

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/319651569>

# Genomic and transcriptomic analysis of the toluene degrading black yeast Cladophialophora immunda

Article in *Scientific Reports* · December 2017

DOI: 10.1038/s41598-017-11807-8

---

CITATIONS  
31

READS  
715

6 authors, including:



Barbara Blasi  
University of Veterinary Medicine, Vienna

22 PUBLICATIONS 582 CITATIONS

[SEE PROFILE](#)



Caroline Poyntner  
University of Innsbruck

28 PUBLICATIONS 267 CITATIONS

[SEE PROFILE](#)



Ksenija Lopandic  
University of Natural Resources and Life Sciences Vienna

124 PUBLICATIONS 1,403 CITATIONS

[SEE PROFILE](#)



Katja Sterflinger  
Akademie der Bildenden Künste Wien

341 PUBLICATIONS 6,189 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:

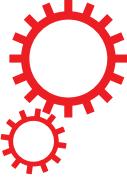


Yeast as symbionts, pathogens or commensals: the history and evolution of yeasts colonizing humans [View project](#)



Yeast hybrids [View project](#)

# SCIENTIFIC REPORTS



OPEN

## Genomic and transcriptomic analysis of the toluene degrading black yeast *Cladophialophora immunda*

Received: 6 April 2017

Accepted: 30 August 2017

Published online: 12 September 2017

Barbara Blasi , Hakim Tafer, Christina Kustor, Caroline Poyntner, Ksenija Lopandic & Katja Sterflinger

*Cladophialophora immunda* is an ascomycotal species belonging to the group of the black yeasts. These fungi have a thick and melanized cell wall and other physiological adaptations that allows them to cope with several extreme physical and chemical conditions. Member of the group can colonize some of the most extremophilic environments on Earth. *Cladophialophora immunda* together with a few other species of the order Chaetothyriales show a special association with hydrocarbon polluted environments. The finding that the fungus is able to completely mineralize toluene makes it an interesting candidate for bioremediation purposes. The present study is the first transcriptomic investigation of a fungus grown in presence of toluene as sole carbon and energy source. We could observe the activation of genes involved in toluene degradation and several stress response mechanisms which allowed the fungus to survive the toluene exposure. The thorough comparative genomics analysis allowed us to identify several events of horizontal gene transfer between bacteria and *Cladophialophora immunda* and unveil toluene degradation steps that were previously reported in bacteria. The work presented here aims to give new insights into the ecology of *Cladophialophora immunda* and its adaptation strategies to hydrocarbon polluted environments.

Aromatic compounds like benzene, toluene, ethylbenzene and xylene isomers, collectively known as BTEX, are one of the major contributors to environmental pollution, globally threatening natural environments, ground-water reservoirs<sup>1,2</sup>, agricultural sustainability and, as a consequence, food safety. They naturally occur in coal, oil and gas deposit and are also one of the most abundantly produced chemicals in the world. Their release into the environment happens through natural events, like forest fires or volcanic eruptions, and human activities, such as vehicle traffic and the coal, gas and oil exploitation chain. The majority of BTEX are released into the atmosphere, but can also accumulate in ground waters and soils located close to natural and anthropogenic BTEX sources<sup>1,2</sup>.

Long-term BTEX exposure can have deleterious effects on the liver, kidneys, central nervous system, lungs and cause endocrine disruption<sup>3</sup>. As a consequence, the production and release of BTEX are strongly regulated by national, supranational and international policies. While technologies for air pollution control or soil remediation are available, these compounds are difficult to remove, since they remain toxic at concentration levels where chemical or physical removal is not economically sustainable<sup>4</sup>. Promising and cost-effective alternative to physico-chemical environmental remediation are slowly emerging. In case of air pollution control, biofilter, where bacterial or fungal communities degrade pollutants from contaminated gases, gained a lot of attention<sup>5</sup>. Even though bacteria-based biofilter are more widely used<sup>6</sup>, fungal biofilters are more efficient upon dehydration and acidification of the media<sup>7,8</sup>.

The most extensively studied fungal isolates in the field of bioremediation belong to the genera *Exophiala*, *Cladophialophora*, *Aspergillus*, *Phanerochaete*, *Cladosporium*, *Paecilomyces*, *Trichoderma* and *Trametes*<sup>9–12</sup>. In recent decades the group of the black yeasts, where *Exophiala* and *Cladophialophora* are classified, have begun to be recognized for their bioremediation potential<sup>12</sup>. Some of these fungi have in fact been isolated from different polluted sources<sup>13</sup>, such as industrial spills, car gasoline tanks, railway sleepers<sup>14</sup> and air biofilters<sup>15</sup>. The common

Department of Biotechnology, VIBT-EQ Extremophile Center, University of Natural Resources and Life Sciences, 1190, Vienna, Austria. Barbara Blasi and Hakim Tafer contributed equally to this work. Correspondence and requests for materials should be addressed to B.B. (email: [barbara.blasi@boku.ac.at](mailto:barbara.blasi@boku.ac.at))

characteristic of these fungal group members is the presence of melanin, constitutively expressed and deposited at the cell wall level. In particular, the genera *Exophiala* and *Cladophialophora* (order Chaetothyriales), and *Pseudallescheria* (order Microascales) have a high potential to grow in polluted environments and to metabolize hydrocarbons as the sole source of carbon and energy<sup>12</sup>. Especially *Cladophialophora immunda* was characterized for its ability to degrade up to 65% of the toluene supplied<sup>12</sup>.

Although most of the known *Cladophialophora immunda* strains were sampled from environmental polluted sources, the fungus is also an opportunistic human pathogen, causing subcutaneous phaeohyphomycosis in immunosuppressed patients<sup>16</sup>. It is currently hypothesized that this dual ecology may arise from the hydrocarbon degrading pathways, which would allow the fungus to use the neurotransmitters for their own energy metabolism<sup>17</sup>. While light has been shed on the mechanisms behind black yeasts ecology and pathogenicity<sup>18–20</sup>, the biological processes responsible for the xenobiotics-degrading ability of these fungi have not yet been studied with next-generation sequencing approaches. In order to get an insight into the mechanisms of toluene tolerance and degradation, we present the first genomic and transcriptomic analysis of *Cladophialophora immunda* upon growth with toluene as sole carbon and energy source.

The strain used in this experiment (CBS 110551) has been isolated from an industrial biofilter operated with toluene air stream that was previously inoculated with soil from a gasoline station<sup>13</sup>. The fungus was inoculated in liquid mineral media and exposed to vapors of toluene through hand-made test tubes in sealed Erlenmeyer flask. In the control experiment, the fungus was grown in the same media with 2% glucose as classical carbon source for one week. At the end of the incubation times, the biomass was collected through filtration and used for RNA isolation and sequencing.

We performed a comparative genomics analysis of *Cladophialophora immunda* with other toluene-degrading and pathogenic species. Cytochrome P450 and fungal specific transcription factors were among the most over-represented domains compared to the other genomes. Genes horizontally transferred from bacteria were found to play an important role in *C. immunda* toluene degradation pathway. The transcriptome analysis revealed that in this fungus toluene triggers the expression of genes involved in the putative fungal toluene degradation pathway, oxidative stress as well as DNA repairs and molecular chaperones.

## Method

**Fungal growth conditions.** *Cladophialophora immunda* (CBS 110551), a strain isolated from a toluene-charged air biofilter inoculated with gasoline-polluted soil, was ordered from Centraalbureau voor Schimmelcultures. The fungus was cultured in Malt extract agarose media (2% Malt extract, 2% D-glucose, 0.1% Bacto-peptone and 2% Agar). For the RNA-seq experiments, *Cladophialophora immunda* was grown in liquid culture in a modified Hartmans mineral<sup>21</sup> media with 2% glucose or 1.35 mM toluene as sole carbon source, 1 ml of vitamin solution and 0.02% of yeast extract as nitrogen source and to boost the fungal growth by shortening the lag phase<sup>22</sup>. For the exact recipe of the mineral media and vitamin solution consult the Supplementary materials. The toluene was supplied through the air of a sealed flask in a 5% solution in dibutylphthalate. Depending on the carbon source, the experiments duration was set to 90 days for toluene and one week for glucose, both at temperature of 22 °C and 100 rpm on an orbital shaker. Both experiments were performed in triplicates. At the end of the experiments the biomass was collected by centrifugation (5000g per 15 minutes at 4 °C), washed with RNase free water, frozen in liquid nitrogen and stored at –80 °C until use.

**RNA isolation and sequencing.** Total RNA was extracted from 100 mg of fungal biomass with FastRNA PRO™ RED KIT (MP Biomedicals) according to the manufacturer instructions. The mRNA was isolated with the Dynabeads®mRNA DIRECT™ Micro Kit (Ambion by Life Technologies) and the subsequent transcriptome library preparation was performed with the Ion Total RNA-Seq Kit v2 (Life Technologies). Total RNA, isolated mRNA and the final cDNA libraries were all qualitatively and quantitatively evaluated with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). The transcriptome sequencing was performed by the Ion Proton™ sequencer (Life Technologies) with the Ion PI Chip v2 (Life Technologies).

**Bioinformatics.** *Genome annotation.* The genome sequence used was previously published in ref. 23. The denovo non-coding RNA (ncRNA) annotation was done with Infernal<sup>24</sup> and the Rfam database<sup>25</sup>, tRNAscan<sup>26</sup> and SNOSTRIP<sup>27</sup>. Overlapping ncRNAs annotations were merged manually.

For the protein coding genes, the reads from all 6 RNA sequencing runs were assembled with Trinity 2.0.6<sup>28</sup> and Cufflinks 2.2.1<sup>29</sup>. Both sets of transcripts were merged with the help of PASA<sup>30</sup>. Transdecoder<sup>30</sup> was then used to predict protein coding transcripts. Intron/Exon from these coding transcripts were used to train Augustus<sup>31</sup>. The prediction from Augustus and PASA were subsequently merged into a set of protein coding loci with EvidenceModeler<sup>32</sup>.

PASA-Transcripts that were not considered coding by Transdecoder and by CPAT<sup>33</sup> (p-value > 0.01), that did not show significant (p-value > 0.001) sequence homology against SWISSPROT when searched with Blastn<sup>34</sup> and that did not contain Pfam-domain<sup>35</sup> (p-value > 0.001) when searched with HMMER<sup>36</sup> were classified as non-coding. The transcripts that could not be unequivocally classified as either coding or non-coding were classified as transcript with unknown functions. Samtools<sup>37</sup> were used to process the mapped reads. Bedtools<sup>38</sup> were used to look for annotation overlaps.

Finally conserved coding and non-coding elements were detected with RNAAcde<sup>39</sup> (p-value < 0.01) and RNAz<sup>40</sup> (probability P > 0.9), respectively. To this aim a multiple genomes alignment of *Aspergillus niger*, *Aspergillus nidulans*, *Aspergillus fumigatus*, *Cladophialophora immunda*, *Exophiala dermatitidis*, *Fusarium oxysporum*, *Fusarium solani*, *Neurospora crassa*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Trichophyton rubrum* was generated with the multiz pipeline<sup>41</sup> implemented in snakemake<sup>42</sup>.

**Comparative genomics.** Gene sequences and annotations for *Aspergillus niger*, *Aspergillus nidulans*, *Aspergillus fumigatus*, *Cladophialophora bantiana*, *Exophiala dermatitidis*, *Fusarium oxysporum*, *Fusarium solani*, *Neurospora crassa*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Trichophyton rubrum* were fetched from ensembl and ncbi. For *Exophiala mesophila*, our own genome assembly and annotation were used<sup>43</sup>.

The protein-coding genes of all studied species were functionally annotated with a local install of interproscan-5.19–58.0<sup>44</sup> with an E-value cut-off of  $1.10^{-3}$ . Homologies with the transporter classification database (TCDB)<sup>45</sup>, the peptidase database (MEROPS)<sup>46</sup> and the carbohydrate-active enzymes database (CAZY)<sup>47</sup> were assessed with Blastp<sup>48</sup> (E-value < 1e-3). Genes homology was assessed with Proteinortho-5.13<sup>49</sup>. Horizontally transferred genes were detected with the latest version of HGTFinder (2016)<sup>50</sup>.

**Genes expression.** All reads were mapped with STAR 2.5.0<sup>51</sup> with `--alignIntronMin15 - alignIntronMax2000 - chimSegmentMin12 - chimJunctionOverhangMin12 - alignSJDBoverhangMin10` against the *Cladophialophora immunda* genome. The mapped reads were then used to find differentially expressed genes and detect chimeric RNAs. The raw and mapped reads were used to assemble the transcriptome of *Cladophialophora immunda*. Count of mapped reads on annotation elements was done with featureCounts v1.4.6p2<sup>52</sup>. The identification of differentially expressed genes was done with edgeR<sup>53</sup>, while the functional enrichment of the significantly regulated genes was done with GoStat<sup>54</sup> and Kobas<sup>55</sup>. Revigo<sup>56</sup> was used to summarize the list of Gene Ontology terms. Expression level analysis was done with Salmon 0.7.1<sup>57</sup> with 30 bootstraps, sleuth and wasabi.

## Results

**Highly complete genome and genes set.** The genome assembly of *Cladophialophora immunda* used for this study was previously published by our group<sup>23</sup>. It has a size of 42 Mb and includes 464 contigs (>500 bp). A total of 15228 protein-coding genes (17887 proteins) were predicted with a combination of gene annotation tools using sequence homology and strand-specific sequencing of RNA (RNAseq, see Method). 15214 were supported by RNAsq (TPM  $>= 0$ ). 95.56% of ultra-conserved genes found in CEGMA<sup>58</sup>, were recovered in *Cladophialophora immunda*. Among the basal small RNAs, 52 tRNAs (28 spliced), 46 snoRNAs, all snRNAs of the major spliceosome as well as RNase-P, RNase-MRP, 5S and U3 were found (See Methods). Another isolate of the same species of *Cladophialophora immunda* (CBS 83496) was recently sequenced and contains 12879 genes (14033 proteins)<sup>59</sup>. From the 3854 supplementary proteins in *Cladophialophora immunda* (CBS110551), 1815 had no homologues in CBS 83496 at an E-value threshold of  $1.10^{-3}$ . Among these genes, 964 had homologues in the non-redundant protein sequences database (nr), resulting in 851 genes being specific to *Cladophialophora immunda* (CBS110551). Eukaryotic rRNA sequences were found on contig JSEJ01000356, while annotation corresponding to mitochondrial tRNAs and rRNAs were detected mainly on contigs JSEJ01000378 and JSEJ01000388 (See Supplementary Table S1).

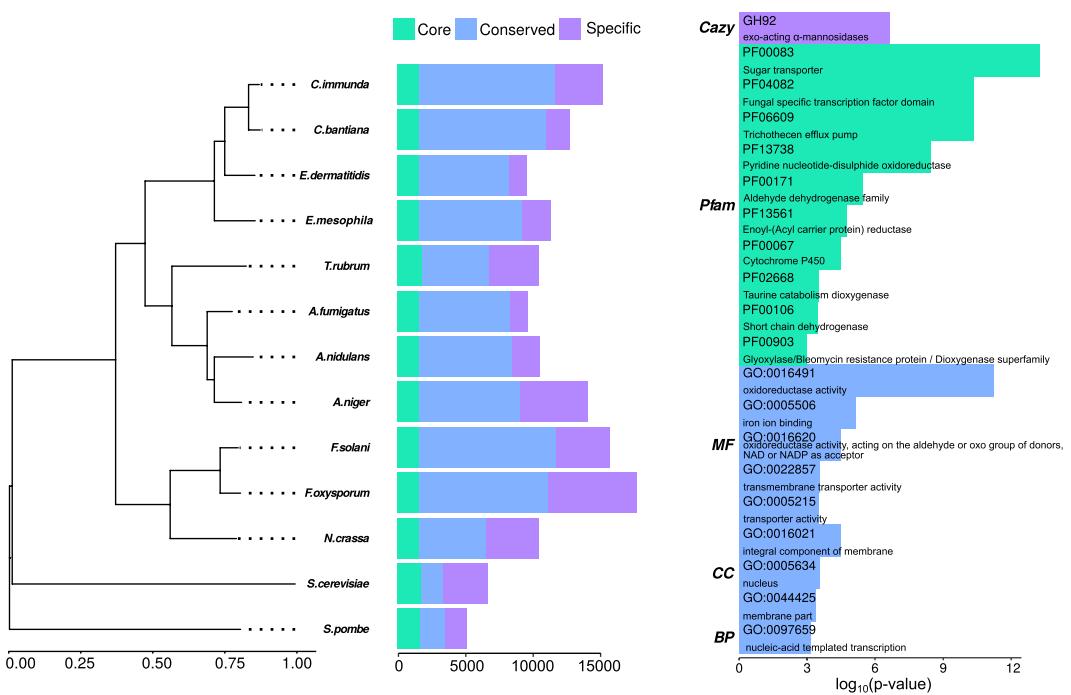
A total of 5300 non-coding transcripts (lncRNAs) from 4507 genomic loci were assembled. 10178 transcripts with a non-negligible protein-coding potential and coming from 9301 loci were assembled.

The set of conserved coding and non-coding elements were detected with RNAz and RNACode<sup>39, 40, 60</sup> on a multiple genomes alignment containing *Cladophialophora immunda* with 10 other fungal genomes (*Aspergillus fumigatus*, *Aspergillus nidulans*, *Aspergillus niger*, *Exophiala dermatitidis*, *Fusarium oxysporum*, *Fusarium solani*, *Neurospora crassa*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Fusarium solani*). The resulting coverage of the *Cladophialophora immunda* genome ranged from 4.8% for *Saccharomyces cerevisiae* to 39.6% for *Exophiala dermatitidis*. The *Cladophialophora immunda*-centered multiple alignment featured a coverage of 79% of the fungal genome. RNAz returned a total of 852 structurally conserved regions with a P-score  $> 0.9$ , while RNACode detected 49508 genomic regions with conserved coding potentials (p-value  $< 0.01$ ) 122 RNACode regions overlapped with 117 RNAz loci. Non-coding PASA transcripts overlapped with 111 RNAz loci and 44 Rfam candidates, among them tRNAs, snRNAs, snoRNAs or rRNAs.

**Genes conservation and gene family evolution.** The gene conservation pattern between *Cladophialophora immunda*, the 10 genomes used for the multiple genomes alignments, *Cladophialophora bantiana* and *Exophiala mesophila* was studied with Proteinortho<sup>49</sup> (see Methods). 1191 genes had homologues in all 13 species, 9742 genes had at least one homologue in any of the species and 5491 genes were specific to *Cladophialophora immunda* (See Fig. 1). 385 paralogous groups, corresponding to 796 genes, were found to be specific to *Cladophialophora immunda*. The significantly enriched protein domains of the *Cladophialophora immunda*-specific paralogues were related to families of transporters, transcription factors, trichothecene efflux pump, cytochrome P450, taurine dioxygenase and short chain dehydrogenase. GO terms were enriched in oxidoreductase activity, iron ion binding, transporter activity, nucleic-acid templated transcription, membrane part and nucleus (see Fig. 1 and Supplementary Table S1). The phylogenetic tree obtained from the alignment of the 771 one-to-one paralogues of 13 fungal species was consistent with the taxonomic tree (See Fig. 1).

The enrichment or depletion of Cazy, Merops and Pfam protein domains in *Cladophialophora immunda* against the other 12 genomes was studied. According to Cazy, the most overrepresented protein domains related to carbohydrates metabolism belong to two families of glycoside hydrolases: GH16 and GH92 (both with fdr  $< 0.001$ ). Families GH16 and GH92 are glucanase/galactosidases and exo-acting  $\alpha$ -mannosidases, respectively.

29 Pfam terms were overrepresented, with six Pfam families related to oxygenase, dehydrogenase, transcription, AMP-binding,  $\beta$ -oxidation and transport (See Fig. 2). Interestingly, *Cladophialophora immunda* exhibits 65 trichothecene efflux pumps, a much higher number than what is found in fungi like *Fusarium oxysporum* that metabolize this compound<sup>61</sup>. This is also seen in two other black yeasts, *Exophiala mesophila*, an aromatic compounds degrader, and *Cladophialophora bantiana*, a lethal neurotropic pathogen<sup>12, 17</sup>. In fact, a recent publication<sup>59</sup> showed that trichothecene efflux pumps probably expanded in the precursor of the black yeast group.



**Figure 1.** Phylogenetic relationship, gene conservation of *Cladophialophora immunda* compared to 12 other genomes and functional enrichment of *Cladophialophora immunda*-specific paralogues. l.h.s Phylogenetic tree constructed based on the concatenated mafft alignment of 717 one-to-one paralogues. Species phylogeny was inferred with iqtree using genes with one-to-one relationships across all 13 genomes and with 1000 bootstraps. center The orthologue conservation for all 13 species studied is shown. Core genes, i.e. genes found in all genomes, are shown in dark blue. The number of conserved genes, i.e. genes found in at least two species, is shown in blue, while species specific genes are shown in light-blue. r.h.s Barplot of the log<sub>10</sub>(fdr) of significantly (fdr < 0.001) enriched protein domains, GO terms and Cazy families for the *Cladophialophora immunda*-specific paralogues.

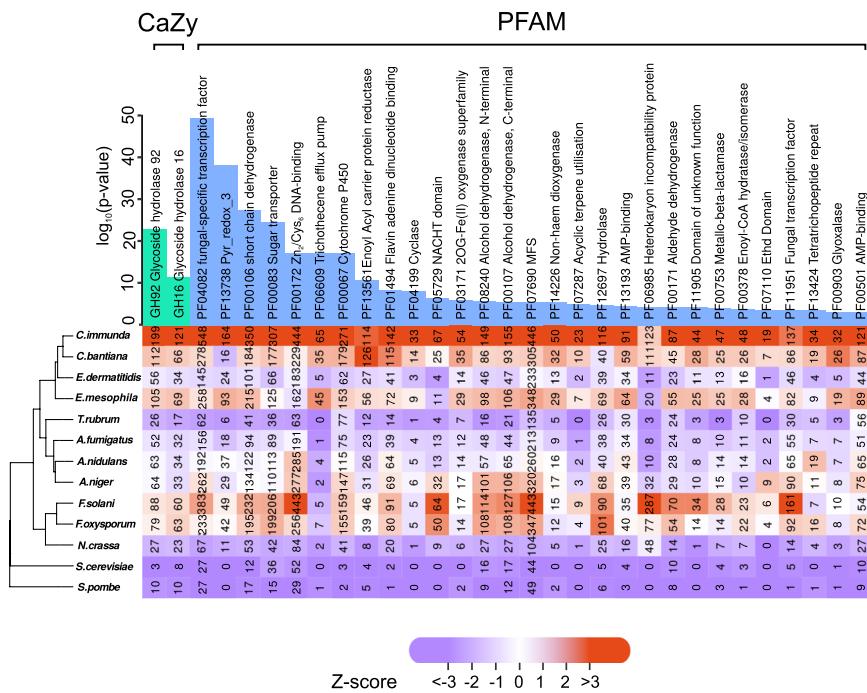
Further enriched protein families were cyclase, acyclic terpene utilization, hydrolase, heterokaryon incompatibility protein, protein with domain of unknown function, metallo-beta-lactamase and proteins having tetratrichopeptide repeats.

In order to better understand which specific genes might be responsible for the toluene degradation activity of *Cladophialophora immunda* we looked at which genes were specific to *Cladophialophora immunda*, *Cladophialophora bantiana*, *Exophiala mesophila*, *Fusarium oxysporum* and *Fusarium solani* which are, on the exception of *C. bantiana*, all (poly-)aromatic degraders<sup>17</sup>. A total of 64 genes were found to be specific to these 5 species (See supplementary Table S1). In *Cladophialophora immunda* these genes were enriched in 30 Pfam families with an fdr < 0.1. Among them PF14832 (fdr = 0.04), 4-Oxalocrotonate tautomerase, is interesting in the context of toluene degradation, as it was shown to be part of a bacterial metabolic pathway that oxidatively catabolizes toluene, o-xylene, 3-ethyltoluene, and 1,2,4-trimethylbenzene into intermediates of the citric acid cycle<sup>62</sup>.

**Horizontal gene transfer.** HGTFinder<sup>50</sup> was used to look for horizontally transferred bacterial genes. 109 genes in *Cladophialophora immunda* were found to be horizontally transferred from bacteria (q-value < 0.001) (See Supplementary Table 1). Functional enrichment analysis on this set of genes indicates that they are preferentially involved in carboxylic acid metabolism, hydrolase activity, pectine and cellulose degradation as well as reductase activity (See Supplementary Table S1).

The occurrence of clusters of horizontally transferred genes was also studied. Two genes were considered to belong to the same cluster if they were separated by at most 4 genes. A total of 10 gene clusters were found to be horizontally transferred from bacteria. Gene name, function and q-value of transfer (qT) are listed in Supplementary Table S1.

**RNA sequencing of *Cladophialophora immunda*.** The transcriptional landscape of *Cladophialophora immunda* was assessed during growth with either glucose or toluene as sole carbon source. Both experiments were sequenced in triplicates on the Ion Proton platform, yielding a total of 496 million reads with an average length per run between 130 and 159 nucleotides. The mapping rate, depending on the sequencing run, ranged between 50% and 83% of the reads, yielding a total of 320 million reads mapped (See Supplementary Table S1 for more details). The majority of the unassigned reads could be mapped against rRNAs, indicating that the poly(A) enrichment protocol did not completely discard rRNAs. Despite the rRNA contamination, sample expression similarity assessed with principal component analysis fits well with the expectation from the experimental design (See Supplementary Figure S1).



**Figure 2.** Plot of the overrepresented functional annotations in *Cladophialophora immunda* compared to 12 other genomes. The upper part of the figure shows the fdr of the overrepresented categories as bars in  $\log_{10}$  scale. PFAM categories are in blue, while CaZy categories are in turquoise. For each annotation element, the number of genes annotated with the categories in each genome is shown. Purple indicates a depletion with respect to the average number of annotated element found in the 13 genomes, while orange represents an enrichment with respect to the average. The shade of each entry is proportional to the deviation from the average number of annotation for the 13 genomes.

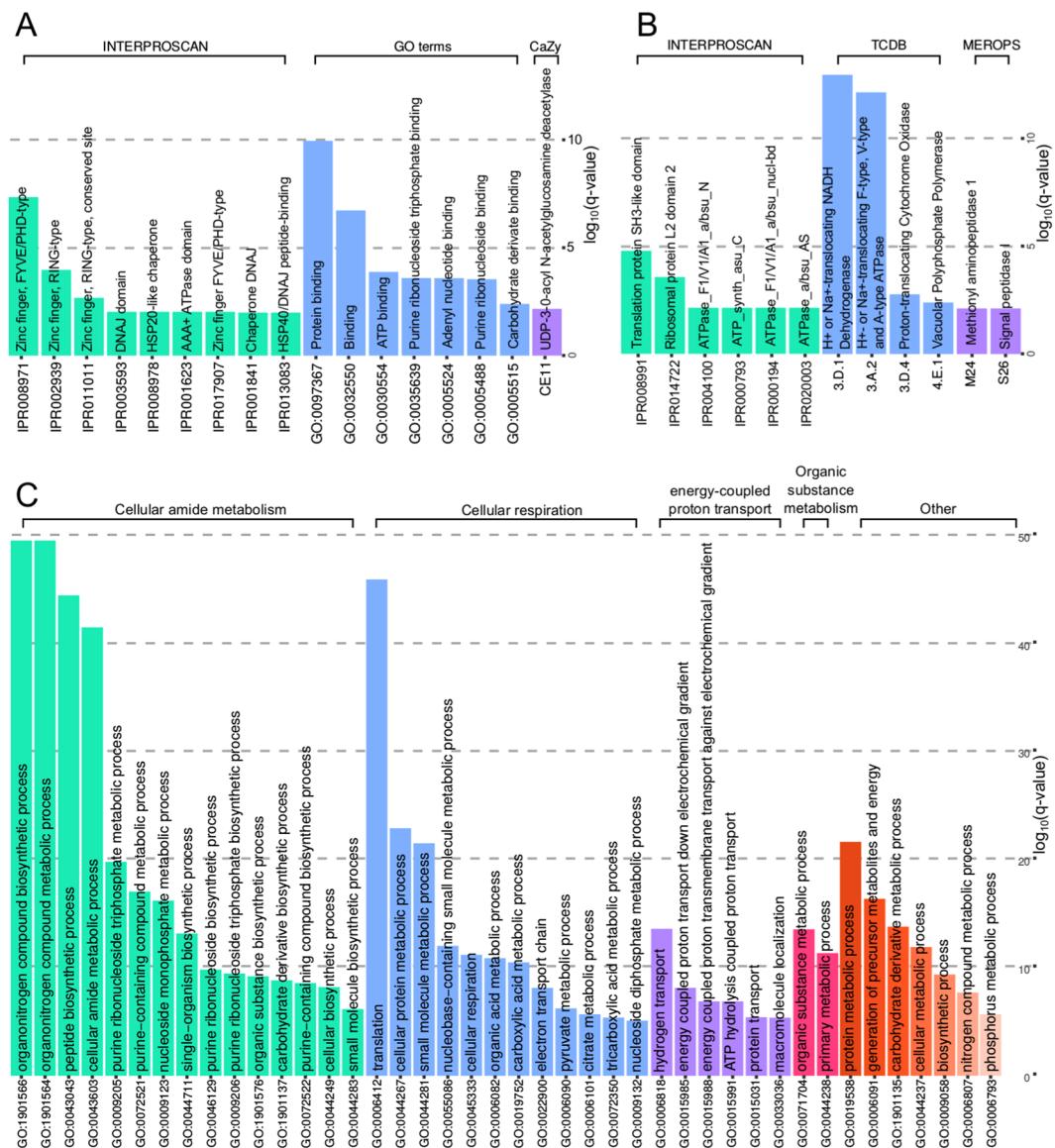
Functional enrichment for the 100 most highly expressed coding genes in the toluene and glucose samples was analyzed. In the glucose experiment, genes related to translation ( $fdr = 9.87 \cdot 10^{-15}$ ), ion transport ( $fdr = 3.69 \cdot 10^{-5}$ ) and energy metabolism ( $fdr = 1.64 \cdot 10^{-5}$ ) are overrepresented, while in the toluene experiment, genes involved in translation initiation ( $fdr = 5.05 \cdot 10^{-5}$ ) and protein folding (HSP20, ClpA/B, HSP40, HSP70) are overrepresented.

**Differential expression.** The differential expression of Rfam annotated ncRNAs in *Cladophialophora immunda* were computed in the toluene experiment compared to the glucose experiment. 15 ncRNAs, exclusively snoRNAs, were upregulated with a  $fdr < 1.00 \cdot 10^{-3}$  in the toluene experiment. On the other hand, 17 tRNAs, 1 snRNA (U6), 10 5S rRNAs and SRP were downregulated in the toluene experiment. Among the assembled non-coding transcripts, 268 were upregulated and 199 were downregulated with a  $fdr < 1.00 \cdot 10^{-3}$ .

543 and 355 protein-coding genes were up- and downregulated ( $fdr < 1.00 \cdot 10^{-3}, |log2FC| > 3$ ), respectively between the toluene and glucose experiments. In the set of upregulated genes, protein domains linked to protein folding (HSP40, DNAJ, HSP20, cis-trans isomerase, DSBA-like thioredoxin)<sup>63</sup> and protein-protein interactions (Ring-Type Zinc finger) are overrepresented (See Fig. 3 and Supplementary Table S1). Further 11 out of 47 proteins involved in the transport of misfolded proteins (TCDB family 3.A.16,  $fdr = 6.83 \cdot 10^{-2}$ ) and 5 out of 16 genes related to damaged DNA binding (GO:0003684,  $fdr = 3.84 \cdot 10^{-2}$ ) were upregulated in the toluene experiment (See Supplementary Table S1). Genes involved in the degradation of xenobiotics and detoxification processes<sup>64</sup> like Glutathione S-transferase (CLAIMM\_13540 x96, CLAIMM\_02727 x4), CLAIMM\_13949 (x28, Glyoxylase II), glutathione-dependent formaldehyde-activating enzyme (CLAIMM\_10595 x51, CLAIMM\_02133 x12, CLAIMM\_10984 x15) are also upregulated in the toluene experiment. The expression of two antioxidants, ascorbate peroxidase (CLAIMM\_03119 x12) and carotenoid oxygenase (CLAIMM\_10918 x12) was also triggered when *Cladophialophora immunda* is exposed to toluene.

The set of downregulated genes covers a broader spectrum of functions than that of the upregulated genes. The GO terms enriched can be roughly grouped into amide metabolism, cellular respiration, proton transport and translation (See Fig. 3 and Supplementary Table S1).

At a cellular level, one of the main effect of toluene is the increase in membrane fluidity, which can cause leakage and damages of proteins and cellular organelles, as observed in bacteria<sup>65</sup>. For this reason, the presence of significantly regulated genes related to cell-wall biosynthesis and lipid metabolism was investigated. From the 64 genes that are homologues to cell-wall biosynthesis genes in *Exophiala dermatitidis*<sup>18</sup>, 16 are downregulated in the toluene experiment while one is upregulated (See Supplementary Table S1). The repressed genes are involved in chitin synthase, regulation, modification, degradation as well as 1,3-β-glucan synthesis and processing. The fungal ergosterol biosynthetic pathway exhibit three downregulated genes (Erg2, Erg5, Erg10) while Erg12,



**Figure 3.** Functional and protein domain enrichment analysis of the genes significantly regulated between the toluene and glucose experiment. The height of each bar represents the  $\log_{10}$  of the fdr of the enrichment of each category. (A) Interproscan, Cazy and GO terms enrichment for the genes upregulated in the toluene experiment. (B) Protein domain enrichment for the downregulated genes. (C) GO term enrichment for the downregulated genes. The GO terms were clustered with REVIGO. GO terms with the same color belong to the same cluster. The most representative term of each cluster is on the top. For the cluster named Other, each bar represents a cluster of its own.

mevalonate kinase, is upregulated (See Supplementary Table S1). Among the 11 genes related to membrane lipids metabolisms (e.g. di- and tri- acyl glicerols synthesis), two exhibited a significantly increase in expression, while none were downregulated (See Supplementary Table S1).

**Potential for secondary metabolites.** A total of 152 distinct secondary metabolite biosynthesis gene clusters (SMBGCs) were predicted by antiSMASH<sup>66</sup> and SMURF<sup>67</sup>. antiSMASH reported that 145 gene clusters had similarity in other species and 8 gene clusters were similar to known SMBGCs (See Supplementary Table 1 and Table 1). Monacoline K is the product of the SMBGC with the highest similarity to a known gene cluster (22% of genes show similarity). *Cladophialophora immunda* further has clusters with similarity to SMBGCs related to antibiotics, antifungal, and toxins (methylstreptimidone, nivalenol). Nivalenol is a toxin from the group of trichothecene, a compound that seems to play an important role in *Cladophialophora immunda* as this fungus possesses 65 trichothecene efflux pumps.

In the toluene experiment, three backbone genes from three distinct SMBGCs were downregulated, while none were upregulated. CLAIMM 01034, a  $\beta$ -ketoacyl synthase, was downregulated by a factor of 17,

Location	antiSMASH type	Most Similar Cluster	Known Cluster
JSEJ01000002.1:442102–524450	cf_putative	<i>A. gypseum</i> 23%	Huperzine A 7%
JSEJ0100004.1:495938–622933	cf_saccharide	<i>S. usitatus</i> 6%	Pneumocandin 6%
JSEJ01000024.1:167780–303648	cf_putative	<i>T. tonsurans</i> 17%	Nivalenol/deoxynivalenol/3-acetyldeoxynivalenol 9%
JSEJ01000031.1:173557–325662	cf_putative	<i>C. psammophila</i> 40%	Griseofulvin 9%
JSEJ01000071.1:105157–208601	cf_putative	<i>A. terreus</i> 11%	Isoflavipucine 12%
JSEJ01000084.1:10697–57480	cf_putative	<i>C. psammophila</i> 46%	1,9-methylstreptimidone 6%
JSEJ01000096.1:7213–92198	cf_putative	<i>L. maculans</i> 27%	Penicillin 12%
JSEJ01000138.1:62903–96712	t1pks	<i>C. epimyces</i> 18%	Monacolin K 22%

**Table 1.** Secondary metabolite biosynthesis gene clusters found in *Cladophialophora immunda* with at least one gene similar to a known SMBGC. The first column contains the position of the cluster in *Cladophialophora immunda*, the second column shows the type of the secondary metabolite, the fourth column shows the species with the cluster containing the highest amount of similarity to the cluster found in *Cladophialophora immunda*. The last column shows the secondary metabolite produced by the most similar known SMBGC and the percentage of genes showing similarity.

CLAIMM\_06122, an other b-ketoacyl synthase, was downregulated by a factor 5.53, while CLAIMM\_09644, an AMP-dependent synthetase and ligase, was downregulated by a factor 6.14.

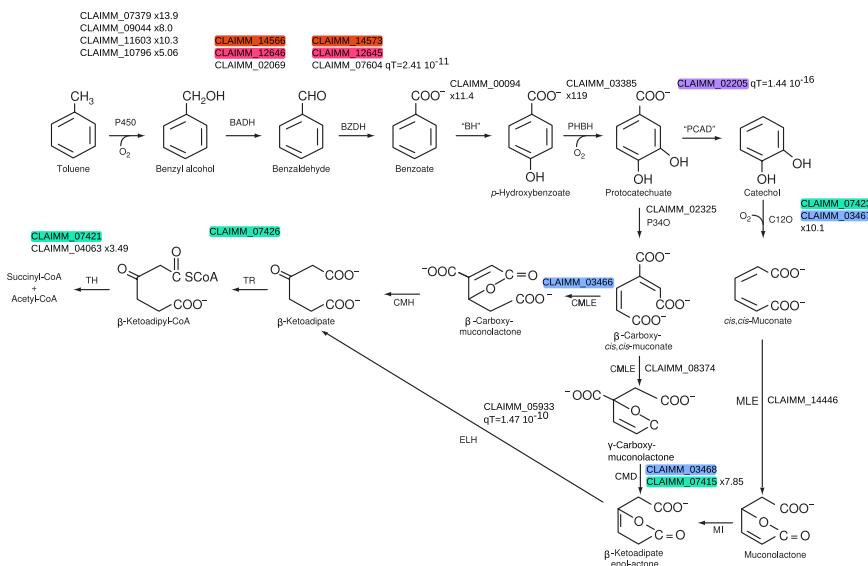
Due to the presence of trichothecene efflux pumps and a cluster similar to Nivalenol/deoxynivalenol/3-acetyldeoxynivalenol SMBGC, the ability of *Cladophialophora immunda* to produce trichothecene and trichothecene-derived toxins was studied. To this aim homologues of the genes belonging to the core trichothecene clusters of *Fusarium graminearum*, *Fusarium sporotrichoides*, *Trichoderma arundinaceum*, *Trichoderma brevicompactum*, *Stachybotrys chartarum* and *Stachybotrys chloronata* were searched in *Cladophialophora immunda*. Out of the 17 genes found in the core trichothecene cluster of these fungi<sup>68</sup> (see Supplementary Table S1), 12 were found in the genome of *Cladophialophora immunda*. The presence of Tri4, Tri5 and Tri11 indicates that the fungus may be able to synthesize molecules with the typical trichothecene structure, while the absence of Tri3, Tri7 might hinder *Cladophialophora immunda* from synthesizing trichothecene-derived mycotoxins (See Supplementary Figure S2). Interestingly, one trichothecene pump, CLAIMM\_06732, is upregulated by a factor of 23 when *Cladophialophora immunda* is exposed to toluene. Recently a paper on comparative genomics of black yeasts<sup>59</sup> showed that the trichothecene efflux pump (TEP) underwent an expansion in black yeasts. Our strain of *Cladophialophora immunda* (CBS 110551) exhibits the highest number of trichothecene efflux pumps among the black yeasts studied in this paper. Interestingly in ref. 59, a strain of *Cladophialophora immunda* (CBS 83496) was studied and had a total of 22 annotated TEPs. This large difference might be explained in part by the findings of a recent paper that showed that changes in environment triggers accelerated adaptation by increasing copy number variation in yeast<sup>69</sup>.

We further looked at genes involved in melanin production in *Cladophialophora immunda* with the data provided in ref. 18 for *Exophiala dermatitidis*. *Cladophialophora immunda* possesses three melanin biosynthesis pathways: the DHN-melanin pathway, the DOPA-melanin pathway and the L-tyrosine degradation pathway. Toluene significantly downregulated the DHN-melanin and DOPA-melanin pathways (see Supplementary Table S1). Still hppD, from the L-Tyrosine pathway, an enzyme that converts p-hydroxyphenylpyruvate to homogentisate (HGA), sees its expression increased by a factor of 79 when *Cladophialophora immunda* exposed to toluene. Besides the role of HGA as precursor of pyomelanin, it is also found in the styrene degradation pathway, where HGA is processed to fumarate and acetoacetate before entering the krebs cycle. The enzymes involved in this process are either not significantly regulated (Homogentisate 1,2-dioxygenase, Malelactoacetate isomerase) or downregulated (Fumarylacetoacetate hydrolase, CLAIMM\_05428,  $\log_2 FC = -2.09$   $fdr = 3.6 \cdot 10^{-6}$ ). As such HGA seems to be rather involved in the melanin than in energy production.

**Sensing.** Homologues to genes involved in the stress-activated MAPK pathway from *Saccharomyces cerevisiae*, *Neurospora crassa*, *Schizosaccharomyces pombe* and *Cryptococcus neoformans* were searched in *Cladophialophora immunda*<sup>70</sup>. Twelve out of 14 genes of the *Schizosaccharomyces pombe* stress-activated MAPK pathway had homologues in *Cladophialophora immunda*, while this proportion dropped to 41% (7/17) for *Saccharomyces cerevisiae*. *Cladophialophora immunda* homologues to Mak1, Mak2, Pap1, Wis1 and Atf1 (*Schizosaccharomyces pombe*) and to Pbs2 (*Saccharomyces cerevisiae*) are all significantly upregulated when toluene is used as sole carbon-source (See Supplementary Table S1).

The use of toluene as sole-carbon source upregulates the AcuK and Snf1 homologues, two genes whose expression is increased under low glucose concentration in fungi<sup>71</sup>. In contrast genes involved in the gluconeogenesis and  $\beta$ -oxidation were significantly downregulated.

**Toluene degradation.** The genome of *Cladophialophora immunda* was examined for the presence of genes belonging to the fungal toluene degradation pathways as described in ref. 72 (See Fig. 4 and Supplementary Table S1) by looking at sequence and functional homologies. The initial oxidation of the methyl-group of toluene is catalyzed by a membrane-bound cytochrome P450<sup>73</sup>. *Cladophialophora immunda* has a total of 195 membrane bound Cytochrome P450 (See Supplementary Table S1). Among them, five were upregulated in *Cladophialophora immunda* (see Fig. 4). CLAIMM\_00094, one of the five P450 candidates, is probably a benzoate hydroxylase, since it is homologue to *Aspergillus niger* CP53\_ASPPN (E-value = 0.0), a benzoate hydroxylase.



**Figure 4.** Representation of the pathways related to toluene-degradation found in *Cladophialophora immunda*. For each step in the pathway gene candidates are listed. Information about differential expression and putative horizontal gene transfer are also indicated. The colors highlight genes belonging to the same cluster. Reaction in the teal and brown frame are reported in *P. putida mt2* and *P. mendocina*, respectively.

Benzylalcohol dehydrogenase (aryl-dehydrogenase, BADH) was searched by homology to the aryl-dehydrogenase (xylB) from *P. putida mt-2*. The best hit is CLAIMM\_14566 (E-value =  $3 \cdot 10^{-98}$ ), a gene whose homologue in *Fusarium oxysporum* (E-value =  $7.2 \cdot 10^{-107}$ ) is described as aryl-dehydrogenase. ProteinOrtho predicts that CLAIMM\_14566 has two paralogs, CLAIMM\_12646 and CLAIMM\_02069. None of the BADH was found to be significantly regulated in the toluene experiment.

Similarly, benzaldehyde dehydrogenase (BZDH) was searched by looking for homologues of xylC in *Cladophialophora immunda*. Three genes, CLAIMM\_14573, CLAIMM\_12645 and CLAIMM\_07604 (E-values  $5 \cdot 10^{-68}$ ,  $1 \cdot 10^{-69}$  and  $1 \cdot 10^{-67}$ , respectively) showed both sequence and functional homologies to the bacterial BZDH. Homologues to CLAIMM\_14573 and CLAIMM\_12645 were found among our set of genomes solely in *Cladophialophora bantiana* (CLAIMM\_14573) and *Fusarium oxysporum* (CLAIMM\_12645, CLAIMM\_14573) while CLAIMM\_07604 had no homologues. In fact, CLAIMM\_07604 was significantly ( $qT$ -value  $2.41 \cdot 10^{-11}$ ) horizontally transferred from the bacterial kingdom to the fungal one. This is further confirmed by the distance tree of its closest homologues found with Blastp against the nr database (Fig. 5). It should be noted that two putative enzymes involved in BADH degradation (CLAIMM\_14566, CLAIMM\_12646) are located in close genomic vicinity to two of the BZDH-degrading enzymes (CLAIMM\_14573, CLAIMM\_12645).

Benzoate is converted to *p*-hydroxybenzoate by the benzoate hydroxylase (BH), CLAIMM\_00094, which is homologue to *Aspergillus niger* CP53 ASPNG (E-value = 0.0) and is upregulated (x11.4) in the toluene experiment. *p*-hydroxybenzoate is oxidized to protocatechuate by *p*-hydroxybenzoate hydroxylase (PHBH). In *Cladophialophora immunda*, five genes (CLAIMM\_02622 CLAIMM\_03385 CLAIMM\_08446 CLAIMM\_09063 CLAIMM\_11189) have the PF07976 domains, which was shown to be involved in the hydroxylation of *p*-hydroxybenzoate<sup>74</sup>. Among those genes only CLAIMM\_03385 was upregulated (x119) when *Cladophialophora immunda* was exposed to toluene.

Protocatechuate is then degraded by two distinct pathways to  $\beta$ -ketoadipate. In one branch, the aromatic ring is opened by the dioxygenase (P340), CLAIMM\_02325, an homologue of CCT72286 (*Fusarium fujikuroi*, E-value =  $5 \cdot 10^{-127}$ ).  $\beta$ -carboxy-*cis,cis*-muconate is then processed by  $\beta$ -carboxy-*cis,cis*-muconate lactonizing enzyme (CMLE). *Cladophialophora immunda* has one CMLE homologue to *Neurospora crassa* CMLE (CLAIMM\_03466, E-value =  $4 \cdot 10^{-176}$ ) and one homologue to *Acinetobacter baylyi* CMLE (CLAIMM\_08374, E-value =  $2 \cdot 10^{-93}$ ). Both enzymes have the same substrate, but delivers  $\beta$  or  $\gamma$ -carboxy-muconolactone, respectively.  $\beta$ -carboxy-muconolactone is processed by  $\beta$ -carboxy-muconolactone hydrolase (CMH), an enzyme for which no candidate could be found in fungi. However  $\gamma$ -carboxy-muconolactone is processed by carboxy-muconolactone decarboxylase (CMD) to  $\beta$ -ketoadipate enol-acitone. A total of 13 genes in *Cladophialophora immunda* have a CMD-like protein domain (IPR003779), with CLAIMM\_07415 being upregulated (x7.85) in the toluene experiment.

In the second branch of the  $\beta$ -ketoadipate pathway, PCAD converts protocatechuate to catechol. *Cladophialophora immunda* contains 4 genes (CLAIMM\_04693, CLAIMM\_14616, CLAIMM\_09412, CLAIMM\_02205) belonging to the UbiD decarboxylase family (IPR002830), whose members are involved among others in the decarboxylation of protocatechuate<sup>75</sup>. Based on HGTfinder, CLAIMM\_02205 is of bacterial origin *Cladophialophora immunda* ( $qT$ -value =  $1.44 \cdot 10^{-16}$ , see Fig. 5). The subsequent oxidation of catechol by catechol 1,2-dioxygenase (C12O,  $\times 10.1$  in toluene, CLAIMM\_03467, homologue to *Aspergillus niger* An13g02000 E-value =  $2 \cdot 10^{-145}$ ) opens the aromatic ring and converts it to *cis,cis*-muconate. Six other genes



**Figure 5.** Distance trees of the Blastp hits of CLAIMM\_07604 (BZDH), CLAIMM\_02205 (PCAD), CLAIMM\_05933 (ELH) and CLAIMM\_02199 (xylJ) against the nr protein database. The tip labels are colored based on the Class the species belong to. The position of *Cladophialophora immunda* in the tree is marked with a star. Blastp homologues were found in three fungal classes (Eurotiomycetes, Sordariomycetes, Dothideomycetes) and 11 bacterial classes.

in *Cladophialophora immunda* have catechol 1,2-dioxygenase activity (See Supplementary Table S1) but none of them are significantly regulated. *Cis,cis*-muconate lactonizing enzymes (CLAIMM\_14446) ( $E\text{-value } 4 \cdot 10^{-104}$  MLE *Trichosporon cutaneum*) yields muconolactone that is processed by muconolactone isomerase (MI). While this enzyme has been reported to be present in fungi, no corresponding fungal sequence could be found in the literature.  $\beta$ -ketoadipate-enol-actone hydrolase (CLAIMM\_05933, ELH from *Cladophialophora carrionii*,  $E\text{-value} = 9 \cdot 10^{-134}$ ), is of bacterial origin ( $qT = 1.47 \cdot 10^{-10}$ , see Fig. 5) and delivers  $\beta$ -ketoadipate.

$\beta$ -ketoadipate is then processed to  $\beta$ -ketoadipyl-CoA by  $\beta$ -ketoadipate succinyl-CoA transferase (CLAIMM\_7426,  $E\text{-value } 4 \cdot 10^{-56}$  from *P. putida* pca).  $\beta$ -ketoadipate succinyl-CoA is converted to Succinyl-CoA and Acetyl-CoA by  $\beta$ -ketoadipyl-CoA thiolase. In *Cladophialophora immunda* 10 genes have a thiolase active site, among which one is upregulated (CLAIMM\_04063  $\times 3.49$ ) and one is located close to genes all involved in the degradation of toluene (CLAIMM\_07421).

Besides the genes involved in the toluene degradation, the presence of functional and sequence homologues of bacterial genes responsible for toluene degradation was investigated. *xylL* is involved in the conversion of toluate cis-dihydrodiol dehydrogenase in *P. putida*<sup>72</sup>. It has 33% sequence similarity to CLAIMM\_00812 (Blastp E-value =  $2 \cdot 10^{-32}$ ), similar length and protein domains (PS00061, PR00080, PR00081, SSF51735, PF00106, IPR002347, IPR020904, IPR016040) (See Supplementary Figure S3). Similarly *xylJ*, 2-hydroxypent-2,4-dienoate hydratase, shows 34% sequence similarity to CLAIMM\_02199, as well as identical domain annotation (IPR011234, SSF56529, G3D 3.90.850.10). CLAIMM\_02199 is interestingly also classified as horizontally transferred by HGTfinder ( $qT = 1.1 \cdot 10^{-5}$ , see Fig. 5) and is located close to CLAIMM\_02205 (PCAD).

CLAIMM\_07216 is a homologue of yeast YNR064C, a gene that belongs to the *xylF* family of esterase, whose members are responsible for the conversion of 2-hydroxymuconic semialdehyde to 2-oxopent-4-eneoate (See Supplementary Figure S3).

In *Pseudomonas mendocina*, *pcuAB* is a dimer responsible for the oxydation of *p*-cresol to *p*-Hydroxy-benzaldehyde. CLAIMM\_11938, a gene upregulated (x18.03) upon toluene exposure, is homologue to *PcuB* (E-value  $3 \cdot 10^{-113}$ ) and harbours the same IPR annotations. Interestingly members of the *Cladophialophora* genus were reported to degrade *p*-cresol<sup>76</sup>. *pcuC* is homolog to CLAIMM\_09141 (E-value =  $4 \cdot 10^{-73}$ ) and CLAIMM\_08951 (E-value  $8 \cdot 10^{-73}$ ), the latter having probably been horizontally transferred from bacteria to fungi ( $qT$ -value =  $3 \cdot 10^{-3}$ ). *pcuC* is responsible for the oxydation of *p*-Hydroxy-benzaldehyde to *p*-Hydroxy-benzoate.

A look at the genomic organization of the genes involved in the toluene degradation indicates that 13 of them are grouped into 5 clusters. The first two clusters contain four genes corresponding to BADH and BZDH. Seven genes involved in catechol and protocatechuate degradation are distributed into two clusters respectively (see Fig. 4). The last cluster contains the genes PCAD and the homologue of *xylJ*, which were both horizontally transferred (Figs 4 and 5).

## Discussion

In this study, for the first time, the transcriptome of a fungus growing on toluene<sup>12</sup> was sequenced. With the help of the transcriptomic data and a comparative genomics analysis, genes previously reported to play a role in metabolizing toluene in other fungi and bacteria<sup>72</sup> were identified and their expression levels during the toluene exposure were assessed. For all but two enzymes in the fungal toluene-degrading pathway described in ref. 72, at least one orthologue was found. The genomic organization of these coding loci in *Cladophialophora immunda* is highly interesting since five gene clusters were identified. While functional clusters are common in fungi<sup>77</sup>, it is the first time that this genomic organization is reported for clusters of genes related to hydrocarbon degradation.

The comparison of the fungal and bacterial toluene-degradation pathways allowed to identify 8 genes in *Cladophialophora immunda* that are responsible for toluene degradation steps in *P. putida*<sup>78</sup> and *P. mendocina*<sup>72</sup>. Among those 8 genes, four were classified as horizontally transferred from bacteria by HGTfinder, further underlining the close interplay between bacteria and the studied *Cladophialophora*. While HGTs are common in fungi, they were only recently reported in the group of black yeasts, i.e. for *Exophiala dermatitidis*<sup>18</sup>. Especially interesting are CLAIMM\_02205 (PCAD) and CLAIMM\_02199 (*xylJ*) (See Fig. 5). CLAIMM\_02205 is the only gene that converts protocatechuate to catechol while CLAIMM\_02199 is the only gene that catalyzes the hydration of 2-oxopent-4-eneoate into 4-hydroxy-2-oxovalerate. The fact that the homologues to bacterial enzymes partially cover additional toluene-degradation pathway might be an evidence that in a toluene-contaminated environment *Cladophialophora immunda* and bacterial species are cooperating in metabolizing toluene, similar to what was shown for polycyclic aromatic hydrocarbons<sup>79</sup>.

The genomic analysis of significantly enriched protein domains lead to some interesting observations about the ecology of *Cladophialophora immunda*. The fact that proteins involved in complex carbohydrates degradation and sugar transport are overrepresented might suggest that the fungus evolved in an oligotrophic and polluted environment. The overrepresentation of trichothecene efflux pumps and the upregulation of one them by a factor 23 in the toluene experiment, might signal that this secretion system is involved in the excretion of toluene or secondary metabolites.

Despite the ability of *Cladophialophora immunda* to degrade toluene, its presence drives the fungus into an energy-saving state. Core processes like cellular amid metabolism, cellular respiration, energy-coupled proton transport, organic substance metabolism are negatively affected by the xenobiotic. Translation is one of the most enriched GO term ( $fdr = 1.17 \cdot 10^{-46}$ ) in the set of downregulated genes. This fits well with the repression of 17 tRNAs during toluene exposure. In contrast, snoRNAs, which are responsible for the rRNA maturation<sup>80</sup>, represent the only type of basal ncRNAs that are significantly upregulated in the toluene experiment. In *Escherichia coli*, toluene triggers the rapid disaggregation of ribosomes<sup>81</sup>. Here the increased expression of snoRNAs might be interpreted as response to the destabilizing effect of toluene on the maturation process of rRNAs.

Chaperones HSP20, HSP40 and HSP70, which are markers of cellular stress<sup>82</sup>, are either highly expressed or highly upregulated in the toluene -experiment. This fact, together with the upregulation of genes involved in the transport of misfolded proteins, might indicate that toluene is having a negative impact on the protein-folding process, as recently reported in prokaryotes<sup>83</sup>. Toluene exposure also leads to the overexpression of genes harboring protein domains linked to DNA repair, like ku70/ku80 and rad14, indicating that the fungal DNA is being damaged. This is in line with the recognized mutagenic properties of aromatic and other poly-aromatic compounds<sup>84</sup>.

Toluene further triggers the expression of antioxidants as well as genes involved in cell detoxification, like Glyoxylase II, Glutathione S-transferase, Ascorbate peroxidase and Carotenoid oxygenase. HppD, a key enzyme in the production of Pyomelanin is upregulated by a factor 79 when *Cladophialophora immunda* is exposed to toluene. This melanin type was shown to be overexpressed when *Exophiala dermatitidis* is grown on skin<sup>20</sup> and to protect against hydrogen peroxide stress<sup>85</sup> in the soil-borne bacteria *Ralstonia solanacearum*.

Six genes belonging to the stress-induced MAPK pathway in fungi are upregulated in the toluene experiment. This signaling pathway is related to various kind of stress, among others oxidative stress<sup>70</sup>. This, together with the damaged-DNA repair response and melanin production, suggests that toluene is causing an oxidative burden on *Cladophialophora immunda*, something previously reported in prokaryotes<sup>64</sup>.

In conclusion, the thorough genome annotation of *Cladophialophora immunda* paired with the transcriptome data allows us to get a better insight into the mechanisms used by this fungus to protect itself from toluene and degrade it. Even though this fungus is one of the most efficient toluene degrader<sup>12</sup>, this xenobiotic is sensed by *Cladophialophora immunda* as a toxic compound that triggers a strong stress reaction and the downregulation of various basal metabolism.

## References

1. Johnston, C., Rayner, J., Patterson, B. & Davis, G. Volatilisation and biodegradation during air sparging of dissolved btex-contaminated groundwater. *Journal of Contaminant Hydrology* **33**, 377–404 (1998).
2. Huesemann, M. H. Guidelines for land-treating petroleum hydrocarbon-contaminated soils. *Soil and Sediment Contamination* **3**, 299–318 (1994).
3. Bolden, A. L., Kwiatkowski, C. F. & Colborn, T. New look at btex: are ambient levels a problem? *Environmental science & technology* **49**, 5261–5276 (2015).
4. Stomińska, M., Król, S. & Namieśnik, J. Removal of btex compounds from waste gases; destruction and recovery techniques. *Critical reviews in environmental science and technology* **43**, 1417–1445 (2013).
5. Azubuike, C. C., Chikere, C. B. & Okpokwasili, G. C. Bioremediation techniques-classification based on site of application: principles, advantages, limitations and prospects. *World Journal of Microbiology and Biotechnology* **32**, 180 (2016).
6. De Lorenzo, V. Recombinant bacteria for environmental release: what went wrong and what we have learnt from it. *Clinical Microbiology and Infection* **15**, 63–65 (2009).
7. Agathos, S. N. et al. Gas-phase methyl ethyl ketone biodegradation in a tubular biofilm reactor: microbiological and bioprocess aspects. *Biodegradation* **8**, 251–264 (1997).
8. Auria, R., Frere, G., Morales, M., Acuna, M. & Revah, S. Influence of mixing and water addition on the removal rate of toluene vapors in a biofilter. *Biotechnology and Bioengineering* **68**, 448–455 (2000).
9. Qi, B., Moe, W. & Kinney, K. Biodegradation of volatile organic compounds by five fungal species. *Applied microbiology and biotechnology* **58**, 684–689 (2002).
10. Wang, C., Xi, J.-Y., Hu, H.-Y. & Wen, X.-H. Biodegradation of gaseous chlorobenzene by white-rot fungus phanerochaete chrysosporium. *Biomedical and Environmental Sciences* **21**, 474–478 (2008).
11. Garca-Peña, I., Ortiz, I., Hernandez, S. & Revah, S. Biofiltration of btex by the fungus paecilomyces variotii. *International Biodeterioration & Biodegradation* **62**, 442–447 (2008).
12. Blasi, B. et al. Pathogenic Yet Environmentally Friendly? Black Fungal Candidates for Bioremediation of Pollutants. *Geomicrobiol. J.* **33**, 308–317 (2016).
13. Prenafeta-Boldú, F. X. et al. Isolation and characterisation of fungi growing on volatile aromatic hydrocarbons as their sole carbon and energy source. *Mycological Research* **105**, 477–484 (2001).
14. Yazdanparast, S. et al. Consistent high prevalence of exophiala dermatitidis, a neurotropic opportunist, on railway sleepers. *Journal de Mycologie Médicale/Journal of Medical Mycology* (2017).
15. Seyedmousavi, S. et al. Exophiala sideris, a novel black yeast isolated from environments polluted with toxic alkyl benzenes and arsenic. *Fungal biology* **115**, 1030–1037 (2011).
16. Badali, H. et al. Biodiversity of the genus cladophialophora. *Studies in Mycology* **61**, 175–191 (2008).
17. Prenafeta-Boldú, F. X., Summerbell, R. & De Hoog, G. S. Fungi growing on aromatic hydrocarbons: biotechnology's unexpected encounter with biohazard? *FEMS Microbiology Reviews* **30**, 109–130 (2006).
18. Chen, Z. et al. Comparative genomic and transcriptomic analysis of Wangiella dermatitidis, a major cause of phaeohyphomycosis and a model black yeast human pathogen. *G3 (Bethesda)* **4**, 561–578 (2014).
19. Blasi, B., Tafer, H., Tesei, D. & Sterflinger, K. From glacier to sauna: Rna-seq of the human pathogen black fungus exophiala dermatitidis under varying temperature conditions exhibits common and novel fungal response. *PLoS one* **10**, e0127103 (2015).
20. Poynther, C. et al. The transcriptome of exophiala dermatitidis during ex-vivo skin model infection. *Frontiers in cellular and infection microbiology* **6** (2016).
21. Hartmans, S. & Tramper, J. Dichloromethane removal from waste gases with a trickle-bed bioreactor. *Bioprocess Eng.* **6**, 83–92 (1991).
22. Okon, Y., Albrecht, S. L. & Burris, R. Methods for growing spirillum lipoferum and for counting it in pure culture and in association with plants. *Applied and Environmental Microbiology* **33**, 85–88 (1977).
23. Sterflinger, K. et al. Draft genome of *cladophialophora immunda*, a black yeast and efficient degrader of polyaromatic hydrocarbons. *Genome announcements* **3**, e01283–14 (2015).
24. Nawrocki, E. P. & Eddy, S. R. Infernal 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–5 (2013).
25. Nawrocki, E. P. et al. Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Research* **43**, D130–D137 (2015).
26. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE snoScan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* **33**, W686–W689 (2005).
27. Bartschat, S., Kehr, S., Tafer, H., Stadler, P. F. & Hertel, J. snoStrip: a snoRNA annotation pipeline. *Bioinformatics* **30**, 115–6 (2014).
28. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
29. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
30. Haas, B. J. et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654–5666 (2003).
31. Stanke, M. et al. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**, 435–439 (2006).
32. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
33. Wang, L. et al. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Research* **41**, 1–7 (2013).
34. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
35. Finn, R. D. et al. Pfam: the protein families database. *Nucleic Acids Research* **42**, D222–30 (2014).
36. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
37. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
38. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
39. Washietl, S. et al. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA* **17**, 578–594 (2011).

40. Gruber, A. R., Findeiß, S., Washietl, S., Hofacker, I. L. & Stadler, P. F. RNAz 2.0: improved noncoding RNA detection. *Pac. Symp. Biocomput.* **69–79** (2010).
41. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
42. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2 (2012).
43. Tafer, H., Lopandic, K., Blasi, B., Poyntner, C. & Sterflinger, K. Draft Genome Sequence of Exophiala mesophila, a Black Yeast with High Bioremediation Potential. *Genome Announc.* **3**, e00203–15 (2015).
44. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
45. Saier, M. H. *et al.* The transporter classification database (tcdb): recent advances. *Nucleic Acids Research* **44**, D372–D379 (2016).
46. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **42**, D503–9 (2014).
47. Cantarel, B. L. *et al.* The carbohydrate-active enzymes database (cazy): an expert resource for glycogenomics. *Nucleic Acids Research* **37**, D233–D238 (2009).
48. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).
49. Lechner, M. *et al.* Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC bioinformatics* **12**, 124 (2011).
50. Nguyen, M., Ekstrom, A., Li, X. & Yin, Y. HGT-finder: A new tool for horizontal gene transfer finding and application to Aspergillus genomes. *Toxins* **7**, 4035–4053 (2015).
51. Dobin, A. *et al.* RNA-STAR: ultrafast universal spliced sequences aligner: Supplementary materials. *Bioinformatics* **1–7** (2012).
52. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
53. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40 (2010).
54. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
55. Xie, C. *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research* **39**, W316–22 (2011).
56. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
57. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. *bioRxiv* (2016).
58. Parra, G., Bradnam, K. & Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
59. Teixeira, M. M. *et al.* Exploring the genomic diversity of black yeasts and relatives (chaetothyriales, ascomycota). *Studies in mycology* **86**, 1–28 (2017).
60. Washietl, S., Hofacker, I. L. & Stadler, P. F. Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci. USA* **102**, 2454–9 (2005).
61. Mirocha, C. J., Abbas, H., Kommedahl, T. & Jarvis, B. Mycotoxin production by fusarium oxysporum and fusarium sporotrichoides isolated from baccharis spp. from brazil. *Applied and environmental microbiology* **55**, 254–255 (1989).
62. Chen, L. H. *et al.* 4-oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer. *Journal of Biological Chemistry* **267**, 17716–17721 (1992).
63. Wang, P. & Heitman, J. The cyclophilins. *Genome biology* **6**, 226 (2005).
64. Domnguez-Cuevas, P., González-Pastor, J.-E., Marquéz, S., Ramos, J.-L. & de Lorenzo, V. Transcriptional tradeoff between metabolic and stress-response programs in pseudomonas putida kt2440 cells exposed to toluene. *Journal of Biological Chemistry* **281**, 11981–11991 (2006).
65. Sikkema, J., De Bont, J. & Poolman, B. Mechanisms of membrane toxicity of hydrocarbons. *Microbiological reviews* **59**, 201–222 (1995).
66. Weber, T. *et al.* antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* **43**, W237–43 (2015).
67. Khaldi, N. *et al.* SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* **47**, 736–41 (2010).
68. Semeiks, J., Borek, D., Otwinowski, Z. & Grishin, N. V. Comparative genome sequencing reveals chemotype-specific gene clusters in the toxicogenic black mold Stachybotrys. *BMC Genomics* **15**, 590 (2014).
69. Hull, R. M., Cruz, C., Jack, C. V. & Houseley, J. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biology* **15**, e2001333 (2017).
70. Bahn, Y.-S. *et al.* Sensing the environment: lessons from fungi. *Nature Reviews Microbiology* **5**, 57–69 (2007).
71. Fleck, C. B., Schöbel, F. & Brock, M. Nutrient acquisition by pathogenic fungi: nutrient availability, pathway regulation, and differences in substrate utilization. *International Journal of Medical Microbiology* **301**, 400–407 (2011).
72. Parales, R. E., Parales, J. V., Pelletier, D. A. & Ditty, J. L. Diversity of microbial toluene degradation pathways. *Adv. Appl. Microbiol.* **64**, 1–73, 2 p following 264 (2008).
73. Luykx, D. M. A. M., Prenafeta-Boldú, F. X. & de Bont, J. A. M. Toluene monooxygenase from the fungus Cladosporium sphaerospermum. *Biochem. Biophys. Res. Commun.* **312**, 373–379 (2003).
74. Chang, H.-K. & Zylstra, G. J. Examination and expansion of the substrate range of m-hydroxybenzoate hydroxylase. *Biochemical and biophysical research communications* **371**, 149–153 (2008).
75. He, Z. & Wiegel, J. Purification and characterization of an oxygen-sensitive reversible 4-hydroxybenzoate decarboxylase from clostridium hydroxybenzoicum. *The FEBS Journal* **229**, 77–82 (1995).
76. Atagana, H. I. Biodegradation of phenol, o-cresol, m-cresol and p-cresol by indigenous soil fungi in soil contaminated with creosote. *World Journal of Microbiology and Biotechnology* **20**, 845–849 (2004).
77. Hittinger, C. T., Rokas, A. & Carroll, S. B. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14144–9 (2004).
78. Otenio, M. H., Silva, M. T. L. D., Marques, M. L. O., Roseiro, J. C. & Bidoia, E. D. benzene, toluene and xylene biodegradation by pseudomonas putida ccm1 852. *Brazilian Journal of Microbiology* **36**, 258–261 (2005).
79. Peng, R.-H. *et al.* Microbial biodegradation of polycyclic aromatic hydrocarbons. *FEMS microbiology reviews* **32**, 927–955 (2008).
80. Watkins, N. J. & Bohnsack, M. T. The box c/d and h/aca snrnp: key players in the modification, processing and the dynamic folding of ribosomal rna. *Wiley Interdisciplinary Reviews: RNA* **3**, 397–414 (2012).
81. Jackson, R. W. & DeMoss, J. Effects of toluene on escherichia coli. *Journal of bacteriology* **90**, 1420–1425 (1965).
82. Lindquist, S. & Craig, E. The heat-shock proteins. *Annual review of genetics* **22**, 631–677 (1988).
83. Volkers, R. J., Snoek, L. B., Ruijssenaars, H. J. & de Winde, J. H. Dynamic response of pseudomonas putida s12 to sudden addition of toluene and the potential role of the solvent tolerance gene trgi. *PloS one* **10**, e0132416 (2015).
84. Muñoz, B. & Albores, A. *DNA damage caused by polycyclic aromatic hydrocarbons: mechanisms and markers* (INTECH Open Access Publisher, 2011).
85. Ahmad, S. *et al.* Genetic determinants for pyomelanin production and its protective effect against oxidative stress in ralstonia solanacearum. *PloS one* **11**, e0160845 (2016).

## Acknowledgements

The computational results presented were achieved, in part, through the use of the Vienna Scientific Cluster (VSC). This research was financed by the Austrian Science Fund (FWF) project (P 25119-B16). The equipment of the VIBT-Extremophile Center used in this study was financed by BOKU-Equipment GesmbH.

## Author Contributions

K.S.G.: Experimental design. B.B. and C.P.: Performed the experimental work. H.T. and C.K.: Data Analysis. B.B. and H.T. and C.K. and C.P. and K.L. interpretation of the data. All authors discussed the results, wrote and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-11807-8](https://doi.org/10.1038/s41598-017-11807-8)

**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.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017