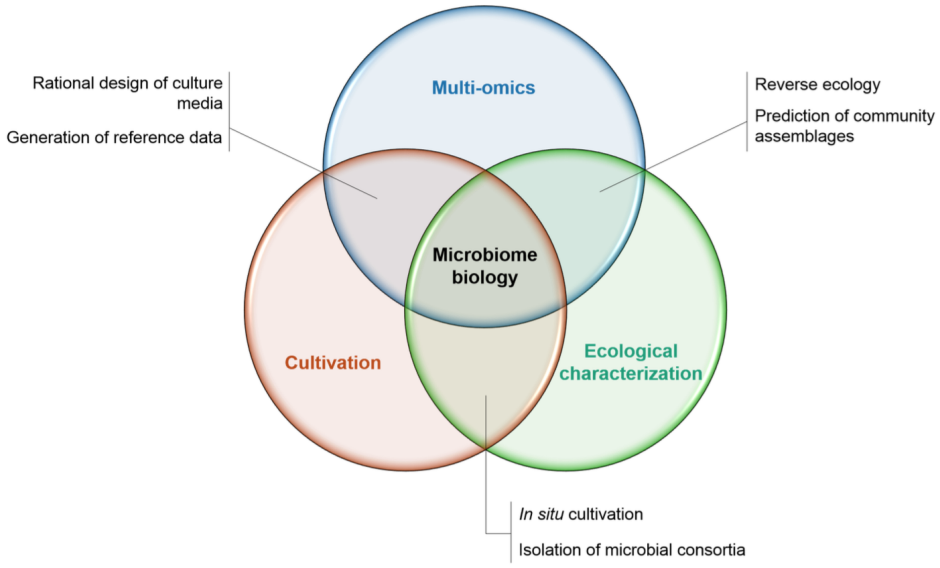


FIGURE 2: Microbiome biology as the interphase among multi-omic data, ecological characterization, and cultivation-dependent techniques.

2009). Therefore, we view microbiome biology not as a simple combination of multi-omic data, but as an emerging crossroad arising from the interphase among multi-omics, cultivation and ecological characterization (Figure 2).

A recent calculation based on scaling laws suggests that a trillion microbial species are yet to be discovered (Locey and Lennon, 2016). Studying those species is one of the greatest challenges of microbiology, and we argue that strategies transcending the information of multi-omics are essential for actually unveiling the composition and the ecology of such incredibly complex microbial communities. Simply combining layers of high-throughput biological data will result in improved databases and methodologies for the discovery of myriad of unknown genes, microbial species, or metabolites. But their *biology* will remain obscure as far as biodiversity repositories are considered as biological enlightenment, rather than what they actually are: impressive yet raw sources of future knowledge, which is something not to be confused with knowledge itself[1].

[1]The main conclusions of this epilogue have been published in:
Vilanova C and Porcar M. 2016. Are multi-omics enough?. *Nat Microbiol* **1** doi:10.1038/nmicrobiol.2016.101

Conclusions

In this thesis, the microbial communities present in a range of environments have been analyzed through multi-level bioprospecting. The conclusions reached for each particular environment are described in detail in the corresponding publications. The general conclusions emerging from our work are listed below:

- Traditional metagenomic analysis allowed the description of the microbial communities of previously totally undescribed environments: natural (resin and insect-associated) ones and artificial (solar panels and the leach tray of coffee machines) environments. This was achieved from the taxonomic and functional point of view. As a result, a range of microbial species and genes with potential applications in biotechnology have been identified.

- The microbial communities analyzed in this thesis displayed different degrees of diversity and specialization. The communities associated to *R. resinella* and solar panels were highly specialized, with the dominant species accounting for metabolic capabilities or stress resistance abilities according to their environment. In contrast, the communities associated to *B. crini*, *H. euphorbiae*, and the coffee leach of several coffee machines were less diverse, and also dominated by apparently generalist bacteria.

- Cultivation in artificial media has been used as a complementary approach to metagenomic sequencing, in order to isolate and characterize individual strains of particular interest. Some of the isolates have a direct application in biotechnological processes, such as bioremediation (*Pseudomonas abietaniphila* strain PS, *Enterococcus casseliflavus* strain He), agroalimentary processes (*P. abietaniphila* strain PS for fungal growth inhibition, or coffee leach isolates such as *Pseudomonas* sp., *Sphingobium* sp., and *Acinetobacter* sp. for biodecaffeination); as well as strains suitable for antioxidant biomolecule production (carotenoid-producing isolates from solar panels such as *Planomicrobium glaciei* or *Micrococcus xinjiangensis*).

- A new method for the detection of microbial consortia from time-series metagenomic data has been set up. This method is based on the calculation of statistically-significant correlations and interactions, and allows, for the first time, the direct visualization of the predicted associations in a single, interactive, 3D-plot.

- Several bacterial associations were predicted for the time-series experiments in resin-containing media and coffee leach. The predictions of the algorithm could not be experimentally tested, but could be trustable according to previous studies on natural microbial consortia.

- In connection with the multi-level bioprospecting techniques used in this thesis, a new, holistic approach has been proposed for microbiome studies in which data from multi-omic technologies, cultivation, and ecological characterization need to be overlapped.

Resum en valencià

Introducció

La bioprospecció és la cerca d'elements biològics (gens, espècies, metabolits, etc.) amb alguna utilitat pràctica. Els éssers humans han practicat la bioprospecció durant mil.lenis, però no va ser fins el segle XIX quan el descobriment dels primers bacteris i fongs va obrir un nou escenari que explorar: el increïblement divers món dels microorganismes. Des de l'aparició de les primeres formes de vida a la Terra, en forma de comunitats microbianes, els microorganismes han evolucionat per a desenvolupar un ampli ventall de recursos per tal d'adaptar-se a pràcticament tots els ambients del planeta, fins i tot els més hostils. Per tant, els microorganismes, d'entre tots els éssers vius, constitueixen la font més extensa d'eines potencialment útils per als humans.

L'aprofitament biotecnològic dels microorganismes, però, no seria possible sense el desenvolupament de potents tècniques d'aïllament i identificació. En aquest sentit, el desenvolupament de tècniques moleculars per a la identificació de microorganismes ha suposat una autèntica revolució en els estudis de bioprospecció, ja que ha sobrepassat les limitacions de les tradicionals tècniques microbiològiques basades en el cultiu en medi artificial. L'últim gran pas en qüestió de tecnologies de bioprospecció han sigut les anomenades "òmiques" que, basades també en tècniques moleculars, permeten l'anàlisi massiu dels microorganismes i els gens (metagenòmica), proteïnes (proteòmica), i metabolits (meta-metabolòmica) que contenen, a partir de qualsevol mostra ambiental.

A dia d'avui, són ja milers els ambients estudiats mitjançant metagenòmica i altres òmiques. Però, aquesta tesi es centra en ambients encara inexplorats amb evidents particularitats que desafien la vida microbiana i que els converteixen en fonts de microorganismes amb potencials aplicacions biotecnològiques.

Objectius

El principal objectiu d'aquesta tesi és la bioprospecció d'ambients naturals i artificials amb aproximacions multi-nivell, és a dir, dirigides a elements biològics de diferent complexitat: gens individuals, espècies, i consorcis microbians. Per tal d'assolir aquest objectiu principal, es plantegen els següents sub-objectius:

- Analitzar les comunitats microbianes associades a diferents ambients naturals i artificials amb tècniques de seqüenciació massiva per tal d'identificar espècies microbianes i gens amb utilitat biotecnològica.
- Caracteritzar espècies microbianes d'interés mitjançant el seu cultiu en medi artificial i la caracterització de la seua activitat biològica.
- Identificar consorcis microbians a una selecció dels ambients estudiants mitjançant un nou mètode d'anàlisi metagenòmica.

Metodologia

Com es desprén dels objectius de la tesi, són vàries les metodologies que s'integren per a la bioprospecció dels ambients seleccionats.

Seqüenciació metagenòmica En primer lloc, es va dur a terme l'anàlisi metagenòmica de les mostres, consistent en l'aïllament del DNA total i la seqüenciació massiva del mateix, amb diferents tecnologies. Posteriorment, es va realitzar l'anàlisi bioinformàtica de les seqüències obtingudes, per a identificar les espècies microbianes presents i el seu contingut gènic.

Cultiu en medi artificial En segon lloc, i com a eina complementària a la metagenòmica, es van aïllar microorganismes d'interés mitjançant el seu cultiu en medi artificial. Aquestos aïllats van ser també caracteritzats mitjançant tècniques microbiològiques i mol.leculars.

Anàlisi bioinformàtic de sèries temporals Finalment, es va desenvolupar una metodologia per a la detecció de consorcis microbians a partir de dades metagenòmiques. Aquesta metodologia es basa en el seguiment dinàmic de comunitats microbianes, per tal de detectar associacions microbianes. Com a resultat, s'obtenen una relació de correlacions i interaccions estadísticament significatives, que es mostren en una única representació tridimensional.

Altres metodologies Amb la fi de completar la informació obtinguda mitjançant les tècniques anteriors, altres aproximacions van ser utilitzades en cada estudi particular. Entre aquestes, destaca la microscopia electrònica, i l'anàlisi transcriptòmica i metaproteòmica.

Resultats i Discussió

La caracterització preliminar de la microbiota associada a diferents ambients naturals (resina, gal.les de l'insecte *Retinia resinella*, i tub digestiu dels insectes *Brithys crini* i *Hyles euphorbiae*) i artificials (plaques solars i residus de café) va ser posible gràcies a la seqüenciació metagenòmica. Com a resultat d'aquesta metodologia, es va tindre accés a la informació taxonòmica de les comunitats microbianes, i per tant, a dades de composició i diversitat de les mateixes. D'aquesta manera es van identificar espècies microbianes amb aplicacions biotecnològiques particulars, com ara la bioremediació o la producció de metabolits d'alt valor industrial. A més a més, en els casos en els quals es va tindre accés als metagenomes complets (i no només al gen 16S/18S rDNA), es va poder analitzar el contingut gènic de les comunitats microbianes, permetent així la identificació d'elements gènics d'interés com ara gens o rutes metabòliques implicades en la degradació de determinats compostos tòxics.

Com a aproximació complementària, es van aïllar soques d'especial interés biotecnològic mitjançant tècniques de cultiu tradicional en medi artificial. Aquesta metodologia va ser essencial per a la bioprospecció dels ambients estudiants, ja que va fer posible la comprovació experimental de les capacitats metabòliques o la resistència de les soques a condicions adverses més enllà de les prediccions bioinformàtiques. D'aquesta manera, va ser posible determinar característiques de soques microbianes difícilment detectables mitjançant tècniques clàssiques de seqüenciació.

Finalment, la posada a punt d'una metodologia per a la detecció de consorcis microbians a partir de dades metagenòmiques dinàmiques va fer posible la identificació d'associacions i interaccions estadísticament significatives en les comunitats microbianes associades a dos dels ambients estudiats (resina de pi i residus de café). Aquesta aproximació està centrada en consorcis microbians en comptes de espècies microbianes individuals, i representa per tant una innovadora eina per a la descripció d'un escenari més realista de les comunitats microbianes, així com del "nucli microbià" relacionat amb els bioprocessos centrals d'un determinat ambient.

Conclusions

Tot i que les tècniques tradicionals d'anàlisi bioinformàtica i de cultiu microbià són fortament informatives per als estudis de bioprospecció microbiana, són encara diversos els aspectes de les comunitats microbianes que s'escapen del punt de mira d'aquestes metodologies. Com a conclusió d'aquesta tesi, es proposa una estratègia holística per a la bioprospecció i l'estudi del microbioma de qualsevol mostra ambiental, que combina les poderoses tecnologies òmiques, amb les noves tècniques de cultiu i la caracterització ecològica de les comunitats en base a les associacions microbianes que s'hi estableixen.

Bibliography

Abate C, Callieri D, Rodríguez E, Garro O. 1996. Ethanol production by a mixed culture of flocculent strains of Zymomonas mobilis and Saccharomyces sp. *Appl Microbiol Biotechnol* **45**:580–3.

Abendroth C, Vilanova C, Günther T, Luschnig O, Porcar M. 2015. Eubacteria and archaea communities in seven mesophile anaerobic digester plants in Germany. *Biotechnol Biofuels* **8**:87.

Adams RI, Bateman AC, Bik HM, Meadow JF. 2015. Microbiota of the indoor environment: a meta-analysis. *Microbiome* **3**:49.

Agapakis CM, Boyle PM, Silver PA. 2012. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat Chem Biol* **8**:527.

Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH, Nielsen PH. 2015. Back to Basics–The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities. *PLoS One* **10**:e0132783.

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**:533–8.

Alcalde M. 2015. Engineering the ligninolytic enzyme consortium. Trends Biotechnol 33:155–62.Antoniou E, Fodelianakis S, Korkakaki E, Kalogerakis N. 2015. Biosurfactant production from marine hydrocarbon-degrading consortia and pure bacterial strains using crude oil as carbon source. *Front Microbiol* **6**:274.

Arias-Cordero E, Ping L, Reichwald K, Delb H, Platzer M, Boland W, et al. 2012. Comparative Evaluation of the Gut Microbiota Associated with the Below- and Above-Ground Life Stages Larvae and Beetles of the Forest Cockchafer, Melolontha hippocastani Oliveira. *PLoS One* **7**:e51557.

Bader J, Mast-Gerlach E, Popović MK, Bajpai R, Stahl U. 2010. Relevance of microbial coculture fermentations in biotechnology. *J Appl Microbiol* **109**:371–387.

Bao M, Wang L, Sun P, Cao L, Zou J, Li Y. 2012. Biodegradation of crude oil using an efficient microbial consortium in a simulated marine environment. *Mar Pollut Bull* **64**:1177–85.

Belda E, Pedrola L, Peretó J, Martínez-Blanch JF, Montagud A, Navarro E, et al. 2011. Microbial diversity in the midguts of field and lab-reared populations of the European corn borer Ostrinia nubilalis. *PLoS One* **6**:e21751.

Berlanga M, Llorens C, Comas J, Guerrero R, Moya A, Peretó J, et al. 2016. Gut Bacterial Community of the Xylophagous Cockroaches Cryptocercus punctulatus and Parasphaeria boleiriana Brusetti. *PLoS One* **11**:e0152400.

Berlemont R, Pipers D, Delsaute M, Angiono F, Feller G, Galleni M, et al. Exploring the Antarctic soil metagenome as a source of novel cold-adapted enzymes and genetic mobile elements. *Rev Argent Microbiol* **43**:94–103.

Bier RL, Bernhardt ES, Boot CM, Graham EB, Hall EK, Lennon JT, et al. 2015. Linking microbial community structure and microbial processes: an empirical and conceptual overview. *FEMS Microbiol Ecol* **91**:fiv113.

Boissière A, Tchioffo MT, Bachar D, Abate L, Marie A, Nsango SE, et al. 2012. Midgut Microbiota of the Malaria Mosquito Vector Anopheles gambiae and Interactions with Plasmodium falciparum Infection Vernick, KD ed. *PLoS Pathog* **8**:e1002742.

Bolaños LM, Rosenblueth M, Castillo-Ramírez S, Figuier-Huttin G, Martínez-Romero E. 2016. Species-specific diversity of novel bacterial lineages and differential abundance of predicted pathways for toxic compound degradation in scorpion gut microbiota. *Environ Microbiol* **18**:1364–1378.

Boucias DG, Cai Y, Sun Y, Lietze V-U, Sen R, Raychoudhury R, et al. 2013. The hindgut lumen prokaryotic microbiota of the termite Reticulitermes flavipes and its responses to dietary lignocellulose composition. *Mol Ecol* **22**:1836–1853.

Brenner K, You L, Arnold FH. 2008. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol* **26**:483–489.

Bright KR, Boone SA, Gerba CP. 2010. Occurrence of bacteria and viruses on elementary classroom surfaces and the potential role of classroom hygiene in the spread of infectious diseases. *J Sch Nurs* **26**:33–41.

Brinkmann N, Martens R, Tebbe CC. 2008. Origin and diversity of metabolically active gut bacteria from laboratory-bred larvae of Manduca sexta Sphingidae, Lepidoptera, Insecta. *Appl Environ Microbiol* **74**:7189–96.

Brock TD. 1999. Robert Koch: a life in medicine and bacteriology. *American Society of Microbiology* doi:10.1128/9781555818272.

Broderick NA, Raffa KF, Goodman RM, Handelsman J. 2004. Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. *Appl Environ Microbiol* **70**:293–300.

Cameron DE, Bashor CJ, Collins JJ. 2014. A brief history of synthetic biology. *Nat Rev Microbiol* **12**:381–390.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci* **108**:4516–4522.

Carlson RH. 2010. Biology is technology: the promise, peril, and new business of engineering life. Harvard University Press.

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD, et al. 2009. Defining DNA-Based Operational Taxonomic Units for Microbial-Eukaryote Ecology. *Appl Environ Microbiol* **75**:5797–5808.

Carrasco P, Pérez-Cobas AE, van de Pol C, Baixeras J, Moya A, Latorre A. 2014. Succession of the gut microbiota in the cockroach Blattella germanica. *Int Microbiol* **17**:99–109.

Carter K. 1987. Essays of Robert Koch. Greenwood Press: Connecticut.Ceja-Navarro JA, Vega FE, Karaoz U, Hao Z, Jenkins S, Lim HC, et al. 2015. Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. *Nat Commun* **6**:7618.

Chandler JA, Morgan Lang J, Bhatnagar S, Eisen JA, Kopp A. 2011. Bacterial Communities of Diverse Drosophila Species: Ecological Context of a Host–Microbe Model System Malik, HS ed. *PLoS Genet* **7**:e1002272.

Chater KF, Biró S, Lee KJ, Palmer T, Schrempf H. 2010. The complex extracellular biology of Streptomyces. *FEMS Microbiol Rev* **34**:171–98.

Cheng C, Xu L, Xu D, Lou Q, Lu M, Sun J. 2016. Does cryptic microbiota mitigate pine resistance to an invasive beetle-fungus complex? Implications for invasion potential. *Sci Rep* **6**:33110.

Cheng J, Pinnell L, Engel K, Neufeld JD, Charles TC. 2014. Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. *J Microbiol Methods* **99**:27–34.

Chistoserdova L. 2010. Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol Lett* **32**:1351–1359.

Colin P-Y, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, et al. 2015. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat Commun* **6**:10008.

Costa D, Mercier A, Gravouil K, Lesobre J, Delafont V, Bousseau A, et al. 2015. Pyrosequencing analysis of bacterial diversity in dental unit waterlines. *Water Res* **81**:223–231.

Daims H, Taylor MW, Wagner M. 2006. Wastewater treatment: a model system for microbial ecology. *Trends Biotechnol* **24**:483–489.

Danhorn T, Young CR, DeLong EF. 2012. Comparison of large-insert, small-insert and pyrosequencing libraries for metagenomic analysis. *ISME J* **6**:2056–66.

Daniel R. 2005. The metagenomics of soil. *Nat Rev Microbiol* **3**:470–478.

de Bruin OM, Birnboim HC, Niemz A, Ferguson T, Boyle D, Yuan S, et al. 2016. A method for assessing efficiency of bacterial cell disruption and DNA release. *BMC Microbiol* **16**:197.

De Vuyst L, Van Kerrebroeck S, Harth H, Huys G, Daniel H-M, Weckx S. 2014. Microbial ecology of sourdough fermentations: Diverse or uniform? *Food Microbiol* **37**:11–29.

Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, et al. 2011. Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol* **77**:1315–24.

Delong EF, Pace NR. 2001. Environmental Diversity of Bacteria and Archae. The three domains and bacterial evolution. *Syst Biol* **50**:470–478.

Diekmann N, Burghartz M, Remus L, Kaufholz A-L, Nawrath T, Rohde M, et al. 2013. Microbial communities related to volatile organic compound emission in automobile air conditioning units. *Appl Microbiol Biotechnol* **97**:8777–8793.

Dillon RJ, Dillon VM. 2004. The gut bacteria of insects: Nonpathogenic Interactions. *Annu Rev Entomol* **49**:71–92.

Dorado-Morales P, Vilanova C, P Garay C, Martí JM, Porcar M. 2015. Unveiling Bacterial Interactions through Multidimensional Scaling and Dynamics Modeling. *Sci Rep* **5**:18396.

Douglas AE. 2013. Microbial brokers of insect-plant interactions revisited. *J Chem Ecol* **39**:952–61.

Eiteman MA, Lee SA, Altman R, Altman E. 2009. A substrate-selective co-fermentation strategy with Escherichia coli produces lactate by simultaneously consuming xylose and glucose. *Biotechnol Bioeng* **102**:822–827.

Ekkers DM, Cretoiu MS, Kielak AM, Elsas JD van. 2012. The great screen anomaly–a new frontier in product discovery through functional metagenomics. *Appl Microbiol Biotechnol* **93**:1005–20.

Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci* **109**:11002–11007.

Faust K, Raes J. 2012. Microbial interactions: from networks to models. *Nat Rev Microbiol* **10**:538–550.

Flores GE, Bates ST, Caporaso JG, Lauber CL, Leff JW, Knight R, et al. 2013. Diversity, distribution and sources of bacteria in residential kitchens. *Environ Microbiol* **15**:588–596.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, et al. 2008. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci* **105**:3805–3810.

Funaro CF, Kronauer DJC, Moreau CS, Goldman-Huertas B, Pierce NE, Russell JA. 2011. Army Ants Harbor a Host-Specific Clade of Entomoplasmatales Bacteria. *Appl Environ Microbiol* **77**:346–350.

Gabaldón T, Peretó J, Montero F, Gil R, Latorre A, Moya A. 2007. Structural analyses of a hypothetical minimal metabolism. *Philos Trans R Soc Lond B Biol Sci* **362**:1751–62.

Genta FA, Dillon RJ, Terra WR, Ferreira C. 2006. Potential role for gut microbiota in cell wall digestion and glucoside detoxification in Tenebrio molitor larvae. *J Insect Physiol* **52**:593–601.

Gerba CP, Wuollet AL, Raisanen P, Lopez GU, Anderson G, Palombo E, et al. 2016. Bacterial contamination of computer touch screens. *Am J Infect Control* **44**:358–360.

Ghosh S, Chowdhury R, Bhattacharya P. 2016. Mixed consortia in bioprocesses: role of microbial interactions. *Appl Microbiol Biotechnol* **100**:4283–4295.

Gibbons SM, Schwartz T, Fouquier J, Mitchell M, Sangwan N, Gilbert JA, et al. 2015. Ecological succession and viability of human-associated microbiota on restroom surfaces. *Appl Environ Microbiol* **81**:765–73.

Gil R, Silva FJ, Peretó J, Moya A. 2004. Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* **68**:518–37.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, et al. 2008. Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities Ahmed, N ed. *PLoS One* **3**:e3042.

Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, et al. 2016. Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**:94–103.

Gosalbes MJ, Lamelas A, Moya A, Latorre A. 2008. The striking case of tryptophan provision in the cedar aphid Cinara cedri. *J Bacteriol* **190**:6026–9.

Goyal G, Tsai S-L, Madan B, DaSilva NA,

Chen W, Demain A, et al. 2011. Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome. *Microb Cell Factories* **10**:124–154.

Goyal G, Nam Y. 2011. Neuronal micro-culture engineering by microchannel devices of cellular scale dimensions. *Biomed Eng Lett* **1**:89–98.

Hayleeyesus SF, Manaye AM. 2014. Microbiological quality of indoor air in university libraries. *Asian Pac J Trop Biomed* **4**:S312-7.

He Y, Xie K, Xu P, Huang X, Gu W, Zhang F, et al. 2013. Evolution of microbial community diversity and enzymatic activity during composting. *Res Microbiol* **164**:189–198.

Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, et al. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**:463–7.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. 2016. A new view of the tree of life. *Nat Microbiol* **1**:16048.

Hulcr J, Rountree NR, Diamond SE, Stelinski LL, Fierer N, Dunn RR. 2012. Mycangia of Ambrosia Beetles Host Communities of Bacteria. *Microb Ecol* **64**:784–793.

Jagmann N, Brachvogel HP, Philipp B. 2010. Parasitic growth of Pseudomonas aeruginosa in co-culture with the chitinolytic bacterium Aeromonas hydrophila. *Environ Microbiol.* **12**:1787–1802.

Jagmann N, Philipp B. 2014. Design of synthetic microbial communities for biotechnological production processes. *J Biotechnol* **184**:209–218.

Jain A, Bediako SH, Henson JM. 2016. Correlation analysis of enzyme activities and deconstruction of ammonia-pretreated switchgrass by bacterial-fungal communities. *Bioresour Technol* **218**:1082–1089.

Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* **158**:1005–1015.

Jung D, Seo E-Y, Epstein SS, Joung Y, Han J, Parfenova V V, et al. 2014. Application of a new cultivation technology, I-tip, for studying microbial diversity in freshwater sponges of Lake Baikal, Russia. *FEMS Microbiol Ecol* **90**:417–23.

Kandel CE, Simor AE, Redelmeier DA. 2014. Elevator buttons as unrecognized sources of bacterial colonization in hospitals. *Open Med* **8**:e81-6.

Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic Escherichia coli. *Nat Rev Microbiol* **2**:123–40.

Kato K, Kozaki S, Sakuranaga M. 1998. Degradation of Lignin Compounds by Bacteria from Termite Guts. *Biotechnol Lett* **20**:459–462.

Kembel SW, Meadow JF, O'Connor TK, Mhuireach G, Northcutt D, Kline J, et al. 2014. Architectural design drives the biogeography of indoor bacterial communities. *PLoS One* **9**:e87093.

Kim H, Rha E, Seong W, Yeom S-J, Lee D-H, Lee S-G. 2016. A Cell–Cell Communication-Based Screening System for Novel Microbes with Target Enzyme Activities. *ACS Synth Biol* acssynbio.5b00287.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**:e1.

Kohler T, Dietrich C, Scheffrahn RH, Brune A. 2012. High-Resolution Analysis of Gut Environment and Bacterial Microbiota Reveals Functional Compartmentation of the Gut in Wood-Feeding Higher Termites Nasutitermes spp. *Appl Environ Microbiol* **78**:4691–4701.

Korenblum E, Jiménez DJ, van Elsas JD. 2016. Succession of lignocellulolytic bacterial consortia bred anaerobically from lake sediment. *Microb Biotechnol* **9**:224–34.

Kotik M, Davidová A, Voříšková J, Baldrian P. 2013. Bacterial communities in tetrachloroethene-polluted groundwaters: A case study. *Sci Total Environ* **454**:517–527.

Krishnan M, Bharathiraja C, Pandiarajan J, Prasanna VA, Rajendhran J, Gunasekaran P. 2014. Insect gut microbiome - An unexploited reserve for biotechnological application. *Asian Pac J Trop Biomed* **4**:S16-21.

Kuhnigk T, König H. 1997. Degradation of dimeric lignin model compounds by aerobic bacteria isolated from the hindgut of xylophagous termites. *J Basic Microbiol* **37**:205–211.

Lagier J-C, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. 2012. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* **18**:1185–93.

Lagkouvardos I, Pukall R, Abt B, Foesel BU, Meier-Kolthoff JP, Kumar N, et al. 2016. The Mouse Intestinal Bacterial Collection miBC provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat Microbiol* **1**:16131.

Lam KN, Cheng J, Engel K, Neufeld JD, Charles TC. 2015. Current and future resources for functional metagenomics. *Front Microbiol* **6**:1196.

Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A. 2011. Serratia symbiotica from the Aphid Cinara cedri: A Missing Link from Facultative to Obligate Insect Endosymbiont Richardson, PM ed. *PLoS Genet* **7**:e1002357.

Larssen P, Dai Y, Collart FR. 2015. Predicting bacterial community assemblages using an artificial neural network approach. *Methods Mol Biol* **1260**:33–43.

Lazcano A. 2011. Natural history, microbes and sequences: shouldn't we look back again to organisms? *PLoS One* **6**:e21334.

Leung MHY, Wilkins D, Li EKT, Kong FKF, Lee PKH. 2014. Indoor-Air Microbiome in an Urban Subway Network: Diversity and Dynamics. *Appl Environ Microbiol* **80**:6760–6770.

Levy R, Borenstein E. 2012. Reverse Ecology: from systems to environments and back. *Adv Exp Med Biol* **751**:329–45.

Li L-L, McCorkle SR, Monchy S, Taghavi S, van der Lelie D. 2009. Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol Biofuels* **2**:10.

Li Q, Zhang B, He Z, Yang X. Distribution and Diversity of Bacteria and Fungi Colonization in Stone Monuments Analyzed by High-Throughput Sequencing. *PLoS One* **11**:e0163287.

Liang J, Peng X, Yin D, Li B, Wang D, Lin Y. 2015. Screening of a microbial consortium for highly simultaneous degradation of lignocellulose and chlorophenols. *Bioresour Technol* **190**:381–7.

Liang X, Fu Y, Tong L, Liu H. 2014. Microbial shifts of the silkworm larval gut in response to lettuce leaf feeding. *Appl Microbiol Biotechnol* **98**:3769–76.

Liebl W, Angelov A, Juergensen J, Chow J, Loeschcke A, Drepper T, et al. 2014. Alternative hosts for functional metagenome analysis. *Appl Microbiol Biotechnol* **98**:8099–109.

Lima WC, Van Sluys M-A, Menck CFM. 2005. Non-gamma-proteobacteria gene islands contribute to the Xanthomonas genome. *OMICS* **9**:160–72.

Lin X-L, Kang Z-W, Pan Q-J, Liu T-X. 2015. Evaluation of five antibiotics on larval gut bacterial diversity of Plutella xylostella Lepidoptera: Plutellidae. *Insect Sci* **22**:619–28.

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, et al. 2015. A new antibiotic kills pathogens without detectable resistance. *Nature* **517**:455–9.

Liu W, Kim HJ, Lucchetta EM, Du W, Ismagilov RF. 2009. Isolation, incubation, and parallel functional testing and identification by FISH of rare microbial single-copy cells from multi-species mixtures using the combination of chemistrode and stochastic confinement. *Lab Chip* **9**:2153–62.

Lo I, Denef VJ, VerBerkmoes NC, Shah MB, Goltsman D, DiBartolo G, et al. 2007. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* **446**:537–541.

Locey KJ, Lennon JT. 2016. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci* 201521291.

Lodge-Ivey SL, Rappe MS, Johnston WH, Bohlken RE, Craig AM. 2005. Molecular analysis of a consortium of ruminal microbes that detoxify pyrrolizidine alkaloids. *Can J Microbiol* **51**:455–465.

Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, et al. 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun* **6**:6372.

Luo F, Devine CE, Edwards EA. 2015. Cultivating Microbial Dark Matter in Benzene-Degrading Methanogenic Consortia. *Environ Microbiol* **18**:2923-36.

Ma L, Kim J, Hatzenpichler R, Karymov MA, Hubert N, Hanan IM, et al. 2014. Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in Human Microbiome Project's Most Wanted taxa. *Proc Natl Acad Sci U S A* **111**:9768–73.

Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, et al. 2011. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480**:368–371.

Madyastha KM, Sridhar GR. 1998. A Novel Pathway for the Metabolism of Caffeine by a Mixed Culture Consortium. *Biochem Biophys Res Commun.* **249**:178–181.

Madyastha KM, Sridhar GR, Vadiraja BB, Madhavi YS. Purification and Partial Characterization of Caffeine Oxidase—A Novel Enzyme from a Mixed Culture Consortium. *Biochem Biophys Res Commun.* **263**:460-4.

Malacrinò A, Schena L, Campolo O, Laudani F, Mosca S, Giunti G, et al. 2016. A Metabarcoding Survey on the Fungal Microbiota Associated to the Olive Fruit Fly. *Microb Ecol.* doi:10.1007/s00248-016-0864-z.

Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG, et al. 2007. Dissecting biological dark matter with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci* **104**:11889–11894.

Martinson VG, Danforth BN, Minckley RL, Ruepell O, Tingek S, Moran NA. 2011. A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol Ecol* **20**:619–628.

Mason CJ, Raffa KF. 2014. Acquisition and structuring of midgut bacterial communities in gypsy moth Lepidoptera: Erebidae larvae. *Environ Entomol* **43**:595–604.

Mateo N, Nader W, Tamayo G. 2011. Encyclopedia of biodiversity. In: Levin, SE ed, Academic Press.McCutcheon JP, McDonald BR, Moran NA. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci* **106**:15394–15399.

McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* **2**:708–18.

McInerney MJ, Struchtemeyer CG, Sieber J, Mouttaki H, Stams AJM, Schink B, et al. 2008. Physiology, Ecology, Phylogeny, and Genomics of Microorganisms Capable of Syntrophic Metabolism. *Ann N Y Acad Sci* **1125**:58–72.

Meadow JF, Altrichter AE, Green JL. 2014. Mobile phones carry the personal microbiome of their owners. *PeerJ* **2**:e447.

Menzel P, Gudbergsdóttir SR, Rike AG, Lin L, Zhang Q, Contursi P, et al. 2015. Comparative Metagenomics of Eight Geographically Remote Terrestrial Hot Springs. *Microb Ecol* **70**:411–24.

Michael GB, Schwarz S. 2016. Antimicrobial resistance in zoonotic nontyphoidal Salmonella: An alarming trend? *Clin Microbiol Infect.* doi:10.1016/j.cmi.2016.07.033.

Milner RJ. 1994. History of Bacillus thuringiensis. *Agric Ecosyst Environ* **49**:9–13.

Mizrahi-Man O, Davenport ER, Gilad Y, Pace N, Woese C, Fox G, et al. 2013. Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs White, BA ed. *PLoS One* **8**:e53608.

Mohr KI, Tebbe CC. 2006. Diversity and phylotype consistency of bacteria in the guts of three bee species Apoidea at an oilseed rape field. *Environ Microbiol* **8**:258–272.

Morales SE, Holben WE. 2011. Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? *FEMS Microbiol Ecol* **75**:2–16.

Moser MJ, DiFrancesco RA, Gowda K, Klingele AJ, Sugar DR, Stocki S, et al. 2012. Thermostable DNA Polymerase from a Viral Metagenome Is a Potent RT-PCR Enzyme Tavis, JE ed. *PLoS One* **7**:e38371.

Moya A, Gil R, Latorre A, Peretó J, Pilar Garcillán-Barcia M, de la Cruz F. 2009. Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol Rev* **33**:225–35.

Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* **9**:218–29.

Nichols D, Cahoon N, Trakhtenberg EM,

Pham L, Mehta A, Belanger A, et al. 2010. Use of ichip for high-throughput in situ cultivation of 'uncultivable' microbial species. *Appl Environ Microbiol* **76**:2445–50.

Nishino T, Tanahashi M, Lin C-P, Koga R, Fukatsu T. 2016. Fungal and bacterial endosymbionts of eared leafhoppers of the subfamily Ledrinae Hemiptera: Cicadellidae. *Appl Entomol Zool* **51**:465–477.

Pace NR, Bull AT, Goodfellow M, Slater JH, Whittaker RH, Zuckerkandl E, et al. 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**:734–40.

Pace NR. 2009. Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev* **73**:565–76.

Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. 2016. Uncovering Earth's virome. *Nature* **536**:425–30.

Paramithiotis S, Gioulatos S, Tsakalidou E, Kalantzopoulos G. 2006. Interactions between Saccharomyces cerevisiae and lactic acid bacteria in sourdough. *Process Biochem* **41**:2429–2433.

Pereira da Fonseca TA, Pessôa R, Felix AC, Sanabani SS. 2016. Diversity of Bacterial Communities on Four Frequently Used Surfaces in a Large Brazilian Teaching Hospital. *Int J Environ Res Public Health* **13**:152.

Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, et al. 2006. A small microbial genome: the end of a long symbiotic relationship? *Science* **314**:312–3.

Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, Garcia JA, et al. 2007. Targeted Access to the Genomes of Low-Abundance Organisms in Complex Microbial Communities. *Appl Environ Microbiol* **73**:3205–3214.

Ponomarova O, Patil KR. 2015. Metabolic interactions in microbial communities: untangling the Gordian knot. *Curr Opin Microbiol* **27**:37–44.

Porcar M, Danchin A, de Lorenzo V. 2015. Confidence, tolerance, and allowance in biological engineering: The nuts and bolts of living things. *BioEssays* **37**:95–102.

Poulsen M, Sapountzis P. 2012. Behind every great ant, there is a great gut. *Mol Ecol* **21**:2054–2057.

Prakash T, Taylor TD. 2012. Functional assignment of metagenomic data: challenges and applications. *Brief Bioinform* **13**:711–27.

Qian M, Tian S, Li X, Zhang J, Pan Y, Yang X. 2006. Ethanol Production From Dilute-Acid Softwood Hydrolysate by Co-Culture. *Appl Biochem Biotechnol* **134**:273–284.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**:59–65.

Qin Y, Druzhinina IS, Pan X, Yuan Z. 2016. Microbially Mediated Plant Salt Tolerance and Microbiome-based Solutions for Saline Agriculture. *Biotechnol Adv.* doi:10.1016/j.biotechadv.2016.08.005.

Rabaey K, Boon N, Siciliano SD, Verhaege M, Verstraete W. 2004. Biofuel Cells Select for Microbial Consortia That Self-Mediate Electron Transfer. *Appl Environ Microbiol* **70**:5373–5382.

Raddadi N, Cherif A, Daffonchio D, Neifar M, Fava F. 2015. Biotechnological applications of extremophiles, extremozymes and extremolytes. *Appl Microbiol Biotechnol* **99**:7907–7913.

Rainford JL, Mayhew PJ. 2015. Diet Evolution and Clade Richness in Hexapoda: A Phylogenetic Study of Higher Taxa. *Am Nat* **186**:777–791.

Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, et al. 2005. Community proteomics of a natural microbial biofilm. *Science* **308**:1915–20.

Rao Q, Rollat-Farnier P-A, Zhu D-T, Santos-Garcia D, Silva FJ, Moya A, et al. 2015. Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly Bemisia tabaci. *BMC Genomics* **16**:226.

Rasmussen LD, Zawadsky C, Binnerup SJ, Oregaard G, Sorensen SJ, Kroer N. 2008. Cultivation of hard-to-culture subsurface mercury-resistant bacteria and discovery of new merA gene sequences. *Appl Environ Microbiol* **74**:3795–803.

Rattray RM, Craig AM. 2007. Molecular Characterization of Sheep Ruminal Enrichments that Detoxify Pyrrolizidine Alkaloids by Denaturing Gradient Gel Electrophoresis and

Cloning. *Microb Ecol* **54**:264–275.

Reid NM, Addison SL, Macdonald LJ, Lloyd-Jones G. 2011. Biodiversity of Active and Inactive Bacteria in the Gut Flora of Wood-Feeding Huhu Beetle Larvae Prionoplus reticularis. *Appl Environ Microbiol* **77**:7000–7006.

Revetta RP, Gomez-Alvarez V, Gerke TL, Santo Domingo JW, Ashbolt NJ. 2016. Changes in bacterial composition of biofilm in a metropolitan drinking water distribution system. *J Appl Microbiol* **121**:294–305.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**:431–7.

Rodriguez-Valera F, Martin-Cuadrado A-B, López-Pérez M. 2016. Flexible genomic islands as drivers of genome evolution. *Curr Opin Microbiol* **31**:154–160.

Roh JY, Choi JY, Li MS, Jin BR, Je YH. 2007. Bacillus thuringiensis as a specific, safe, and effective tool for insect pest control. *J Microbiol Biotechnol* **17**:547–59.

Ross AA, Neufeld JD, Porter J, Evans C, Smith W, Johnston E, et al. 2015. Microbial biogeography of a university campus. *Microbiome* **3**:66.

Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM, et al. 2012. High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* **13**:53.

Russell JA, Moreau CS, Goldman-Huertas B, Fujiwara M, Lohman DJ, Pierce NE. 2009. Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proc Natl Acad Sci* **106**:21236–21241.

Salem H, Kreutzer E, Sudakaran S, Kaltenpoth M. 2013. Actinobacteria as essential symbionts in firebugs and cotton stainers Hemiptera, Pyrrhocoridae. *Environ Microbiol* **15**:1956–1968.

Sandmann G. 2015. Carotenoids of biotechnological importance. *Adv Biochem Eng Biotechnol* **148**:449–67.

Santero E, Floriano B, Govantes F. 2016. Harnessing the power of microbial metabolism. *Curr Opin Microbiol* **31**:63–69.

Savage AM, Hills J, Driscoll K, Fergus DJ, Grunden AM, Dunn RR. 2016. Microbial diversity of extreme habitats in human homes. *PeerJ* **4**:e2376.

Schauer C, Thompson C, Brune A. 2014. Pyrotag sequencing of the gut microbiota of the cockroach Shelfordella lateralis reveals a highly dynamic core but only limited effects of diet on community structure. *PLoS One* **9**:e85861.

Schloss PD, Handelsman J. 2005. Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Appl Environ Microbiol* **71**:1501–1506.

Schloss PD, Westcott SL. 2011. Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. *Appl Environ Microbiol* **77**:3219–3226.

Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* **4**. doi:10.1038/ncomms3304.

Shin H-D, McClendon S, Vo T, Chen RR. 2010. Escherichia coli Binary Culture Engineered for Direct Fermentation of Hemicellulose to a Biofuel. *Appl Environ Microbiol* **76**:8150–8159.

Simon C, Daniel R. 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biotechnol* **85**:265–76.

Sloan DB, Moran NA. 2012. Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids. *Mol Biol Evol* **29**:3781–3792.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. 2006. Microbial diversity in the deep sea and the underexplored rare biosphere. *Proc Natl Acad Sci U S A* **103**:12115–20.

Solden L, Lloyd K, Wrighton K. 2016. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr Opin Microbiol* **31**:217–226.

Staley JT, Konopka A. 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39**:321–46.

Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, et al. 2013.

Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota Von Mering, C ed. *PLoS Comput Biol* **9**:e1003388.

Sudakaran S, Salem H, Kost C, Kaltenpoth M. 2012. Geographical and ecological stability of the symbiotic mid-gut microbiota in European firebugs, Pyrrhocoris apterus Hemiptera, Pyrrhocoridae. *Mol Ecol* **21**:6134–6151.

Sul WJ, Park J, Quensen JF, Rodrigues JLM, Seliger L, Tsoi T V., et al. 2009. DNA-Stable Isotope Probing Integrated with Metagenomics for Retrieval of Biphenyl Dioxygenase Genes from Polychlorinated Biphenyl-Contaminated River Sediment. *Appl Environ Microbiol* **75**:5501–5506.

Tang X, Freitak D, Vogel H, Ping L, Shao Y, Cordero EA, et al. 2012. Complexity and variability of gut commensal microbiota in polyphagous lepidopteran larvae. *PLoS One* **7**:e36978.

Triadó-Margarit X, Veillette M, Duchaine C, Talbot M, Amato F, Minguillón MC, et al. 2016. Bioaerosols in the Barcelona subway system. *Indoor Air.* doi:10.1111/ina.12343.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. *Nature* **449**:804–10.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**:37–43.

Valenzuela L, Chi A, Beard S, Orell A, Guiliani N, Shabanowitz J, et al. 2006. Genomics, metagenomics and proteomics in biomining microorganisms. *Biotechnol Adv* **24**:197–211.

Vanwonterghem I, Evans PN, Parks DH, Jensen PD, Woodcroft BJ, Hugenholtz P, et al. 2016. Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat Microbiol* **1**:16170.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**:66–74.

Vilanova C, Marco G, Domínguez-Escribà L, Genovés S, Sentandreu V, Bataller E, et al. 2012. Bacteria from acidic to strongly alkaline insect midguts: Potential sources of extreme cellulolytic enzymes. *Biomass and Bioenergy* **45**:288–294.

Vriesekoop F, Russell C, Alvarez-Mayorga B, Aidoo K, Yuan Q, Scannell A, et al. 2010. Dirty Money: An Investigation into the Hygiene Status of Some of the World's Currencies as Obtained from Food Outlets. *Foodborne Pathog Dis* **7**:1497–1502.

Wang Y, Gilbreath TM, Kukutla P, Yan G, Xu J. 2011. Dynamic Gut Microbiome across Life History of the Malaria Mosquito Anopheles gambiae in Kenya Leulier, F ed. *PLoS One* **6**:e24767.

Wanner G, Vogl K, Overmann J. 2008. Ultrastructural Characterization of the Prokaryotic Symbiosis in Chlorochromatium aggregatum. *J Bacteriol* **190**:3721–3730.

Ward C, Nolan AM, O'Hanlon K, McAree T, Barron N, McHale L, et al. 1995. Production of ethanol at 45ºC on starch-containing media by mixed cultures of the thermotolerant, ethanol-producing yeast Kluyveromyces marxianus IMB3 and the thermophilic filamentous fungus Talaromyces emersonii CBS 814.70. *Appl Microbiol Biotechnol* **43**:408–411.

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. 2016. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J.* doi:10.1038/ismej.2016.45.

Woese CR. 1987. Bacterial evolution. *Microbiol Rev* **51**:221–71.

Wong CNA, Ng P, Douglas AE. 2011. Low-diversity bacterial community in the gut of the fruitfly Drosophila melanogaster. *Environ Microbiol* **13**:1889–1900.

Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H, et al. 2009. Assembling the Marine Metagenome, One Cell at a Time. *PLoS One* **4**:1034–1039.

Wright ES, Vetsigian KH. 2016. Inhibitory interactions promote frequent bistability among competing bacteria. *Nat Commun* **7**:11274.

Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, et al. 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**:1056–1060.

Xiang H, Wei G-F, Jia S, Huang J, Miao X-X, Zhou Z, et al. 2011. Microbial communities in the larval midgut of laboratory and field populations of cotton bollworm Helicoverpa armigera. *Can J Microbiol.* doi: abs/10.1139/w06064.V01Uw-mLS72.

Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J, et al. 2011. Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J* **5**:414–26.

Yugi K, Kubota H, Hatano A, Kuroda S. 2016. Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple Omic Layers. *Trends Biotechnol* **34**:276–90.

Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen W-H, et al. 2009. Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326**:1263–8.

Zhou J, Huang H, Meng K, Shi P, Wang Y, Luo H, et al. 2009. Molecular and biochemical characterization of a novel xylanase from the symbiotic Sphingobacterium sp. TN19. *Appl Microbiol Biotechnol* **85**:323–33.

Zomorrodi AR, Islam MM, Maranas CD. 2014. d-OptCom: Dynamic multi-level and multi-objective metabolic modeling of microbial communities. *ACS Synth Biol* **3**:247–57.

Appendix A

Original publication reprints

OPEN ACCESS Freely available online

PLOS ONE

Selecting Microbial Strains from Pine Tree Resin: Biotechnological Applications from a Terpene World

Cristina Vilanova[1], Maria Marín[1], Joaquín Baixeras[1], Amparo Latorre[1,2], Manuel Porcar[1,3]*

1 Universitat de València (Cavanilles Institute of Biodiversity and Evolutive Biology), Valencia, Spain, **2** Unidad Mixta de Investigación en Genómica y Salud, Centro Superior de Investigación en Salud Pública (CSISP), València, Spain, **3** Fundació General de la Universitat de València, València, Spain

Abstract

Resin is a chemical and physical defensive barrier secreted by many plants, especially coniferous trees, with insecticidal and antimicrobial properties. The degradation of terpenes, the main components accounting for the toxicity of resin, is highly relevant for a vast range of biotechnological processes, including bioremediation. In the present work, we used a resin-based selective medium in order to study the resin-tolerant microbial communities associated with the galls formed by the moth *Retinia resinella*; as well as resin from *Pinus sylvestris* forests, one of the largest ecosystems on Earth and a yet-unexplored source of terpene-degrading microorganisms. The taxonomic and functional diversity of the cultivated, resin-tolerant fraction of the whole microbiota were unveiled by high-throughput sequencing, which resulted in the detection of more than 40 bacterial genera among the terpene-degrading microorganisms, and a range of genes involved in the degradation of different terpene families. We further characterized through culture-based approaches and transcriptome sequencing selected microbial strains, including *Pseudomonas* sp., the most abundant species in both environmental resin and *R. resinella* resin-rich galls, and three fungal species, and experimentally confirmed their ability to degrade resin and also other terpene-based compounds and, thus, their potential use in biotechnological applications involving terpene catabolism.

Citation: Vilanova C, Marín M, Baixeras J, Latorre A, Porcar M (2014) Selecting Microbial Strains from Pine Tree Resin: Biotechnological Applications from a Terpene World. PLOS ONE 9(6): e100740. doi:10.1371/journal.pone.0100740

Editor: Kostas Bourtzis, International Atomic Energy Agency, Austria

Received December 11, 2013; **Accepted** May 29, 2014; **Published** June 27, 2014

Funding: Financial support was provided by grant BFU2012-39816-C02-01 (Ministerio de Ciencia e Innovación, Spain) co-financed by FEDER funds to Amparo Latorre and CGL2008-00605 (Ministerio de Ciencia e Innovación) to Joaquín Baixeras. Cristina Vilanova is a recipient of a FPU fellowship from the Ministerio de Educación (Spain). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Strain PS, deposited in the Spanish Type Culture Collection (CECT) under reference number CECT 8327, has been found by the authors to hold not only for scientific publication, but also for patenting (Application number P201300612 at Spanish Office of Patents and Trademarks, OEPM). The authors have prepared the patent and the registration in collaboration with the Research Transfer Office (OTRI) of the University of Valencia (contact person, Silvia Bort: silvia.bort@uv.es). This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

* Email: manuel.porcar@uv.es

Introduction

Coniferous resin is a complex mixture of secondary metabolites. Resin protects injured tree tissues from phytophagous insects [1] and plant pathogens [2], [3], [4], [5]. Terpenoids, flavonoids, and fatty acids are the main components of resin [6], [7]. Among these, terpenes (terpenoids containing a variable number of complete repetitions of isoprene) are the best characterized metabolites because they can be easily identified with techniques such as gas chromatography. Monoterpenes, sesquiterpenes, and diterpenes (containing two, three, and four repetitions of isoprene, respectively) are the most abundant families of terpenes in pine tree resin [6]. Since most of these *de novo* synthesized compounds display antibacterial and antifungal properties, they are considered to be phytoalexins accounting for the toxicity of resin [8], [9], [10], [11].

Many terpenes display interesting features for the chemical industry, specifically in the production of fragrances, essential oils, and food additives [12]. They are also valuable molecules in medicine because of their cytotoxic, cardiotonic, and anti-inflammatory properties [13], [14]. However, terpenes are one of the main pollutants in effluents of pulp mill industries [15], [16], and terpene-based materials such as tire rubber or latex account for tons of solid waste per year. Thus, these molecules are among the main chemical targets for bioremediation [17]. Given all their applications in biotechnology, bioprospection aiming to identify single genes, gene networks, and microorganisms able to transform or catabolize these molecules is a key starting point in the development of a range of biotechnological applications. The toxicity of terpenes and complexity of their chemical structure hinder their degradation by microorganisms, and few studies describe the ability of microorganisms to use individual components of resin as a sole carbon source [18] or to biotransform particular terpene molecules [19]. The ability of some insects to overcome tree's defensive compounds [20] and terpenes in particular [21] has been previously attributed to their association with microorganisms.

A yet unexplored source of potential terpene-degrading microorganisms is related to the insect *Retinia resinella* Linnaeus. The larvae of *R. resinella* feed on young twigs of the Scotch pine *Pinus sylvestris* [22], [23]. They cause small wounds that induce the secretion of resin, which is manipulated by the larva to construct a nodule-like resin capsule –commonly known as ''resin gall''- as a hard protective cocoon. A single larva develops inside this resin blister for nearly two years, completely isolated from the external environment by its terpene-rich shelter [24]. *R. resinella*'s gut microbiota has not been studied to date, and represents a potential reservoir of resin-tolerant microorganisms. This work, however,

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 June 2016
doi: 10.3389/fmicb.2016.01005

The Generalist Inside the Specialist: Gut Bacterial Communities of Two Insect Species Feeding on Toxic Plants Are Dominated by *Enterococcus* sp.

Cristina Vilanova[1,2], Joaquín Baixeras[1], Amparo Latorre[1,2,3] and Manuel Porcar[1,2]**

[1] Cavanilles Institute of Biodiversity and Evolutionary Biology, Universitat de València, Valencia, Spain, [2] Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, Valencia, Spain, [3] Unidad Mixta de Investigación en Genómica y Salud, Centro Superior de Investigación en Salud Pública, Valencia, Spain

OPEN ACCESS

Edited by:
Mark Alexander Lever, ETH Zürich, Switzerland

Reviewed by:
Virginia Helena Albarracín, CONICET, Argentina
Jeremy Dodsworth, California State University, San Bernardino, USA

***Correspondence:**
Manuel Porcar
manuel.porcar@uv.es;
Amparo Latorre
amparo.latorre@uv.es

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *01 April 2016*
Accepted: *13 June 2016*
Published: *28 June 2016*

Citation:
Vilanova C, Baixeras J, Latorre A and Porcar M (2016) The Generalist Inside the Specialist: Gut Bacterial Communities of Two Insect Species Feeding on Toxic Plants Are Dominated by Enterococcus sp. Front. Microbiol. 7:1005. doi: 10.3389/fmicb.2016.01005

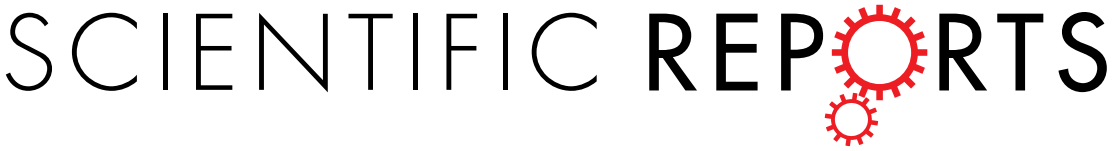
Some specialist insects feed on plants rich in secondary compounds, which pose a major selective pressure on both the phytophagous and the gut microbiota. However, microbial communities of toxic plant feeders are still poorly characterized. Here, we show the bacterial communities of the gut of two specialized Lepidoptera, *Hyles euphorbiae* and *Brithys crini*, which exclusively feed on latex-rich *Euphorbia* sp. and alkaloid-rich *Pancratium maritimum*, respectively. A metagenomic analysis based on high-throughput sequencing of the 16S rRNA gene revealed that the gut microbiota of both insects is dominated by the phylum Firmicutes, and especially by the common gut inhabitant *Enterococcus* sp. *Staphylococcus* sp. are also found in *H. euphorbiae* though to a lesser extent. By scanning electron microscopy, we found a dense ring-shaped bacterial biofilm in the hindgut of *H. euphorbiae*, and identified the most prominent bacterium in the biofilm as *Enterococcus casseliflavus* through molecular techniques. Interestingly, this species has previously been reported to contribute to the immobilization of latex-like molecules in the larvae of *Spodoptera litura*, a highly polyphagous lepidopteran. The *E. casseliflavus* strain was isolated from the gut and its ability to tolerate natural latex was tested under laboratory conditions. This fact, along with the identification of less frequent bacterial species able to degrade alkaloids and/or latex, suggest a putative role of bacterial communities in the tolerance of specialized insects to their toxic diet.

Keywords: : lepidoptera, gut communities, metagenomics, *Enterococcus* sp., secondary metabolites

INTRODUCTION

Plants have biochemical and molecular mechanisms to defend themselves from insects attack. Among those, plants produce a vast range of secondary metabolites with anti-herbivore effects, which are produced either constitutively or in response to tissue damage (War et al., 2012). Some plant biochemicals are toxic, repellent, or antinutritive for herbivores. Among these compounds,

www.nature.com/scientificreports



OPEN

A highly diverse, desert-like microbial biocenosis on solar panels in a Mediterranean city

Received: 04 January 2016
Accepted: 13 June 2016
Published: 05 July 2016

Pedro Dorado-Morales[1,*], Cristina Vilanova[1,2,†,*], Juli Peretó[1,2,3,†], Francisco M. Codoñer[4], Daniel Ramón[5] & Manuel Porcar[1,2,†]

Microorganisms colonize a wide range of natural and artificial environments although there are hardly any data on the microbial ecology of one the most widespread man-made extreme structures: solar panels. Here we show that solar panels in a Mediterranean city (Valencia, Spain) harbor a highly diverse microbial community with more than 500 different species per panel, most of which belong to drought-, heat- and radiation-adapted bacterial genera, and sun-irradiation adapted epiphytic fungi. The taxonomic and functional profiles of this microbial community and the characterization of selected culturable bacteria reveal the existence of a diverse mesophilic microbial community on the panels' surface. This biocenosis proved to be more similar to the ones inhabiting deserts than to any human or urban microbial ecosystem. This unique microbial community shows different day/night proteomic profiles; it is dominated by reddish pigment- and sphingolipid-producers, and is adapted to withstand circadian cycles of high temperatures, desiccation and solar radiation.

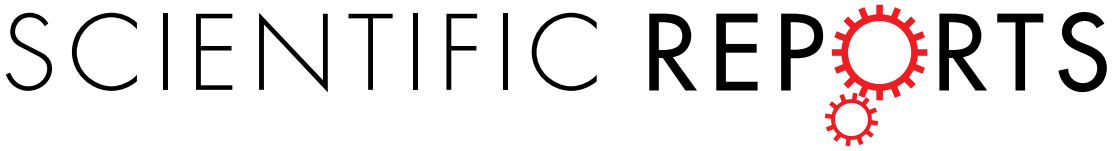
Today, photovoltaic panels cover around 4000 square kilometers, and are forecasted to be the world's main electricity source by 2050 (http://www.epia.org). Solar panels are unique biotopes characterized by a smooth flat glass or glass-like surface, minimum water retention capacity and maximum sunlight exposure, all of which determine circadian and annual peaks of irradiation, desiccation and heat. Extreme natural habitats such as thermal vents, mountain plateaus or hyper arid deserts are known to host microbial biocenoses adapted to those particular selection pressures[1–3]; and artificial or humanized environments, such as industrial reactors[4], radioactive waste[5] or oil spills[6] are also colonizable by specialized microorganisms. Among artificial environments, indoor biomes constitute around 0.5% of the ice-free land area, a surface comparable to the subtropical coniferous forest[7]. Geography and building type, among other factors, structure indoor microbes, which are inoculated by human skin and outdoor air[8] and even by indoor plants[9]. One of the recently studied indoor habitats is the subway. A recent study of the aerosol microbial communities of the New York City subway platforms[10] revealed that such communities were a mixture of soil, environmental water, and human skin commensal bacteria. In a similar work in the Hong Kong subway networks[11], researchers found that each subway line harbored a different phylogenetic community, depending on architectural characteristics, nearby (outdoor) microbiomes, as well as connectedness with other lines.

Most of the so-called built environment studies have so far focused in the indoor biomes but less attention has been paid to artificial, outdoor environments. A recent example of those is an ambitious city-scale metagenomics screening on the bacterial diversity of New York (including both indoor and outdoor biomes), which found a very high diversity of microorganisms -roughly 1700 bacterial taxa-, half of which did not match any known organism[12]. However, many artificial outdoor environments remains unexplored to date.

In the present work, we aimed at studying solar panels, an outdoor artificial environment whose role as biotope had not previously been reported. The goal of our study was to identify possible microbial communities thriving in the harsh conditions of the panels' surfaces, and to determine to which extent this highly irradiated

[1]Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Paterna 46980, Spain. [2]Darwin Bioprospecting Excellence, SL., Parc Cientific Universitat de València, Paterna, Valencia, Spain. [3]Departament de Bioquímica i Biologia Molecular, Universitat de València, Burjassot 46100, Spain. [4]Lifesequencing S.L., Parc Científic Universitat de València, Paterna, Spain. [5]Biopolis S.L., Parc Cientific Universitat de València, Paterna, Valencia, Spain. [†]Present address: Institute for Integrative Systems Biology (I2SysBio, Universitat de València-CSIC). Parc Cientific Universitat de València, Paterna, Valencia, Spain. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.P. (email: manuel.porcar@uv.es)

www.nature.com/scientificreports



OPEN

Unveiling Bacterial Interactions through Multidimensional Scaling and Dynamics Modeling

Received: 06 May 2015
Accepted: 17 November 2015
Published: 16 December 2015

Pedro Dorado-Morales[1], Cristina Vilanova[1], Carlos P. Garay[3], Jose Manuel Martí[3] & Manuel Porcar[1,2]

We propose a new strategy to identify and visualize bacterial consortia by conducting replicated culturing of environmental samples coupled with high-throughput sequencing and multidimensional scaling analysis, followed by identification of bacteria-bacteria correlations and interactions. We conducted a proof of concept assay with pine-tree resin-based media in ten replicates, which allowed detecting and visualizing dynamical bacterial associations in the form of statistically significant and yet biologically relevant bacterial consortia.

There is a growing interest on disentangling the complexity of microbial interactions in order to both optimize reactions performed by natural consortia and to pave the way towards the development of synthetic consortia with improved biotechnological properties[1,2]. Despite the enormous amount of metagenomic data on both natural and artificial microbial ecosystems, bacterial consortia are not necessarily deduced from those data. In fact, the flexibility of the bacterial interactions, the lack of replicated assays and/or biases associated with different DNA isolation technologies and taxonomic bioinformatics tools hamper the clear identification of bacterial consortia. We propose here a holistic approach aiming at identifying bacterial interactions in laboratory-selected microbial complex cultures. The method requires multi-replicated taxonomic data on independent subcultures, and high-throughput sequencing-based taxonomic data. From this data matrix, randomness of replicates can be verified, linear correlations can be visualized and interactions can emerge from statistical correlations. The whole procedure can be summarized as follows:

1. Taxonomic data from multi-replicated, independent assays is obtained.
2. Fluctuation scaling of replicates, i.e. slope of Taylor's law, is validated for the samples of the first time step against the expectation of a Poisson-distributed selection.
3. Linear correlation coefficients are determined, converted into distances and displayed by multidimensional scaling.
4. Interaction matrix is inferred from the correlation matrix using the discrete Lotka-Volterra model with relative abundances[3].

As a proof of concept, we chose to analyze independent laboratory cultures grown on a natural, recalcitrant compound. A range of recalcitrant substrates, from synthetic dyes to polycyclic aromatic hydrocarbons, polychlorinated biphenyls, and other organic pollutants can be efficiently degraded by mixed microbial cultures combining catabolic enzyme activities of individual consortium members[4]. Therefore, a carbon source that requires complex pathways for degradation is expected to shape the structure of the microbial community and behave as a strong selection pressure towards the establishment of microbial consortia with biodegradation properties.

In a previous work, we characterized the cultivable microbial communities present in coniferous resin, and detected a rather diverse microbial community, including several fungal and bacterial strains with potential use in bioremediation as deduced from their ability for the degradation of different terpenic compounds[5]. In the proof of concept we present here, we used pine-tree resin as the main carbon source of a resin-rich semi-synthetic medium (prepared as described in our previous report[5]), which we inoculated with environmental resin in ten independent subcultures. We designed this multi-replicate experimental evolution assay to address three main issues: i) Time-course variation in biodiversity fate (does sub-culturing in a recalcitrant compound lead

[1]Cavanilles Institute of Biodiversity and Evolutionary Biology (Universitat de València), 46020 Valencia, Spain. [2]Fundació General de la Universitat de València, Spain. [3]Instituto de Física Corpuscular, CSIC-UVEG, 46071, Valencia, Spain. Correspondence and requests for materials should be addressed to M.P. (email: manuel.porcar@uv.es)

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

The coffee-machine bacteriome: biodiversity and colonisation of the wasted coffee tray leach

Received: 12 June 2015
Accepted: 16 October 2015
Published: 23 November 2015

Cristina Vilanova[1], Alba Iglesias[1] & Manuel Porcar[1,2]

Microbial communities are ubiquitous in both natural and artificial environments. However, microbial diversity is usually reduced under strong selection pressures, such as those present in habitats rich in recalcitrant or toxic compounds displaying antimicrobial properties. Caffeine is a natural alkaloid present in coffee, tea and soft drinks with well-known antibacterial properties. Here we present the first systematic analysis of coffee machine-associated bacteria. We sampled the coffee waste reservoir of ten different Nespresso machines and conducted a dynamic monitoring of the colonization process in a new machine. Our results reveal the existence of a varied bacterial community in all the machines sampled, and a rapid colonisation process of the coffee leach. The community developed from a pioneering pool of enterobacteria and other opportunistic taxa to a mature but still highly variable microbiome rich in coffee-adapted bacteria. The bacterial communities described here, for the first time, are potential drivers of biotechnologically relevant processes including decaffeination and bioremediation.

Caffeine (1,3,7-trimethylxanthine) is a natural alkaloid with anti-herbivorous properties produced by *Coffea arabica* and *Coffea canephora*, which is present in a wide range of beverages including coffee, tea and soft –mainly cola- drinks. Caffeine is a well-known bioactive compound with stimulating effects on the central nervous system, along with a range of other potentially positive effects on human health. Such effects range from enhancing long-term memory[1], improving sports performance[2], inactivating breast cancer-associated myofibroblasts[3], reducing the risk of type 2 diabetes mellitus[4], or even reducing the risk of mortality among coffee consumers[5]. However, it should also be noted that caffeine intake habits are often linked to living styles and thus it is difficult to draw epidemiological conclusions linking caffeine intake to health.

Caffeine may be an environmental pollutant[6], and has also been proposed as an easily detectable marker for untreated wastewater[7]. Indeed, the presence of caffeine in the natural environment is one of the best indicators of anthropogenic contamination. Caffeine bioactivity on human health and the environment has led to the development of processes to remove caffeine, either to yield decaffeinated products or to degrade environmental caffeine. Decaffeination, namely the caffeine removal process, is used industrially to produce low-caffeine beverages and can also be implemented for environmental remediation. One intriguing option is to use microorganisms to perform decaffeination processes. Some microorganisms have been reported to degrade caffeine, such as *Aspergillus tamarii*[8], *Trichosporon asahii*[9], *Pseudomonas* sp.[10,11] or *P. putida*[12].

In this work, we report a diversity analysis aiming to characterise bacterial communities growing on coffee leach waste, using high througput sequencing, culturing, and electron microscocopy techniques. To achieve this goal, we have chosen one of the most widespread coffee preparation systems, Nespresso, due to its popularity and standard nature. In fact, Nespresso-compatible machines are highly standardized coffee making devices (same capsule type, same basic design, same pressure: 19 bars), and they represent a unique oportunity for a massive biological screening. Here we present the first attempt to do so.

[1]Universitat de València (Cavanilles Institute of Biodiversity and Evolutionary Biology), 46020 Valencia, Spain. [2]Fundació General de la Universitat de València, Spain. Correspondence and requests for materials should be addressed to M.P. (email: manuel.porcar@uv.es)

PUBLISHED: 26 JULY 2016 | ARTICLE NUMBER: 16101 | DOI: 10.1038/NMICROBIOL.2016.101

comment

Are multi-omics enough?

Cristina Vilanova and Manuel Porcar

Multi-omic techniques are often seen as the future of microbiome studies. We argue that recent strategies for simplifying complex omic-derived data will need to be combined with improved cultivation techniques to pave the way towards a more targeted approach for understanding microbial communities.

In 1665, Robert Hooke was the first human to observe a microorganism. Two centuries elapsed between these first sightings of microbial cells and the isolation of bacterial species in pure culture with artificial media. The identification and study of microbial species relied then on their cultivation and phenotypic characterization for centuries, even though it was soon discovered that only ~1% of the microorganisms present in environmental samples were cultivable[1]. In 2002, the birth of metagenomic sequencing facilitated an explosion of microbiome studies, enabling much of a microbial community to be identified in a single experiment. A few years later, other omic technologies arose (metatranscriptomics, metaproteomics, meta-metabolomics, etc.) to complement metagenomics, expanding the landscape of tools available for the high-throughput analysis of complex biomes.

Even though the integration of multi-omic data into the 'trans-omic' pipeline[2] is able to generate unprecedentedly complete results, the analysis of such datasets often ignores the search for ecologically relevant conclusions, and focuses, instead, on getting increasingly exhaustive catalogues of species, expressed genes, or metabolites. As a consequence, multi-omics has the risk of increasing the complexity it is supposed to address. As an example of 'trees hidden by the forest', biological interactions among members of a microbial community often remain buried beneath the massive multi-omic datasets. It would be wrong to assume that because multi-omics is used, relevant biological interactions will emerge. It is known that microorganisms are naturally assembled into interacting communities, and that these community structures are directly linked to microbial processes. Therefore, the identification of key players in a taxonomically complex sample is necessary to understand the ecology of a particular habitat. This is especially true when it comes to the study of biotechnologically relevant microbial consortia, such as those present in the biogas industry[3], where engineers tend to consider their fermenters as 'black boxes' that produce biogas. Omic approaches certainly help to shed light on the taxonomic or functional complexity of a fermenting biomass, but such a strategy might fail to identify the ecological and economical core of the process. In plain words, the challenge is to reduce complexity to improve understanding.

We argue that a core of ecological conclusions has to emerge beyond the combination of the complex information obtained through multi-omics studies: multi-omic analyses should yield more than the sum of their parts, as suggested by some ecologists[4]. Even though there is still a long way to go, the relatively recent birth of so-called reverse ecology might prove helpful for the prediction of interactions among species and for improving our understanding of metabolic networks in the context of their natural habitats[5]. Moreover, alternative approaches to predict ecologically relevant information have emerged in the last couple of years. For instance, artificial neural networks have been proposed for modelling microbial communities as functions of environmental parameters and intra-microbial interactions[6]; the dynamics and composition of microbial consortia have been unveiled by measuring temporal variations in interspecies metabolic interactions[7]; and new approaches for visualizing microbial consortia through mathematical modelling and multi-dimensional scaling have been recently reported[8]. Nevertheless, the integration of experimental multi-omic data with predictive mathematical models based on mechanistic understanding is still considered a missing link in microbial ecology[9].

We envision a bright future in microbial ecology, where multi-omic databases will soon be analysed with approaches that are able to condense the gigabytes of information into simpler, ecologically relevant, conclusions. Does this mean that we will never again need to culture microorganisms? We strongly believe that culturing is needed more today than ever before. In fact, the mere identification and characterization of bacterial species or consortia through multi-omics may not be enough when

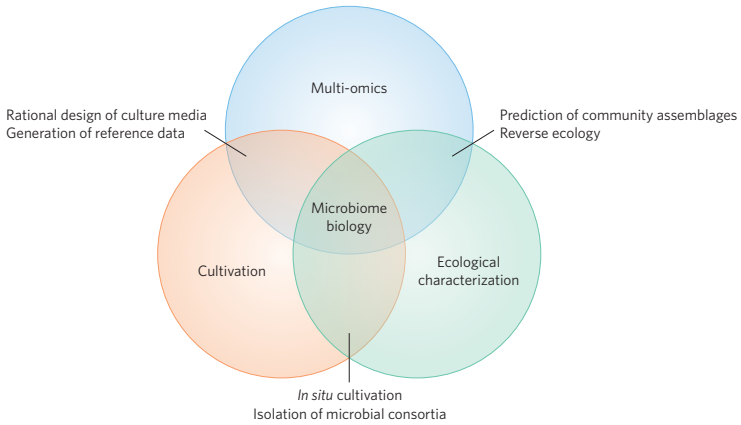


Figure 1 | Microbiome biology as the interphase among multi-omic data, ecological characterization, and cultivation-dependent techniques.

Appendix B

Supplementary tables and figures

Publication I

Supplementary Methods

18S rDNA amplification and sequencing

The DNA from fungal isolates was extracted with the Power Soil DNA Isolation kit (MO BIO Laboratories) according to the manufacturer's instructions and directly used as PCR template. A 800 bp fragment of the 18S rDNA genes was amplified by PCR with universal primers EF3 (5′-TCCTCTAAATGACCAAGTTTG-3′) and EF4 (5′-GGAAGGG[G/A]TGTATTTATTAG-3′) (Smit *et al.*, 1999). The PCR program was as follows: an initial denaturing step at 95 ºC for 300 s, followed by 35 cycles of denaturing, annealing and extension (95 ºC, 30 s; 48 ºC, 30 s; and 72 ºC, 90 s) and a final extension step at 72 ºC for 480 s. PCR amplicons were purified by the High Pure PCR Product Purification Kit (Roche Diagnostics GmbH, Mannheim, Germany) and sequencing was carried out with the ABI PRISM BigDye Terminator v3.1 system (Applied Biosystems) on an ABI 3730 automated sequencer. PCR products were sequenced in both senses with the EF3 and EF4 primers. Sequences were verified and both strands assembled using the STADEN package. Sequence taxonomy was attributed with BLASTN searches against the RefSeq database of the NCBI.

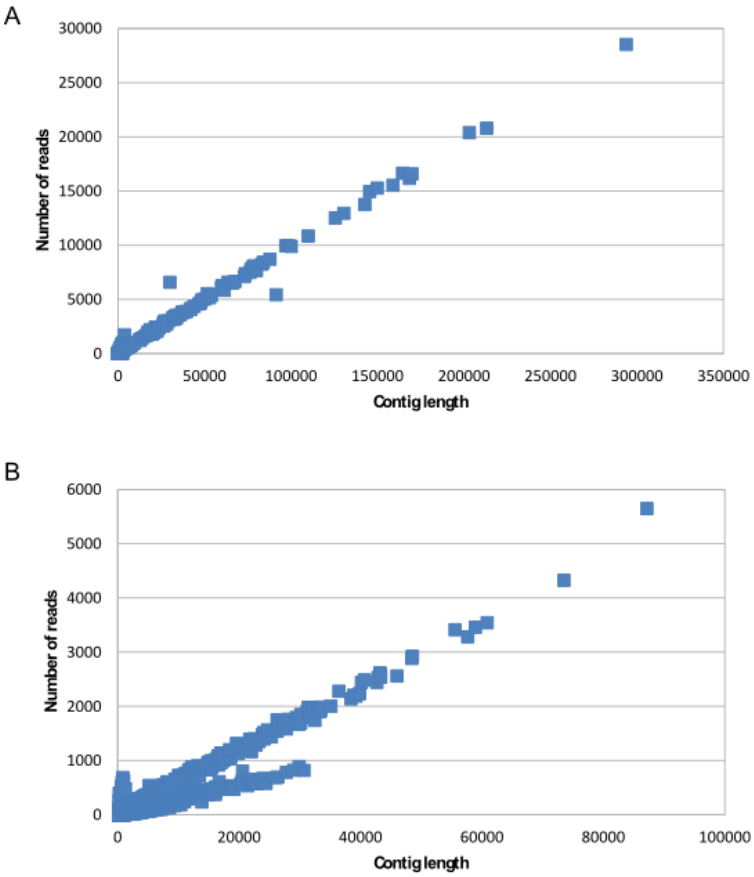
Quantification of resin degradation by fungal isolates and PS

Fungal isolates able to grow on RM were identified through 18S rDNA sequencing as described above. Each isolate was grown at 30 ºC in sterile Erlenmeyer 100 mL flasks with 50 mL of RM broth containing 0.1% w/v resin and shaken at 250 rpm. Water-insoluble resin formed a white colloidal suspension in the RM liquid medium. As fungi degraded resin particles, the medium became more transparent, in such a way that resin degradation was assessed by measuring the optical density (OD) of the medium at 600 nm. As fungi degraded resin particles, the medium became more transparent, in such a way that resin degradation was assessed by measuring the optical density (OD) of the medium at 600 nm. Fungal hyphae did not influence the OD of the culture, since they were strongly aggregated forming mycelium spheres, which quickly settled as a consequence of biofilm production and orbital shaking (Pirt, 1966).

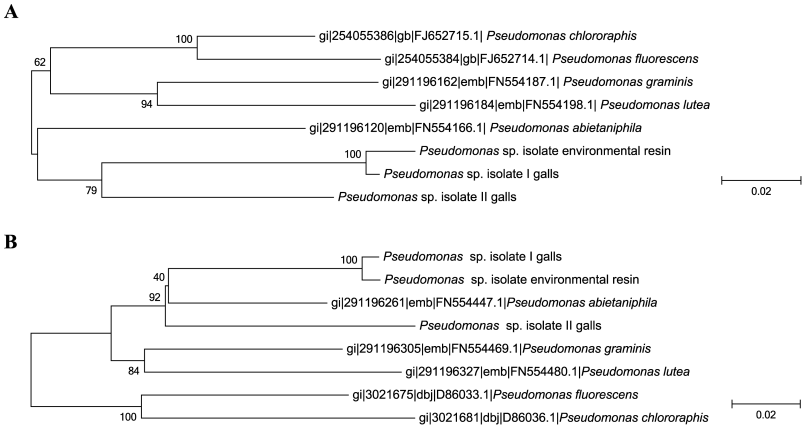
The resin degradation ability of *Pseudomonas* sp. could not be determined with the method described above because bacterial cells in suspension interfered with the OD measurements, and co-sedimented with resin particles when a centrifugation step was performed. Therefore, resin degradation was estimated with an indirect method. The resin content of the medium was estimated by calculating the difference between the total dry weight of the culture and the dry weight of the cellular fraction. This latter value was calculated from the number of viable bacteria (directly measured by plate counting) after establishing a correlation between the number of colony forming units and the corresponding bacterial pellet dry weight in LB broth. This correlation was assumed to be the same for LB and RM cultures of the same strain.

Preparation of other terpene-based culture media

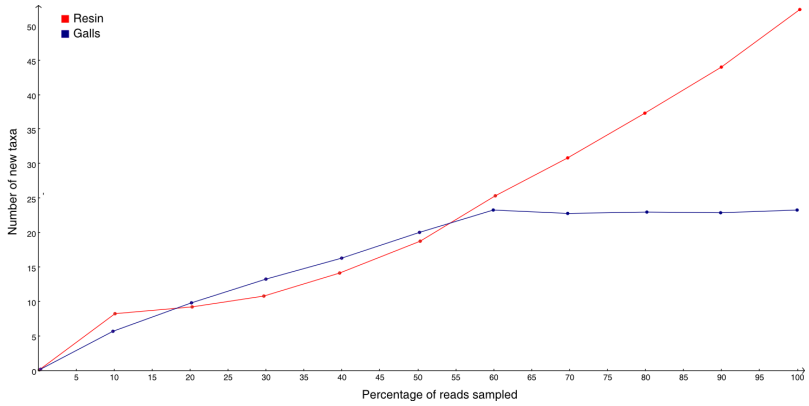
The ability of the isolates described in this work to degrade terpene-based polymers was tested in culture media containing latex and rubber as sole carbon sources. To prepare the latex-containing medium, standard laboratory gloves made of natural latex were ground along with dry ice until small latex laminas of 1-3 mm were obtained. A minimal saline solution, like



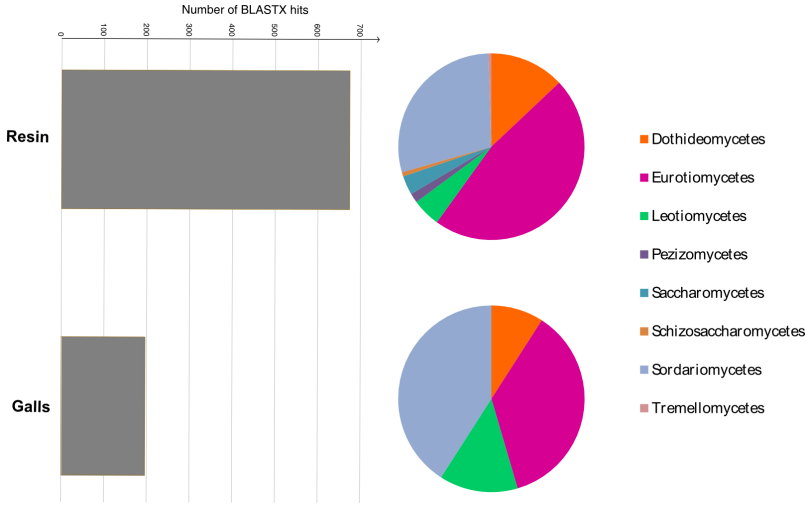
SUPPLEMENTARY FIGURE I. 1: Correlation between contig length and number of reads. Data corresponding to environmental resin (A) and galls (B) assemblies.



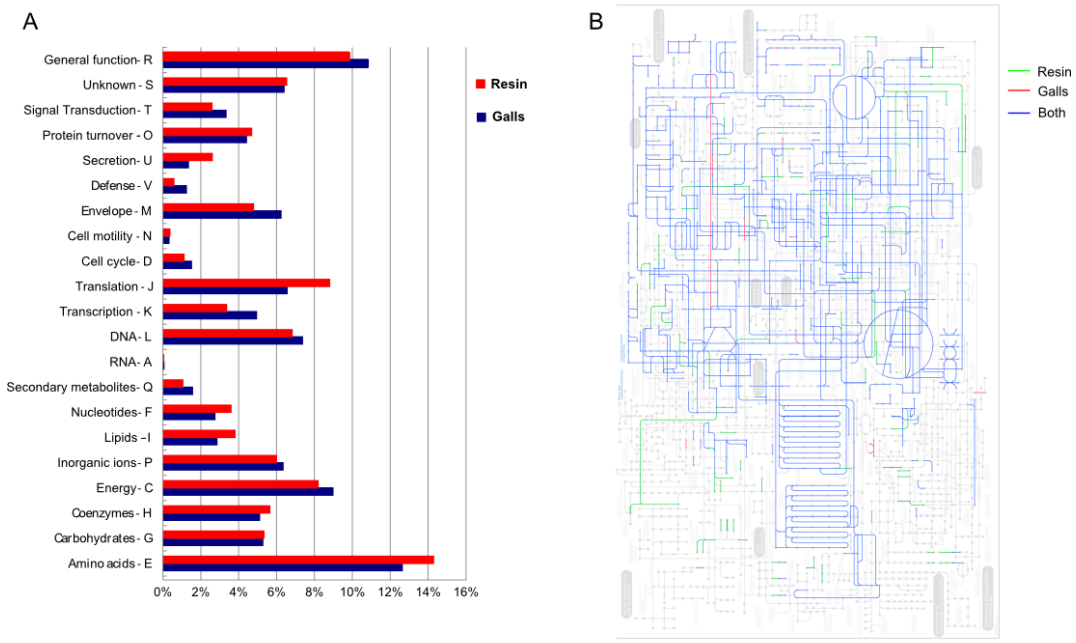
SUPPLEMENTARY FIGURE I. 2: Phylogenetic analysis of house-keeping genes corresponding to the genus Pseudomonas found in galls and environmental resin. Sequences from a range of Pseudomonas species were retrieved from the NCBI Nucleotide database, and Neighbor Joining trees for *gyrB* (A) and *rpoD* (B) nucleotide sequences were obtained with software MEGA.



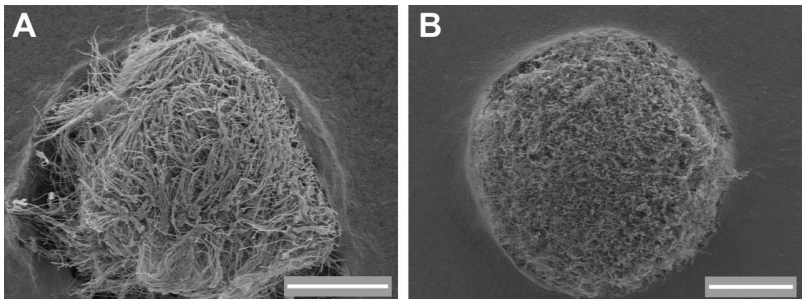
SUPPLEMENTARY FIGURE I. 3: Rarefraction curves obtained for the sequencing data. The analysis was performed for both environmental resin (red) and galls (blue) after processing the BLASTX results with the software MEGAN, as described in Materials and Methods.



SUPPLEMENTARY FIGURE I. 4: Abundance and distribution of fungal taxa in environmental resin and gall-associated microbial communities. The absolute abundance of fungal sequences (expressed as number of BLASTX hits matching fungal sequences) and the relative distribution of fungal taxa are shown.



SUPPLEMENTARY FIGURE I. 5: Functional reconstruction of the cultivable microbial communities associated with environmental resin and galls. A) Distribution of annotated genes according to COG functional categories for environmental resin- (red) and gall- (blue) cultivated communities. B) Schematic representation of the KEGG Pathways shared by both samples (blue); and of those found exclusively in environmental resin (green) or galls (red) samples.



SUPPLEMENTARY FIGURE I. 6: Spheres were obtained after growing (A) F1 and (B) a resin sample in RM medium for 10 days. A scale bar is 200 μm; B scale bar is 500 μm.

SUPPLEMENTARY TABLE I. 1: List of genes involved in the degradation of different terpene families detected in resin and gall metagenomes.

	Gene ID	Description	Counts in resin*	Counts in galls*
Pinene[a]	K00492	α-Pinene monooxygenase	23	1
	K00155	cis-2-Methyl-5-isopropylhexa-2,5-dienal dehydrogenase	6	1
	K01913	cis-2-Methyl-5-isopropylhexa-2,5-dienoate-CoA ligase	2	0
	K01692	cis-2-Methyl-5-isopropylhexa-2,5-dienoyl-CoA hydro-lyase	33	9
	R06405	3-Hydroxy-2,6-dimethyl-5-methylene-heptanoyl-CoA dehydrogenase	1	0
	K00680	2,6-Dimethyl-5-methylene-3-oxo-heptanoyl-CoA C-acetyltransferase	29	4
	K01076	3-Isopropylbut-3-enoyl-CoA thioesterase	14	3
	K00517	α-Pinene dehydrogenase	10	0
Diterpenes[b]	ditI	Dehydrogenase/reductase	2	1
	ditA2	β Subunit of the ring-hydroxylating dioxygenase	3	1
	ditA1	α Subunit of the ring-hydroxylating dioxygenase	2	1
	ditH	Isomerase/decarboxylase	2	1
	ditG	Dehydrogenase/reductase	2	1
	ditF	Sterol carrier-like protein	3	1
	ditR	IclR-type transcription regulator	3	1
	ditE	Permease of the major facilitator superfamily	2	1
	ditD	Isomerase/decarboxylase	2	1
	ditC	Extradiol cleavage dioxygenase	3	1
	ditB	Dehydrogenase/reductase	3	1
	ditA3	Ferredoxin component of ring-hydroxylating dioxygenase	3	1
	ditJ	CoA ligase	3	1
	ditK	Transcriptional regulator, TetR family	3	1
	ditL	Hypothetical protein	2	1
	ditM	Hydrolase	3	1
	ditN	3-hydroxyacyl CoA dehydrogenase	3	1
	ditO	Thiolase	2	1
	ditP	Conserved hypothetical protein	0	1
	ditQ	Cytochrome P450	3	1
	ditR	Transcriptional regulator, IcLR family	3	1
Acyclic terpenes[c]	atuA	3-hydroxy-3isohexenylglutaryl-CoA:acetate lyase	0	2
	atuB	Citronellol and citronellal dehydrogenase	1	0
	atuC	Geranyl-CoA carboxylase carboxyl transferase subunit	0	1
	atuD	Citronellyl-CoA dehydrogenase	1	0
	atuE	Isohexenylglutaconyl-CoA hydratase	0	0
	atuF	Geranyl-CoA carboxylase biotin-containing subunit	0	1

* Only hits with an e-value$\leq 10^{-5}$ and an identity percentage$\geq$ 60% were considered

[a] KEGG Pathway 00903

[b] Martin and Mohn, 2000; Smith *et al.*, 2007

[c] Forster-Fromme and Jendrossek, 2006

SUPPLEMENTARY TABLE I. 2: Summary of the sequencing and assembly statistics for F1 transcriptomes.

	Isolated	**Confrontated**
Megabases generated	3.23	2.18
Number of reads	20,893	15,319
Average read length (bp)	189.93	228.27
Number of isotigs	472	348
Average isotig length (bp)	382.18	339.98
Average number of reads per contig	25.83	58.74
Maximum number of reads per contig	699	701
Number of protein-coding isotigs	299	179

SUPPLEMENTARY TABLE I. 3: Identification of a selection of fungal strains isolated from environmental resin samples according to 18S rDNA sequence similarity.

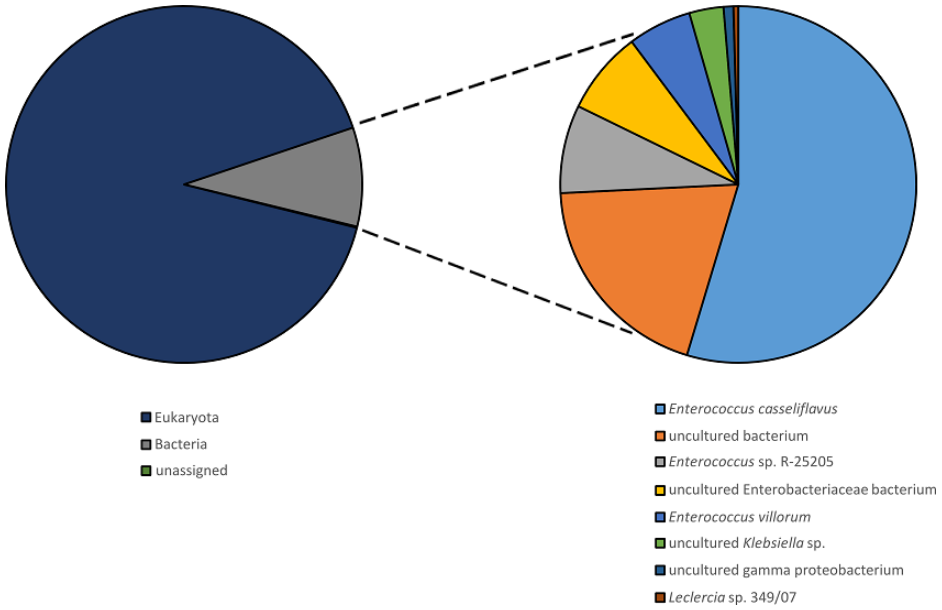
Isolate name	**Best Match**	**NCBI Accesion**	**% identitity**
F1	*Aspergillus terreus* strain HDJZ-ZWM-18	GU227345.1	99
F8	*Aspergillus flavus* strain TPID12	EU263602.1	100
F9	*Penicillium decumbens* isolate MMH 89-p1	FR774046.1	99

the one used in the resin-containing medium (RM), was supplemented with 40 g/L of latex laminas and used as broth. In the case of the rubber-containing medium, non-vulcanized rubber was kept at -80 ºC for 30 min and then grated to obtain a fine powder. This rubber powder (20g) was added to 1 L of minimal saline solution. Isolates were grown at 30 ºC in sterile Erlenmeyer 100 mL flasks with 50 mL of these media and shaken at 250 rpm until required.

Scanning electron microscopy of microbial growth on terpene-based materials

Little pieces of latex, rubber, or simply fragments of mycelium obtained from cultures of the different isolates were fixed by immersion into paraformaldehyde 2% - glutaraldehyde 2.5% for more than two hours, then lightly washed with water and preserved in ethanol 70%. These pieces were placed inside microporous specimen capsules (30 μm pore size, available from Ted Pella Inc. product number 4619) immersed in absolute ethanol, following critical point drying in an Autosamdri 814 (Tousimis). The fragments so obtained were then arranged on SEM stubs by silver conducting paint TAAB S269. Pieces were always manipulated under a stereomicroscope Leica MZ9.5 with Dumont forceps number 5. Stubs were examined under a scanning electron microscope Hitachi S-4100. Images were edited with Photoshop CS3 (Adobe).

Publication II



SUPPLEMENTARY FIGURE II. 1: Taxonomic classification of the metagenomic sequences from the biofilm ring. The species-level assignations of sequences belonging to the bacterial domain are depicted on the right.

SUPPLEMENTARY TABLE II. 1: Number of reads and diversity estimates for the samples analyzed.

Sample*	Raw reads	Reads after trimming	Observed Species	1/Simpson index
Bc_MG_I	6186	6167	73	1,14
Bc_MG_II	11831	11821	50	1,05
Bc_MG_III	6861	6825	108	1,26
Bc_HG_I	18712	18659	78	1,12
Bc_HG_II	21693	21591	79	1,05
Bc_HG_III	31313	31209	114	1,14
He_MG_I	220	213	16	7,56
He_MG_II	15921	15559	259	3,57
He_MG_III	3971	3848	209	3,74
He_HG_I	2073	2056	56	2,90
He_HG_II	7660	7542	222	3,51
He_HG_III	4118	4090	62	1,76

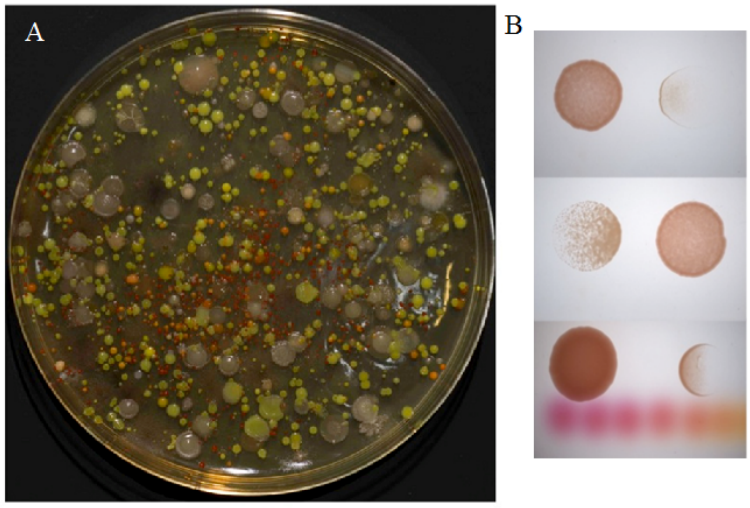
* Bc: *B. crini*; He: *H. euphorbiae*; MG: midgut; HG: hindgut

Publication III

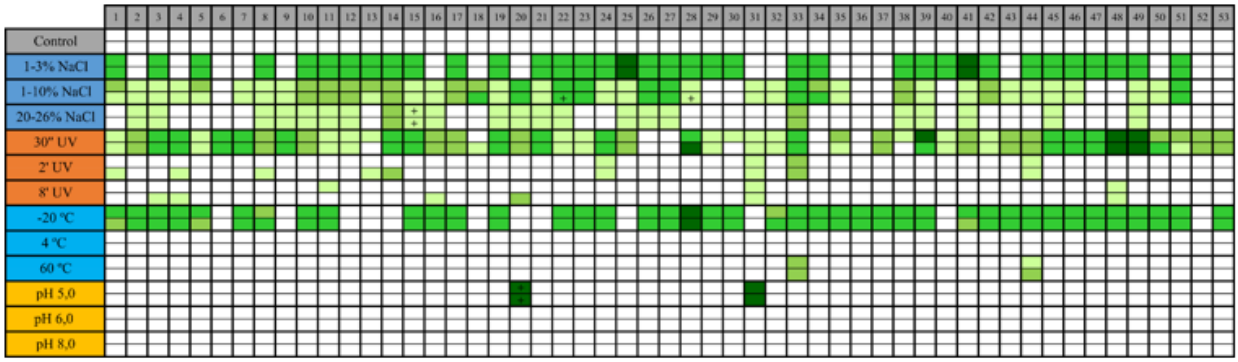
SUPPLEMENTARY TABLE III. 1: Taxonomical classification of the bacterial species isolated from solar panel samples according to 16S rDNA sequence similarity.

	Isolate ID	Best BLAST hit		Isolate ID	Best BLAST hit
1	S10-LB-01	*Planomicrobium glaciei*	28	S13-LB-03	*Kocuria rosea*
2	S10-LB-02	*Kocuria rosea*	29	S13-LB-04	*Planomicrobium chinense*
3	S10-LB-03	*Kocuria rosea*	30	S13-LB-05	*Planomicrobium koreense*
4	S10-LB-04	*Microbacterium paraoxydans*	31	S13-LB-06	*Pseudomonas psychrotolerans*
5	S10-LB-05	*Planomicrobium glaciei*	32	S13-MAR-01	*N.A.**
6	S10-LB-06	*Arthrobacter agilis*	33	S13-MAR-02	*Bacillus aquimaris*
7	S10-LB-07	*Kocuria sediminis*	34	S13-MAR-03	*N.A.*
8	S10-LB-09	*Kocuria rosea*	35	S13-MAR-04	*Planococcaceae bacterium*
9	S11-LB-01	*Kocuria rosea*	36	S13-MAR-05	*N.A.*
10	S11-LB-02	*Micrococcus xinjiangensis*	37	S13-MAR-06	*Planomicrobium okeanokoites*
11	S11-LB-03	*Planomicrobium glaciei*	38	S14-LB-01	*Planococcaceae bacterium*
12	S11-LB-04-01	*Bacillus infantis*	39	S14-LB-02	*N.A.*
13	S11-LB-04-02	*Bacillus infantis*	40	S14-LB-03	*Planomicrobium chinense*
14	S11-LB-05	*N.A.*	41	S14-LB-04	*Kocuria rosea*
15	S12-LB-01	*Kocuria rosea*	42	S14-MAR-01	*Planomicrobium okeanokoites*
16	S12-LB-02	*Cellulosimicrobium cellulans*	43	S14-MAR-02	*Planomicrobium okeanokoites*
17	S12-LB-03	*Planomicrobium glaciei*	44	S14-MAR-03	*N.A.*
18	S12-LB-04	*Enterobacter cloacae*	45	S15-LB-01	*Kocuria rosea*
19	S12-LB-05	*Kocuria rosea*	46	S16-LB-01	*Kocuria rosea*
20	S12-LB-06	*Bacillus flexus*	47	S16-LB-02	*Arthrobacter agilis*
21	S12-LB-07	*N.A.*	48	S16-LB-03	*N.A.*
22	S12-LB-08	*Planomicrobium glaciei*	49	S17-LB-01	*Kocuria rosea*
23	S12-LB-09	*Planomicrobium glaciei*	50	S18-MAR-01	*N.A.*
24	S12-LB-10	*Cellulosimicrobium cellulans*	51	S18-MAR-02	*Planomicrobium chinense*
25	S12-LB-11	*Kocuria rosea*	52	S18-MAR-03	*Domibacillus robiginosus*
26	S13-LB-01	*Planomicrobium okeanokoites*	53	S19-MAR-01	*Domibacillus robiginosus*
27	S13-LB-02	*Planomicrobium okeanokoites*			

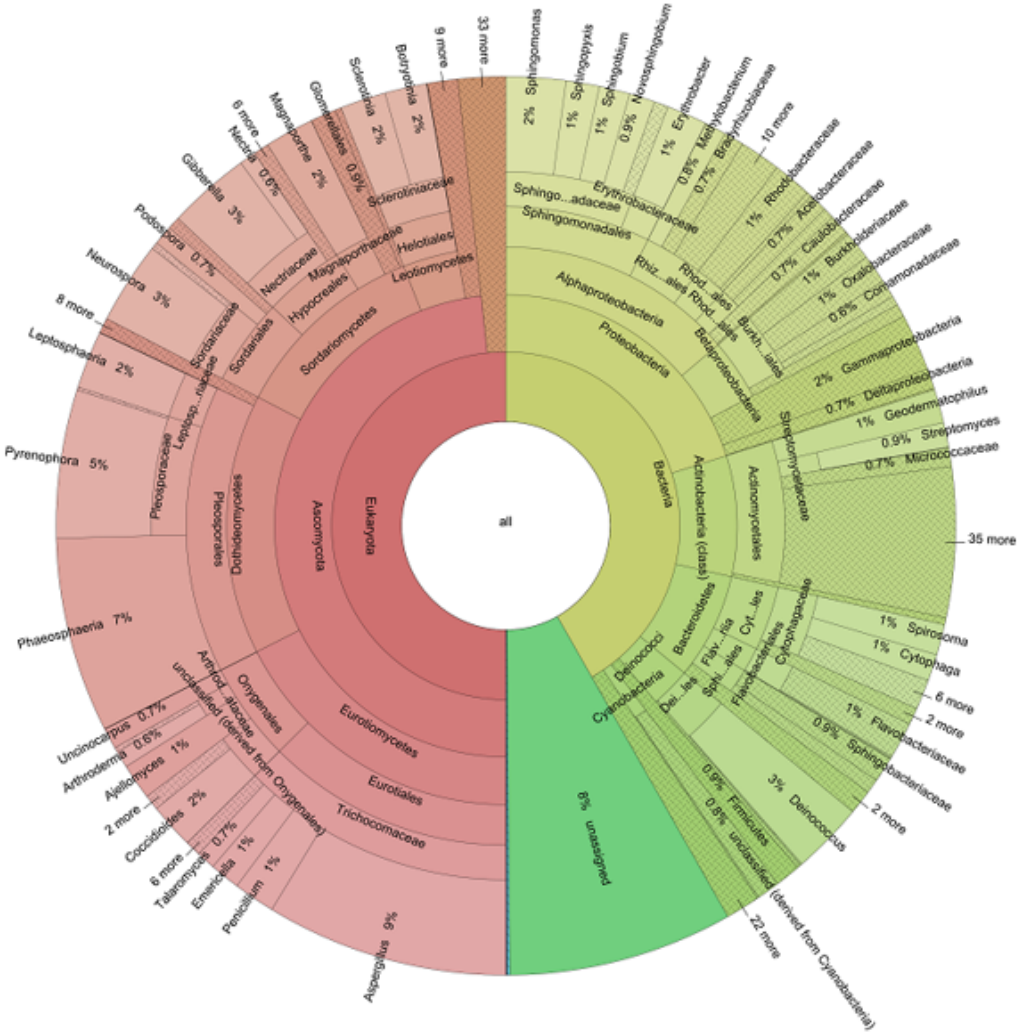
* N.A.: Non-assigned. Low quality sequences or assignations



SUPPLEMENTARY FIGURE III. 1: Microbial colonies growing on LB incubated at room temperature for two weeks (A). Case example of growth restoration in nearby-grown isolates under conditions of extreme pH. A local buffering of the pH of the plate is observed (B).



SUPPLEMENTARY FIGURE III. 2: Stress tests results data matrix. The growth of the 53 strains isolated from the solar panels under particular stress conditions was compared to that of a control strain (XL1-Blue *E. coli* strain). Green colour (from light to dark) indicates better growth than the control (from slightly to strongly better growth). Symbols '+' indicate cases of growth restoration by another isolate. Two independent experiments were performed for each test. Strains are numbered according to Table S1.



SUPPLEMENTARY FIGURE III. 3: Taxonomic diversity of one of the panels (panel 1) sampled in the summer solstice of 2014 as deduced from shotgun metagenomic sequencing.

SUPPLEMENTARY TABLE III. 2: Summary of sequencing statistics from the 16S/18S profile analysis performed on the solar panel pool and the individual panels.

	Pool (2013)	Pool (2013)	Panel 1 (2014)	Panel 1 (2014)	Panel 2 (2014)	Panel 2 (2014)	Panel 3 (2014)	Panel 3 (2014)
	16S	**18S**	**16S**	**18S**	**16S**	**18S**	**16S**	**18S**
# Sequences obtained	485094	105781	47013	4566	48633	5446	25971	18260
Average length	582	574	495	456	498	478	505	507
Total Mb	282,32	60,71	23,28	2,08	24,26	2,61	13,13	9,27
# Sequences after trimming	136194	6463	25184	796	23628	1722	14804	6220
Averag. length after trimming	549	551	473	475	481	478	486	490
# Genera (>0.01% abundance)	160	40	119	43	168	34	138	28
# Species (>0.01% abundance)	303	49	249	78	343	63	271	61
Shannon index	4,94	4,01	5,64	3,74	6,23	4,89	5,31	4,62

SUPPLEMENTARY TABLE III. 3: Summary of sequencing statistics from the shotgun metagenomic sequencing of solar panels 1 and 3 in 2014.

	Solar panel 1 (2014)	**Solar panel 3 (2014)**
Number of reads obtained	18596346	19220250
Average length	186,315	186,345
Total Mb	3464,783	3581,649
Average quality	32,855	32,995
Number of sequences after trimming	18240690	18716970
Average length after trimming	188,19	188,81
Total Mb after trimming	3432,688	3533,922
Average quality after trimming	33,075	33,275
Number of contigs after assembly	1486634	717211
N50	229	335
Maximum contig length	6408	13796
Average contig length	262	405
Number of predicted ORFs	104311	189676
Number of genera (>0.01% abundance)[1]	254	453
Number of species (>0.01% abundance)[1]	487	669
Shannon index[2]	4,39	5,57

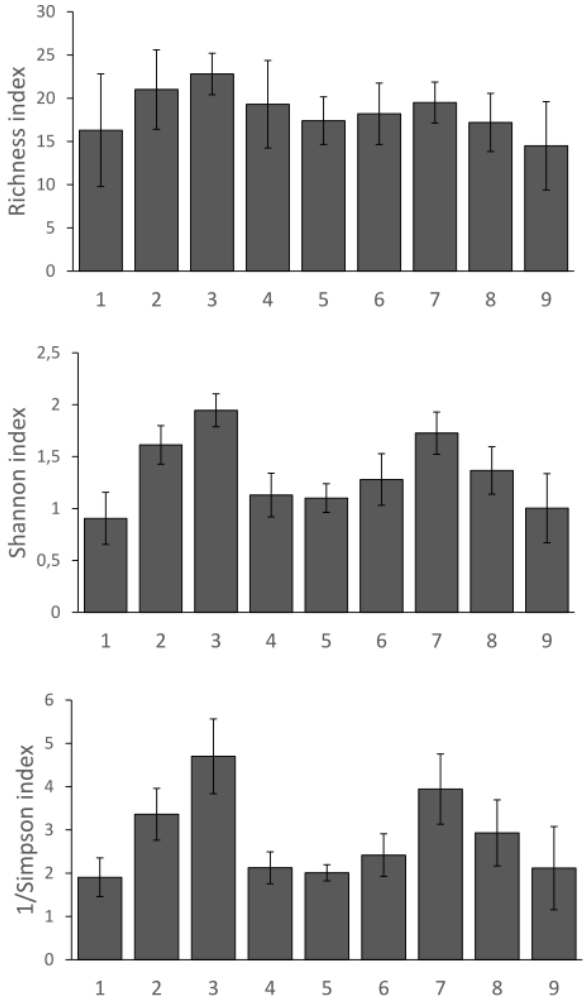
[1] Estimated from BLASTP searches of the ORFs against the NCBI nr database (e-value < 1e-5)

[2] Calculated from species-level assignations

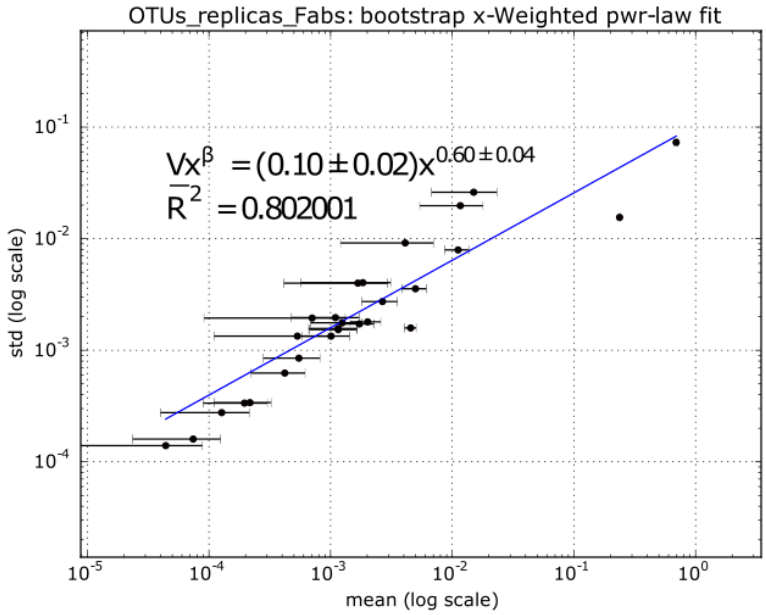
SUPPLEMENTARY TABLE III. 4: Differentially expressed proteins between day- and night-collected samples.

		Day/ Night t-value	p-value	Fold Change Day/Night	Fold Change Night/Day
gi|504554779	cold-shock protein [Geodermatophilaceae]	3,122	0,035	**25,986**	0,038
gi|656266558	molecular chaperone GroEL [Arsenicicoccus bolidensis]	4,361	0,012	**13,665**	0,073
gi|503987093	50S ribosomal protein L7/L12 [Niastella koreensis]	3,041	0,038	**9,944**	0,101
gi|502016435	hypothetical protein [Deinococcus deserti]	3,625	0,022	**9,373**	0,107
gi|517310793	ATP-1 H+-transporting ATP synthase [Fusarium fujikuroi IMI 58289]	-3,263	0,031	0,473	**2,114**
gi|656340823	S-layer protein [Deinococcus sp. RL]	-5,991	0,004	0,346	**2,890**
gi|636360381	major outer membrane lipoprotein 1 [Klebsiella pneumoniae MGH 64]	-2,847	0,047	0,189	**5,280**
gi|657196340	50S ribosomal protein L7 [Acidiphilium angustum]	-3,031	0,039	0,188	**5,326**
gi|499563314	F0F1 ATP synthase subunit beta [Synechococcus elongatus]	-2,968	0,041	0,150	**6,686**
gi|493585871	diguanylate cyclase [Frankia sp. EUN1f]	-2,884	0,045	0,116	**8,630**
gi|398394263	isocitrate dehydrogenase [NAD] sub 1 [Zymoseptoria tritici IPO323]	-16,097	0,000	0,066	**15,147**
gi|518290483	molecular chaperone GroEL [Roseomonas]	-9,372	0,001	0,057	**17,594**
gi|618851123	membrane-bound proton-translocating pyrophosphatase [Clostridium tetanomorphum DSM665]	-3,028	0,039	0,024	**40,965**

Publication IV



SUPPLEMENTARY FIGURE IV. 1: Evolution of biodiversity indexes throughout the experiment. Numbers in the horizontal axis correspond to each subculturing step.



SUPPLEMENTARY FIGURE IV. 2: Taylor law. We show the standard deviation as a function of the mean relative abundances for the 26 OTUs in the ten replica at the initial time sampled. Data follow a power law with a fluctuating scaling compatible with 0.5 (slope in log-log scale), as expected by Poisson-distributed replica. Variability of the Poisson sampling, including all sources up to taxonomic classification, is 8%, which can be interpreted as the expected variability as the number of taxa tends to one.

Publication V

SUPPLEMENTARY TABLE V. 1: Summary of sequencing statistics from the 16S rRNA gene amplicon pool.

Number of sequences obtained	966,699
Average length	527.187
Total Mb	510.121
Averag. number of sequences per sample	17,901.833
Number of sequences after trimming	503,544
Averag. length after trimming	520.231
Averag. number of sequences per sample after trimming	15,258.909
Number of bacterial genera (>0.01% abundance)	59

Appendix C

Patent

(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 525 398**

(21) Número de solicitud: 201300612

(51) Int. CI.:

C12N 1/20 (2006.01)
C12R 1/38 (2006.01)

(12) PATENTE DE INVENCIÓN CON EXAMEN PREVIO B2

(22) Fecha de presentación:

21.06.2013

(43) Fecha de publicación de la solicitud:

22.12.2014

Fecha de la concesión:

06.07.2015

(45) Fecha de publicación de la concesión:

13.07.2015

(73) Titular/es:

UNIVERSITAT DE VALÈNCIA (100.0%)
Av. Blasco Ibañez, 13
46010 Valencia (Valencia) ES

(72) Inventor/es:

PORCAR MIRALLES , Manuel ;
VILANOVA SERRADOR, Cristina ;
LATORRE CASTILLO, Amparo y
BAIXERAS ALMELA, Joaquín

(54) Título: **Cepas de Pseudomonas sp. y usos de las mismas**

(57) Resumen:
Cepas de Pseudomonas sp. y usos de las mismas.
La presente invención hace referencia a la cepa Pseudomonas sp. CECT8327 y/o mutantes de la misma, así como a su uso para la degradación de terpenos cíclicos y acíclicos y/o derivados terpérnicos cíclicos y acíclicos, a su uso como agente antifúngico, además de como agente productor de surfactantes. La presente invención describe también una composición que comprende dicha cepa, sola o en combinación con al menos un segundo agente antifúngico. Así, la presente invención puede englobarse en el campo del medio ambiente, aplicándose particularmente en técnicas de biorremediación, degradación de residuos y metabolismo de terpenos en todo tipo de reacciones industriales.

ES 2 525 398 B2

DESCRIPCIÓN

Cepas de Pseudomonas sp. y usos de las mismas.

CAMPO DE LA INVENCIÓN

La presente invención hace referencia a una cepa aislada de Pseudomonas sp. (en adelante cepa CECT8327 o cepa de la invención), y a su uso para la degradación de terpenos en general y derivados terpénicos; así como a su uso como agente antifúngico o como agente productor de surfactantes. La presente invención se engloba en el sector del medio ambiente, aplicándose particularmente en técnicas de biorremediación, degradación de residuos ambientales de terpenos y derivados terpénicos en todo tipo de reacciones industriales.

ESTADO DE LA TÉCNICA

La resina de coníferas es una mezcla compleja de metabolitos secundarios de dichas plantas. La resina protege los tejidos dañados de los árboles frente a insectos fitófagos y patógenos de las plantas. Los principales componentes de la resina son los terpenos cíclicos, los flavonoides y los ácidos grasos. Los terpenos son los metabolitos de la resina mejor caracterizados porque pueden ser fácilmente identificados con técnicas tales como la cromatografía de gases. Los terpenos cíclicos más abundantes en la resina de pino son los monoterpenos, sesquiterpenos y diterpenos (TRAPP S y CROTEAU R. Defensive resin biosynthesis in conifers. Annu. Rev. Plant. Phys. 2001. Vol. 52, páginas: 689-724). Dado que muchos de los terpenos de la resina muestran propiedades antibacterianas y antifúngicas, son considerados fitoalexinas que explican la toxicidad de la misma (GRAYER RJ y HARBORNE JB. A survey of antifungal compounds from higher plants, 1982-1993. Phytochemistry 1994, Vol. 37, páginas: 19-42).

Adicionalmente muchos terpenos exhiben características interesantes para la industria química, especialmente en la producción de fragancias, aceites esenciales y aditivos alimentarios; así como en medicina por sus propiedades citotóxicas, cardiotónicas y antiinflamatorias. No obstante, desde un punto de vista ambiental los terpenos y los derivados terpénicos que no son biodegradables bajo condiciones ambientales naturales, representan uno de los principales contaminantes ambientales. Dichos compuestos se encuentran contaminando el agua efluente de plantas industriales tales como el agua de las fábricas de pasta de papel (SUNTIO LR. et al. A review of the nature and properties of chemicals present in pulp mill effluents. Chemosphere. 1988, Vol. 17, páginas: 1249-1290), representando toneladas de residuos anuales procedentes de los materiales basados en terpenos.

Hasta la fecha, en el estado de la técnica se conocen muy pocos microorganismos con capacidad para degradar terpenos. Además entre los microorganismos conocidos con esta capacidad, o bien degradan únicamente terpenos cíclicos (MOHN WW. et al. Physiological and Phylogenetic Diversity of Bacteria Growing on Resin Acids. Syst. Appl. Microbiol. 1999, Vol. 22, páginas: 68-78), o bien degradan solo terpenos acíclicos (LINOS A. et al. Biodegradation of cis-1, 4-polyisoprene rubbers by distinct actinomycetes: microbial strategies and detailed surface analysis. Applied and environmental microbiology. 2000, Vol. 66, páginas: 1639-45). Por todo ello, el reciclaje es hasta la presente invención la única alternativa posible ante la acumulación de este tipo de materiales (ABRAHAM E. et al. Recent Developments in Polymer Recycling. 2011; Paginas: 47-100. ISBN: 978-81-7895-524-7; BOONDAMNOEN O. et al. Recycling waste natural rubber latex by blending with polystyrene: characterization of mechanical properties. Advanced Materials Development and Performance (AMDP2011). International Journal of Modern Physics: Conference Series. 2012, Vol. 6, páginas 391-396).

En definitiva, no existe un tratamiento eficaz para la degradación de los distintos tipos de terpenos, cíclicos y acíclicos, así como de los derivados terpénicos, sin necesidad de discriminar el tipo de terpeno concreto. Por ello, los terpenos en general, y los derivados terpénicos de cualquier tipo, continúan siendo un problema, como contaminantes medio-

ambientales, por su difícil degradación. A la vista de lo anterior, existe una necesidad permanente de desarrollar métodos para la degradación y eliminación de los terpenos y derivados terpénicos cíclicos y acíclicos, incluyendo la biorremediación bacteriana.

DESCRIPCIÓN DE LA INVENCIÓN

Definiciones

A los efectos de la presente invención, la expresión "derivado terpénico" se refiere a materiales, sustratos y sustancias cuya estructura química este total o parcialmente basada en terpenos tanto cíclicos como acíclicos.

A los efectos de la presente invención por el término "degradación" se entiende el proceso por el que un terpeno o derivado terpénico pierde uno o varios átomos de carbono, o por el que un terpeno o derivado terpénico se descompone en moléculas más simples.

A efectos de la presente invención, se entiende por el término "degradación", la transformación de la estructura compleja del compuesto, preferentemente, terpenos o derivados terpénicos, en otra de estructura más sencilla.

Por la expresión "tratamiento de compuesto hidrofóbico" se entiende el aumento de su biodisponibilidad en procesos de biorremediación.

En la presente invención, las expresiones "terpeno acíclico" y "terpeno lineal" se usan indistintamente y hacen referencia a todo terpeno que no tiene una estructura cíclica.

En la presente invención por las expresiones "cepa de la invención" y "cepa CECT8327" se entiende la cepa CECT8327, cepa aislada de Pseudomonas sp y mutantes de la misma.

A efectos de la presente invención, los términos "mutante", "cepas mutantes" y "mutantes de la misma", se refieren específicamente a las cepas bacterianas que son generadas u obtenidas mediante técnicas de mutación espontanea o inducida, a partir de la cepa perteneciente a Pseudomonas sp CECT8327, que se describe y ejemplifica en el presente documento. Las cepas mutantes descritas en el presente documento y obtenidas a partir de la cepa CECT8327 como material de partida, mantienen o mejoran una o varias de las propiedades de la cepa de la invención, que incluye al menos su capacidad para degradar terpenos cíclicos y acíclicos y derivados terpénicos; pudiendo incluir adicionalmente, su capacidad antifúngica, su capacidad de producir agentes surfactantes en presencia de compuestos hidrofóbicos, su utilización como biofactorías y su capacidad para degradar imazalil. La persona experta en la técnica es capaz de realizar las técnicas apropiadas para verificar si las cepas mutantes obtenidas presentan dichas propiedades. Posibles métodos para determinar dichas propiedades se recogen en el apartado Ejemplos.

La obtención de cepas mutantes a partir de la cepa CECT8327 puede llevarse a cabo por medio de la aplicación de uno o varios métodos de mutagénesis. El método se selecciona, sin limitación, de la lista que comprende mutagénesis química, mutagénesis por radiaciones y mutagénesis por elementos transponibles. Las mutaciones espontaneas pueden ocurrir debido a la acción de las radiaciones naturales e incluso durante la replicación del ADN debido a errores en la lectura de las bases. La frecuencia de mutación puede aumentarse significativamente con el uso de métodos de mutación inducida. Tanto Las mutaciones espontaneas como las inducidas se producen como resultado de cambios estructurales en el genoma como por ejemplo, pero sin limitación, cambio en el número de cromosomas, cambio en el orden de uno o varios genes dentro del cromosoma o cambio en la secuencia de bases dentro de un gen (mutación puntual). Como agentes que provocan mutagénesis química se usan, sin limitación, agentes que reaccionan con el ADN, aunque el ADN no se esté replicando, ocasionando cambios químicos en las bases que provocan un apareamiento incorrecto y agentes intercalantes o análogos de bases. Entre agente que reacciona con el ADN, sin

limitación, están el ácido nitroso, hidroxilamina, agente alquilante (etil metano sulfonato [EMS], metil metano sulfonato [MMS], dietil sulfato [DES], diepoxi butano [DEB], N-metil-N-nitro-N-nitrosoguanidina [NTG], N-metil- N-nitroso urea (gas mostaza), agente intercalante (acridina, bromuro de etidio o dihidroetidio).

La mutagénesis por radiaciones puede ser producida, pero sin limitarse, a radiaciones ultravioleta o radiaciones ionizantes como los rayos X o los rayos gamma.

La mutagénesis por elementos transponibles puede ser producida, pero sin limitarse, a la inserción de una secuencia de inserción o de un transposón en una secuencia de uno o varios genes de la cepa bacteriana de la presente invención.

A efectos de la presente invención, el término "similitud" se refiere al grado de similitud de secuencia entre dos moléculas de ácido nucleico comparadas mediante la alineación de sus secuencias. El grado de similitud entre dos secuencias de ácidos nucleicos que se comparan es una función del número nucleótidos idénticos que se localizan en posiciones comparables. El porcentaje de similitud de dos secuencias de ácidos nucleicos, a efectos de la presente invención, se determina mediante los programas informáticos ClustalW, BLAST, FASTA o Smith-Waterman.

A efectos del presente documento y haciendo referencia a la "Lista de Secuencias" que lo acompaña, se incluye la traducción al castellano de las palabras que aparecen en ingles en dicha "Lista de Secuencias". Así, "source": fuente; "mol_type": tipo de molécula; "unassigned DNA": ADN no asignado; y "organism": organismo.

BREVE DESCRIPCIÓN DE LA INVENCIÓN

La presente invención soluciona el problema planteado en el apartado anterior mediante la identificación y uso de una cepa aislada de Pseudomonas sp. , la cepa CECT8327 y mutantes de la misma.

Por lo tanto, el primer aspecto de la presente invención hace referencia a la cepa aislada perteneciente a Pseudomonas sp. CECT8327 y cepas mutantes de la misma, donde las cepas mutantes son obtenidas a partir de la cepa de la invención como material de partida y donde las cepas mutantes mantienen una o varias de las propiedades de la cepa de la invención, propiedades que incluye al menos su capacidad para degradar terpenos cíclicos y acíclicos y derivados terpénicos; pudiendo incluir adicionalmente, su capacidad antifúngica y su capacidad de producir agentes surfactantes en presencia de compuestos hidrofóbicos, además de su capacidad para degradar imazalil y como biofactoría para la producción de sustancias o compuestos a escala industrial partiendo de resina rica en terpenos y/o derivados terpénicos, como fuente de carbono.

En la presente invención por las expresiones "cepa de la invención" y "cepa CECT8327" se entiende la cepa CECT8327, cepa aislada de Pseudomonas sp. , .y mutantes de la misma, como se han definido anteriormente.

Sorprendentemente, los inventores de la presente invención, han aislado y caracterizado la cepa de Pseudomonas sp. CECT8327 en comunidades microbianas asociadas a una fuente de microorganismos no considerada hasta la fecha en el estado de la técnica como posible fuente de microorganismos, en concreto la resina de Pinus sylvestris (P. sylvestris, o pino silvestre, en adelante), que incluye la resina asociada a las agallas generadas por el insecto Retinia resinella Linnaeus, grupo del orden Lepidoptera (NIEUKERKEN EJ. et al. Animal biodiversity: An outline of higher level classification and survey of taxonomic richness. Order Lepidoptera Linnaeus, 1758.En: Zhang, ZQ (ed). Magnolia Press: New Zealand. 2011, paginas: 212-221).

Es conocido en el estado de la técnica que las larvas de Retinia causan pequeñas heridas en el P. sylvestris, que inducen la secreción de resina, la cual es manipulada por la larva para

construir una cápsula de resina a modo de una dura capa protectora. También es conocido que las agallas son nódulos construidos con la resina exudada por el pino, en el interior de las cuales viven las larvas de Retinia. Cada larva se desarrolla dentro de la agalla de resina durante cerca de dos años, completamente aislada del ambiente externo por su refugio rico en terpenos.

La identificación de la cepa CECT8327 es un hallazgo inesperado por varios motivos. En primer lugar porque la fuente de la que se ha aislado la cepa, la resina, como es bien conocido en el estado de la técnica, es mayoritariamente rica en terpenos cíclicos, y estos, a priori , de acuerdo a los conocimientos generales, inhiben el crecimiento de los microorganismos (WANG J. et al. Robust antimicrobial compounds and polymers derived from natural resin acids. Chem Commun. 2012. Vol: 48, paginas: 916-8), por lo que no era esperable aislar de la resina, la cepa con las propiedades descritas en la presente invención. Pero, además, sorprendentemente, la cepa de la presente invención es capaz de degradar otro tipo de terpenos no presentes en la resina en la que se ha aislado la cepa Pseudomonas sp. CECT8327, los terpenos acíclicos.

A la vista de lo anterior, en concreto, la sorprendente capacidad de la cepa CECT8327 y mutantes de la misma, para la degradación tanto de terpenos cíclicos como acíclicos y derivados terpénicos, dichas cepas pueden ser utilizadas industrialmente en la degradación de los mismos. Consecuentemente, la presente invención puede aplicarse en técnicas de biorremediación.

El segundo aspecto de la presente invención hace referencia a una composición que comprenda al menos la cepa CECT8327, y/o posibles mutantes de la misma, tal y como se definen en el primer aspecto de la invención.

Por ello, siguiendo con lo explicado anteriormente en relación a las posibles aplicaciones de la cepa CECT8327, el tercer aspecto de la presente invención hace referencia al uso de la cepa CECT8327 y/o mutantes de la misma, o a una composición que comprenda una cantidad efectiva de la cepa de la invención y/o de mutantes de la misma, según se ha definido anteriormente, para la degradación de terpenos, tanto acíclicos como cíclicos y derivados terpénicos.

La degradación de terpenos por la cepa CECT8327 y/o por los mutantes de la misma, o por una composición que comprenda una cantidad efectiva de la cepa de la invención y/o de mutantes de la misma, posibilita el aprovechamiento energético de los mismos con fines biotecnológicos, en concreto como biofactoría para la producción de sustancias o compuestos a escala industrial partiendo de los terpenos y/o derivados terpénicos como fuente de carbono.

El cuarto aspecto de la presente invención es el uso de la cepa CECT8327 y/o de los mutantes de la misma, o de una composición que comprenda una cantidad efectiva de la cepa de la invención y/o de mutantes de la misma, como biofactoría en la producción de sustancias o compuestos de interés empleando resina como fuente de carbono. Esta aplicación se lleva a cabo mediante la inclusión del gen o los genes de interés que codifican la sustancia o compuesto deseado en el genoma de la cepa CECT8327.

La persona experta en la técnica conoce perfectamente como insertar el gen o los genes de interés en el genoma de la cepa de la invención manteniendo las propiedades sorprendentes de la cepa de la invención usada como material de partida. Entre ellas, y sin carácter limitante, están la transformación con vectores integrativos o la transformación con vectores episomales.

Sorprendentemente, además, la cepa CECT8327 y/o los mutantes de la misma, o una composición que comprenda una cantidad efectiva de la cepa de la invención y/o de mutantes de la misma, muestran acción antifúngica al inhibir el crecimiento de hongos. Esta capacidad antifúngica de la cepa de la invención o de una composición que comprenda una cantidad

efectiva de la misma, permite su uso en el control de hongos fitopatógenos, permitiendo un ahorro en el uso de fungicidas y reduciendo su impacto ambiental.

Por ello, el quinto aspecto de la presente invención es el uso de la cepa CECT8327 y/o de los mutantes de la misma, o de una composición que comprenda una cantidad efectiva de la cepa de la invención y/o de mutantes de la misma, como fungicida. Como se ha mencionado previamente, la composición descrita en la presente invención, puede comprender además, al menos otro agente antifúngico, siendo preferido el imazalil. Por lo tanto, la presente invención describe también, el uso de una composición que comprende una cantidad efectiva de la cepa de la invención y/o de mutantes de la misma en combinación con otro agente antifúngico, preferentemente el imazalil, para el control de bongos fitopatógenos.

La capacidad inesperada de la cepa CECT8327 y/o de los mutantes de la misma, para combinar las actividades previamente citadas, de degradación de terpenos cíclicos y acíclicos, derivados terpénicos y su acción antifúngica, da lugar a aplicaciones industriales adicionales de la cepa o de una composición que comprenda una cantidad efectiva de la misma, debido a la presencia de terpenos y derivados terpénicos en biomasas, al permitir la degradación de terpenos y derivados terpénicos al mismo tiempo que muestra propiedades antifúngicas.

A la vista de lo anterior el sexto aspecto de la presente invención hace referencia al uso de la cepa CECT8327 y/o a los mutantes de la misma, o de una composición que comprenda al menos la cepa CECT8327, y/o posibles mutantes de la misma, para la degradación de terpenos cíclicos y acíclicos y derivados terpénicos y como antifúngico, para inhibir el crecimiento de hongos.

Adicionalmente, y de modo inesperado, la cepa de la invención es capaz de crecer en presencia del fungicida imazalil como única fuente de carbono, de ahí su uso o de una composición que comprenda una cantidad efectiva de la misma, para la descontaminación de fluidos que contienen imazalil.

Otro efecto inesperado de la cepa CEC18327 y/o de los posibles mutantes de la misma, o de una composición que comprenda una cantidad efectiva de la cepa de la invención y/o de mutantes de la misma, es su capacidad para crecer en presencia de medios líquidos que contengan compuestos hidrofóbicos. Este efecto inesperado se basa en la producción de agentes surfactantes que facilitan la emulsión de los compuestos hidrofóbicos, para su posterior tratamiento. Por ello, el séptimo aspecto de la presente invención hace referencia al uso de la cepa CECT8327 y/o de los mutantes de la misma, o de una composición que comprenda una cantidad efectiva de la cepa de la invención y/o de mutantes de la misma, para la producción de agentes surfactantes que faciliten la degradación o tratamiento de compuestos hidrofóbicos.

DESCRIPCIÓN DE LOS FIGURAS

Figura 1. Peso seco total (T) y peso seco bacteriano (C) (expresados en g/mL, según se muestra en el eje vertical de la gráfica) de los cultivos en medio RM (medio basado en resina) de las cepas ensayadas y evolución de la cantidad estimada de resina (R) en el medio. El eje horizontal representa el tiempo de cultivo expresado en días.

Figura 2. Microscopía electrónica de cultivos con láminas de látex no inoculados (A) e inoculados (B, C y D) con la cepa CECT8327, obtenidas tras 15 días de crecimiento a 30°C.

Figura 3. Imágenes de medio de cultivo mínimo suplementado con diesel no inoculado (A) e inoculado (B) con la cepa de la invención CECT8327.

Figura 4. Diámetro de colonia (d, expresado en mm) de tres hongos aislados de la resina (Aspergillus terreus, Aspergillus flavus y Penicillium decumbens) en ensayos de confrontación con la cepa CECT8327, en medio rico (graficas B, D y F) y en medio mínimo con resina como

única fuente de carbono (graficas A, C y E). El eje horizontal representa el tiempo de cultivo medido en días. En cada grafica C corresponde al hongo confrontado con la cepa CECT8327; y A corresponde al hongo no confrontado con la cepa CECT8327. Las graficas A y B corresponden a ensayos llevados a cabo con Aspergillus terreus, mientras que las graficas C y D corresponden a Aspergillus flavus y las graficas E y F a Penicillium decumbens

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

La presente invención se basa en la cepa CECT8327, del género Pseudomonas sp., y cualquiera de sus mutantes, según se define a lo largo del presente documento, identificada a partir de muestras de resina de P. sylvestris y de agallas del lepidóptero R. resinella. Sorprendentemente, y tras la caracterización de la cepa de la invención a nivel de secuencia (SEQ ID NO: 1 y SEQ ID NO: 2) y a nivel fenotípico, se identificaron un conjunto de características metabólicas y ecológicas, mencionadas anteriormente, que convierten a la cepa CECT8327 en una cepa esencial en distintos procesos biotecnológicos que impliquen la degradación de terpenos cíclicos y acíclicos, y derivados terpénicos, así como la actividad antifúngica o la generación de surfactantes para el tratamiento posterior de compuestos hidrofóbicos.

La cepa CECT8327 tolera y degrada cantidades significativas de resina. La Figura 1 muestra que la cepa de la presente invención degrada 1,5 gramos de resina por litro de medio mínimo, con resina como única fuente de carbono, después de cuatro días de crecimiento exponencial en dicho medio. El protocolo usado para determinar el contenido de resina tuvo únicamente en cuenta la cantidad de resina convertida en biomasa, contribuyendo al incremento del peso seco de la bacteria. Por ello, la cantidad de resina degradada es una estimación a la baja, siendo la cantidad real de resina degradada muy superior.

En una realización preferida de la presente invención, la cepa CECT8327 y/o mutantes de la misma, degradan los terpenos cíclicos seleccionados del grupo que comprende diterpenos, monoterpenos y sesquiterpenos.

La cepa CECT8327 y/o los mutantes de la misma, son capaces de degradar los monoterpenos monocíclicos, limonoides, presentes en una amplia gama de alimentos funcionales y fitonutrientes naturales. En una realización preferida de la invención, la cepa CECT8327 degrada limoneno, presente, por ejemplo, en las peladuras de cítricos. Por ello, un aspecto de la invención es su uso en el tratamiento de biomasas ricas en limonenos como paso previo a su fermentación por parte de microorganismos sensibles a estos terpenos, como las levaduras. En concreto la degradación de limoneno de biomasas procedentes del procesamiento de cítricos para la mejora en la producción de etanol, mediante fermentación alcohólica de las biomasas pre-tratadas mediante levaduras, a partir de cascara de naranja.

El contenido génico de la cepa CECT8327 analizado a partir de las secuencias metagenómicas, detectó un grupo o clúster de genes correspondientes a la SEQ ID NO: 1. La presencia de este paquete génico indica el fuerte valor adaptativo de este grupo de genes en un ambiente rico en terpenos. Además, la cepa de la invención comprende la secuencia parcial 16S correspondiente a la SEQ ID NO: 2.

En una realización preferida de la presente invención, la cepa CECT8327 y/o los mutantes de la misma, comprenden la secuencia correspondiente a SEQ ID NO: 1 en su material genómico. En una realización aún más preferida de la invención, la cepa CECT8327 y/o los mutantes de la misma, comprenden una secuencia con una similitud de al menos 70% respecto a la secuencia correspondiente a la SEQ ID NO: 1.

Por otro lado, la cepa CECT8327 y/o los mutantes de la misma, de la presente invención, ventajosa e inesperadamente, tienen capacidad para degradar terpenos totalmente distintos a los presentes en la resina, los terpenos acíclicos. En una realización preferida de la presente

invención, los terpenos acíclicos degradados son el látex, la goma vulcanizada o caucho y la gutapercha.

En este sentido, también es sorprendente la elevada velocidad de degradación mostrada por la cepa CECT8327 al ser capaz de degradar visiblemente látex en dos semanas, observándose al microscopio electrónico grietas en la estructura del mismo (ver Figura 2). Otro aspecto sorprendente de la cepa CECT8327 es la forma de degradar el látex. La cepa CECT8327 actúa de forma local mediante el contacto íntimo entre la misma y el sustrato, al cual se fija incluso bajo vigorosa agitación del medio, observándose una perfecta adhesión y produciéndose una excavación en forma de nicho u hornacina debido a la desintegración del látex en las inmediaciones de cada célula adherida al mismo (ver Figura 2). Así, la cepa CECT8327 puede usarse en el tratamiento de materiales acumulados a gran escala, tales como guantes u otros residuos sólidos, actuando directamente sobre su superficie, por lo que el tratamiento de estos materiales a escala industrial no requeriría de un pre-tratamiento para solubilizarlos en medio líquido.

Además, la cepa CECT8327 y/o los mutantes de la misma, sorprendentemente se caracterizan por mostrar un efecto combinado al ser capaces de degradar eficazmente terpenos cíclicos y acíclicos y derivados terpénicos, al mismo tiempo que son antifúngicos. Un aspecto preferido de la presente invención hace referencia a un método combinado para la degradación de terpenos cíclicos y acíclicos y derivados terpénicos y para la inhibición del crecimiento de hongos, que comprende la puesta en el medio de una cantidad efectiva de la cepa CECT8327 o de una composición que la comprenda.

En una realización preferida de la invención relativa a la acción antifúngica de la cepa CECT8327 y/o los mutantes de la misma, el uso de inhibir el crecimiento de hongos es sobre hongos pertenecientes a los géneros Aspergillus y Penicillium. En resumen, la cepa de la invención es también un agente de control biológico, especialmente para la industria agroalimentaria.

En una realización preferida de la presente invención, la cepa CECT8327 se usa como fungicida en tratamientos post-cosecha. En la presente invención, por tratamiento post-cosecha se entiende el proceso de almacenaje de las cosechas y los procesos necesarios para la transformación y la adecuada conservación de los alimentos al estado natural o fresco. En otra realización preferida de la presente invención, la cepa CECT8327 se usa en el control de hongos en la industria agroalimentaria, en concreto en los productos para la alimentación humana y/o animal procesados, y más particularmente en ensaladas de cuarta gama, mejorando la seguridad de los mismos.

La cepa de la presente invención es capaz de crecer en presencia de agentes antifúngicos. La Tabla 1 muestra la comparación del crecimiento mediante conteo de Unidades Formadoras de Colonia (UFC) de la cepa CECT8327 en medio de cultivo sólido LB y medio de cultivo sólido LB suplementado con distintas concentraciones de un fungicida comúnmente usado en la industria agroalimentaria, el fungicida imazalil. En una realización preferente de la invención, la cepa CECT8327 se usa como fungicida en combinación con el imazalil, típico de la industria agroalimentaria.

A la vista de lo anterior, en una realización preferida, la presente invención hace referencia a una composición que comprende al menos la cepa CECT8327 y otro agente antifúngico. En una realización aún más preferida, el otro agente antifúngico es imazalil.

Tabla 1

Medio de cultivo	UFC/mL
LB	3.84E+07
LB+500ppm imazalil	8.80E+06
LB+1500ppm imazalil	5.60E+06

Como se ha mencionado anteriormente, la cepa CECT8327 es capaz de crecer en un medio con compuestos hidrofóbicos, En una realización preferida, la cepa CECT8327 y/o mutantes de la misma, son capaces de crecer en presencia de hidrocarburos, mediante la producción de agentes surfactantes que posibilitando el tratamiento posterior de dicho hidrocarburo. En una realización aún mas preferida de la invención, esta capacidad surfactante se aplicaría al tratamiento de medios líquidos que contengan diesel, (ver Figura 3) y/u otros compuestos hidrofóbicos, como, y sin carácter limitativo, aceites, parafinas y grasas.

DEPÓSITO DE MICROORGANISMOS

De acuerdo con lo establecido en el Tratado de Budapest, la cepa de la invención del género Pseudomonas fue depositada en la Colección Española de Cultivos Tipo (CECT), situada en el Parc Cientific de la Universitat de Valencia, calle Catedrático Agustín Escardino, 9, 46980 Paterna, Valencia, (España).

La fecha de depósito es 22 de abril de 2013 y el número de depósito correspondiente es CECT8327.

EJEMPLOS

Ejemplo 1. Colección de muestras.

Se recogieron muestras de resina ambiental asociada al pino albar (Pinus sylvestris) (obtenidas de secreciones naturales del pino) y de resina asociada a las agallas inducidas por R. resinella del bosque "Fuente del Tajo" (Mora de Rubielos, Teruel, España) en septiembre del 2011. Se extrajeron las larvas de las agallas. Tanto las agallas como la resina fueron molidas individualmente y suspendidas en tampón PBS estéril (NaCl 8 g/L, KCl 0.2 g/L, Na_2HPO_4 1.44 g/L, KH_2PO_4 0.24 g/L, pH ajustado a 7.4).

Ejemplo 2. Medios de cultivo y condiciones de crecimiento.

Se estableció un medio mínimo selectivo con resina de pino como la única fuente de carbono. Como factor de selección, se obtuvo una solución concentrada al 10% (peso/volumen) de resina mediante la disolución de las muestras de resina mencionadas en el apartado anterior en etanol. Se eliminaron los restos de plantas y otras partículas insolubles en suspensión mediante centrifugación (2,000xg, 5 min). El sobrenadante se filtró y esterilizó a través de un filtro de 0.2µm de diámetro de poro. Se preparó, un medio mínimo (2 g/L $NaNO_3$, 1 g/L K_2HPO_4, 0.5 g/L $MgSO_4$, 0.5 g/L KCl, 0.2 g/L pectona bacteriológica; y 15 g/L de agar para el medio sólido). Una vez esterilizado con autoclave, este medio mínimo se mantuvo a 80°C y se mezcló con la solución resina:etanol 10% (peso/volumen), obteniendo un medio selectivo (RM) con concentraciones crecientes de resina (desde 0.5% a 8% v/v). Tanto las muestras de resina del pino, como las muestras de resina asociadas a las agallas, fueron sembradas en placas con RM e incubadas a 30°C durante 14 días.

Ejemplo 3. Extracción de ADN.

Las colonias de microorganismos observadas en las placas después de la incubación se lavaron con PBS estéril. Se aisló el ADN metagenómico de la suspensión microbiana usando el kit Power Soil DNA Isolation (MO BIO Laboratories).

Ejemplo 4. Secuenciación de ADN, ensamblado y análisis bioinformático.

El ADN metagenómico fue pirosecuenciado y las lecturas obtenidas fueron posteriormente ensambladas de nova con el programa informático NEWBLER (454 LifeSciences Roche). A partir del ensamblado, se realiza una predicción de genes con la herramienta MetaGeneAnnotator y se identificó el origen filogenético y la función de dichos genes mediante búsquedas BLASTX en la base de datos nr (no redundante) del NCBI. Para obtener una reconstrucción global del metabolismo a partir de los resultados del BLASTX, se utiliza la herramienta KAAS (KEGG Automatic Annotation System).

Ejemplo 5. Análisis de similitud de secuencia y sintenia del clúster de genes de degradación de diterpenos.

La secuencia del genoma de la cepa CECT8327 correspondiente al clúster de genes de degradación de diterpenos fue comparada con otras secuencias similares descritas para otras especies del mismo género. Para ello, se realizaron alineamientos múltiples con ClustalW y se comparó el orden génico dentro del clúster con el paquete GenoPlotR del software R.

Ejemplo 6. Caracterización de la capacidad de la cepa CECT8327 para degradar terpenos presentes en la resina.

La cepa CECT8327 encontrada en el análisis metagenómico fue aislada y crecida en un medio con resina de pino como única fuente de carbono. Para cuantificar la capacidad de degradar resina, se estime la masa de resina en el medio de cultivo a varios tiempos por diferencia entre la masa seca total y la masa seca correspondiente a las bacterias crecidas en dicho medio. Esta última masa se estimó a partir de un cultivo en medio rico LB, donde puede ser cuantificada de forma exacta por no existir componentes no solubles en el medio de cultivo. El contenido en resina durante la fase exponencial del cultivo en medio RM decreció de 5.4 g/L a 4.1 g/L después de 4 días, indicando una degradación cercana al 25% de la resina presente originalmente en el medio (ver Figura 1). El crecimiento de la cepa de la invención en medio RM se acompañó por la secreción de un sideróforo fluorescente, probablemente pioverdina, normalmente producida por diferentes especies de Pseudomonas bajo condiciones de ausencia de hierro.

Ejemplo 7. Caracterización de la capacidad de la cepa CECT8327 para degradar un material basado en terpenos lineales: el látex.

La cepa CECT8327 creció además en medio con látex como única fuente de carbono produciendo un sideróforo fluorescente. Se observe una unión celular al látex en los cultivos de la cepa CECT8327, y el crecimiento de esta cepa se correlacione con la aparición de fisuras en la superficie de muchas de las partículas de látex. Además, a partir de los 15 días, se observó una notable degradación del látex a nivel local, con la producción de hoyos u hornacinas asociadas a la interacción célula-látex (Figura 2).

Ejemplo 8. Ensayos de crecimiento en medio con compuestos hidrofóbicos, tipo hidrocarburos.

La cepa CECT8327 fue inoculada en medio mínimo (2 g/L $NaNO_3$, 1 g/L K_2HPO_4, 0.5 g/L $MgSO_4$, 0.5 g/L KCl, 0.2 g/L pectona bacteriológica) suplementado con diesel, y se cultivo durante 7 días a 30°C en condiciones de agitación. Los cultivos se dejaron en reposo tras los 7 días y se observó el nivel de emulsionado de la capa superficial de diesel. El medio inoculado

con la mencionada cepa presentó un emulsionado mayor que el medio no inoculado (ver Figura 3), debido a la secreción de agentes surfactantes por parte de Pseudomonas CECT8327.

Ejemplo 9. Ensayos de tolerancia a biomasas ricas en terpenos y a fungicidas.

Para evaluar la tolerancia de la cepa a otros terpenos y a fungicidas, se suplemento un medio de cultivo estándar como el LB (Luria-Bertani) con peladuras de naranja y con distintas concentraciones de imazalil (un agente fungicida ampliamente utilizado en la industria alimentaria), respectivamente. En el caso del medio suplementado con imazalil, se cuantifica el crecimiento mediante conteo de unidades formadoras de colonias en placa (ver Tabla 1), y no se observaron diferencias significativas en el crecimiento de la cepa de la invención en estos medios con respecto al medio control, no suplementado con ningún compuesto terpénico ni fungicida, por lo que la cepa demostró ser totalmente tolerante a estos compuestos.

Ejemplo 10. Ensayos de inhibición del crecimiento de hongos.

Las propiedades antifúngicas de la cepa CECT8327 fueron testadas con distintos aislados fúngicos de los géneros Aspergillus y Penicillium en medio mínimo con resina como única fuente de carbono y en medio rico LB (medio Luria-Bertani) mediante ensayos de confrontación. Estos ensayos se basaron en co-cultivar diferentes aislados fúngicos en medios sólidos en presencia o en ausencia de la cepa de la invención. Las esporas fúngicas fueron inoculadas en puntos situados en el centro de las placas Petri, mientras que, en los casos correspondientes, la cepa de la invención fue inoculada en líneas horizontales paralelas situadas a 2cm del hongo. El diámetro de las colonias fúngicas fue medido durante los 7 días siguientes. En todos los casos se observó una parada irreversible en el crecimiento fúngico entre el primer y el tercer día de crecimiento en presencia de Pseudomonas sp CECT8327, como se muestran en La Figura 4.

REIVINDICACIONES

1. Cepa aislada perteneciente a Pseudomonas sp. CEC18327 y mutantes de la misma donde las cepas mutantes son obtenidas a partir de la cepa CECT8327 como material de partida y mantienen o mejoran una o varias de las propiedades biológicas de la cepa CECT8327, que incluye al menos su capacidad para degradar terpenos cíclicos y acíclicos y/o derivados terpénicos cíclicos y acíclicos.

2. Cepa, según la reivindicación 1, caracterizada por comprender una secuencia con una similitud de al menos 70% respecto a la secuencia correspondiente a la SEQ ID NO: 1.

3. Cepa, según una cualquiera de las reivindicaciones 1-2, caracterizada porque su genoma comprende al menos un gen, introducido artificialmente, donde el al menos gen codifica para una sustancia o compuesto para su producción a escala industrial.

4. Composición que comprende la cepa de una cualquiera de las reivindicaciones 1-2.

5. Composición según la reivindicación 4 que comprende al menos un segundo agente antifúngico.

6. Composición según la reivindicación 5, donde el al menos segundo agente antifúngico es imazalil.

7. Uso de la cepa definida en una cualquiera de las reivindicaciones 1-2 o de la composición definida en la reivindicación 4, para la degradación de terpenos cíclicos y/o acíclicos y/o derivados terpénicos cíclicos y/o acíclicos.

8. Uso según la reivindicación 7, caracterizado porque los terpenos cíclicos se seleccionan del grupo que comprende: diterpenos, monoterpenos y sesquiterpenos.

9. Uso según la reivindicación 8, donde el monoterpeno en un limonoide.

10. Uso según la reivindicación 9, donde el limonoide es limoneno.

11. Uso según la reivindicación 10 en el tratamiento de biomasas para la producción de etanol por fermentación alcohólica.

12. Uso según la reivindicación 7, para la degradación de terpenos acíclicos seleccionados del grupo que comprende: látex, goma vulcanizada, goma no vulcanizada o gutapercha.

13. Uso de la cepa definida en una cualquiera de las reivindicaciones 1-2 o de la composición definida en cualquiera de las reivindicaciones 4-6, como antifúngico.

14. Uso según la reivindicación 13 en tratamientos post-cosecha.

15. Uso de la cepa definida en una cualquiera de las reivindicaciones 1-2 o de la composición definida en la reivindicación 4, para la degradación de imazalil.

16. Uso de la cepa definida en una cualquiera de las reivindicaciones 1-2 o de la composición de la reivindicación 4, como biofactoría en la producción de sustancias o compuestos industriales empleando la resina como fuente de carbono.

17. Uso de la cepa definida en una cualquiera de Las reivindicaciones 1-2 o de la composición de la reivindicación 4, como productor de agentes surfactantes en el tratamiento de un compuesto hidrofóbico.

18. Uso según la reivindicación 17, donde el compuesto hidrofóbico es un hidrocarburo.

19. Uso, según la reivindicación 18, donde el hidrocarburo es diesel.

Appendix D

Other publications

Vilanova *et al. Journal of Biological Engineering* (2015) 9:17
DOI 10.1186/s13036-015-0017-9

JOURNAL OF BIOLOGICAL ENGINEERING

LETTERS TO THE EDITOR Open Access

Standards not that standard

Cristina Vilanova[1†], Kristie Tanner[1†], Pedro Dorado-Morales[1], Paula Villaescusa[1], Divya Chugani[1], Alba Frías[1], Ernesto Segredo[1], Xavier Molero[1], Marco Fritschi[1], Lucas Morales[1], Daniel Ramón[2], Carlos Peña[3], Juli Peretó[1,4] and Manuel Porcar[1,5*]

Abstract

There is a general assent on the key role of standards in Synthetic Biology. In two consecutive letters to this journal, suggestions on the assembly methods for the Registry of standard biological parts have been described. We fully agree with those authors on the need of a more flexible building strategy and we highlight in the present work two major functional challenges standardization efforts have to deal with: the need of both universal and orthogonal behaviors. We provide experimental data that clearly indicate that such engineering requirements should not be taken for granted in Synthetic Biology.

Keywords: Synthetic biology, Biobrick parts, Standardization, Orthogonality

Letter to the editor

Synthetic Biology, as an engineering approach to biotechnology, requires standard biological parts in order to overcome the limitations of assay-and-error strategies widely used in regular biotechnology. Indeed, tinkering may be sophisticated enough for successfully accomplishing simple genetic modifications, but metabolic engineering, let alone genome "programming", require a basic toolbox of reliable standard biological parts to be combined into progressively increasing levels of complexity. In two recent letters published in this journal, concerns on the constraints of the Registry of Standard Biological parts associated to the limitations of 3A assembly methods have been highlighted [1, 2]. The Registry is indeed a valuable tool for synthetic biologists as a comprehensive catalog of biological parts, which can be physically obtained from it, combined *in silico* with the aid of ad-hoc software tools (http://sbolstandard.org/), and finally assembled to yield complex biological circuits with, in principle, predictable behaviors. However, as an analysis of the use of the Registry by iGEM participants demonstrates, there is a surprisingly limited reuse of biological parts [3].

In the present letter, we want to contribute to this debate on the challenges of biological standards by critically revising engineering assumptions in *E. coli* bioengineering (http://2014.igem.org/Team:Valencia_Biocampus). Of those assumptions, there are two key concepts linked to standardization that are often incorrectly taken for granted: universality and orthogonality. The first notion refers to the standard behavior (when "standard" is used as an adjective it is usually synonymous to "universal"); in other words, biological parts are expected to display the same or very similar outputs independently of the biological system they are placed into. The second concept, orthogonality, relates to the independent behavior of biological parts.

We bench-tested these engineering pillars in the simplest scenarios: standardization was studied by introducing six DNA constructions (see Additional file 1: Table S1) built from commonly-used Biobrick parts in six different laboratory strains of *E. coli* and measuring their output under the same experimental conditions, whereas orthogonality was tested by co-transforming one of the strains (XL1-Blue) with a couple of these constructions (a green fluorescent protein placed under the control of a constitutive promoter, and a red fluorescent protein controlled by the same promoter) and measuring their output with flow cytometry techniques. Under our experimental conditions, significant differences in terms of expression levels were found among all the strains in five out of six constructions (Fig. 1) regardless the promoter type (constitutive or inducible) and the reporter protein

* Correspondence: manuel.porcar@uv.es
†Equal contributors
[1]Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, C. Catedràtic José Beltrán 2, 46980 Paterna, Spain
[5]Fundació General de la Universitat de València, Valencia, Spain
Full list of author information is available at the end of the article

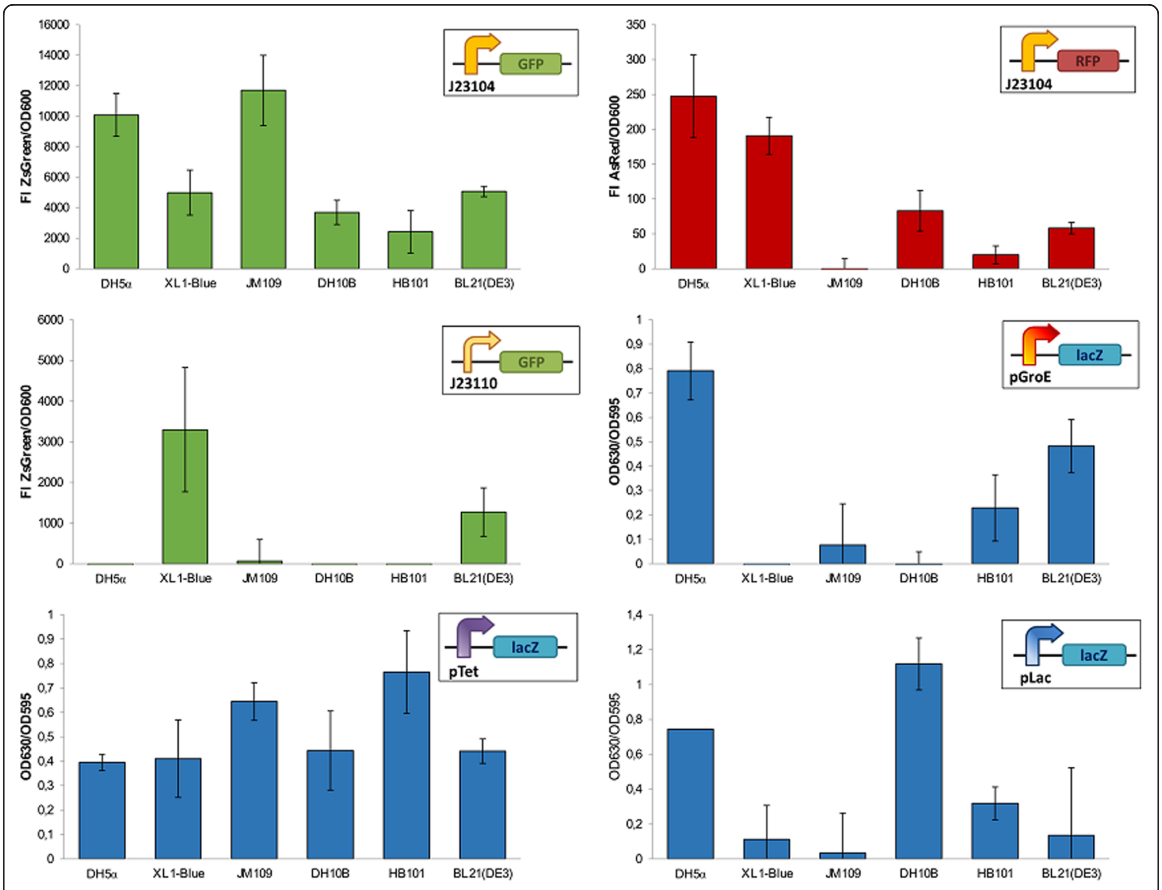


Fig. 1 Behaviour of a set of Biobrick parts in different *E. coli* host strains. The output of DNA constructs consisting of a promoter coupled to a reporter protein was measured under the same experimental conditions for each strain. Note that both constitutive and inducible promoters were tested, and different reporters (fluorescence proteins and coloured compounds) were used. All measurements were normalized by the OD600 value of each culture, and corrected by the basal output observed in control strains transformed with an empty plasmid. Error bars show the standard deviation of three independent biological replica. See Additional file 1 for further experimental details

(fluorescent proteins or β-galactosidase), and double transformants did not exhibit a 1:1 red:green fluorescent phenotype (Fig. 2a). The lack of orthogonality of these two biological parts between them was in contrast with the stability of *E. coli* as a chassis, as we tested through a proteomics approach. Figure 2b shows the proteomic profile of a transformed *E. coli* strain with a GFP-containing plasmid and of two control strains (one non-transformed and one containing the empty plasmid), which reveals a minor impact of GFP and/or antibiotic resistance expression on the global bacterial proteomic architecture. *E. coli* is thus –at least in our conditions– a solid, orthogonal system respect to the heterologous protein expression shuttle it hosts.

The cellular phenomena underlying the lack of standard and orthogonal behavior of the Biobrick parts we tested might range from differences in protein maturation times [4] and impact on biosynthetic burden [5] to context – upstream and downstream sequences effects– dependencies [6] as well as to stochastic effects or intrinsic and extrinsic noise [7]; and references therein].

The fact that the tested genetic modifications proved weak standards in terms of universality and orthogonality does not necessarily imply the impossibility of engineering a particular strain in a predictable way, but it poses enormous difficulties in engineering strains *on the basis of* the work done on other strains, particularly taking into account that only a fraction of the genome is shared by all E. coli strains [8]. This suggests the need of a strain-by-strain both modelling and experimental previous effort. On the other hand, taking advantage of biological flexibility can be used in order to set up more robust devices, such as the use of bacterial haemoglobin to enhance production of foreign fluorescent proteins

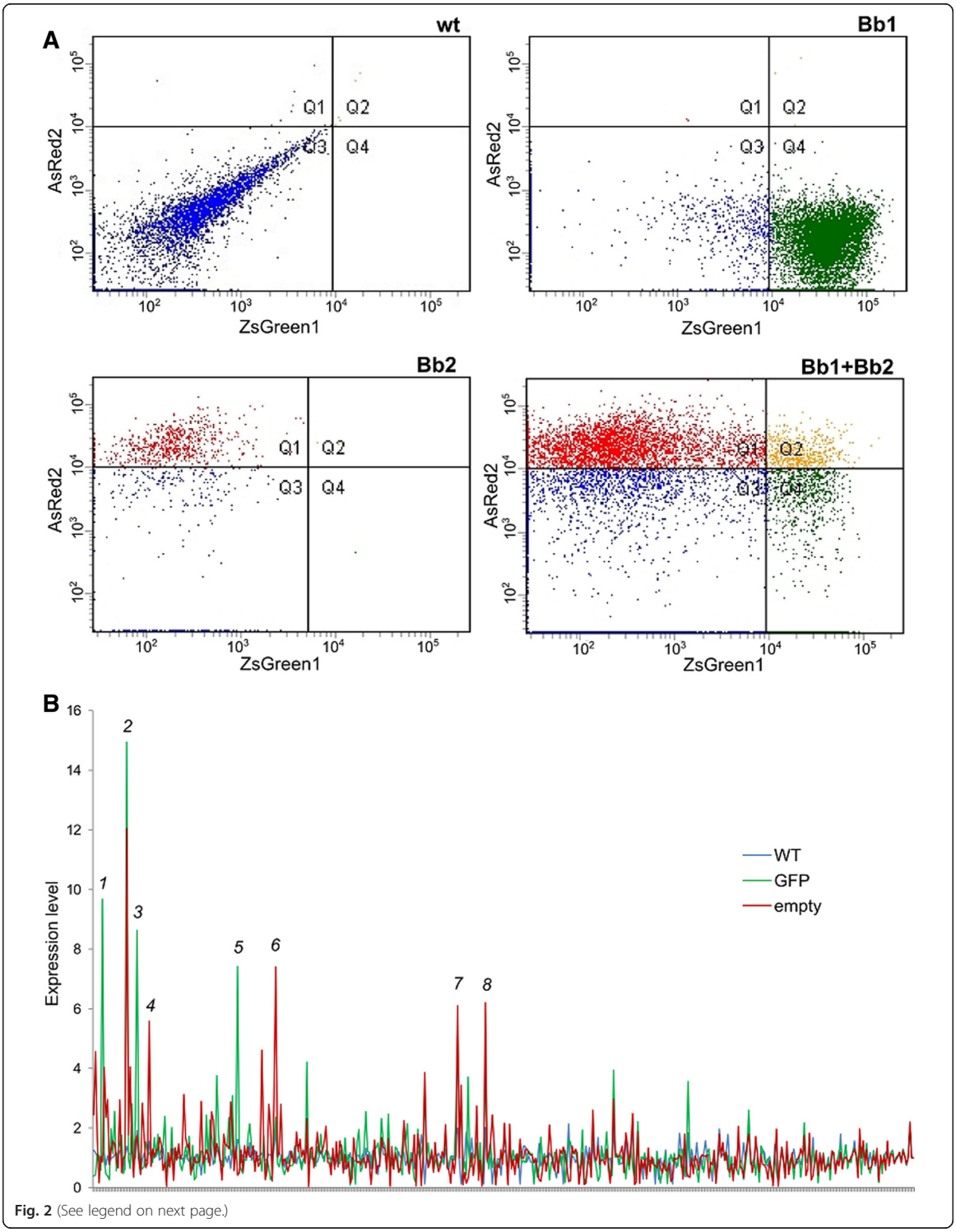


Fig. 2 (See legend on next page.)

(See figure on previous page.)
Fig. 2 Orthogonality tests performed on a simple combination of two Biobrick parts. **a** Fluorescence output displayed by *E. coli* XL1 strain transformed with a single plasmid containing a constitutive promoter coupled to a green fluorescent protein (*Bb1*), a single plasmid containing the same promoter coupled to a red fluorescent protein (*Bb2*), and a combination of both plasmids. Plots showing flow cytometry measurements performed on individual cells (*dots*). **b** Comparison of the proteomic profile of an *E. coli* strain constitutively expressing a green fluorescent protein (*green lines*) with that of the same strain carrying an empty plasmid (*red lines*) and the control non-transformed strain (*blue lines*). Proteins showing a statistically significant change in expression are numbered according to Additional file 1: Table S2

[9], tuning intracellular physical distances between the regulator source and the target promoter for selecting a given level of noise in Synthetic Biology constructs [10], or designing synthetic constructs imposing a minimal burden to the host cells [5].

There is a general assent in the Synthetic Biology community on the need of collections of biological parts, in which engineering features (universality, stability, orthogonality, among others; [11]) should be checked and unambiguously quantified as a basic prerequisite for obtaining predictable and scalable designs. Systematic failures and difficulties to meet engineering standards might not constitute the most prized result in terms of publication purposes, but we strongly believe that a comprehensive view on the standardization failures of today is the strongest path towards the development of fully standard and orthogonal biological parts in the future.

Additional file

Additional file 1: Table S1. Biobrick parts (Bb1 to Bb6) included in the different DNA constructions tested in this work. **Table S2.** Proteins displaying statistically significant alterations in their expression levels, as detected by iTRAQ analysis (DOCX 18 kb).

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MP conceived and coordinated the design of the study and supervised laboratory work. CV and KT performed the experiments related to orthogonality and proteomics, whereas PDM and AF performed the measurements related to standard behavior. PV and DC supported all the experimental work. ES, MF, XM, LM and CP contributed to data analysis. JP and DR technically supported this work and strongly participated in data interpretation and discussion. MP, CV and KT wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
Valencia-Biocampus iGEM 2014 Team was supported by the European Projects ST-Flow and ERA-SynBio, Biopolis SL (Parc Científic de la Universitat de València) and the University of València through the International Excellence Campus and Col·legi Major Rector Peset. We are especially indebted to Rafael García-Martínez (OPEX, Oficina per a les Polítiques d'Excel·lència, Universitat de València) and Soledat Rubio (Càtedra de Divulgació de la Ciència de la Universitat de València) for their constant encouragement and support. We also thank the proteomics service (SCSIE) of the Universitat de València for their technical support. CV is a recipient of a FPU fellowship from the MECD Spanish Ministry.

Author details
[1]Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, C. Catedràtic José Beltrán 2, 46980 Paterna, Spain. [2]Biopolis S.L, Parc Científic Universitat de València, Paterna, Valencia, Spain. [3]Instituto de Física Corpuscular (IFIC), CSIC - Universitat de València, Burjassot 46100, Spain. [4]Departament de Bioquímica i Biologia Molecular, Universitat de València, Burjassot 46100, Spain. [5]Fundació General de la Universitat de València, Valencia, Spain.

Received: 6 August 2015 Accepted: 16 September 2015
Published online: 01 October 2015

References
1. Alnahhas RN, Slater B, Huang Y, Mortensen C, Monk JW, Okasheh Y, et al. The case for decoupling assembly and submission standards to maintain a more flexible registry of biological parts. J Biol Eng. 2014;8:28.
2. Azizi A, Lam W, Phenix H, Tepliakova L, Roney IJ, Jedrysiak D, et al. No training required: experimental tests support homology-based DNA assembly as a best practice in synthetic biology. J Biol Eng. 2015;9:8.
3. Vilanova C, Porcar M. iGEM 2.0–refoundations for engineering biology. Nat Biotechnol. 2014;32:420–4.
4. Hebisch E, Knebel J, Landsberg J, Frey E, Leisner M. High variation of fluorescence protein maturation times in closely related Escherichia coli strains. PLoS One. 2013;8:e75991.
5. Ceroni F, Algar R, Stan G-B, Ellis T. Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat Methods. 2015;12:415–8.
6. Pasotti L, Politi N, Zucca S, Cusella De Angelis MG, Magni P. Bottom-up engineering of biological systems through standard bricks: a modularity study on basic parts and devices. PLoS One. 2012;7:e39407.
7. Viney M, Reece SE. Adaptive noise. Proc R Soc B. 2013;280:20131104. doi:10.1098/rspb.2013.1104.
8. Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced *Escherichia coli* genomes. Microb Ecol. 2010;60:708–20.
9. Kang DG, Kim YK, Cha HJ. Comparison of green fluorescent protein expression in two industrial *Escherichia coli* strains, BL21 and W3110, under co-expression of bacterial hemoglobin. Appl Microbiol Biotechnol. 2002;59:523–8.
10. Goñi-Moreno A, Benedetti I, Kim J, de Lorenzo V: Deconvolution of Gene Expression Noise into Physical Dynamics of Cognate Promoters. bioRxiv doi: http://dx.doi.org/10.1101/019927
11. Lucks JB, Qi L, Whitaker WR, Arkin AP. Toward scalable parts families for predictable design of biological circuits. Curr Opin Microbiol. 2008;11:567–73.

disclosure requirements, an opportunity now exists to focus the SEC on using biotech-specific guidelines to specify those disclosures that are relevant to the biotech companies from a reasonable investor's point of view and eliminate the disclosure that has grown over the years and is no longer relevant.

Biotech companies would also benefit if the SEC acknowledged that an investor in the biotech sector should be presumed to have some knowledge of the regulatory factors affecting biotech business, such as the US Food and Drug Administration approval process, and a rudimentary understanding of patents whose ownership is a vital component of the value of a drug and reflected in the market value of a biotech company without marketed products. The SEC also could revise its rules to recognize that institutional investors and not individual investors contribute actively to the price-setting mechanism of the public capital markets. Thus, the regulatory framework should take into account that any investor in a specialized field like biotech has to take the responsibility for understanding the industry and some of the fundamental science necessary to appreciate a particular company's business. Disclosure has become voluminous as too much emphasis has been placed on the biotech company having to explain industry basics to presumably uninformed investors.

The acceptance of the premise that an investor has a basic level of familiarity with an industry and its business fundamentals might contribute to the SEC ameliorating the extent of risk factor disclosures currently provided to investors, which can run to 20 or more pages. Comments on company filings by the SEC staff are only one source of pressure to include extensive risk factors. The major driver of the expanded lists of risk factors has been the perceived protection they provide against future securities law litigation. The "bespeaks-caution doctrine" adopted by the US courts and "safe harbors" for forward-looking statements support the idea that a company can protect itself by not skimping on risk factors. That results in the inclusion of risk factor disclosures that should be obvious to even a reasonably educated investor. The SEC encouraged inclusion of fewer risk factors as part of its plain-English initiative in the late 1990s. Even so, companies and their lawyers resisted that pressure without some assurance from the SEC that they would not be giving up what has been considered to be liability protection. Similarly, if disclosure reform results in companies not being required to disclose information that had been included as a matter of course in the past, the SEC will need to protect companies that avail themselves of reduced disclosure enacted by the SEC.

Although the SEC staff report mandated by the JOBS Act calls for a comprehensive review of the disclosure requirements rather than incremental steps to ease the disclosure burden, this should not preclude the SEC from making immediate changes in the right direction. Recently, SEC staff has taken some minor steps in this regard by reducing the disclosures it has been requiring in IPO prospectuses on the valuation methods the company has used to determine the exercise prices of stock options granted preceding the IPO. Over the years, what began as SEC staff comment letters requesting details to ensure appropriate accounting for stock options (and related so-called 'cheap stock' issues) grew into several pages of disclosure in IPO prospectuses detailing the company's valuation of its own stock for options-granting purposes.

Along with the opportunity it presents, disclosure reform poses risks to the biotech industry. When regulators make changes, even if focused primarily on reducing disclosure overload, they are likely to include increased requirements in certain areas. One area of potential pressure could be toward faster disclosure. Other countries have continuous disclosure regimes, whereas the underlying premise of US securities laws is that, absent a duty to disclose, no disclosure is required. The periodic disclosure requirements—such as Form 10-Ks, Form 10-Qs and proxy statements, as well as the Form 8-K mandating current reporting obligations of material events—impose limits on that premise, while still providing meaningful protection to companies in many circumstances. Weakening that approach or mandating that companies report new developments more quickly than currently required would place additional burdens on companies to increase staff for this disclosure framework and could result in liability in more situations than is already the case.

Many factors have contributed to our current state of US public company disclosure. The SEC's current efforts to consider disclosure reform that could reduce the burdens of that disclosure are welcome. Biotech companies and BIO have an opportunity to contribute to the success of that effort and to help shape the reforms in ways that particularly benefit biotech companies and their investors. This is an opportunity to make their voices heard and help shape their disclosure destiny.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Robert Evans III & Mark Kessel

Shearman & Sterling LLP, New York, New York, USA.
e-mail: revans@shearman.com or mark.kessel@shearman.com

iGEM 2.0—refoundations for engineering biology

To the Editor:

The international Genetically Engineered Machine (iGEM) student competition is both a workbench and a showcase for synthetic biology. The competition is based on a simple idea: synthetic biology engineering principles of standardization, abstraction and modularity can be applied to biotech to make engineering new functions in life systems less intimidating, more accessible and more predictable. This year, iGEM will have been running for a decade, and the organization will celebrate the event with a 'giant jamboree' involving as many as 300 teams. The competition has reached a peak in terms of media impact (with a considerable number of Internet searches and a clear seasonal search pattern fitting the competition calendar (http://www.google.es/trends/explore#q=igem, accessed 17 December 2013), attendance and expectations. As former participants in iGEM (C.V. was a student attendee for three years and M.P. was a team supervisor and judge for six years), we have conducted an analysis of iGEM projects presented over the past 10 years. Our analysis reveals several challenges that the competition must face if it is to remain a flagship of synthetic biology.

iGEM takes place in a pedagogic setting and within a time frame of less than 1 year, in such a way that even undergraduate students without prior training in biology, but with reasonable technical and theoretical support, can participate[1]. It has been described as a test bed of synthetic biology projects; as

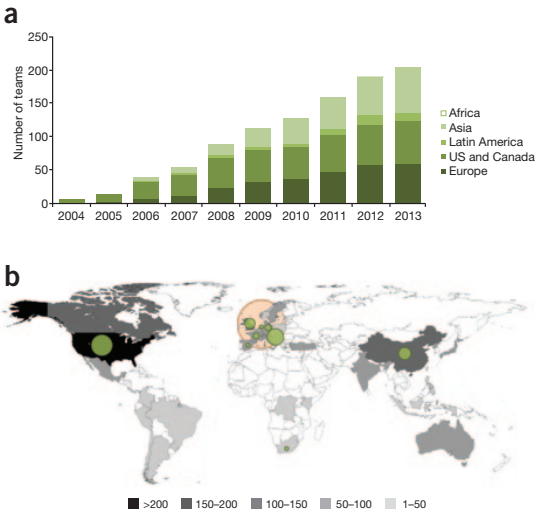


Figure 1 iGEM attendee analysis. (**a**) Number of teams attending iGEM by region. (**b**) Regional origin of attendees. Shading indicates the cumulative number of teams representing each region since 2004. The distribution of awards is represented by circles, whose size is proportional to the number of finalist teams representing each country (smallest circles represent one finalist team, whereas the biggest one represents 14 finalist teams since 2006). For European teams, awards are represented by country (green circles) as well as for the whole continent (orange circle).

an example of engineering ingenuity[2]; as a framework for increasing interest in 'human practices' (the term used in iGEM referring to ethical, legal and social implications (ELSI)) approaches, such as biosecurity and biosafety[3] and intellectual property[4]; and as a challenge providing leadership to the field[5]. iGEM seeks to not only educate young students in synthetic biology but also foster other personal abilities, such as self-confidence, creativity and effort. Multicultural and interdisciplinary exchange of knowledge, teamwork, networking and information sharing via public wikis are also part of the additional values of the iGEM experience.

The peak of the iGEM competition is the jamboree, where the teams present their synthetic biology projects conducted during the summer. Since 2011, the competition has been organized in regional jamborees that take place in October, and only a percentage (around one-third) of the teams go to the world championship, which is held the first week of November at the Massachusetts Institute of Technology (MIT) in Cambridge, Massachusetts.

Over the past decade, participation in the competition has expanded from an early group of US projects to >200 teams distributed worldwide (**Fig. 1**). Even so, the geographical distribution of iGEM participants remains biased toward North America, Eastern Asia and Europe. Every year each of these regions account for around 25% of the participating teams; however, groups from Latin America have been more involved in recent years (**Fig. 1a**). In terms of awards, Europe has been the most successful continent (**Fig. 1b**); indeed, there have been three competitions in which all the finalists were of European origin. One team, from Ljubljana, Slovenia, has reached the finalists pool (comprising three teams) and been awarded the grand prize five and three times, respectively.

In terms of judges, geographical distribution is predominantly local in the regional jamborees (i.e., judges come from where the regional jamboree is based) and international in the world phase. Even so, the diversity is lower among judges than among teams in the world jamboree; overall, there is a strong geographic bias in favor of US and European judges over Asian ones (see http://igem.org/Judge_List and http://2013.igem.org/Judging/Transparency_at_iGEM).

In terms of funding, an average team attending the regional phase (ten students and two advisors) spends a minimum of $10,000 just for team and individual fees, travel and lodging (this does not include the additional fees when teams advance to the world championship). Added to this is the cost of the rather expensive materials and technologies used to perform often ambitious wet lab experiments in iGEM (such costs can range from thousands to tens of thousands of dollars, depending on the project). In the absence of publicly available data, we estimate (very roughly) that the average cost per team is around $20,000–50,000 per year, suggesting that the cost for all participants in the 2013 competition overall was approximately $4 million–10 million.

Although the total price tag for each year's iGEM is similar to that of a medium-to-large cooperative scientific project, which would be expected to yield important scientific publications and/or patents, it is important to stress that iGEM is an educational program. As such, its success should not be measured in terms of scientific publications or patents. In fact, each iGEM competition typically yields very few scientific publications or intellectual property and, to the best of our knowledge, fewer than half of the finalist projects have been published so far. The rate of published iGEM projects has not risen in line with the maturity of the competition (**Table 1**).

A set of specialized volunteer judges choose medal and award winners and select which teams will advance from regional competitions to the world championship. In the last world jamboree, 52 judges were in charge of assessing the performance of 146 teams. The majority of judges (76.9%) came from academia (many of which were also team instructors); 9.6% originated from government departments; 5.8% were from companies; and the remaining 7.7% were from the committee (iGEM organizers; see http://igem.org/Judge_List).

Whereas in regional competitions judges consider a team's overall project on the wiki, presentation, modeling of the problem, submission and use of BioBrick standards, in the world championship only four aspects of the project are assessed (overall project, wiki, presentation and modeling). Each judge casts votes that are converted into a numerical score in an online-based rubric. There is a double award system. First, medals (gold, silver and bronze) are awarded on completion of a list of requisites, including the construction of new BioBrick parts, the submission of these parts to the Registry of Standard Biological Parts (http://parts.igem.org/Main_Page) and the assessment of the project in terms of safety and bioethics. Second, prizes are awarded by judges to the winner, first runner-up and second runner-up (http://2013.igem.org/Judging/Awards). Only about one-third of the teams advance to the world championship at MIT, and this rate might become even more competitive in the coming years if attendance continues to rise.

The iGEM competition and the Registry of Standard Biological Parts are two branches of the same tree. In fact, one requisite to earn a medal is to submit at least one biological part, either natural or engineered, to the registry. To prepare a BioBrick part from raw DNA, students have to 'stick' specific prefix and suffix short adapters, including restriction enzyme cutting sites, to the desired DNA sequences to make them suitable for the registry and thus, theoretically, standardizable and module ready. The average number of parts submitted

Table 1 Finalist projects in the iGEM competition, 2006–2013

	Team	Project	Reference
1	Slovenia 2006	'Engineered human cells: say NO to sepsis'	
2	Imperial 2006	'The *E.coli* reporter'	11
3	Princeton 2006	'Programmed differentiation of mouse embryonic stem cells using artificial signaling pathways'	
4	Peking 2007	'Towards self-differentiated bacterial assembly lines'	12
5	University of California, Berkeley 2007	'Bactoblood'	
6	Paris 2007	'Synthetic multicellular bacterium'	
7	Ljubljana 2007	'Virotrap': a synthetic biology approach against HIV	13
8	University of California, San Francisco 2007	'Directing biology through synthetic assemblies and organelles'	14
9	University of Science and Technology of China 2007	'Extensible logic circuit in bacteria'	15
10	Slovenia 2008	'Immunobricks'	16
11	Freiburg 2008	'Modular synthetic receptor system'	
12	Caltech 2008	'Engineering multi-functional probiotic bacteria'	
13	Cambridge 2009	'E. chromi'	
14	Heidelberg 2009	'Spybricks'	17
15	Valencia 2009	'iGEM Lighting Cell Display'	18
16	Slovenia 2010	'DNA coding beyond triplets'	19
17	Peking 2010	'Biodecontamination kit'	
18	Bristol Cathedral Choir School-Bristol 2010	'AgrE.coli'	
19	Washington 2011	'Make it or break it: diesel production and gluten destruction'	20
20	Imperial College 2011	'AuxIn'	
21	Zhejiang University, China 2011	'Rainbofilm'	
22	Groningen 2012	'Food warden' or in any of the project names	
23	Paris Bettencourt 2012	'bWARE'	
24	Slovenia 2012	'Switch IT': inducible therapeutics	
25	Heidelberg 2013	'The Philosopher's Stone'	
26	Technical University Munich 2013	'PhyscoFilter'	
27	Imperial College 2013	'Plasticity'	
28	Paris Bettencourt 2013	'Fight tuberculosis with modern weapons'	
29	Bielefeld 2013	'Ecolectricity'	
30	Sun Yat-sen University, China 2013	'iPS cells safeguard'	

to the registry, around 10 per team, has remained relatively stable throughout the history of the competition, although award-winning teams tend to submit many more, up to hundreds (**Fig. 2a**). To date, iGEM teams have collectively submitted >12,000 parts to the registry (**Fig. 2b**). Only around 40% of those have been checked satisfactorily to ensure that they work as expected upon submission (**Fig. 2c**). It must be stressed that the growing number of parts, the diversity of assembly methods used and the difficulties of performing quality control on a continuous basis (most of which still relies on the controls made by the teams) require iGEM organizers and the registry to undertake characterization, preparation and delivery work of titanic proportions.

The link between iGEM and the registry also works the other way around: iGEM teams do not only submit parts but are also encouraged to use the BioBrick standards submitted by other teams in previous editions at their convenience, which are already present in the registry. But an analysis of the *de novo* versus registry-issued parts (according to the case-by-case information in the wiki of each team) reveals that iGEM teams that have been successful in terms of awards tend to avoid the uncertainties of the parts designed and/or characterized by others and choose to use new, ad hoc DNA parts for a specific purpose (**Fig. 2d**). As previously mentioned, a cautious attitude toward standard DNA parts seems to be common among participants[6,7].

In evaluating the success of iGEM as a didactic endeavor, an interesting comparison can be made with similar efforts, such as the FIRST robotics competition (http://www.usfirst.org/), devoted to promoting engineering and technology skills among young students (most of whom are in high school). Although the number of students trained during the first decade of the FIRST competition is much higher than in iGEM, the latter has reached—with substantially lower overhead costs —an unprecedented geographical spread in a shorter time. Similarly to that of FIRST, the main outcome of iGEM is educational. The main goal of iGEM is to educate students in synthetic biology, so that they might contribute to transformational advances at some time in the future. Funding in iGEM is thus a long-term educational investment. As stressed by Randy Rettberg, general iGEM coordinator and president of the iGEM foundation, during the 2013 closing ceremony, iGEM aims to foster effort, accomplishment, excellence, respect, cooperation and integrity.

From our analysis of iGEM over the past ten years, we believe that the competition has to adapt if it is to maintain its status as a pillar of synthetic biology and as an example of an exciting and dynamic scientific competition. For it to do so, we believe greater focus should be placed on the quality rather than the quantity of parts in the registry. The increasing number of parts

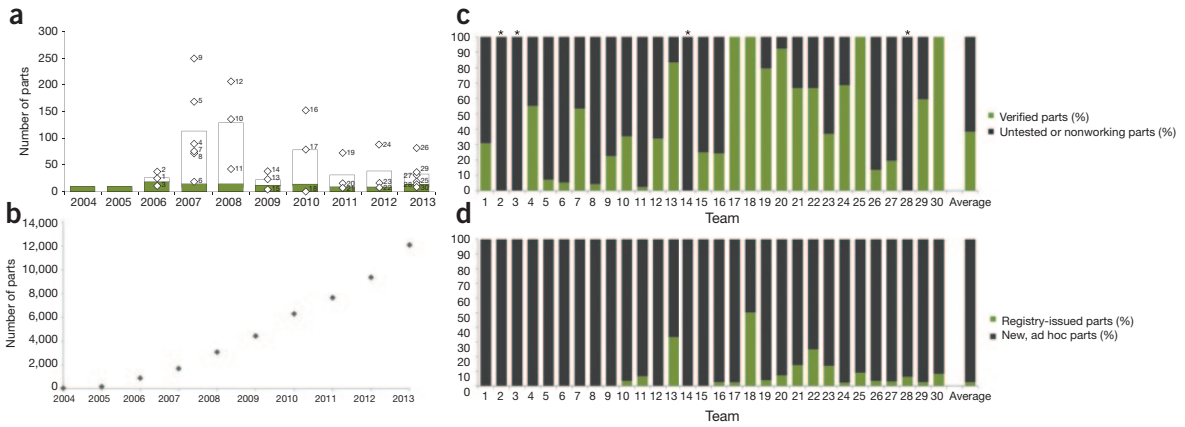


Figure 2 iGEM and the Registry of Standard Biological Parts. (**a**) Average number of parts submitted per team (green bars) and per finalist team (white bars) during 2004–2013. Diamonds (numbered in accordance with **Table 1**) represent finalist teams. (**b**) Cumulative number of iGEM parts in the registry since 2004. (**c**) Proportions of verified parts (labeled as 'working' in the registry (http://parts.igem.org/Main_Page)) and untested or nonworking parts submitted to the registry, as stated by iGEM finalist teams in 2006–2013. Finalist teams are numbered according to **Table 1**. Ambiguous data were found for teams marked with asterisks. (**d**) Proportion of registry-issued and new parts used by iGEM finalist teams in 2006–2013, as reported on each team's wiki pages.

and the worrying trend of adding new ones rather than using standard ones deserve deep reflection by participants and organizers alike. An intelligent strategy is already in place: iGEM teams get more 'points' if they improve the characterization of an existing part. We would suggest an additional improvement strategy consisting of selecting a relatively small number (no more than 100) of parts every year and asking teams to improve their characterization and/or test their performance in a range of hosts and conditions. This would yield a smaller but improved pool of more reliable parts.

From another perspective, one might ask whether the strong linking of iGEM to the BioBrick biological standard is necessary. Is this the only and/or best standard possible? Should molecular cloning–issued biological devices be a requisite for a team to attend the competition? This question is particularly pertinent in an era where not only adaptor-based standard cloning systems, but also zinc-finger nucleases, transcription activator–like endonucleases and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas-based methods[8] are playing an increasing part in synthetic biology. In the context of continuously falling costs of DNA chemical synthesis[9], the cost of chemically synthesizing all the BioBrick parts used by an iGEM team is relatively affordable: considering 1-kilobase constructs (containing a small promoter, an average-length prokaryotic gene and a terminator) and a low synthesis price (around $0.28 per base pair), an iGEM team could have ten BioBrick standards ready to transform for $2,800. This cost is ~10 times higher than that of a BioBrick parts assembly kit, but it is still substantially lower than the other expenses an iGEM team must cover (a regional jamboree registration for the whole team, for instance). Furthermore, if molecular cloning were not needed, students would receive presynthesized constructs quickly (in fewer than two weeks), so more time could be dedicated to characterizing the parts in depth. This fact leads us to suggest that, however radical this initiative may be, any synthetic DNA part, even those lacking BioBrick-ready adapters, should be acceptable in the competition.

Regardless of technical developments, iGEM organizers should also reflect on the open-source nature of biological parts, especially if competition projects are to find use in the biotech sector, an area traditionally based on patents and trade secrets. Ownership and sharing in synthetic biology oscillates between widespread gene patenting and open source, in what has been described as a 'diverse ecology'[10]. iGEM is inspired by engineering and thus by open-source software and distributed innovation. Because the competition focuses on developing particular synthetic biology applications as well as fundamental tools in a scenario of thousands of building parts, open source might be the most logical choice. As Drew Endy stressed in an interview in 2007, "My hope is that by giving things away I will get more back in the long run." One successful example of this philosophy is the BioBrick Public Agreement (BPA) from the BioBricks Foundation, which is a free-to-use "legal tool that allows individuals, companies, and institutions to make their standardized biological parts free for others to use" (http://biobricks.org/bpa/).

However, a clarification of the registry's legal status is desirable to ensure transparency and ensure that transfer to companies can take place in situations where a project shows applications of particular industrial promise[4].

Last, we feel that now is an opportune moment for organizers to reevaluate the judging criteria used in iGEM. Judging is more than the final stage of the work; judges' preferences are likely to strongly influence team instructors and advisors and shape their choices for future projects. Given the large diversity of research topics, experimental models and technological choices, award-winning projects tend to be imitated. A clear example of this is the multiple projects aiming to develop a biodecontamination kit, which are recurrently found among finalists (see Peking 2010 (http://2010.igem.org/Team:Peking), TU Munich 2013 (http://2013.igem.org/Team:TU-Munich) and Dundee 2013 (http://2013.igem.org/Team:Dundee)). However, although almost all projects have a highly applied purpose, judges are typically pleased by aspects such as the 'originality' of the work (it has not been seen before in iGEM) or its 'roundness' (it tells a story, from a very simple idea to a prototype). One example of this is the Groningen 2012 (http://2012.igem.org/Team:Groningen) project on food spoilage control, which used an original bioprospection strategy to identify and select strong promoter sequences. Beyond the impact this strategy had on the

judges, the immediate application to industry (or the 'responsible research and innovation' (RRI) issues, which emphasize the utility and benefit for society and the environment) should be key factors of a successful project. If they were, final rankings would certainly change and successful teams would send a clear message on the trends to follow. Judging is always, but particularly in iGEM, a bidirectional process: it ranks proposals but it also shows the way for forthcoming ones.

Given the wide range of complexity and immediate industrial applicability among iGEM projects, we suggest that the degree of sophistication (for example, the number of biological parts/devices used, the difficulty of the host organism and the complexity of regulation output) should be formally considered as a ranking criterion for judges. This would help to further increase the competitiveness of the projects.

If the competition is to place more emphasis on translating projects into real industrial applications, then more thought needs to be put into judging criteria that reinforce this aspect. At present, prizes perhaps encourage spectacular and audacious basic research, which is often not built upon; each year many teams set up brand new projects unrelated to past efforts, even award-winning ones. A greater proportion of industrial members on the judging committee would have an immediate effect by redirecting the competition from the 'game phase' (preliminary exploration) to the 'real-world phase'. More judges from government with expertise in regulatory, health, agricultural or defense issues may also expand the diversity of views and decrease academic biases. An increased presence of Asian judges in the world jamboree would also be highly desirable. Another suggestion for improving the quality of judging is standardization of the number of judges per team. Although similar numbers are assigned to each track, judges can cast votes for unassigned teams. As a result, some teams often have many more votes (either positive or negative) than others. Judging has improved a lot during the past few iGEM competitions. The online questionnaire introduced in 2012 to be filled in by judges incorporates some suggestions that arose during the 2012 regional jamborees, particularly in Europe. It is arguable whether the machine-based ranking of teams should be corrected with data such as team budget or number of students, advisors or instructors. Given the educational nature of the competition, we suggest that it should.

A greater involvement of ELSI specialists and, particularly, a focus on reflexivity and RRI would also help to shape competition trends by encouraging teams to define their projects with societal and environmental benefits as major goals, along with one of the central aspects of RRI: transparency. Transparency has always been a guiding principle in iGEM, with an open-source–like Internet-based community that shares data, protocols and DNA samples. The economic resources used in iGEM should not be excluded from such information in future. Detailed data on public and private funding as well as their precise assignment throughout the project should be a requisite for each iGEM team. As stated above, we believe that fair judgment is not possible without taking into account the funding-to-results ratio. Determining this ratio is central to assessing productivity of a particular project and of the competition as a whole. Therefore, for the sake of transparency, we propose that participating teams be asked to make their budget public on their wikis.

In summary, we have proposed a range of suggestions that could improve the quality of standards, increase transparency of funding, foster industrial orientation and redefine and enhance judging of the competition. The experience of a decade of iGEM indicates that such redefinition is imperative for this outstanding competition to meet the great expectations of synthetic biology going forward.

ACKNOWLEDGMENTS

We are indebted to A. Delgado, V. de Lorenzo and J. Peretó for the critical reading of this manuscript. We are very grateful to anonymously interviewed editors of the top-ranked biotechnology and applied microbiology journals who kindly gave us their opinion on synthetic biology and iGEM. We are indebted to Càtedra de Divulgació de la Ciència (UCC+i) and VLC/Campus—Valencia International Campus of Excellence (Universitat de València) for their enthusiastic support. C.V. was funded with a FPU (Formación de Personal Universitario) grant from the Spanish MECD (Ministerio de Educación, Cultura y Deporte). This work was partially funded by the European Project ST-FLOW (FP7-KBBE-2011-5).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Cristina Vilanova[1] & Manuel Porcar[1,2]

[1]Cavanilles Institute of Biodiversity and Evolutionary Biology (Universitat de València), Valencia, Spain. [2]Fundació General de la Universitat de València, Valencia, Spain.
e-mail: manuel.porcar@uv.es

1. Endy, D. *Nature* **438**, 449–453 (2005).
2. Goodman, C. *Nat. Chem. Biol.* **4**, 13 (2008).
3. Guan, Z. *et al. Bioscience* **63**, 25–34 (2013).
4. Campos, L. *Biosocieties* **7**, 115–139 (2012).
5. Materi, W. *Methods Mol. Biol.* **852**, 251–272 (2012).
6. Kwok, R. *Nature* **463**, 288–290 (2010).
7. Porcar, M. & Peretó, J. *Syst. Synth. Biol.* **6**, 79–83 (2012).
8. Gaj, T., Gersbach, C.A. & Barbas, C.F. *Trends Biotechnol.* **31**, 397–405 (2013).
9. Carlson, R.H. *Biology Is Technology* (Harvard University Press, 2010).
10. Calvert, J. *Biosocieties* **7**, 169–187 (2012).
11. Sander, C. *et al. BMC Syst. Biol.* **1**, S9 (2007).
12. Lou, C. *et al. Mol. Syst. Biol.* **6**, 350 (2010).
13. Majerle, A., Pristovsek, P., Mancek-Keber, M. & Jerala, R. *J. Biol. Chem.* **286**, 26228–26237 (2011).
14. Bashor, C.J., Helman, N.C., Yan, S. & Lim, W.A. *Science* **319**, 1539–1543 (2008).
15. Zhan, J. *et al. Mol. Syst. Biol.* **6**, 388 (2010).
16. Mori, J. *et al. Vaccine* **30**, 5856–5863 (2012).
17. Reis, Y. *et al.* European Patent Application EP2311961 (2009).
18. Vilanova, C. *et al. J. Biotechnol.* **152**, 93–95 (2011).
19. Conrado, R.J. *et al. Nucleic Acids Res.* **40**, 1879–1889 (2012).
20. Gordon, S.R. *et al. J. Am. Chem. Soc.* **134**, 20513–20520 (2012).

Sex-ratio-biasing constructs for the control of invasive lower vertebrates

To the Editor:

There are few cost-effective means of controlling the many types of invasive fish, amphibians and reptiles that cause substantial economic and ecological damage worldwide[1]. Notable examples include sea lampreys (*Petromyzon marinus*), common carp (*Cyprinus carpio*), cane toads (*Bufo marinus*), bullfrogs (*Rana catesbeiana*) and the brown tree snake (*Boiga irregularis*)[1]. Genetic strategies based on constructs that heritably reduce female survival or fertility[2–4] could provide a solution. Here we report the first successful trials of such constructs in fish and present models suggesting that their use in combination with other strategies could lead to effective species-specific control and possible long-term eradication of such pests.

We examined two approaches to reduce effective female population sizes: female-specific sterility (FS) and female-specific lethality (FL), focusing on the FL strategy because of the successful application of this approach in insects[4]. A preliminary

VNIVERSITAT DE VALÈNCIA