The genomic study of an environmental isolate of *Scedosporium apiospermum* shows its metabolic potential to degrade hydrocarbons

Laura T. Morales¹, Laura N. González-García², María C. Orozco¹, Silvia Restrepo² and Martha J. Vives¹,³*

Abstract

Crude oil contamination of soils and waters is a worldwide problem, which has been actively addressed in recent years. Sequencing genomes of microorganisms involved in the degradation of hydrocarbons have allowed the identification of several promoters, genes, and degradation pathways of these contaminants. This knowledge allows a better understanding of the functional dynamics of microbial degradation. Here, we report a first draft of the 44.2 Mb genome assembly of an environmental strain of the fungus *Scedosporium apiospermum*. The assembly consisted of 178 high-quality DNA scaffolds with 1.93% of sequence repeats identified. A total of 11,195 protein-coding genes were predicted including a diverse group of gene families involved in hydrocarbon degradation pathways like dioxygenases and cytochrome P450. The metabolic pathways identified in the genome can potentially degrade hydrocarbons like chloroalkane/alkene, chorocyclohexane, and chlorobenzene, benzoate, aminobenzoate, fluorobenzoate, toluene, caprolactam, geraniol, naphthalene, styrene, atrazine, dioxin, xylene, ethylbenzene, and polycyclic aromatic hydrocarbons. The comparison analysis between this strain and the previous sequenced clinical strain showed important differences in terms of annotated genes involved in the hydrocarbon degradation process.

Keywords: *Scedosporium apiospermum*, Genome, Hydrocarbon degradation

Introduction

Accidental spills of oil have risen as a global important problem due to the serious environmental damages caused by soil and water contamination [1]. Whereas oil is a complex mixture of aromatic and aliphatic hydrocarbons of different molecular weights, its removal from the environment is difficult and its permanence is prolonged [2]. These compounds have gained considerable attention because of their harmful features like resistance to degradation, bioaccumulation, and carcinogenic activity. Their persistence in the environment increases with their molecular weight and there is a need to develop technologies or processes able to degrade or to transform these compounds into less toxic molecules [3].

The ability of several organisms, primarily microorganisms (bacteria, fungi and microalgae), to degrade these toxic substances has been extensively studied in recent decades [1, 4–9]. The main goal is to improve the decontamination of the environment via bioremediation, which encompasses technologies that allow the transformation of compounds to less harmful or not harmful forms, with less use of chemicals, energy, and time [10, 11]. Microbial bioremediation is very effective due to the catabolic activity of microorganisms; among these, many species of bacteria, fungi, and microalgae have demonstrated the ability of hydrocarbon degradation. This process involves the breakdown of organic molecules through biotransformation in less complex metabolites, or mineralization to water, carbon dioxide, or methane [3].

Several strategies have been employed to study these microorganisms and to understand the processes carried...
out by them. Within these, genomics have allowed the recognition of promoters, genes, and degradation pathways that influence the construction of more efficient degradative strains relevant in bioremediation processes [12, 13]. Genome sequencing of hydrocarbon-degrader organisms has allowed the identification of several genes involved in metabolism and catabolism of aliphatic, aromatic alcohols, and other similar compounds, as well as some metals resistance genes [14]. However, the number of sequenced genomes of fungal species is lower than in bacteria. To date, there are 103,076 prokaryotic genomes sequenced whereas there are only 4503 genomes from eukaryotes in GenBank database (July 2017).

Scedosporium apiospermum (teleomorph: Pseudallescheria apiosperma [15]) is a fungus belonging to the phylum Ascomycota, which has been isolated from various environments, usually in those influenced by human activity [16]. This fungus was reported as a hydrocarbon-degrading microorganism since 1998 due to its ability to degrade polluting compounds, such as phenol and p-cresol [17]. One year later, its ability to degrade phenylbenzoate and its derivatives was elucidated [18]. In recent years, studies regarding degradation of complex compounds, such as toluene [19], polycyclic aromatic hydrocarbons (PAHs) [20], long-chain aliphatic hydrocarbons, and mixtures of these contaminants (unpublished results from our group) [21] have risen. Additionally, the fungus’ ability to regenerate granular activated carbon once it has been saturated with phenol was shown in our laboratory (unpublished results).

Therefore, Scedosporium apiospermum presents a wide range of opportunities in bioremediation and its genome sequencing can allow the identification of promoters, genes, and degradation pathways of hydrocarbons. Indeed, the genomic analysis of this fungus can improve the understanding of functional dynamics of contaminants microbial degradation and enhance conditions for effective decontamination processes in different environments [2]. On the other hand, this fungus has been recognized as a potent etiologic agent of severe infections in immunocompromised and occasionally in immunocompetent patients [22]. For this reason, in 2014, the genome of an isolate from a cystic fibrosis patient (clinical strain) was sequenced with the aim of gaining knowledge of its pathogenic mechanisms [23].

Thus, our objective was the complete characterization of the genome of the S. apiospermum environmental strain HDO1. In order to analyze the genes and pathways involved in the degradation process and to assess the unique components of its genome compared to the clinical strain and other sister species, we sequenced, assembled, annotated, and fully characterized the environmental strain’s genome.

Organism information
Classification and features
Scedosporium apiospermum environmental strain HDO1 was isolated as a contaminant from assays on bacterial strains able to grow in crude oil (API gravity 33) as the unique carbon source. It was selected for sequencing due to its capability to grow in cultures containing aliphatic hydrocarbons of crude oil, naphthalene, phenanthrene, phenol, and mixtures of these compounds in the laboratory. The fungal isolate was grown on potato dextrose agar (OXOID LTD, Hampshire, UK) plates for a period of 7 days at 30 °C. The optimal growth temperature was 30–37 °C. Identification was based on the following morphological characteristics: obverse and reverse colony color (according to the color chart Küppers, H. [24]), colony texture, size, and presence of diffusible pigments, hyphae characteristics, and conidia arrangement. The morphological characteristics were: colonies with a diameter of 7 cm on PDA at 25 °C after 7 days, cottony textured, greyish-white (N00, C00-A00) with yellowish-white reverse. No diffusible pigment was observed. The mycelium was hyaline, septate, and thin. Unbranched conidiphores with long neck-bottle shaped phialides were observed. Conidia were hyaline, approximately 5 μm in diameter, occurring in basipetal chains leaving long hyaline anellids (Fig. 1). For the molecular characterization, the fungus was grown in Sabouraud broth at 25 °C, 150 rpm for 7 days, and the biomass obtained was lyophilized for at least 12 h. Fungal genomic DNA was extracted from 100 mg of lyophilized and pulverized mycelia conducting the CTAB and Phenol/Chloroform/Isomyllic alcohol method [25]. The universal primers used for amplification of the ITS region, were ITS4 (5′-TCCTCCGCTATTGATATGC-3′) and ITS5 (5′-GGAAGTAAAAGTCGTAACAAGG-3′) [26]. Sanger sequencing was performed by Macrogen (South Korea). Nucleotide sequences obtained were compared with the non-redundant database of the National Center for Biotechnology Information (NCBI) using the tBlastx program (parameters by default), and the ITS region sequences were assigned to the fungus Scedosporium apiospermum with an E-value equal to 0.0, 100% query coverage and 100% identity. The obtained sequence is deposited at the NCBI Genbank nr database with the accession number JQ003882.1.

A phylogenetic analysis was performed using the long subunit rRNA gene, the internal transcribed spacer and the elongation factor 1-α sequences obtained from GenBank. Species from the Microascaceae family were included [27] [28] and are described in the Additional file 1: Table S3. Individual gene regions (LSU, ITS and TEF) were aligned using MAFFT v. 7.187 [29]. Maximum Likelihood analyses were performed using RAxML v.7.6.3 [30] as implemented on the CIPRES portal [31]. The sequence alignment was
Fig. 1 Micrograph of Scedosporium apiospermum. a Optical microscopy of hyphae and conidia from a PDA culture, at 100x total magnification. Lactophenol cotton blue wet mount preparation. b Scanning electron microscopy of hyphae and conidia from a liquid culture grown in minimal salt medium plus crude oil as the sole carbon and energy source.

Fig. 2 Phylogenetic Analysis of S. apiospermum HDO1. Estimated relationships of S. apiospermum HDO1 with S. apiospermum IHEM 14462 and other species from the Microascaceae family. The tree shows the concatenated analysis of the Internal Transcribed Spacer, the Large Subunit and the Elongation factor gene regions. Sequences from reference strains were used (Additional file 1: Table S3). Support values represent Bootstrap support values (Maximum Likelihood).
partitioned into three subsets, each one under a specified model of nucleotide substitution, chosen with PartitionFinder [32]. Estimation of different shapes, GTR rates, and base frequencies for each partition were allowed. The majority rule criterion implemented in RAxML [33] (−autoMRE) was used to assess clade support by bootstrap. The resulting trees were plotted using FigTree v. 1.4.2 [34]. Microascus longirostris and Scopulariopsis brevicaulis were used as outgroups. Environmental strain HDO1 used in this study clustered with the clinical strain IHEM14462 with good support, and they are the sister group of Trichurus spiralis CBS635.78 (Fig. 2). The whole group is contained within the wardamycopsis lineage described by Sandoval-Denis, M. et al. [28].

Summary of the classification and general features of S. apiospermum is given in Table 1.

### Genome sequencing information

#### Genome project history

The Genome of the isolate HDO1 was sequenced by NovoGene Technology Bioinformatics Co., Ltd. (Hong Kong). The whole genome shotgun project of S. apiospermum has been deposited in NCBI database under the accession number MVOQ0000000, belonging to the bioproject PRJNA357602. A summary of the project and information about genome sequence are shown in Table 2.

---

**Table 1** Classification and general features of *Scedosporium apiospermum* strain HDO1

| MIGS ID | Property                      | Term                                | Evidence code |
|---------|-------------------------------|-------------------------------------|---------------|
|         | Classification                | Domain Fungi                        | TAS [71]      |
|         |                               | Phylum Ascomycota.                  | TAS [72]      |
|         |                               | Class Sordariomycetes               | TAS [72]      |
|         |                               | Order Microascaceae                 | TAS [72]      |
|         |                               | Family – Microascaceae              | TAS [72]      |
|         |                               | Genus Scedosporium                  | TAS [73]      |
|         |                               | Species Scedosporium apiospermum    | TAS [73]      |
|         |                               | strain: HDO1                        |               |
|         | Gram stain                    | n/a                                 |               |
|         | Cell shape                    | Mycelium with septae                | IDA           |
|         | Motility                      | non-motile                          | IDA           |
|         | Sporulation                   | Conidia                             | TAS [72]      |
|         | Temperature range             | 25–42 °C                            | TAS [17]      |
|         | Optimum temperature           | 30–37 °C                            | TAS [17]      |
|         | pH range; Optimum             | 5.5–8.5: 7.3                        | TAS [17]      |
|         | Carbon source                 | Glucose, sucrose, urea, ethanol, ribitol, xylitol, L-arabinol, phenol, p-cresol, hydroquinone, 1,2,4-benzenetriol, catechol, 4-hydroxybenzylalcohol, 4-hydroxybenzaldehyde, 4-hydroxybenzoate, protocatechuic, 3-oxoadipate, phenylbenzoate, napthalene, pyrene, phenantrene, crude oil. | TAS [17–21] |
| MIGS-6  | Habitat                       | Soil and polluted water             | TAS [74, 75]  |
| MIGS-6.3| Salinity                      | 1.7–2.8/Lp to 5% in vitro           | TAS [76]      |
| MIGS-22 | Oxygen requirement            | Aerobic/Tolerate low pressure of O_2 | TAS [76, 77]  |
| MIGS-15 | Biotic relationship           | free-living                         | IDA           |
| MIGS-14 | Pathogenicity                 | Pathogenic                          | TAS [22]      |
| MIGS-4  | Geographic location           | Bogotá, Colombia                    | IDA           |
| MIGS-5  | Sample collection             | 20 May 2008                         | IDA           |
| MIGS-4.1| Latitude                      | 4.600659                            | IDA           |
| MIGS-4.2| Longitude                     | −74.065992                          | IDA           |
| MIGS-4.4| Altitude                      | 2658 m                              | IDA           |

*Evidence codes: IDA inferred from direct assay (first time in publication); TAS traceable author statement (i.e., a direct report exists in the literature); NAS non-traceable author statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These codes are from the Gene Ontology project [58]*
**Growth conditions and genomic DNA preparation**
Fungus growth was carried out in liquid culture (YPG: 1% yeast extract, peptone 1% and 2% glucose) at 30 °C for 7 days, followed by vacuum filtration, lyophilization, and maceration to have a homogeneous sample. DNA was extracted by the CTAB and Phenol/Chloroform/Isoamyl alcohol method [25]. DNA quality was analyzed by Nanodrop2000 (Thermo Fisher Scientific, MA, USA) and agarose gel electrophoresis (0.8%). DNA quantity was determined by Qubit2.0 (Invitrogen, CA, USA).

**Genome sequencing and assembly**
Genome sequencing of the strain was performed using high-throughput Illumina technology on a Hiseq2500 and employing two libraries: a 250 bp paired-end library and a 5k pb mate-pair library. Quality trimming of reads was performed using Trimmomatic 0.23 [35] and quality control was performed using FastQC 0.11.2 [36]. Coverage and depth of sequencing was analyzed by mapping the reads using Bowtie2 – 2.2.4 [37], the sam files were converted to bam files for visualization using samtools1.1 [38], and the visualization was made using tablet. 1.15.09.1 [39]. The genome was the novo assembled using Abyss 1.0.9.20 [40] with a kmer size of 64, scaffolds were generated with SSPACE BASIC 2.0 [41], and gaps were reduced using GapFiller 1.1 [42]. Assembly statistics were obtained using Quast 2.3 (Additional file 1: Table S1) [43]. Repetitive elements were identified with RepeatMasker 4.0.5 [44]. The draft genome of _S. apiospermum_ strain HDO1 was assembled from a total of 97,208,043 reads using Abyss [40] assembler. The assembly yielded 178 scaffolds (larger than 500 bp) with a genome size of 44.2 Mbp and a G + C content of 49.91% with a mean depth of 541X. The genome assembly statistics are shown in Table 3. The total number of non-coding repetitions was found using RepeatMasker [44] and was of 1.93%. The majority of repetitions were found to be simple repeats (0.89%) and low complexity regions (0.25%). The complete report of the annotation results for the non-coding repeats sequences can be seen in the Additional file 1: Table S2. The assembly features obtained for the draft sequence were similar to other fungal genome sequence projects [23, 45, 46].

**Genome annotation**
Gene prediction and structure annotation was conducted using Augustus 3.0.3 [47]. Functional annotation was performed using Blast2GO 3.1 [48]. Briefly, a BLASTx against the National Center for Biotechnology Information “nr” database [49] was conducted. Then, results were classified among Gene Ontology categories [50]. Protein classification was made using the COG [51], KOG (Eukaryotic Orthologous Groups) [52] and EggNOG [53] databases using Blast2GO v4.0 platform [48]. Annotated genes were mapped against Kyoto encyclopedia of genes and genomes [54] to its functional analysis and assigned the Enzyme Codes. A total of 11,195 protein-encoding genes were predicted using Augustus [47]. Functional annotation showed a total of 8595 (76.0% of predicted genes) sequences with predicted function using Blastx [49]. Then, InterProScan [55] and Gene Ontology [56] permitted the annotation of 7978 (70.8%) genes with signal peptides 1333 (11.8) Genes with transmembrane helices 2293 (20.3)
predicted in the RNAmmer 1.2 Server [59] and making an alignment with the predicted genes for Neurospora crassa from the database FungiDB [60], same database was used for pseudogenes prediction comparing with pseudogenes predicted for Neurospora crassa. The statistics of the genome annotation are shown in Table 3. A total of 4789 (42.5%) genes were assigned to the KOG [61] categories, most of them (60%) were assigned to one or more functional groups and the rest of genes were assigned to the function unknown group (Table 4). KEGG pathway analysis assigned an enzyme code to 2645 (23.5%) genes and revealed specific genes involved in the pathways of hydrocarbon degradation. These hydrocarbons are chloroalkane/alkene, chlorocyclohexane and chlorobenzene, benzoate, aminobenzoate, fluoro-benzoate, toluene, caprolactam, geraniol, naphthalene, styrene, atrazine, dioxin, xylene, ethylbenzene, and polycyclic aromatic hydrocarbons. Also, the analysis revealed the presence of genes involved in metabolism of xenobiotics by cytochrome P450 and in synthesis and degradation of ketone bodies. These results are shown in Fig. 3.

**Genome properties**

The assembled genome of the strain HDO1 has a size of 44,188,879 pb (distributed in 178 scaffolds) with a G-C content of 49.91%; the genome size and the G-C content was similar to the draft genome reported for the strain IHEM 14462 [22] (Table 5). A total of 11,278 genes were predicted; among these, 11,184 were identified as coding protein genes (representing the 99.16% of the total genes); 92 as RNA genes (0.81%); and 2 as pseudogenes (0.02%) (Table 3). Some other features of the predicted genes are shown in Table 4. The number of chromosomes could not be elucidated.

### Table 4 Number of genes associated with general COG functional categories

| Code | Value | % age | Description                                      |
|------|-------|-------|--------------------------------------------------|
| J    | 234   | 2.09  | Translation, ribosomal structure and biogenesis   |
| A    | 42    | 0.37  | RNA processing and modification                   |
| K    | 163   | 1.46  | Transcription                                    |
| L    | 126   | 1.12  | Replication, recombination and repair             |
| B    | 19    | 0.17  | Chromatin structure and dynamics                  |
| D    | 32    | 0.29  | Cell cycle control, Cell division, chromosome partitioning |
| V    | 32    | 0.29  | Defense mechanisms                                |
| T    | 153   | 1.37  | Signal transduction mechanisms                    |
| M    | 126   | 0.44  | Cell wall/membrane biogenesis                     |
| N    | 2     | 0.02  | Cell motility                                    |
| U    | 241   | 2.15  | Intracellular trafficking and secretion           |
| O    | 294   | 2.63  | Posttranslational modification, protein turnover, chaperones |
| C    | 216   | 1.93  | Energy production and conversion                 |
| G    | 399   | 3.56  | Carbohydrate transport and metabolism             |
| E    | 257   | 2.30  | Amino acid transport and metabolism               |
| F    | 65    | 0.58  | Nucleotide transport and metabolism               |
| H    | 63    | 0.56  | Coenzyme transport and metabolism                 |
| I    | 140   | 1.25  | Lipid transport and metabolism                    |
| P    | 131   | 1.17  | Inorganic ion transport and metabolism            |
| Q    | 186   | 1.66  | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 0     | 0     | General function prediction only                  |
| S    | 1918  | 17.14 | Function unknown                                  |
| –    | 6402  | 57.2  | Not in KOGs                                      |
**Table 5** Genomic features comparison between HDO1 strain and IHEM 14462 strain [22]

| Parameter     | IHEM 14462 | HDO1   |
|---------------|------------|--------|
| Size (Mb)     | 43.44      | 44.19  |
| Content G-C (%) | 50.4       | 49.91  |
| Predicted genes | 10,919    | 11,278 |
| Predicted proteins | 8,375     | 11,184 |

**Insights from the genomic sequence**

**Comparative genomics**

Reads were mapped versus the clinical strain IHEM 14462 using Bowtie2–2.2.4 [26]. The sam files were converted to bam files for the visualization using samtools1.1 [27] and the visualization was made using tablet.1.15.09.1 [28] resulting in an overall alignment of 92.75%. Genomes’ comparison between the environmental strain HDO1 and the clinical strain IHEM 14462 was performed using MAUVE 20150226 [62]. The genome sequence of HDO1 strain aligned with the sequence of IHEM 14462 strain in 88.1% of its length. The MAUVE [62] alignment showed a high level of similarity between the clinical and the environmental strains (Fig. 4). A total of 508 local collinear blocks (LCBs) that correspond to the homologous regions that are shared by the two sequences were found and a few of them were in reverse orientation after eight reordered cycles. From ordered output fasta file obtained with MAUVE [48] a new alignment was made with Nucmer at nucleotide level (maximum gap between two adjacent matches in a cluster of 90 bp and a minimum length of a maximal exact match of 20 bp) and Promer at amino acid level (maximum gap between two adjacent matches in a cluster of 30 amino acids and a minimum length of a maximal exact match of 6 amino acids). Nucmer and Promer alignments were plotted using Mumerplot, the last three mentioned tools from mummer 3.0 suite [63] (Fig. 5).

This analysis revealed that a high number of forward matches are in the greatest scaffolds of HDO1 genome sequence and reverse matches are more common in the smallest scaffolds. These differences and similarities seen for the nucleotides showed the same trends when these were translated to amino acids. These analyses and their corresponding plots also permitted to determine rearrangements, insertions, and deletions between both genomes.

A thorough comparative analysis showed some important differences between the genome draft sequences of the clinical and the environmental strain sequenced here. These differences were evident in the genome size of the assemblies and the number of predicted genes (Table 5). Indeed, our assembly had a total of 783.135 bp (1.77% of genome size) and 276 coding sequences more than the clinical strain. The remarkable difference in the number of annotated genes involved in hydrocarbons degradation pathways could be attributed to the pipeline followed to annotate genes. For the clinical strain the CDSs found were annotated against TrEmbl database [64] that only comprises UniProtKB/Swiss-Prot, while in this study, we used the nr (non-redundant protein sequences) database of NCBI which has a wider coverage because it comprises sequences obtained from another databases like GenPept, TPA, PIR, PRF, PDB, NCBI RefSeq, and UniProtKB/Swiss-Prot [65]. Since the repetitive elements of the genome were estimated as only 1.93%, it is highly probable that the difference in size can be attributed to some of the elements involved in functional categories.

**Genes involved in hydrocarbon biodegradation pathways**

Several genes involved in hydrocarbon biodegradation pathways were annotated in the genome of the environmental strain. In Table 6 the genes previously reported in the clinical strain [23] are shown. Results revealed...
that some genes are involved in several degradation pathways, principally corresponding to aromatic hydrocarbon metabolism (polycyclic aromatic hydrocarbons and phenolic compounds) and cytochrome P450 system. The number of these genes annotated for each strain can also be seen in the table and these values showed a higher number of genes in the environmental strain HDO1. The genes solely found in the draft sequence of the strain HDO1 are reported in Table 7. These genes comprised some genes belonging to the aromatic hydrocarbons degradation pathways completing the pathways in which genes found in both strains are also involved.

Table 6 Annotated genes involved in hydrocarbons degradation pathways

| Gene                                      | Pathway                                  | # of genes in HDO1 | # of genes in IHEM14462 [23] |
|-------------------------------------------|------------------------------------------|--------------------|------------------------------|
| Cytochrome P450 monoxygenase (EC:1.14.13.12) | PAHs degradation, alkane biodegradation [78, 79] | 79                 | 44                           |
| Phenol hydroxylase                        | Phenol degradation [17]                   | 4                  | 4                            |
| Epoxide hydrolase (EC:3.3.2.9)            | PAHs degradation [80]                     | 3                  | 2                            |
| Oxidoreductase                            | Organic compound metabolism [81]         | 13                 | 8                            |
| Salicylate hydroxylase (EC:1.14.13.24)    | Naphthalene degradation [82]              | 13                 | 4                            |
| Laccase (EC:1.10.3.2)                     | PAHs degradation [83]                     | 2                  | 2                            |
| Catechol 1,2-dioxygenase (EC:1.13.11.1)   | Phenol degradation [17]                   |                    |                              |
| 2,4-dichlorophenol 6-monoxygenase (EC:1.14.13.20) | Chlorinated phenols degradation [84]   | 5                  | 5                            |
| 2,3-dihydroxybenzoate deoxygenase (EC:4.1.1.46) | 2,3-dihydroxybenzoate degradation [85] | 1                  | 1                            |
| Carboxy-cis,cis-muconate cyclase           | Phenol degradation [17]                   | 4                  | 2                            |
| Phenyacetate 2-hydroxylase                 | Homogentisate degradation [86]            | 1                  | 2                            |
| 2-nitropropane dioxygenase (EC:1.13.12.16) | Nitroalquene oxidation [87]              | 4                  | 4                            |
| Biphenyl-2,3-diol 1,2-dioxygenase          | Byphenyl degradation [88]                 | 1                  | 2                            |
| Dienelactone hydrolase                     | Chloroaromatic degradation [89]           | 2                  | 7                            |
| Vanillyl-alcohol oxidase (EC:1.1.3.38)     | Aromatic degradation [90]                 | 4                  | 6                            |
| Cyclopentanone 1,2-monoxygenase            | Cyclopentanol degradation [91]            | 2                  | 2                            |
| Tyrosinase                                 | Phenolic compounds degradation [92]       | 1                  | 3                            |
| Lignostilbene dioxygenase (EC:1.13.11.43)  | Lignin degradation [93]                   | 2                  | 1                            |
| Total number of genes                      |                                          | 145                | 103                          |
Genes involved in the degradation of other organic compounds like toluene, lignin, and xylenol were found (Table 6).

The complete annotation of the genome and, particularly, of the genes belonging to a major class of protein families involved in fungal catabolism of organic pollutants was made. We could identify genes coding for proteins that have the ability to oxidize aromatic compounds like dioxygenases or monooxygenases. Among these, we could predict dioxygenases such as 2-nitropropane dioxygenase, extracellular dioxygenase (EC:1.13.11.1), gentisate 1,2-dioxygenase, intradiol ring-cleavage dioxygenase (EC:1.13.11.11), lignostilbene dioxygenase (EC:1.13.11.43), catechol 1,2-dioxygenase (EC:1.13.11.11), biphenyl-2,3-diol 1,2-dioxygenase, aromatic ring-opening dioxygenase, and 4-hydroxyphenylpyruvate dioxygenase (EC:1.13.11.27). These enzymes have great importance because, along with NADH-dependent flavin reductase and [2Fe-2S] redox centers, they catalyze the transformation of several aromatic compounds to dihydrodiols [66], allowing the complete mineralization of these compounds to CO2 and H2O (with the participation of other specific enzymes). Another enzyme family identified among the annotated genes was cytochrome P450. These enzymes have an interesting catabolic potential because they do not have substrate specificity and can catalyze epoxidation and hydroxylation of several organic pollutants like dioxins, nonylphenol, and PAHs [67].

Table 7 Annotated genes found only in the HDO1 strain

| Genes in HDO1 Pathway | # of genes |
|-----------------------|------------|
| 3-oxoadipate enol-lactonase Phenol degradation [94] | 5 |
| 5-carboxymethyl-2-hydroxymuconate isomerase Homoprotocatechuate degradation pathway [95] | 1 |
| Trihydroxytoluene oxygenase 2,4-dinitrotoluene degradation [96] | 1 |
| Benzoate 4-monoxygenase (EC:3.6.1.3) Benzoate degradation [18] | 5 |
| Diphenol oxidase Phenolic compounds degradation [97] | 1 |
| Cyclohexanone monoxygenase (EC:1.14.13.8) Cyclohexane degradation [98] | 4 |
| Gentisate 1,2-dioxygenase PAHs degradation [99] | 2 |
| 2-keto-4-pentenoate hydratase (EC:3.7.1.5) Benzoate degradation [100] | 1 |
| carboxymuconolactone decarboxylase Protocatechuate degradation [101] | 1 |
| 3-(3-hydroxy-phenyl)propionate 3-hydroxycinnamic acid hydroxylase Phenyl propionate degradation [102] | 2 |
| 3-hydroxybenzoate 6-hydroxylase Xylenol [103] and 3-hydroxybenzoate degradation [104] | 1 |
| 3-hydroxyisobutyrate dehydrogenase (EC:1.1.1.44; EC:2.1.1.43) Aromatic compounds metabolism [105] | 5 |
| Total number of genes | 30 |

Conclusions

The draft genome sequence of environmental strain S. apiospermum HDO1 isolated from bacterial bioremediation assays in crude oil was described here. The structural and functional information of the genome sequence of S. apiospermum has allowed advancing in the understanding of the ability of this fungus to degrade several kinds of xenobiotic compounds mainly several hydrocarbons families and offers an opportunity to propose its use or its enzymes in controlled bioremediation or bioaugmentation processes.
Additional file 1: Table S1. Genome assembly statistics reported by Quast [44]. Table S2. Non-coding repeats sequences summary. Table S3. Species and genes (accession numbers) used in the phylogenetic analysis [29]. (DOCX 37 kb)

Abbreviations
API: American Petroleum Institute; CDS: Coding DNA Sequence; COG: Cluster of Orthologous Groups; CTAB: Cetyl Trimethyl Ammonium Bromide; EC: Enzyme Codes; GO: Gene Ontology; ITS: Internal Transcribed Spacer; KEGG: Kyoto encyclopedia of genes and genomes; KOG: Eukaryotic Orthologous Groups; LSU: long subunit rRNA; PAHs: Polycyclic Aromatic Hydrocarbons; PDA: Potato Dextrose Agar; PDB: Protein Data Bank Protein; PIR: The Protein Information Resource; PRF: Protein Research Foundation; TEF: Elongation factor; TPA: Third-Party Annotation Sequences; YPG: Yeast extract, Peptone, Glucose

Acknowledgements
Authors thank Alejandro Grajales for helping with the phylogenetic studies. Special thanks to Juan Pablo Mallarino and the High Performance Computing core facility at Universidad de los Andes at Bogota, Colombia, for providing HPC resources that have contributed to the research results reported within this paper. URL: http://hpc.uniandes.edu.co.

Funding
The School of Sciences, Universidad de los Andes, funded the project “Sequencing of the whole genome of fungal strain capable of degrading hydrocarbons”. Departamento Administrativo de Ciencia, Tecnología e Innovación - Colciencias, financed the project “Hydrocarbons biodegradation using a microbial consortium (fungi-bacteria) immobilized on activated carbon” code 1204-569-34732. Additionally, Colciencias awarded to LTM the beca-pasantía joven investigador from “Young researchers and innovators program 2014”, announcement 645–21 2014. These funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions
MJV isolated the fungal strain, conceived the study, drafted and revised the manuscript. MCD identified the isolate and determined its enzymatic properties. LTM performed the laboratory experiments for the sequencing of the genome, the bioinformatics analyses and wrote the manuscript. LNG performed the bioinformatics analyses. SR conceived the study, drafted and revised the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Centro de Investigaciones Microbiológicas, Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia. 2Laboratorio de Micología y Fitopatología U尼亚ndes, Universidad de los Andes, Bogotá, Colombia. 3Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia.

Received: 19 June 2017 Accepted: 23 November 2017
Published online: 04 December 2017

References
1. Atlas RM. Microbial degradation of petroleum hydrocarbons: an environmental perspective. Microbiol Rev. 1981;45(1):180.
2. Baker SE, et al. Fungal genome sequencing and bioenergy. Fungal Biol Rev. 2008;22(1):1–5.
3. Haritash A, Kaushik C. Biodegradation aspects of polycyclic aromatic hydrocarbons (PAHs): a review. J Hazard Mater. 2009;169(1):1–15.
4. Leahy JG, Colwell RR. Microbial degradation of hydrocarbons in the environment. Microbiol Rev. 1990;54(3):305–15.
5. April TM, Fogh T, Currah R. Hydrocarbon-degrading filamentous fungi isolated from flare pit soils in northern and western Canada Can J Microbiol. 1999;46(1):38–49.
6. Semple KT, Cain RB, Schmidt B. Biodegradation of aromatic compounds by microalgae. FEMS Microbiol Lett. 1999;170(2):291–300.
7. Smith, M.R. The biodegradation of aromatic hydrocarbons by bacteria, in Physiology of Biodegradative Microorganisms. 1991, Springer. p. 191–206.
8. Grady CL Jr. Biodegradation of toxic organics: status and potential. J Environ Eng. 1999;116(5):805–28.
9. Samanta SK, Singh OV, Jain RK. Polyyclic aromatic hydrocarbons: environmental pollution and bioremediation. Trends Biotechnol. 2002;20(6):243–8.
10. Providenti MA, Lee H, Trevors JT. Selected factors limiting the microbial degradation of recalcitrant compounds. J Ind Microbiol. 1993;12(6):379–95.
11. Morgan P, Atlas RM. Hydrocarbon degradation in soils and methods for soil biotreatment. Crit Rev Biotechnol. 1988(8(4)):305–33.
12. Hickey WJ, Chen S, Zhao J. The phn island: a new genomic island encoding catabolism of polynuclear aromatic hydrocarbons. Front Microbiol. 2012;1:25.
13. Hickey WJ. Development of tools for genetic analysis of phenanthrene degradation and nanopod production by Delftia sp. C5-4. Front Microbiol. 2011;2:187.
14. Desal C, Pathak H, Madarmwar D. Advances in molecular and “omics” technologies to gauge microbial communities and bioremediation at xenobiotic/anthropogen contaminated sites. Biosour Technol. 2010;101(6):1558–69.
15. Gilgado F, Gené J, Cano J, Guarro J. Heterothallism in Scedosporium apiospermum and description of its teleomorph Pseudallescheria apiosperma sp. nov. Med Mycol. 2010;48(1):122–8.
16. Kaltesis J, Rainer J, De Hoog GS. Ecology of Pseudallescheria and Scedosporium species in human-dominated and natural environments and their distribution in clinical samples. Med Mycol. 2009;47(4):398–405.
17. Claußen M, Schmidt S. Biodegradation of phenol and p-cresol by the phycomycete Scedosporium apiospermum. Res Microbiol. 1998;149(6):399–406.
18. Claußen M, Schmidt S. Biodegradation of phenylbenzoate and some of its derivatives by Scedosporium apiospermum. Res Microbiol. 1999;150(6):413–20.
19. García-Perla EJ, HERNÁNDEZ S, Favela-Torres E, AURIA R, REVASH S. Tolueno biotransformation by the fungus Scedosporium apiospermum sp. Biotechnol Bioeng. 2001;76(1):161–9.
20. Reyes-César A, Absalón A, Fernández F, Gómez-Cortés, Espinosa. Biodegradation of a mixture of PAHs by non-ligninolytic fungal strains isolated from crude oil-contaminated soil. World J Microbiol Biotechnol. 2014;30(3):999–1009.
21. Sandoval, C, Pilarán J, C. Vives-Florez, M. Remoción de hidrocarburos por hongos filamentosos. Universidad de los Andes (Master’s Thesis ). 2012.
22. Guarro J, Kantarcioğlu A, Hörn R, Rodríguez-Tudela J, Cuenca M, Berenguer J, De Hoog G. Scedosporium apiospermum: changing clinical spectrum of a therapy-refractory opportunist. Med Mycol. 2006;44(4):295–327.
23. Vandeputte P, et al. Draft genome sequence of the pathogenic fungus Scedosporium apiospermum. Genome Announcements. 2014;2(5):e00988–14.
24. Kümper, H. Atlas of the colures. 3 Edición. Barcelona, Blume, S.A., 1996.ISBN 84–87535–37–2.
25. Vasco MF, Cepero M, Restrepo S, Vives-Florez M. Recovery of mitosporic fungi actively growing in soils after bacterial bioremediation of oily sludge and their potential for removing recalcitrant hydrocarbons. Int Biodeterior Biodegrad. 2011;65(4):649–55.
26. White TJ, Bruns T, Lee D, Taylor SB, J. W. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetic. PCR Protocols. 1990(181):1315–22.
27. Sandoval-Denis M, Sutton DA, Fothergill AW, Cano-Lira JF, Gené J, Decock C, De Hoog GS, Guarro J. Scopulariopsis, a poorly known opportunistic fungus: spectrum of species in clinical samples and in vitro responses to antifungal drugs. J Clin Microbiol. 2013;51(12):3937–43.
28. Sandoval-Denis M, et al. Phylogeny and taxonomic revision of Microascaceae with emphasis on synnematous fungi. Stud Mycol. 2016;83:193–233.
29. Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics. 2010,26(15):1899–900.
30. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90.
88. Yang X, Xie F, Zhang G, Shi Y, Qian S. Purification, characterization, and substrate specificity of two 2, 3-dihydroxybiphenyl 1, 2-dioxygenase from Rhodococcus sp. RD4, showing their distinct stability at various temperature. Biochimie. 2008;90(10):1530–8.
89. Schöllmann M, Schmidt E, Knackmuss H. Different types of dienelactone hydrolase in 4-fluorobenzoate-utilizing bacteria. J Bacteriol. 1990;172(9):5112–8.
90. Fraaije MW, Pikkemaat M, Van Berkel W. Enigmatic gratuitous induction of the covalent Flavoprotein Vanillyl-alcohol Oxidase in Penicillium simplicissimum. Appl Environ Microbiol. 1997;63(2):435–9.
91. Iwaki H, Hasegawa Y, Wang S, Kayser MW, Lau PC. Cloning and characterization of a gene cluster involved in cyclopentanol metabolism in Comamonas sp. strain NCIMB 9872 and biotransformations effected by Escherichia Coli-expressed cyclopentanone 1, 2-monoxygenase. Appl Environ Microbiol. 2002;68(11):5671–84.
92. Ikehata K, Nicell JA. Characterization of tyrosinase for the treatment of aqueous phenols. Bioresour Technol. 2000;74(3):191–9.
93. Kamoda S, Habu N, Samejima M, Yoshimoto T. Purification and some properties of Lignostilbene-a,β-dioxygenase responsible for the ca—cp cleavage of a Diarylpropane type lignin model compound from pseudomonas sp. TM1009. Agric Biol Chem. 1989;53(10):2757–61.
94. Hanwood CS, Parales RE. The β-ketoadipate pathway and the biology of self-identity. Annu Rev Microbiol. 1996;50(1):553–90.
95. Roper DI, Cooper RA. Purification, some properties and nucleotide sequence of 5-carboxymethyl-2-hydroxymuconate isomerase of Escherichia Coli C. FEBS Lett. 1990;266(1–2):63–6.
96. Johnson GR, Jain RK, Spain JC. Properties of the trihydroxytoluene oxygenase from Burkholderia cepacia R34: an extradiol dioxygenase from the 2, 4-dinitrotoluene pathway. Arch Microbiol. 2000;173(2):86–90.
97. Williamson PR. Biochemical and molecular characterization of the diphenol oxidase of Cryptococcus Neoformans: identification as a laccase. J Bacteriol. 1994;176(3):656–64.
98. Sheng D, Ballou DP, Massey V. Mechanistic studies of cyclohexanone monooxygenase: chemical properties of intermediates involved in catalysis. Biochemistry. 2001;40(37):11156–67.
99. Luo S, Liu DQ, Liu H, Zhou NY. Site-directed mutagenesis of gentisate 1, 2-dioxygenases from Klebsiella Pneumoniae M51 and Raistonia sp. strain U2. Microbiol Res. 2006;161(2):138–44.
100. Perez-Pantoja D, De la Iglesia R, Pieper DH, González B. Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium Cupriavidus Cupriavidus nematis JMP134. FEMS Microbiol Rev. 2008;32(5):736–94.
101. Eulberg D, Lakner S, Golovleva LA, Schöllmann M. Characterization of a protocatechuate catabolic gene cluster from Rhodococcus opacus 1CP: evidence for a merged enzyme with 4-carboxymuconolactone-decarboxylating and 3-oxoacid epoxide-lactone-hydrolyzing activity. J Bacteriol. 1998;180(5):1072–81.
102. Burlingame R, Chapman PJ. Catabolism of phenylpropionic acid and its 3-hydroxy derivative by Escherichia Coli. J Bacteriol. 1983;155(1):113–21.
103. Poh CL, Bayly RC. Evidence for isofunctional enzymes used in m-cresol and 2, 5-sylenol degradation via the gentisate pathway in Pseudomonas Alcaligenes. J Bacteriol. 1980;143(1):59–69.
104. Park M, Jeon Y, Ho Hee J, Hyun-su R, Woodjun P, Madsen EL, Che Ok J. Molecular and biochemical characterization of 3-hydroxybenzoate 6-hydroxylase from Polaromonas naphthalenivorans C22. Appl Environ Microbiol. 2007;73(16):5146–52.
105. Masai E, Katayama Y, Fukuda M. Genetic and biochemical investigations on bacterial catabolic pathways for lignin-derived aromatic compounds. Biosci Biotechnol Biochem. 2007;71(1):1–15.