Comparative pan-genomics reveals divergent adaptations in clinically-relevant members of the *Fusarium solani* species complex

Phillip J.T. Brassington[1,2], Maria Laura Fabre[1#], Anna Zimmermann[2#], Marion Perrier[1,3], Ronny Martin[2], Oliver Kurzai[2,4], Amelia E. Barber[1,3]

[1] Junior Research Group Fungal Informatics. Institute of Microbiology. Friedrich Schiller University. Jena, Germany.
[2] Institute of Hygiene and Microbiology. University of Würzburg. Würzburg, Germany.
[3] Cluster of Excellence Balance of the Microverse. Friedrich Schiller University. Jena, Germany.
[4] National Reference Center for Invasive Fungal Infections (NRZMyk). Leibniz Institute for Natural Product Research and Infection Biology-Hans Knöll Institute. Jena, Germany.

[#] Maria Laura Fabre and Anna Zimmermann contributed equally. Author order was determined alphabetically.

Address correspondence to:
Amelia E. Barber
amelia.barber@uni-jena.de

Abstract

The *Fusarium solani* species complex are a group of dual-kingdom fungal pathogens capable of causing devastating disease on a wide range of host plants and life-threatening infections in humans that are difficult to treat. In this study, we generate highly contiguous genomes for three clinical isolates of *Fusarium keratoplasticum* and three clinical isolates of *Fusarium petroliphilum* and compare them with other genomes of the FSSC from plant and animal sources. We find that human pathogenicity is polyphyletic within the FSSC, including in *F. keratoplasticum*. Pan-genome analysis revealed a high degree of gene presence-absence in the complex, with only 41% of genes (11,079/27,068) found in all samples and the presence of accessory chromosomes encoding isolate- and species-specific genes. We also defined conserved long non-coding RNAs (lncRNAs) between *F. keratoplasticum* and *F. petroliphilum,* revealing that they show a similar low degree of presence-absence variation. Secondary metabolite analysis revealed a conserved core set of biosynthetic gene clusters across the FSSC, as well as a unique cluster potentially linked to keratitis. Transcriptomic analysis under stress conditions showed minimal differential gene expression, indicating that both *F. keratoplasticum* and *F. petroliphilum* are well adapted to human infection-relevant conditions. This study provides valuable insights into the evolutionary dynamics, genomic architecture, and potential pathogenicity mechanisms of the FSSC, with implications for understanding multi-kingdom virulence, of increasing relevance as climate change potentially increases the number of fungal species that can grow at human temperatures.

Introduction

The genus *Fusarium* encompasses a highly diverse range of filamentous fungal species, including environmental saprobes, plant pathogens, and organisms capable of causing disease in animals and humans. Well established as plant pathogens, *Fusarium* spp. are causal agents for devastating worldwide food supply disruptions [1–3], causing significant losses in global yields of wheat, maize, rice, potato, and soybean [4], with outbreaks of *Fusarium* head blight causing over 50% wheat losses in China, Europe, and the Americas [5]. Yet beyond this, *Fusarium* spp are largely unique among fungi in that they are multi-kingdom pathogens capable of also infecting higher eukaryotes. In this capacity, *Fusarium* spp. are playing a role in the decline of threatened animal species [6], and are the causative agents with rising numbers of highly lethal human fusariosis infections [7]. In 2022, the World Health Organization identified the genus as one of the highest priority human fungal pathogens, and one with the most urgent need for additional study based on current knowledge gaps [8]. These include lack of diagnostic options, limited

epidemiologic data, and the intrinsic antifungal resistance, which cumulatively contribute to a high mortality rate of 43-67% [8].

Previous studies have highlighted the exceptional range of genomic diversity and plasticity among *Fusarium* spp. Even within a single species, the genome structure is highly flexible, and strains often possess accessory chromosomes that are variably conserved and, importantly, confer the ability to infect specific plant species [9]. These accessory chromosomes can even be horizontally transferred, resulting in a corresponding transfer of host plant specificity [10–12]. Despite the large number of species in the genus, the majority (reportedly up to 80-90%) of human infections are due to only two species groups: the *Fusarium oxysporum* species complex and the *Fusarium solani* species complex [13–17].

The *Fusarium solani* species complex (FSSC) comprises globally ubiquitous problematic agricultural pathogens and is the most frequent cause of human fusariosis [1,3,13,17,18]. Among the human pathologies it causes, it is particularly problematic in causing corneal infection and disseminated disease. More recently, it has been linked to two outbreaks of meningitis in immunocompetent patients in Mexico due to contaminated anesthetic solutions [19,20]. In the plant kingdom, the FSSC exhibits a broad host range and is capable of causing root rot in multiple agriculturally important crops, including peas, cucurbits, and sweet potato [21–24]. Even within individual species of the complex, there is a broad range of habitats and hosts. For example, *F. keratoplasticum* has been found in diverse habitats ranging from plumbing systems, strawberry roots, Chinese softshell turtle eggs, to human nails, eyes, and endocardium [6,25–29]. This is also likely a non-exhaustive list, as many studies classify isolates as *forma specialis* (*f. sp*). based on the host plant, rather than identifying the exact FSSC species member. The diversity of habitats and pathologies caused by the FSSC in both plants and animals is accompanied by a correspondingly high level of genomic diversity, including the presence of core and accessory chromosomes [6]. However, the genomes of FSSC members remain understudied compared to other *Fusarium* groups like the *Fusarium oxysporum* species complex. Moreover, it is unclear what, if any, genomic features are specific to particular species or hosts (plant or animal).

In this study, we characterized the genomic diversity of six FSSC member species focusing on *F. keratoplasticum* and *F. petroliphilum*, two of the most prominent FSSC species in human keratitis [28,30], with the latter being particularly understudied and lacking even a reference genome. To this end, we generated near-chromosomal genomes of three novel clinical isolates

of *F. keratoplasticum*, and provided the first three genomes for *F. petroliphilum*. We also predict long non-coding RNAs for these species, the first to investigate this class of regulatory genes in the FSSC. Finally, we integrate these genomes with six previously sequenced, nearly gapless, genomes from plant and animal isolates of the FSSC [6] to study the genomic diversity of the complex.

Results

Genome sequencing of novel clinical *Fusarium solani* species complex isolates and description of the study dataset

In this study we used PacBio long read sequencing combined with Illumina short read sequencing to assemble the genomes of six novel FSSC isolates that were collected from keratitis cases by the German National Reference Center for Invasive Fungal Infections (n=3 *F. keratoplasticum* and n=3 *F. petroliphilum*). To facilitate more in-depth comparisons within the species complex, and to include isolates of environmental and animal origin, six additional FSSC isolates with similarly contiguous genomes were incorporated into the study dataset [6]. The published samples included isolates of *F. falciforme*, *F. keratoplasticum, F. solani* FSSC12, *F. vanettennii*, and originated from plant (pea plant and orchid flower) and marine animal (including hawksbill sea turtle, roughtail stingray, and harbor seal) sources (Table 1). The final assemblies of the novel and published assemblies were highly contiguous, with an average of 58 scaffolds (range 14 to 143) and an average N50 of 3.8 Mb ± 0.4 (Table 1). The FSSC genomes analyzed in this study ranged in size from 48.0 Mb for one of the newly sequenced *F. keratoplasticum* isolates to 72.9 Mb for *F. vanettenii* Fs6 and encoded between 14,050 and 16,755 protein coding genes. Parallelling the range of genome sizes, a similar pattern was observed for the fraction of the genome consisting of repetitive regions, including transposable elements, which varied from 4.5% for *F. keratoplasticum* NRZ-2015-035 to 14.2% for *F. vanettennii* Fs6. Using BUSCO as a measure of genome completeness, we found that all genomes were highly complete, recovering on average >98.5% of the expected single-copy orthologs. Despite originating from different studies, there were no significant differences in key genome quality metrics such as N50, the number of protein coding genes, or BUSCO completeness between the isolates sequenced in this study and the public assemblies ($p>0.05$ by Mann-Whitney U test).

Human and plant association is polyphyletic in the FSSC.

119 To determine the phylogenetic relationship of the newly sequenced clinical isolates within the
120 genus *Fusarium* and the FSSC complex, we constructed a whole genome phylogeny from amino
121 acid sequence of conserved single copy orthologs. As expected, the new isolates of
122 *F. keratoplasticum* and *F. petroliphilum* were located within clade 3 of the FSSC in the phylogeny
123 (Figure 1). The *F. petroliphilum* isolates formed a monophyletic group, with *Fusarium* FSSC12
124 being the closest relative in the dataset. The *F. keratoplasticum* isolates belonged to a separate
125 monophyletic group containing *F. solani sensu stricto*, *F. falciforme*, and the two other
126 *F. keratoplasticum* isolates in the dataset. Interestingly, the clinical *F. keratoplasticum* isolates
127 from Germany did not segregate from the marine host-associated isolates from east Asia,
128 indicating a lack of geographic population structure or host specificity in the species. This overall
129 pattern was observed in the entire complex, with virulence in both plant and human hosts being
130 polyphyletic in the FSSC.
131
132 **41% of genes are common to all FSSC isolates, while pan-gene conservation is higher in**
133 **individual species.**
134 We next sought to quantify the genomic diversity of the FSSC by characterizing its pan-genome
135 and delineating the genes common to all genomes analyzed (core genome) and those which
136 show presence-absence variation within the species complex (accessory genome). Of the 27,068
137 orthologous genes identified across the 12 isolates and 6 species, 41% (n=11,079) were core, or
138 present in all isolates (Figure 2a & b). The FSSC pan-genome exhibited a long tail distribution,
139 where 6% of the pan-genome (n=1,666) was present in all but one sample and 7% (n=1,946) of
140 the accessory genome was isolate-specific (7%; n=1,946) (Figure 2b). Despite the low fraction of
141 core genes, Heaps' law indicates that the FSSC pan-genome is closed (alpha = 1.57, where alpha
142 values > 1 indicate a closed pan-genome) and the number of pan-genes beginning to plateau
143 after six genomes (Figure 2c). Protein sequences of core genes were significantly longer than
144 those of accessory genes, averaging 505 amino acids compared to 377 for the latter (Wilcoxon p
145 < 0.0001). Within the scope of single species, the fraction of core genes was notably higher. For
146 *F. keratoplasticum,* 75% of pan-genes were core, or in all 5 isolates (13,275/17,617), while 88%
147 of pan-genes were core in *F. petroliphilum* (14,067/15,923 pan-genes; n=3 isolates) (Figure 2d &
148 e).
149
150 **Diverging patterns in sequence conservation and repetitive elements between core and**
151 **accessory chromosomes**

152 In addition to the high degree of gene presence-absence variation, some of the accessory
153 genome in *Fusarium* species is known to be encoded on discrete accessory chromosomes. These
154 accessory chromosomes are involved in plant virulence and the ability to infect specific host
155 plants, but their role in human infection is unclear. Intriguingly, work in the *F. oxysporum* species
156 complex hints that clinical isolates of this species complex may carry accessory chromosomes
157 that are distinct from plant-pathogenic isolates [10]. To investigate this in the FSSC, we analyzed
158 our study dataset for the conservation of both core and accessory contigs among the isolates
159 from plant and animal origins. We identified 12 core chromosomes, which were present across
160 the isolates in the dataset with a high degree of nucleotide identity and synteny (Figure 3).
161 However, the relative proportion of the core chromosomes compared to the total genome varied
162 considerably within the species complex. Core chromosomes, encompassing a total length of
163 46.8 Mb in *F. keratoplasticum* Fu6, made up 95% of its total genome length. In comparison, core
164 chromosomes comprised only 73% of the total genome length of *F. vanettenii* Fs6, which
165 contained 15 additional large scaffolds or contigs with low identity to any other sequences in the
166 dataset (Supplementary Figure 1). We also observed that all samples in the clade containing
167 *F. petroliphilum* and *F. sp. FSSC12* possessed a substantial chromosomal translocation between
168 the two largest core chromosomes (Supplementary Figure 3).
169
170 In contrast to the high synteny and sequence conservation among the core chromosomes, the
171 accessory chromosomes were largely species- or isolate-specific or had much lower levels of
172 sequence identity for the regions that were present (Figure 4). We detected between three and
173 eight accessory chromosomes or contigs >150kb in the *F. keratoplasticum* isolates, and between
174 five and nine in the three *F. petroliphilum* isolates. Many accessory chromosomes were
175 conserved in all isolates of the same species, but not across species. For example, two of the
176 three accessory chromosomes present in *F. keratoplasticum* Fu6 had high identity with the other
177 *F. keratoplasticum* isolates but low synteny with other FSSC isolates (Figure 4). In other cases,
178 accessory chromosomes were isolate-specific, such as chromosome 13 of *F. keratoplasticum*
179 LHS11, which displayed <52.3% identity with any other contig in the dataset. However, we did
180 not identify any accessory chromosomes or contigs in *F. keratoplasticum* that were specific to the
181 clinical strains. We also observed cases where the distribution of accessory chromosomes or
182 contigs did not mirror the core genome-generated phylogeny. For example, four
183 *F. keratoplasticum* isolates (Fu6, LHS11, 2015-087, and 2016-116) possessed one to three
184 scaffolds whose nucleotide identity was higher to accessory genomic regions of *F. petroliphilum*
185 than to the other strains of its own species (Figure 4a). Conversely, F. *petroliphilum* strains 2014-

186 013 and 2015-035 possessed accessory contigs with higher synteny and nucleotide identity to
187 *F. keratoplasticum* than to their own species (Figure 4b).
188
189 The accessory chromosomes of all six FSSC species possessed a significantly higher fraction of
190 repetitive elements, including transposons, compared to the core chromosomes ($p < 0.05$ for all
191 species by Mann-Whitney U test) (Supplementary Figure 2). For example, the median repeat
192 content of the 12 core chromosomes among the five *F. keratoplasticum* isolates was 6.7%,
193 compared to 24.5% for the accessory chromosomes and contigs. The synteny analysis also
194 revealed that the short regions of lower synteny in the core chromosomes were closely associated
195 with the presence of repetitive elements. Altogether, we defined 12 core chromosomes with a
196 high degree of sequence conservation across the FSSC and an assortment of species- and
197 isolate- specific accessory chromosomes that also had a higher content of repetitive elements.
198
199 **57% of intergenic lncRNAs in *F. keratoplasticum* are conserved in *F. petroliphilum***
200 In addition to variation in protein-coding genes, eukaryotic genomes encode myriad long non-
201 coding transcripts (lncRNAs) that are often involved in gene regulation and evolve faster than
202 coding genes [31]. Previous work on lncRNAs in *Fusarium graminearum* have highlighted their
203 roles in regulating sexual reproduction [32] and carotenoid biosynthesis [33], however the lncRNA
204 landscape of the FSSC has not been explored. To this end, we sequenced the transcriptomes of
205 *F. keratoplasticum* NRZ-2014-069 and *F. petroliphilum* NRZ-2014-079 conidia and of mycelia
206 grown under several environmental conditions to capture the breadth of lncRNAs expressed by
207 these two species. Using this data, we identified 2,632 and 2,222 putative lncRNAs for the two
208 species, respectively. Both species shared a comparable ratio of intergenic to antisense lncRNAs,
209 with the majority of lncRNAs being intergenic and only 16% being antisense in both species
210 (Supplementary Figure 4). Intergenic lncRNAs were significantly shorter than antisense lncRNAs
211 and protein-coding genes, with a median length of 795 bp for *F. keratoplasticum* compared to
212 1,231 bp and 1,317 bp for antisense and for protein coding genes, respectively ($p < 0.001$ for all
213 comparisons by Bonferroni-corrected Mann-Whitney U test). Similarly, intergenic lncRNAs in
214 *F. petroliphilum* had a median length of 831 bp compared to 1,352 bp for antisense lncRNAs and
215 1,345 bp for protein coding genes (Supplementary Figure 4).
216
217 To identify conserved intergenic lncRNAs between the two species, we used a combination of
218 reciprocal BLASTn (local alignment), MUMmer (global alignment), and synteny-based
219 approaches. BLASTn identified 395 conserved lncRNA pairs, representing 18% of the intergenic

220 lncRNAs in *F. keratoplasticum*. MUMmer alignment identified 385 pairs, with a 95% concurrence
221 with BLASTn. As lncRNAs are known to have lower levels of sequence conservation compared
222 to protein-coding genes, we also employed a synteny-based approach as in [34]. Conserved
223 lncRNAs were identified between the two species if they were flanked by at least one up- and
224 downstream orthologous genes. This approach identified 1,647 syntenic genomic blocks,
225 containing lncRNAs (Supplementary Figure 5, Supplementary File 3). Taking the union of the
226 three approaches revealed that 57% of the intergenic lncRNAs in *F. keratoplasticum* (1264/2224)
227 were conserved in *F. petroliphilum*, similar to the conservation rate among protein-coding genes
228 in the two species. Notably, all conserved lncRNAs identified through synteny analysis were
229 located on core chromosomes. However, we detected 51 species-specific lncRNAs in
230 *F. keratoplasticum* and 104 in *F. petroliphilum* that were located on accessory chromosomes.
231 Altogether, we defined the lncRNAome of two FSSC species and found that the degree of their
232 conservation was similar to that of protein coding genes.
233
234 ***Fusarium keratoplasticum* genomes are enriched in branched-chain amino acid and**
235 **metallopeptidases, while all FSSC genomes show features of plant pathogenicity**
236 Given the degree of variation in the pan-genome of the FSSC and the accessory chromosomes
237 within and between species, we next performed gene set enrichment analysis to identify Gene
238 Ontology (GO) categories that were over- or underrepresented in species of the FSSC, with a
239 focus on *F. keratoplasticum* and *F. petroliphilum*. The genomes of *F. keratoplasticum* had an
240 expansion of genes involved in branched-chain amino acid GO terms relative to other FSSC
241 genomes, a pathway which is not-conserved in metazoans, but essential to growth and virulence
242 in several human pathogenic fungi [35,36]. We also observed an enrichment in genes annotated
243 in protein ubiquitination and metallopeptidase activity in this species, which serve roles in fungal
244 growth and pathogenicity [37], including tissue invasion [38] and hyphae formation [39]. The
245 genomes o*f F. petroliphilum* had higher numbers of genes with endonuclease activity and
246 glutathione metabolic processes relative to the other FSSC species. The former plays a strong
247 oxidoprotectant role in fungi [40], and the latter of which is involved in defense against reactive
248 oxygen species and the repair of damaged DNA [41].
249
250 Underscoring their evolutionary trajectory as plant pathogens, all FSSC genomes contained an
251 abundance of secreted effectors and genes with GO annotations related to plant virulence. These
252 included similar levels of genes with GO terms related to chitosanase activity, pectinesterase
253 activity, and cellulase activity, all of which are involved in the degradation of plant cell walls

254 [42,43]. All genomes also had between 477-546 secreted effectors. The fact that these genomic
255 features were present at similar levels in plant isolates and human and marine animal isolates
256 suggests that the latter likely retain some degree of plant pathogenicity or, alternatively, that these
257 factors are also important during interaction with higher eukaryotes. Taken together, the gene
258 functional analysis revealed that all FSSC genomes share features associated with plant
259 virulence, while *F. keratoplasticum* and *F. petroliphilum* possessed differential expansion of
260 genes involved in branched chain amino acid metabolism, oxidative stress protection, and DNA
261 repair.
262
263 **Comparison of secondary metabolite biosynthetic gene clusters reveals a core set of**
264 **secondary metabolites in the FSSC**
265 Fungi produce an extraordinary array of secondary metabolites, which play key roles in their
266 interactions with other environmental organisms. They can be used for microbial communication,
267 nutrient scavenging, and even as antimicrobial weapons. These molecules are typically encoded
268 in biosynthetic gene clusters (BGCs), or organized units of multiple genes all involved in
269 metabolite production [44]. *Fusarium* spp, including FSSC, produce a diverse range of
270 compounds, with 162 novel *Fusarium* metabolites described between the years of 2015-2020
271 alone [45]. However, the range of BGC produced by FSSC isolates from human and animal origin
272 has not been studied. Across the 12 genomes, we predicted 42 BGCs and each isolate encoded
273 a median of 33 BGCs (range 32-37). Of the BGCs identified, 55% (23/42) were present across all
274 study strains (Figure 6). Among the BGCs present in all genomes that matched a known product
275 was ochratoxin A, an agricultural contaminant. Among the other core BGCs with known products
276 were squalestatin S1, fusarubin, metachelin C, tryptoquialanine, and lucilactaene, many of which
277 have antimicrobial activity [46–50]. The BGC encoding the antimicrobial and immunosuppressant
278 cyclosporin C was produced by the phylogenetic subclade encompassing *F. keratoplasticum,*
279 *F. falciforme*, and *F. sp. Ph1*, but absent in the other FSSC clades examined.
280
281 The majority of the detected BGCs (24/42) encoded unknown products. One such example was
282 BGC19 (Supplementary Figure 6), a secondary metabolite cluster of the isocyanide class, which
283 are known for their ability to chelate transition metals. This BGC was found in five of the six
284 keratitis isolates (n=2 *F. petroliphilum* and n=3 *F. keratoplasticum*), but none of the other study
285 isolates, including the non-keratitis *F. keratoplasticum* isolates, hinting that it might play a role in
286 virulence in the eye. Also of interest, *F. sp. Ph1* was predicted to produce gibberellin, a BGC
287 typically restricted to the *Fusarium fujikuroi* species complex. Gibberellin is a plant growth

288 hormone and its production by *Fusarium fujikuroi* is considered a major factor in the development
289 of symptoms in rice bakanae disease, which causes up to 50% crop loss [51,52].
290
291 **Minimal transcriptional changes in *F. keratoplasticum* and *F. petroliphilum* at ocular**
292 **infection temperatures**
293 Given the genomic differences observed between *F. keratoplasticum* and *F. petroliphilum* in this
294 study, we wanted to compare the transcriptomes of these two species under several stressors.
295 To this end, we performed RNAseq of mycelia from *F. keratoplasticum* NRZ-2014-069 and
296 *F. petroliphilum* NRZ-2014-079 grown in complete media at 25ºC, under carbon starvation at this
297 temperature, and at elevated temperature mimicking the surface temperature of the eye (34ºC).
298 Additionally, the transcriptome of conidia from each isolate was also assayed. Growth stage
299 (conidia vs. mycelia) had the largest impact on gene expression with 1,185 and 2,017 differentially
300 expressed genes (DEG) in *F. keratoplasticum* and *F. petroliphilum* respectively (Supplementary
301 File 4). Gene set enrichment analysis of GO terms revealed that conidia from *F. keratoplasticum*
302 had an upregulation of DNA-binding proteins and transporters compared to mycelia (Figure 7,
303 Supplementary File 5), while *F. petroliphilum* showed an upregulation in genes cell wall
304 organization and structure.
305
306 Relative to complete media, mycelia grown under carbon starvation resulted in 357 DEGs in
307 *F. keratoplasticum* and 850 DEGs in *F. petroliphilum* (Supplementary File 4). Both species
308 showed an upregulation in nucleotide binding and ATP binding. However, only F. petroliphilum
309 showed an upregulation in genes with metal ion binding activity (Figure 7, Supplementary File 5).
310 Unexpectedly, elevated temperature (34°C) resulted in very modest change to the transcriptome,
311 with 95 genes differentially expressed in *F. petroliphilum* and zero in *F. keratoplasticum*
312 (Supplementary File 4). This suggests that both species are more than capable of growing at
313 temperatures experienced during mammalian association or infection
314
315 Consistent with earlier studies [32], both intergenic and antisense lncRNAs identified in this study
316 had significantly lower expression relative to protein-coding genes (Mann-Whitney U test with
317 Bonferroni correction, $p < 0.0001$ for both lncRNA classes in *F. keratoplasticum* and
318 *F. petroliphilum* (Supplementary Figure 4b). However, we did identify 169 lncRNAs in
319 *F. keratoplasticum* and 190 in *F. petroliphilum* that were uniquely differentially expressed in
320 conidia relative to mycelia, suggesting they may play a role in asexual development
321 (Supplementary File 4). Overall, the transcriptional changes of *F. keratoplasticum* and

322 *F. petroliphilum* during carbon starvation and elevated temperature were minimal, indicating that
323 neither infection-relevant condition was a significant stressor to either species.
324

325 Discussion

326 In this study we generate the highly contiguous genomes for six isolates of the FSSC from clinical
327 origin, including the first three genomes available for *F. petroliphilum*. We show that German
328 keratitis isolates of *F. keratoplasticum* are not phylogenetically distinct from Asian isolates from
329 marine animal origin. We also reveal a low level of core gene conservation among isolates of the
330 FSSC, including pronounced variation in lncRNAs, secondary metabolites, and the presence of
331 accessory chromosomes with isolate- and species- specific genes and a higher repeat content.
332 Finally, transcriptional profiling of *F. keratoplasticum* and *F. petroliphilum* at temperatures
333 mimicking the human eye revealed minimal changes, suggesting that neither species experiences
334 infection temperatures as a significant stress.
335
336 In line with previous work, our study highlights the considerable genomic flexibility of *Fusarium*
337 spp. However, we find that the level of gene conservation within the FSSC (41%) is even lower
338 than the *Fusarium oxysporum* species complex (FOSC), whose core genome fraction has been
339 reported at 51-53% [53,54]. The proportion of core genes in the FSSC much also lower than
340 environmental human fungal pathogens such as *Aspergillus fumigatus* (69%) and *Cryptococcus*
341 *neoformans* (80%), but similar to another multi-kingdom pathogen, *Aspergillus flavus* (41-58%)
342 [55–58]. This may indicate that the FSSC and other multi-kingdom pathogens occupy a broader
343 niche space and the larger pan-genome has evolved to cope with a more variable set of
344 environmental conditions, as has been found in bacteria [59]. It is also tempting to hypothesize
345 that this expanded genomic repertoire and the resulting flexibility underpin the ability of these
346 organisms to infect multiple kingdoms, with very different metabolic environments and immune
347 systems; a point that requires further investigation. On the other hand, comparing pan-genomes
348 across studies is not without its own limitations [60]. It is complicated by differences in the number
349 of isolates between studies, the pangenome methodology (*e.g.* how orthologous genes are
350 identified and what fraction of isolates must contain a gene for it to be considered 'core'), and
351 whether the study defines taxonomic groups using *forma specialis* or species complex definitions.
352

353 Previous work comparing a phytopathogenic *F. oxysporum* strain with two *F. oxysporum* isolates
354 from human disease isolates from humans suggested that the latter may have unique accessory
355 chromosomes. While all FSSC genomes we examined had accessory chromosomes with higher
356 repeat content, as has been observed for many *Fusarium* species complexes [6,9,61,62], we did
357 not find clear evidence of accessory chromosome content differed from that of plant or other non-
358 clinical sources. We did not find any accessory scaffolds unique to human isolates, with the
359 exception of accessory chromosomes found only in F. *petroliphilum*. However, since all isolates
360 of this species in our dataset were of human origin, they lack a phylogenetically matched
361 comparator. As sequencing becomes even more accessible, we expect that larger datasets of
362 isolates from plant and human disease will help to further investigate the important question of
363 whether there are specific FSSC genomic elements associated with human infection.
364
365 The genomes of *F. keratoplasticum* were enriched in various genes related to branched-chain
366 amino-acid metabolism, carbon-nitrogen hydrolysis, and acetolactate activity, the latter of which
367 catalyzes the first step of the BCAA biosynthetic pathway (BCAA) [63]. BCAA synthesis is
368 conserved in plants, bacteria, and fungi, but not in higher metazoans [64], and has been shown
369 to be essential for fungal growth and virulence [35,36], making it a potential target for antifungal
370 drug development [65]. *F. keratoplasticum* also possessed an enrichment in genes relating to
371 oxidoreductase activity and base-excision repair, both potentially relevant to the prevention and
372 repair of DNA damage as a result of host reactive oxygen species [41]. In comparison,
373 *F. petroliphilum* was enriched in glutathione metabolism and sulphuric ester hydrolysis.
374 Glutathione serves as as a crucial oxidative stress protectant in fungi, the biosynthesis of which
375 has in *F. graminearum* been directly linked to sulphur metabolism [40], while the expansion in
376 helicase and endonuclease genes could be relevant for the repair of oxidatively damaged DNA.
377 Although these gene functional enrichments shed light on the different mechanisms by which
378 these pathogens may cause disease, or resist and repair damage resulting from ROS-mediated
379 host defenses.
380
381 The identification of BGCs for squalestatin, fusarubin, oxyjavanicin, metachelin, and
382 sansalvamide in all isolates is consistent with previous work exploring the secondary metabolite
383 diversity of the FSSC using different species [66], suggesting that these are core secondary
384 metabolites of the complex. However, this study identified the potential siderophores ferricrocin
385 and malonichrome, as additional core secondary metabolites, while these BGCs were not
386 predicted for any isolates in our dataset. Conversely, all isolates analyzed in this study encode

387 ochratoxin A, which was not reported in their study. Also intriguing was the observation that
388 BGC19, an unknown natural product of isocyanide class, was found in five of the six clinical
389 isolates but none of the plant or marine animals isolates. Isocyanides have been described in
390 virulence factors in both bacterial and fungal pathogens of insects and humans [67,68]. These
391 natural products are involved in copper uptake, a micronutrient that is often limited during
392 infection, and isocyanide BGCs are upregulated upon copper starvation [69]. We speculate that
393 this BGC might play a role in copper scavenging during keratitis, but this remains to be
394 experimentally tested.
395
396 lncRNAs are increasingly recognized to play important roles in regulating pathogenicity and
397 virulence in fungal pathogens of plants and humans [70–72]. However, they have not been
398 defined in the FSSC or any *Fusarium* species that infects humans. The list of lncRNAs described
399 in this study will be an important starting point to determine their roles in regulating basic cellular
400 processes and potentially plant and human virulence. The fact that many lncRNAs are located on
401 core chromosomes and are conserved between species indeed suggests that they are likely to
402 have conserved and essential functions. On the other hand, the non-conserved lncRNAs on
403 accessory chromosomes will be exciting to explore as potential virulence factors in plants and
404 animals.
405
406 In closing, the FSSC is an excellent system to study the molecular mechanisms and potential
407 evolutionary trade-offs associated with multi-kingdom pathogenicity. This is increasingly important
408 as climate change is expected to increase the number of environmental fungal species that can
409 grow at human body temperatures, potentially opening the door for other phytopathogens to
410 become human pathogens. The genomes and strains provided here are an excellent resource for
411 unraveling multi-kingdom virulence, such as comparing the transcriptomes of these isolates
412 during plant and mammalian infection to identify *in planta* virulence strategies that are repurposed
413 for human pathogenesis.

414 Materials and methods

415 **Strains analyzed in this study**

416 Six isolates originating from cases of keratitis collected by the German National Reference Centre
417 for Invasive Fungal Infection (NRZMyk) were sequenced in this work. Prior to whole genome
418 sequencing, species identification was performed by sequencing of the TEF-1α gene as in [28].

419 The additional FSSC genomes analyzed in this study originate from plant and aquatic animal
420 hosts and were generated in [6]. The accession numbers of the FSSC genomes analyzed in this
421 study and those included in the phylogeny are available in Supplementary File 1.
422
423 **Nucleic acid isolation and sequencing**
424 For the isolates sequenced in this study, mycelia were grown shaking at 30°C in complete media.
425 Mycelia were ground in liquid nitrogen, heat shocked for 1 minute at 95°C, and cellular debris
426 precipitated using steps and reagents from the MasterPure Yeast DNA Purification Kit (Lucigen).
427 DNA was finally purified using the Monarch Genomic DNA Purification Kit (New England Biolabs)
428 for binding and elution.
429
430 **Genome assembly and annotation, including lncRNA annotation**
431 The genomes generated in this study were assembled using the PacBio long reads and Flye
432 v2.9.1. The draft assemblies then underwent two rounds of polishing using Pilon v1.24 and the
433 strain-matched Illumina data. Assemblies were then scaffolded using RagTag v2.1.0 with
434 minimum confidence score set to 0.5, minimum location confidence score set to 0.35, and the
435 minimum orientation confidence score set to 0.6. The *F. keratoplasticum* genomes were
436 scaffolded using *F. keratoplasticum* LHS11 as a reference as it was the most contiguous genome
437 available for the species. As there was no existing chromosome-level genome available for
438 *F. petroliphilum*, their genomes were scaffolded using LHS14 because it resulted in the highest
439 mean location and orientation confidence scores of the FSSC isolates analyzed in this study.
440 Sequence contamination was identified and removed from genome assemblies using Foreign
441 Contaminant Screen (FSC) v0.4.0.
442
443 Before annotation, assemblies were cleaned to remove contigs shorter than 1kb and those with
444 >95% percent identity and >95% coverage with other contigs in the assembly. A custom FSSC
445 repeat library was generated using RepeatModeler v2.0.4 and the FSSC genome assemblies
446 analyzed in this study. RepeatMasker v4.1.2 and the FSSC repeat library was then used to identify
447 and mask repetitive regions. Genomes were annotated using Funannotate v1.8.9 and all
448 subsequent software referenced uses the versions packaged in this containerized version. Briefly,
449 an *ab initio* FSSC gene prediction model was trained using Augustus and the assembled
450 transcriptome of *Fusarium sp. FSSC11* available from Mycocosm as the RNAseq data generated
451 in our study was not available yet. Gene prediction was then performed using EvidenceModeler
452 incorporating predictions from Augustus, GeneMarkES, Snap, and GlimmerHMM. tRNAs were

453 predicted using tRNAScanSE. When strain-specific RNAseq data became available later, gene
454 models were corrected using the update function of Funannotate, which also added 5' and 3'
455 UTRs to the gene predictions. Functional annotations were assigned using PFAM, Eggnog-
456 mapper, InterProScan, CAZYmes, BUSCO (dikarya_odb9 lineage), SignalP and KEGG.
457 EffectorP 3.0 was used to predict secreted effectors. To ensure consistent genome annotations
458 across the study dataset, the genome assemblies from Hoh *et al.* 2022 [6] were downloaded from
459 NCBI and annotated as above, with the Augustus model trained using the RNAseq data
460 generated in the same publication. Genome completeness in Table 1 was determined using
461 BUSCO v5.2.2 and the hypocreales_odb10 lineage set.
462
463 To annotate lncRNAs de novo, we used the RNA-seq data from *F. ketatoplasticum* NRZ-2014-
464 069 and *F. petroliphilum* NRZ-2014-079 generated in this study. Genome-guided transcriptome
465 assembly was performed using Stringtie v.2.2.1 [73] and the individual transcript assemblies from
466 each species unified into a single transcript assembly using the Stringtie -merge option. To identify
467 novel transcripts, assembled transcripts were compared against protein-encoding transcripts
468 using Gffcompare v0.12.6 and novel transcripts >200 bp selected. Where applicable, the longest
469 isoform was chosen utilizing gtf2gtf from CGAT apps v0.6.0 [73,74]. Coding potential prediction
470 was performed using two software tools: CPC2 v3 [75], which was executed against the UniProt
471 database (https://www.uniprot.org/download), and Feelnc v0.2 [76]. Transcripts were classified
472 as lncRNAs when identified as non-coding transcripts by both tools. lncRNA transcripts were
473 categorized as intergenic (class code 'u') or antisense (class code 'x') based on the Gffcompare
474 output. To evaluate the expression levels of lncRNAs in Supplementary Figure 4, read counts for
475 lncRNA and protein-coding transcripts were calculated using Featurecounts v2.0.1 [77].
476 Subsequently, count data was normalized by transcript length and library size, yielding transcripts
477 per million (TPM) values.
478
479 **Whole-genome phylogeny**
480 The *Fusarium* spp phylogeny was inferred 1,164 single copy orthologs identified in all species
481 genomes using OrthoFinder v2.5.4. Amino acid sequences for each orthogroup were aligned
482 using MUSCLE v3.8, the alignments trimmed using ClipKIT v1.3.0, and then concatenated into
483 an alignment of length 614,941. The phylogeny was constructed using IQ-TREE v2.2.0.3.
484 ModelFinder within IQ-TREE was used to identify JTT+F+I+GR as the best fitting substitution
485 model based on Bayesian information criterion (BIC). Bootstrapping was performed using 1,000
486 ultrafast bootstraps and all branches were supported by values of >98%.

487

488 **Pan-genome analysis and identification of core vs. accessory chromosomes and contigs**

489 GENESPACE v1.3.1, incorporating Orthofinder v2.5.4 and MCScanX v01.11.2022, was used for
490 synteny-guided orthologous gene identification and their presence-absence between strains.
491 Heap's Law calculation was done using the MicroPan v2.1 R package using 500 permutations.
492 Chromosomes and contigs were defined as core and accessory based on alignment using the
493 nucmer tool of MUMmer v4.0.0. For these analyses, *F. keratoplasticum* Fu6, was used as the
494 reference query. Core chromosomes were identified using a cutoff of >92.5% mean identity
495 across a minimum alignment length of 1.75 Mb in all strains and were in agreement with those
496 identified in Hoh et al. 2022 [6]. The remaining chromosomes and contigs, which were not found
497 in all 12 study strains or showed <92.0% identify, were designated as accessory.
498

499 To compare the intergenic lncRNAs of *F. keratoplasticum* and *F. petroliphilum*, we used NRZ-
500 2014-069 and NRZ-2014-079 as representative examples of each species. Cross-species
501 lncRNA conservation was assessed using three approaches: BLASTn (local alignment), MUMmer
502 v.4.0.0 (global alignment) and orthology-based synteny. First, we generated two FASTA files
503 containing the sequences of intergenic lncRNA in *F. keratoplasticum* and *F. petroliphilum*. The
504 pairwise comparison of reciprocal hit using BLASTn and dnadiff tool from the MUMmer4 suite [78]
505 were filtered to get those with e-value > 1e-03 and a coverage > 50%. In parallel, a synteny based
506 orthology approach developed and validated in Pegueroles *et al*. 2019 [34] was performed. We
507 implemented the custom pipeline with modification to fit the analysis for two species instead of
508 four, and it was executed using the parameters '3 3 1'. This configuration involved analyzing three
509 protein-coding genes on either side of a specific lncRNA, requiring at least three common genes
510 for each pairwise comparison between species, and ensuring a minimum of one common gene
511 on each side of the lncRNA. To visualize the alignment coordinates, we used Circos v.0.69 to plot
512 the 1-to-1 coordinates (coordinates available in Supplementary File 3).
513

514 **Functional enrichment annotation and BGC analysis**

515 For determining the functional differences between FSSC genomes, genes were converted into
516 matching ontology terms and summed for each strain. ANOVA and post-hoc Dunnett test were
517 then applied to identify annotations over- or under-represented in a single species, using a
518 Bonferroni-correct p-value cutoff of 0.05. For visualization purposes per-species expression
519 values were normalized to z-scores, representing the number of standard deviations above/below
520 the mean.

521
522 Biosynthetic gene clusters were predicted for each isolate genome using the fungal version of
523 antiSMASH v7.0, and unknown metabolites were grouped together across isolates using BiG-
524 SCAPE v1.1.5. For each predicted BGC cluster, percent identity was calculated via BLASTp.
525
526 **RNA isolation, sequencing, and transcriptome analysis**
527 Two representative keratitis samples, *F. keratoplasticum* NRZ-2014-069 and *F. petroliphilum*
528 NRZ-2014-079 were selected for transcriptomic analysis under a set of five growth conditions:
529 mycelia grown for 12h in complete media (5.0 g/L ammonium sulfate + 10.0 g/L glucose) at 25°C,
530 carbon-starvation media (0.1 g/L glucose), elevated temperature (34°C), and ungerminated
531 conidia harvested after 3 days' growth at 28°C. RNAs from two replicates were isolated using a
532 phenol-chloroform protocol. Briefly, harvested mycelia were pelleted and bead-beaten in a mix of
533 lysis buffer and acid phenol:chloroform:isoamyl alcohol (PCI). Lysate was centrifuged, upper
534 phase transferred, then centrifuged again in fresh PCI. Upper phase was incubated at -20 degrees
535 in sodium acetate and 2-propanol, pelleted, then twice ethanol-washed. The washed pellet was
536 air-dried and resuspended in RNAse-free dH_2O, then stored at -80 degrees.
537
538 Illumina TruSeq stranded library preparation and sequencing was performed by the University of
539 Würzburg Core Unit Systems Medicine on the NextSeq platform. Raw FastQ data files were
540 trimmed via TrimGalore v0.6.10, quality threshold $\leq$20. Filtered reads were mapped using STAR
541 v2.7.11b, counted using RSEM v1.3.3 and differentially expressed genes were identified using
542 the intersection of DEGs identified by DESeq2 v1.40.2, Limma v3.56.2, and EdgeR v3.42.4, with
543 DEGs that met the threshold of FDR p-adj < 0.01 for all callers being retained.
544
545 For the functional analysis of differentially expressed genes, gene set enrichment analysis was
546 conducted using the eponymous software GSEA (v.4.3.2) from Broad Institute. GSEA was run in
547 preranked mode, analyzing a gene list of differentially-expressed genes ranked by expression
548 log-fold change, against a background of every gene ontology term in the genome. Enriched or
549 depleted terms were filtered by FDR < 0.25, as recommended by Broad Institute in their
550 documentation.
551
552 **Acknowledgements**
553 Computational work in this study was performed, in part, on the HPC cluster of the Friedrich
554 Schiller University.

Funding

This project was supported by The Federal Ministry for Education and Science (Bundesministerium für Bildung und Forschung, BMBF, bmbf.de) within the framework of InfectControl 2020 project FINAR 2.0 (grant no. 03ZZ0834) and by the Interdisciplinary Center for Infection Research (IZKF; ukw.de) at the University Hospital Würzburg (project D-420). AEB and MP are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, dfg.de) under Germany's Excellence Strategy – EXC 20151 – Project-ID 390813860). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data and regent availability.

The strains sequenced in this study are deposited in and publicly available from the Jena Microbial Resource Collection (JMRC). The PacBio, Illumina WGS, and RNAseq data generated in this study are deposited in the NCBI SRA under BioProject PRJNA1137914. The additional *Fusarium* genomes analyzed in this study are publicly available under the accession numbers listed in Supplementary File 1.

Author Contributions

AEB, RM, and OK conceptualized the study. AZ and RM performed experimental investigations. PJTB, LMF, MP, and AEB performed formal analysis. PJTB, LMF, and AEB were responsible for visualization and writing the original draft. All authors were involved in reviewing and editing the manuscript.

578 # Tables & Figures

	Scaffolds (n)	N50 (Mb)	Genome Size (Mb)	Protein coding genes	BUSCO completeness	Repeat content	Isolation Source	Reference
***Fusarium falciforme* Fu3**	14	3.88	53.6	15,051	99.1%	10.1%	Hawksbill sea turtle	[6]
***Fusarium keratoplasticum* Fu6**	25	4.21	49.5	14,050	99.0%	7.1%	Hawksbill sea turtle	[6]
***Fusarium keratoplasticum* LHS11**	26	3.50	53.5	15,777	98.9%	7.2%	Harbor seal	[6]
***Fusarium keratoplasticum* NRZ-2014-069**	70	4.16	49.6	14,398	98.9%	4.8%	Keratitis (human eye)	This study
***Fusarium keratoplasticum* NRZ-2015-087**	53	4.02	50.6	14,387	98.9%	7.1%	Keratitis (human eye)	This study
***Fusarium keratoplasticum* NRZ-2016-116**	55	4.24	50.5	14,979	98.9%	6.5%	Keratitis (human eye)	This study
***Fusarium petroliphilum* NRZ-2014-013**	120	3.98	51.2	15,060	99.1%	6.7%	Keratitis (human eye)	This study
***Fusarium petroliphilum* NRZ-2015-035**	143	3.34	48.0	14,212	98.5%	4.5%	Keratitis (human eye)	This study
***Fusarium petroliphilum* NRZ-2014-079**	95	3.33	48.5	14,541	99.0%	4.8%	Keratitis (human eye)	This study
***Fusarium sp. FSSC12* LHS14**	35	3.82	56.3	14,512	99.1%	11.2%	Roughtail stingray	[6]
Fusarium sp. Ph1	24	3.30	51.5	14,109	99.0%	8.6%	Orchid flower	[6]
***Fusarium vanettenii* Fs6**	39	3.36	72.9	16,755	99.1%	14.2%	Pea plant	[6]

579 **Table 1: Genomic characteristics of FSSC isolates included in this study.**
580 BUSCO completeness quantified as the percent of BUSCOs identified as single copies using the
581 hypocreales_odb10.2019-11-20 ortholog set.

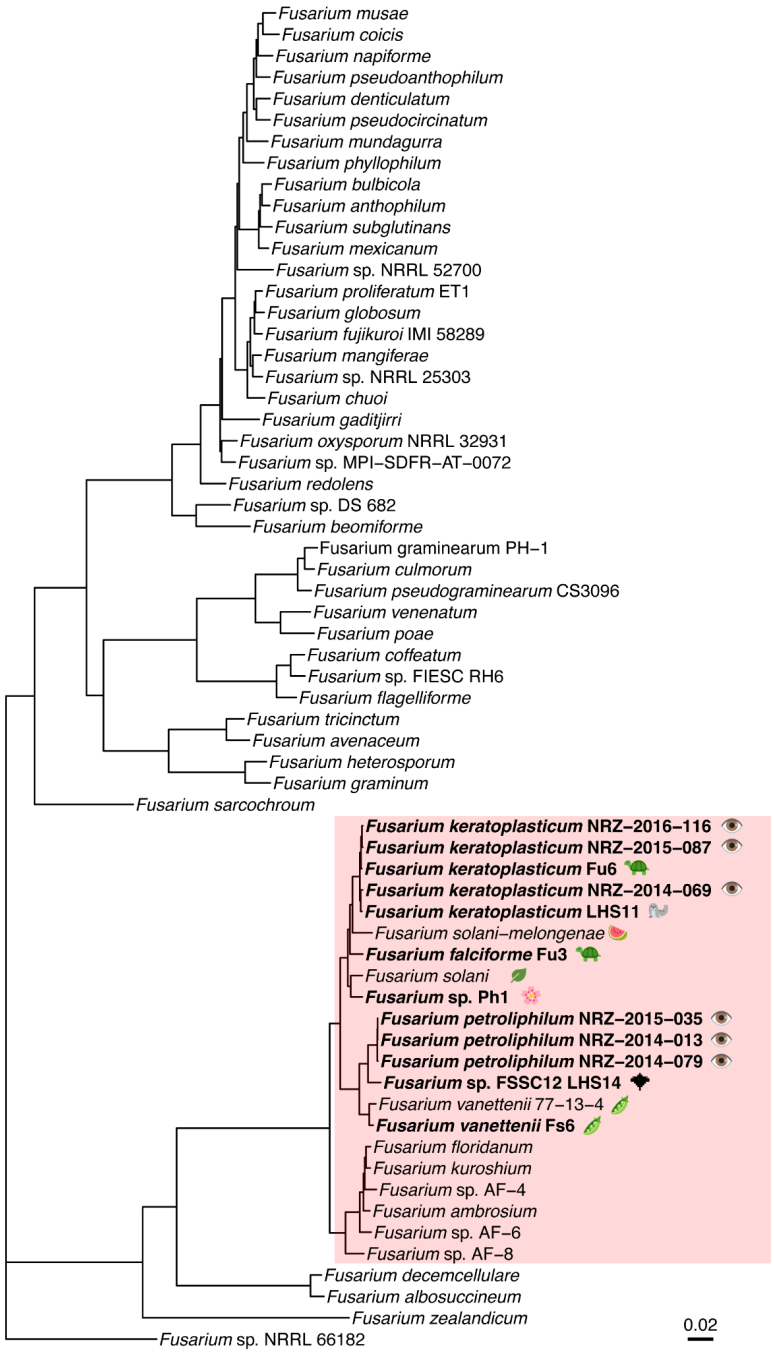


Figure 1. Phylogenetic reconstruction of *Fusarium* spp, including the FSSC isolates analyzed in this study

Phylogeny constructed from an amino acid sequence of 1,866 single copy orthologs. Pink box indicates the *Fusarium solani* species complex. The strains with highly contiguous long-read genome assemblies included in this study are indicated in bold. Emoji indicates the source for the FSSC isolates that belong to the same clade as the study dataset members. All branches supported by ultrafast bootstrap values >98%.

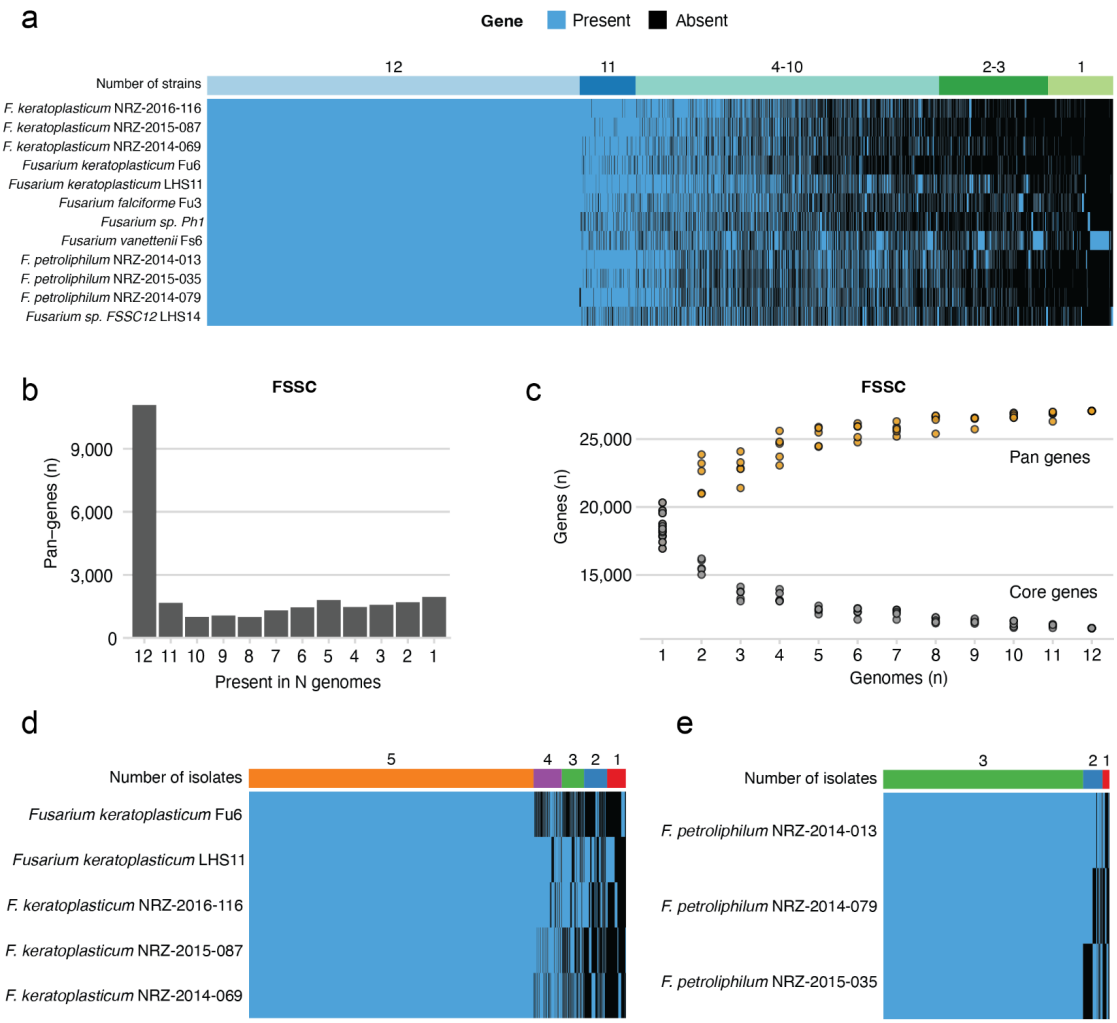


Figure 2. The pan-genome of the FSSC, *Fusarium keratoplasticum*, and *F. petroliphilum*. (a, d, e). Pan-gene presence-absence matrix of FSSC (a), *F. keratoplasticum* (d), and *F. petroliphilum*, where gene presence is indicated with blue and gene absence in black. Colored bars at the top of the matrices indicate the number of isolates each pangene was present in. (b) Histogram of pangene frequency in the FSSC, calculated as the number of genomes in which each pan-gene is present. (c) Pan-gene accumulation plot quantifying the relationship between the number of genomes analyzed and the number of pan-genes identified (orange) and the fraction of which were core (grey). Dots represent the individual values from 24 permutations.

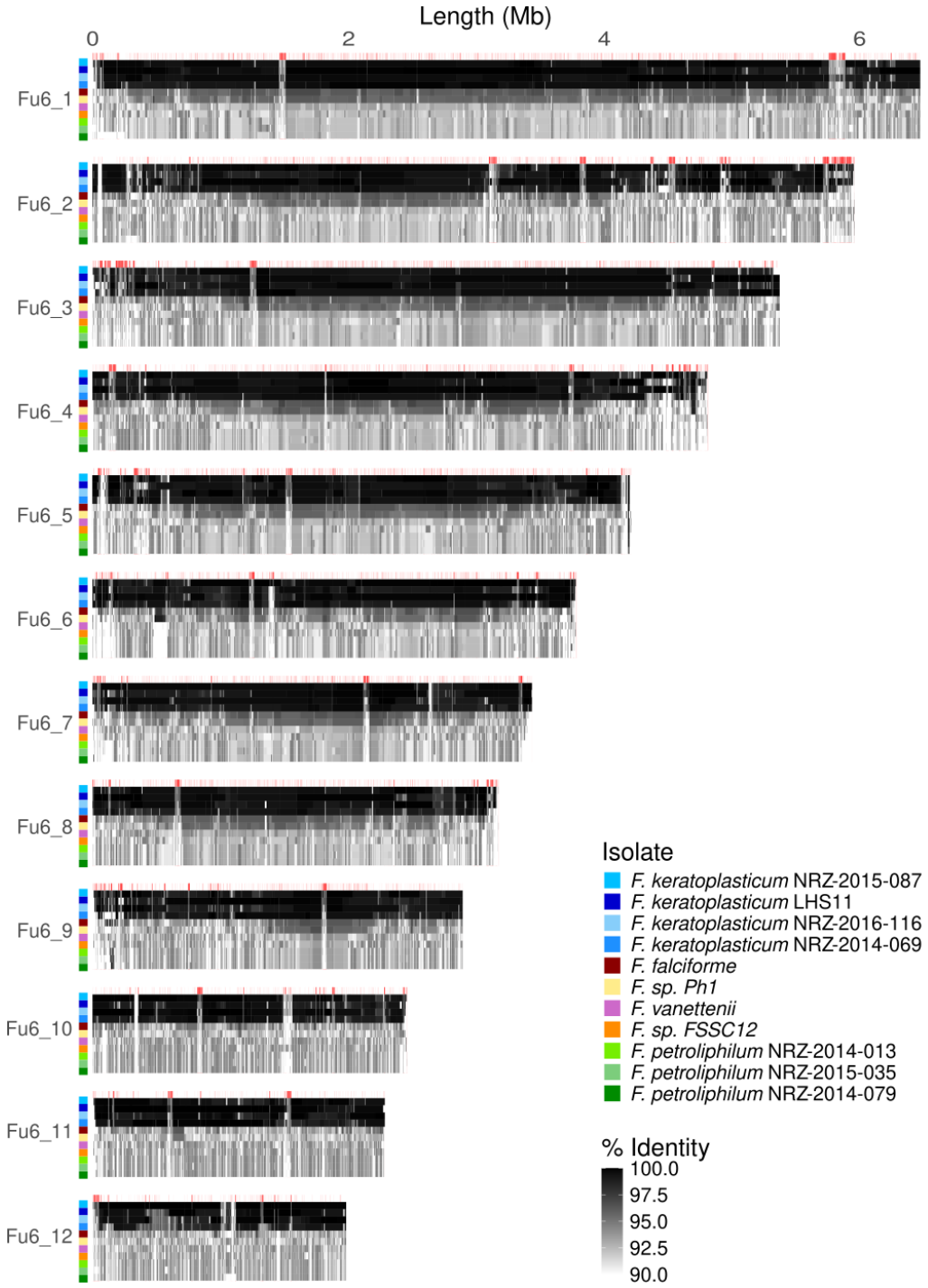


Figure 3. Sequence conservation of core chromosomes of the FSSC. Gray-scale fill indicates the nucleotide percent identity between the core chromosome sequence of *F. keratoplasticum* Fu6 and other FSSC members, which are indicated with the colored boxes. Red lines indicate presence of transposable and other repetitive elements.

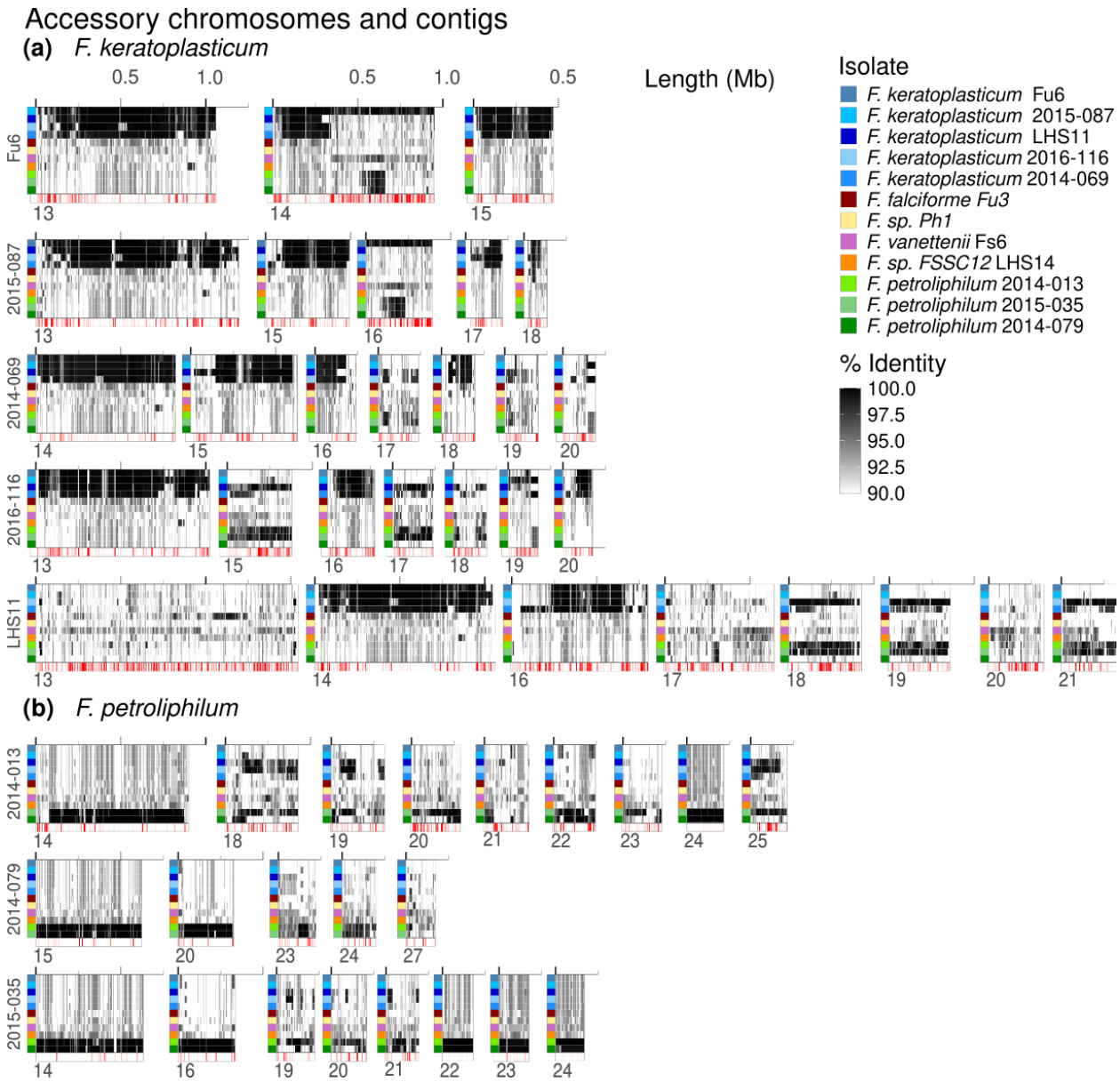


Figure 4. Conservation of accessory chromosomes and contigs from *Fusarium keratoplasticum* and *Fusarium petroliphilum*. (a-b) Accessory chromosomes and contigs of (a) *F. keratoplasticum* or (b) *F. petroliphilum* strains and their sequence conservation with other FSSC isolates. Shading indicates percent identity across the syntenic block between the indicated isolate and accessory chromosome/contig indicated, versus the best matching chromosome or contig from the other isolates. For simplicity, the subject (reference sequence) for each comparison is omitted from the plot as it would be a solid black band. Red lines indicate the location of transposable elements and other repetitive elements. Only genomic fragments >150 kb are displayed.

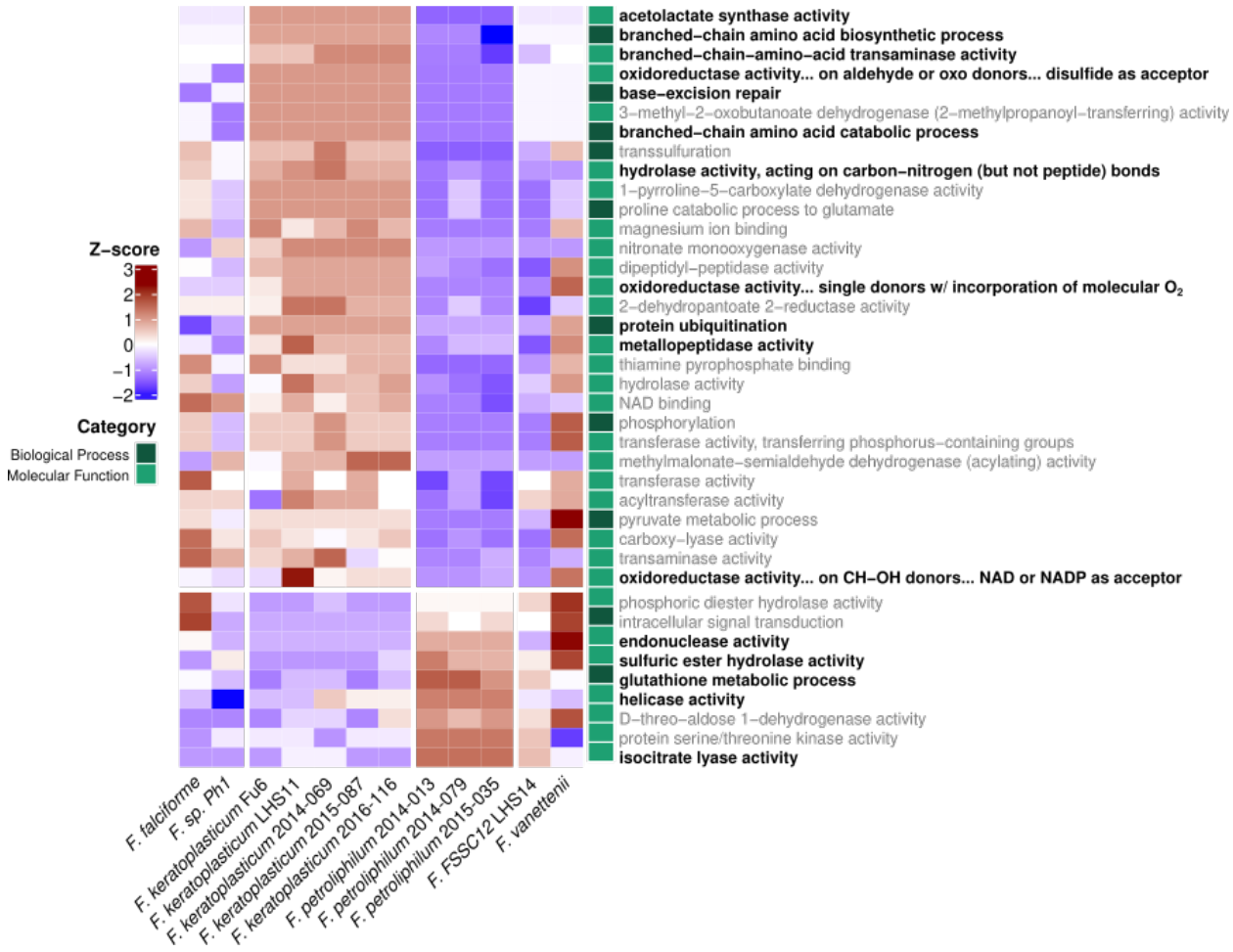


Figure 5. Heatmap of over and underrepresented gene ontology (GO) terms in *Fusarium keratoplasticum* and *petroliphilum* compared to other members of the FSSC. Significantly over- or under-represented GO terms for *F. keratoplasticum* and *F. petroliphilum* identified by t-test and Bonferroni-corrected $p < 0.05$. Heat map fill represents the Z-scaled number of genes per genome annotated with each term. GO terms of particular interest or highlighted in the text are written in bold.

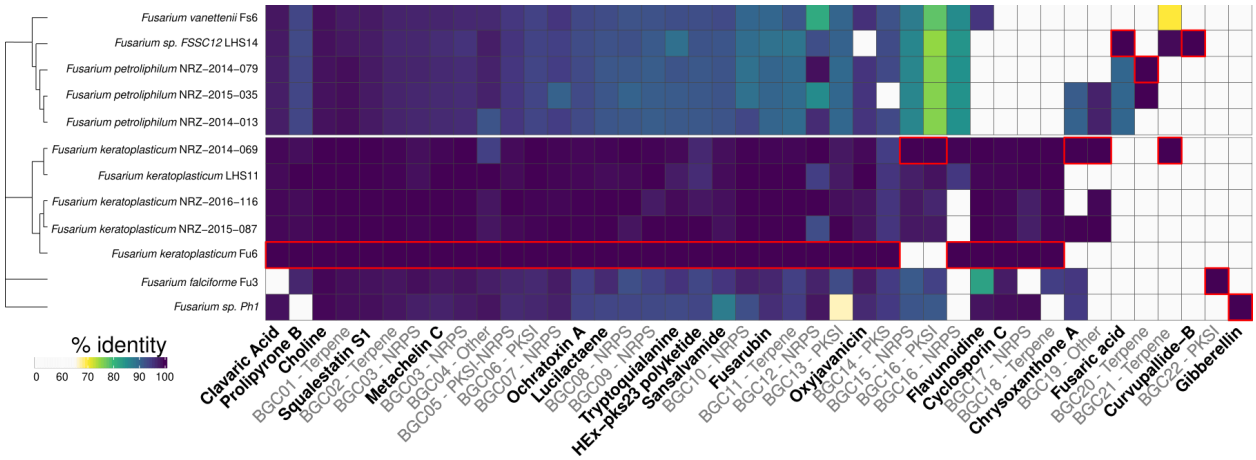


Figure 6. Comparison of secondary metabolite biosynthetic gene clusters reveals a core set of secondary metabolites in the FSSC. BGCs with a known predicted product are indicated in bold and those with unknown products labeled with their natural product class. Percent identity was calculated using amino acid sequence of the backbone enzymes. Red boxes indicate the isolate whose sequence was used as the query when calculating percent identity. The dendrogram on the left represents the phylogenetic relationship of the isolates.

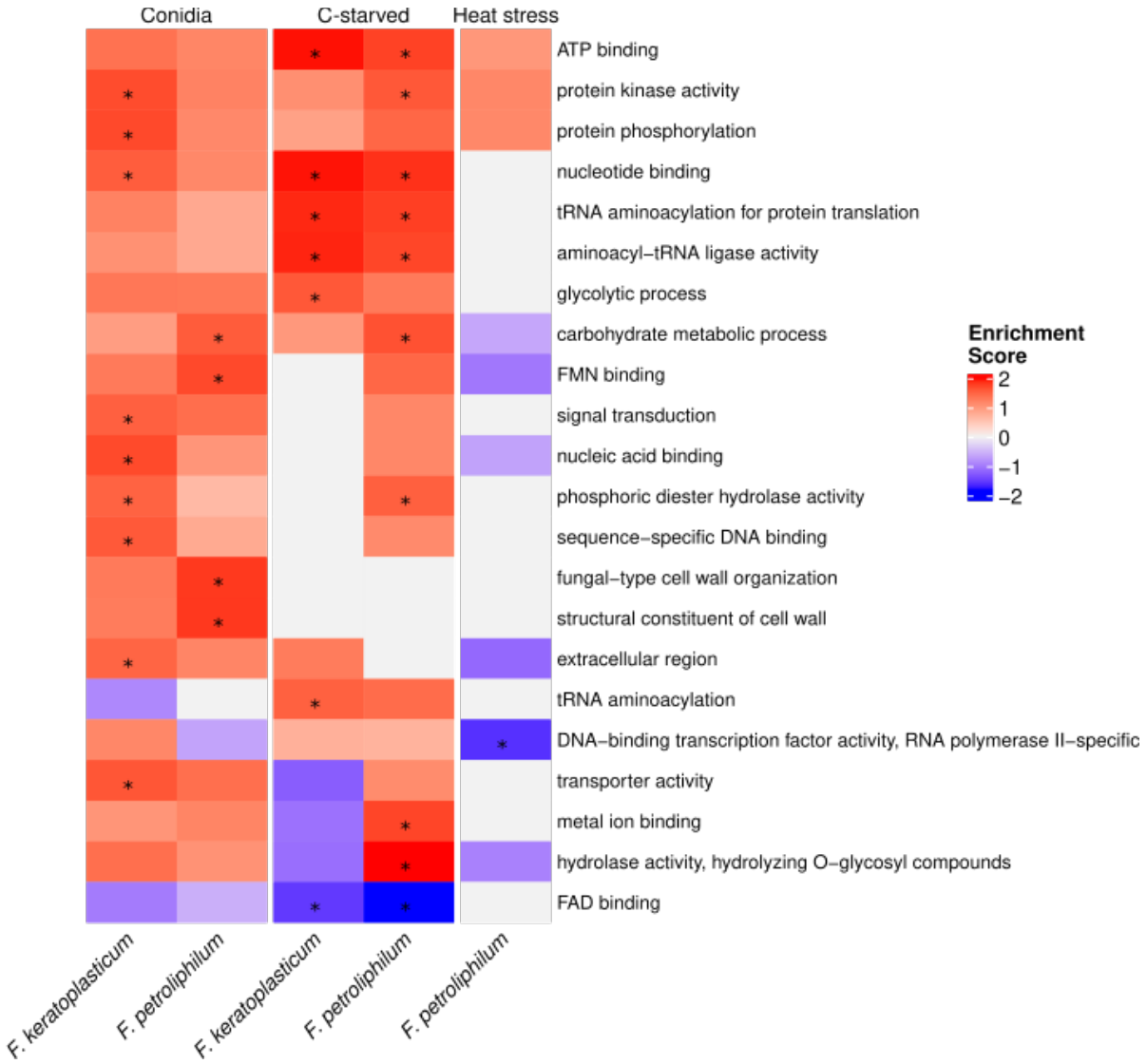
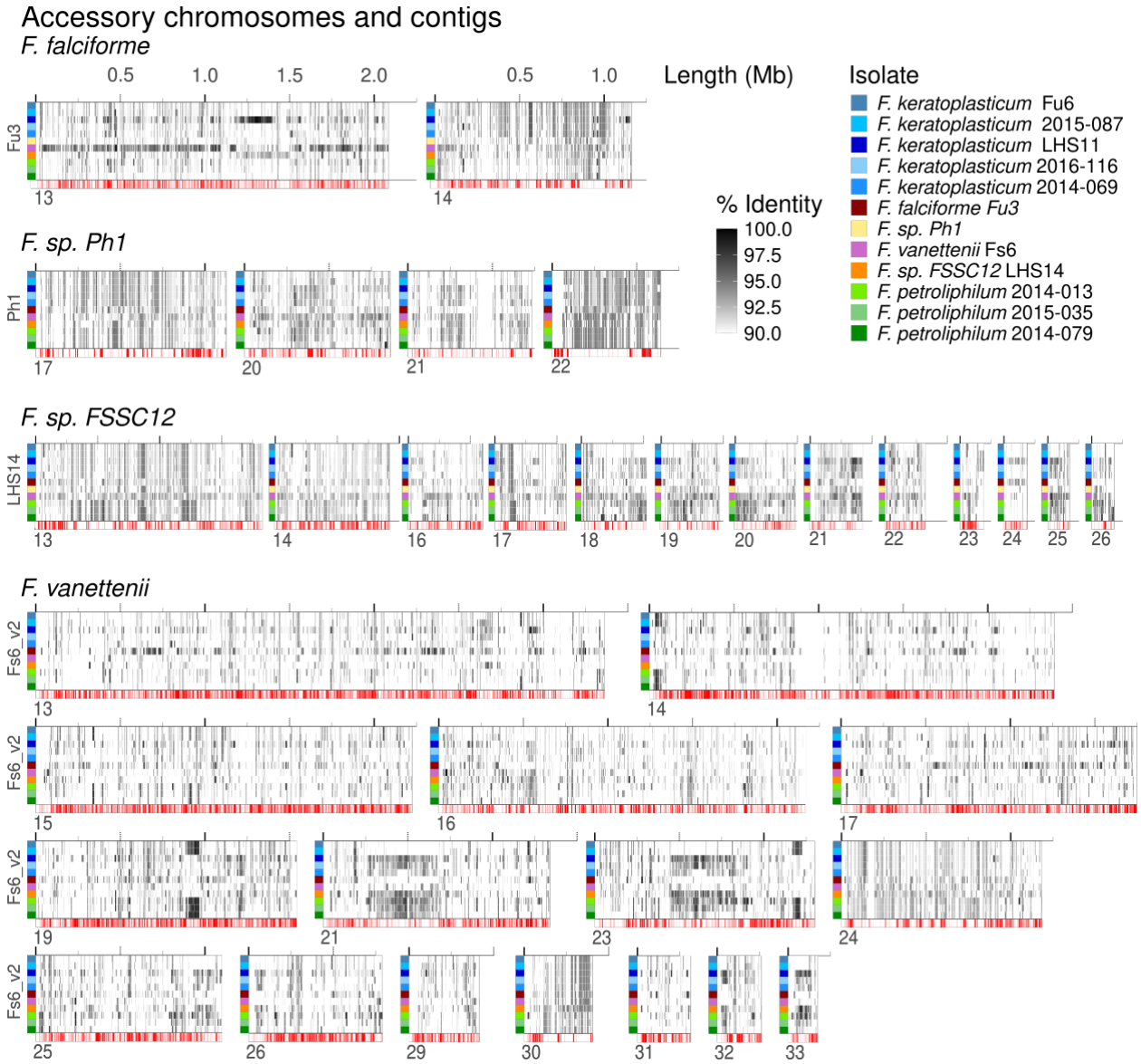
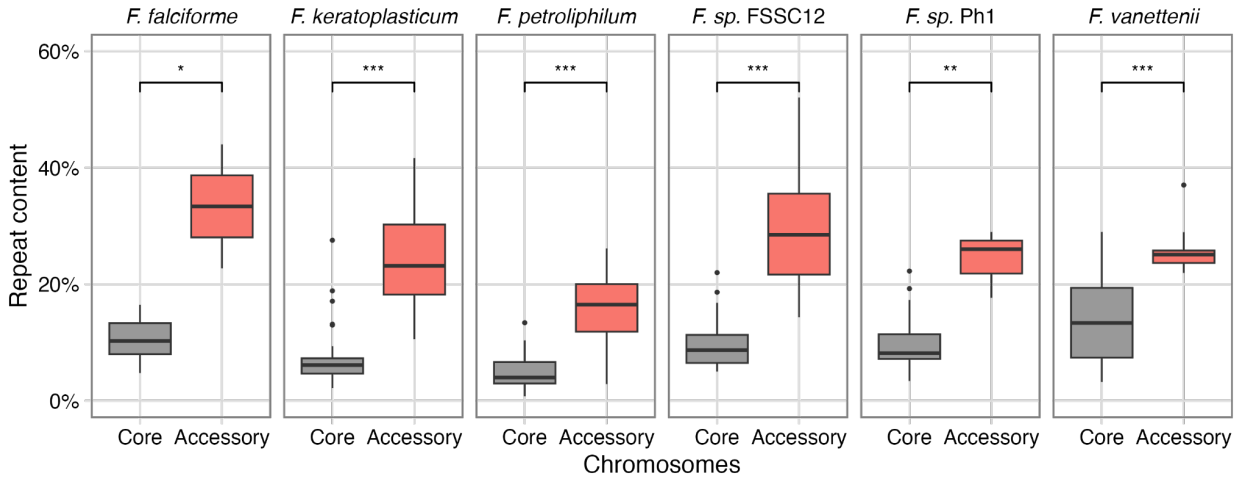


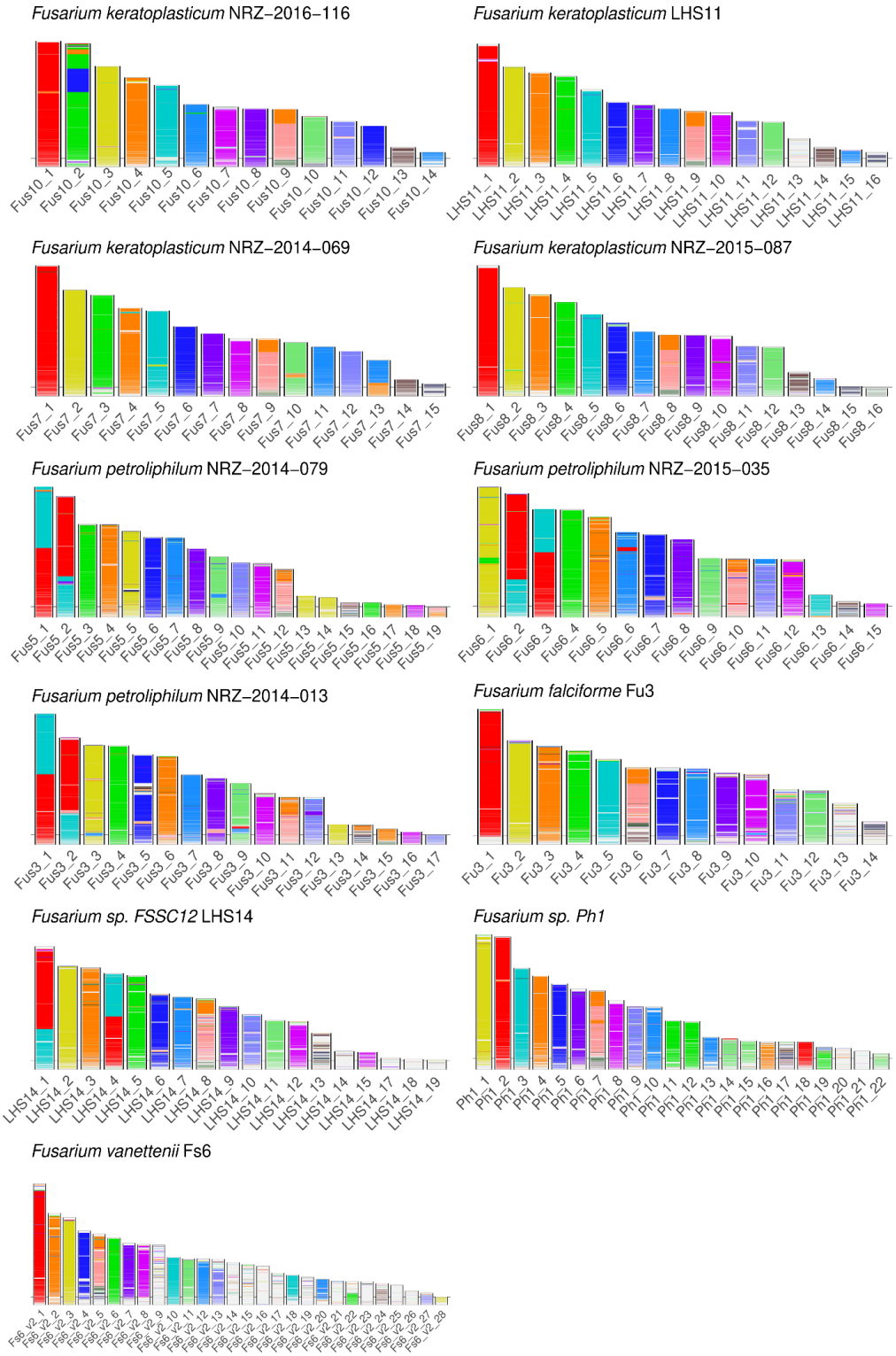
Figure 7. Comparison of enriched and depleted Gene Ontology terms in the transcriptomes of *F. keratoplasticum* and *F. petroliphilum*. Experiments were conducted using NRZ-2014-069 for *F. keratoplasticum* and NRZ-2014-079 for *F. petroliphilum*. Differential expression for the conditions shown was calculated in comparison to mycelia grown in complete media at 25 degrees. Color indicates enrichment score calculated by gene set enrichment analysis. Stars indicate strains for which $q < 0.25$, as recommended by Broad Institute for GSEA.



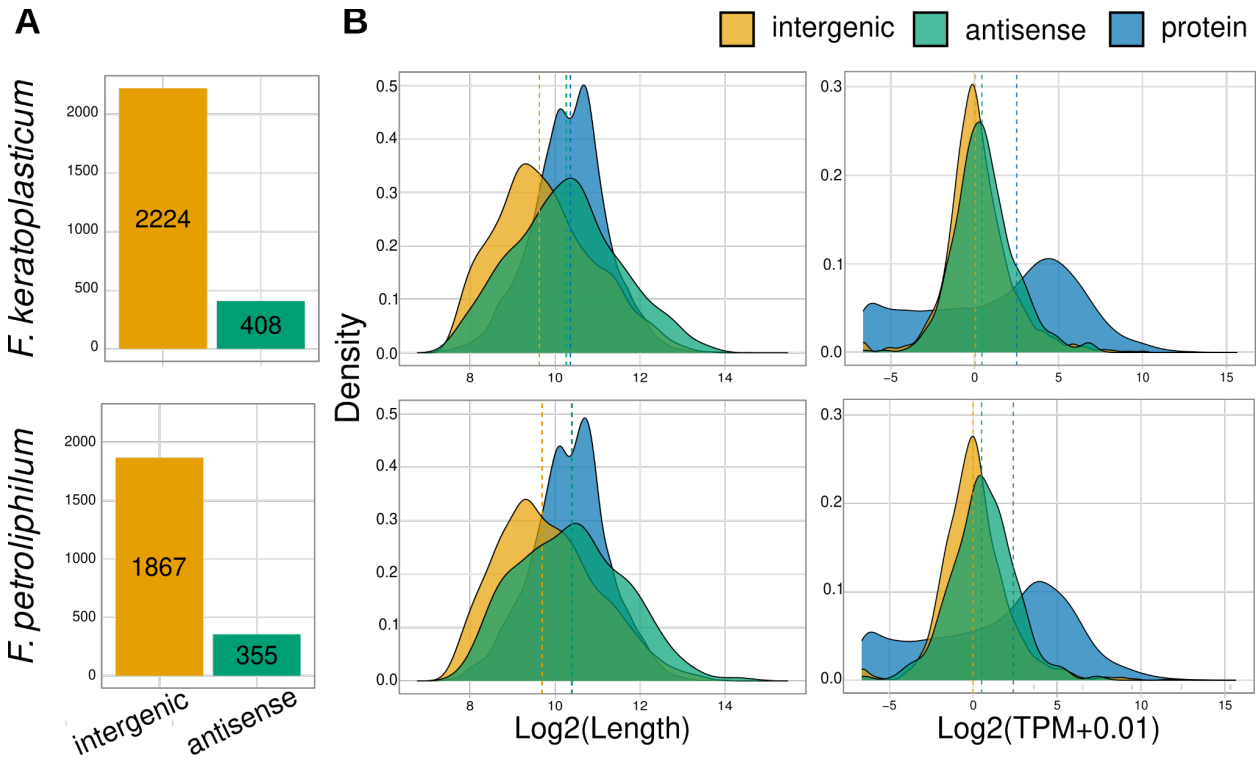
Supplementary Figure 1. Conservation of accessory chromosomes and contigs from *Fusarium falciforme, Fusarium sp. FSS12, Fusarium sp. Ph1 and Fusarium vanettenii*. (a-d) Accessory chromosomes and contigs of (a) *F. sp.* Ph1 (b) *F. falciforme* (c) *F. sp FSS12*, and (d) *F. vanettenii* strains and their sequence conservation with other FSSC isolates. Shading indicates percent identity across the syntenic block between the indicated isolate and accessory chromosome/contig indicated, versus the best matching chromosome or contig from the other isolates. Red lines indicate the location of transposable elements and other repetitive elements.



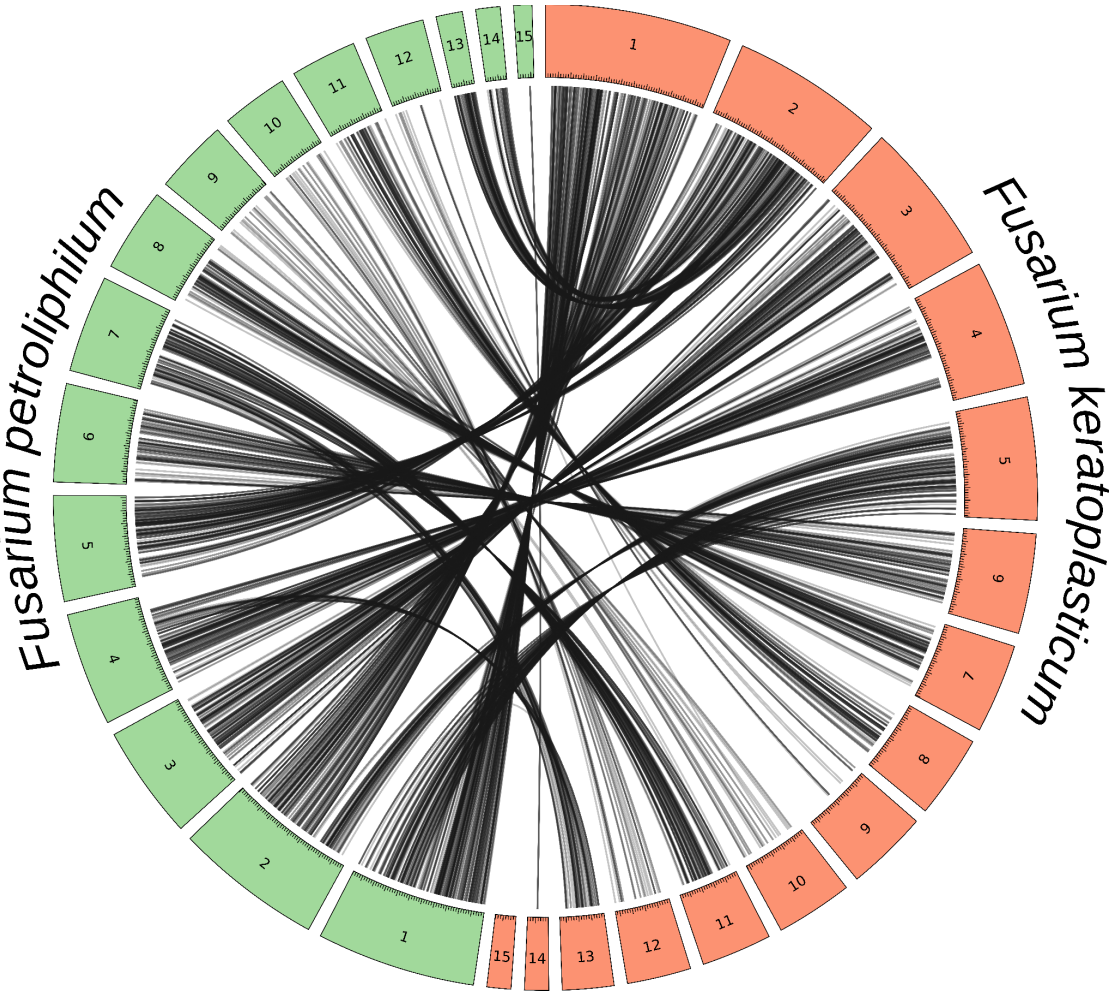
Supplementary Figure 2. Repeat content of core and accessory chromosomes. Repeat content calculated as the number of repetitive bases masked compared to the total chromosome length. Only chromosomes and contigs >150kb were included in the analysis. *F. keratoplasticum* and *F. petrolphilum* boxplots represent aggregated chromosome data from n=3 genomes each; all other species represent the chromosomes from n=1 genome. Significance determined by Mann-Whitney U test. * denotes $p<0.05$, ** $p<0.01$, *** $p<0.001$.



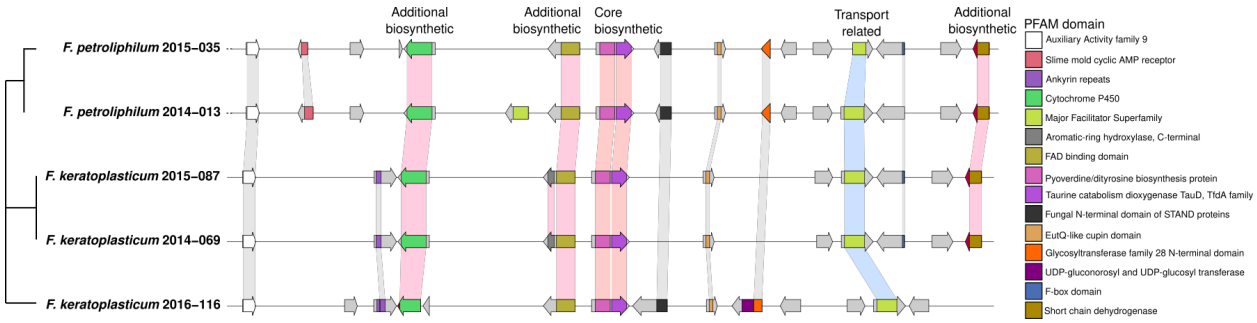
Supplementary Figure 3. Pairwise synteny of *F. keratoplasticum* Fu6 and the other study isolates. Colors indicate synthetic blocks shared between *F. keratoplasticum* Fu6 and the indicated strain, with the individual chromosome colors representing the syntenic block's chromosome number in Fu6. Only genomic chromosomes and contigs larger than 400 kbp are displayed.



Supplementary Figure 4. Identification and molecular properties of lncRNAs in *Fusariarium keratoplasticum and Fusarium petroliphilum*. (a) The number of intergenic and antisense predicted for *F. keratoplasticum* (upper panel) and *F. petroliphilum* (bottom panel). (b) Density plots showing the distribution of sequence length (left) and expression (right) between intergenic lncRNAs, antisense lnRNAs, and protein-coding genes. Dashed lines represent the median. Statistical significance calculated using the Mann-Whitney U test with Bonferroni correction. Intergenic lncRNAs are significantly shorter than protein-coding genes (*F. keratoplasticum*: $p < 0.0001$, *F. petroliphilum*: $p < 0.0001$). Antisense lncRNAs do not differ significantly in length compared to protein-coding genes. Both intergenic and antisense lncRNAs show lower expression levels compared to protein-coding genes ($p < 0.0001$ for both classes of lncRNAs in both species compared to protein coding genes.



Supplementary Figure 5. Synteny analysis of conserved intergenic lncRNAs in *Fusariarium keratoplasticum* and *Fusarium petroliphilum*. Circos plot depicts the syntenic relationships of intergenic lncRNAs between *F. keratoplasticum* NRZ-2014-069 and *F. petroliphilum* NRZ-2014-079. For a lncRNA to be considered syntenic, at least three common genes for each pairwise comparison between species, with a minimum of one common gene on each side of the lncRNA (parameters 3 3 1). Syntenic regions are visualized as black ribbons connecting segments between the two species. Only genomic scaffolds >615 kb are displayed.



Supplementary Figure 6. Synteny of genes in biosynthetic gene cluster 19 (BCG19). Alignment performed using BiG-SCAPE. PFAM domains in individual genes are indicated by color.

Supplementary File List

Supplementary File 1. Accession numbers for the genome assemblies used in the phylogeny, including the publicly available FSSC isolates analyzed.

Supplementary File 2. Gene Ontology terms significantly over- or underrperesneted in the genomes of *F. keratoplasticum* and *F. petroliphilum*. File include the gene counts and z-scored values depicted in Figure 5. Description of genome-wide ontology terms.

Supplementary File 3. Conservation of lncRNAs between *F. keratoplasticum* and *F. petroliphilum* identified through nucleotide alignment of lncRNA sequences and synteny using flanking protein-coding genes.

Supplementary File 4. Differentially expressed genes in conidia, under carbon starvation, and at 34°C relative to mycelia grown in complete media.Description of DEG file.

Supplementary File 5. Geneset enrichment analysis of differentially expressed genes in *F. keratoplasticum* and *F. petroliphilum*.

Supplementary File 6. Newick phylogeny displayed in Figure 1.

Supplementary File 7. Core and accessory genes for the FSSC, *F. keratoplasticum* and *F. petroliphilum*

References

1. Zhang N, O'Donnell K, Sutton DA, Nalim FA, Summerbell RC, Padhye AA, et al. Members of the Fusarium solani species complex that cause infections in both humans and plants are common in the environment. J Clin Microbiol. 2006;44: 2186–2190.

2. O'Donnell K, Sutton DA, Wiederhold N, Robert VAR, Crous PW, Geiser DM. Veterinary Fusarioses within the United States. J Clin Microbiol. 2016 [cited 17 May 2024]. doi:10.1128/jcm.01607-16

3. Coleman JJ. The Fusarium solani species complex: ubiquitous pathogens of agricultural importance. Mol Plant Pathol. 2016;17. doi:10.1111/mpp.12289

4. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. Nature Ecology & Evolution. 2019;3: 430–439.

5. Singh RP, Singh PK, Rutkoski J, Hodson DP, He X, Jørgensen LN, et al. Disease Impact on Wheat Yield Potential and Prospects of Genetic Control. Annu Rev Phytopathol. 2016;54: 303–322.

6. Hoh DZ, Lee H-H, Wada N, Liu W-A, Lu MR, Lai C-K, et al. Comparative genomic and transcriptomic analyses of trans-kingdom pathogen Fusarium solani species complex reveal degrees of compartmentalization. BMC Biol. 2022;20: 1–18.

7. Batista BG, de Chaves MA, Reginatto P, Saraiva OJ, Fuentefria AM. Human fusariosis: An emerging infection that is difficult to treat. Rev Soc Bras Med Trop. 2020;53. doi:10.1590/0037-8682-0013-2020

8. WHO fungal priority pathogens list to guide research, development and public health action. World Health Organization; 25 Oct 2022 [cited 23 Jul 2024]. Available: https://www.who.int/publications/i/item/9789240060241

9. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature. 2010;464: 367–373.

10. Zhang Y, Yang H, Turra D, Zhou S, Ayhan DH, DeIulio GA, et al. The genome of opportunistic fungal pathogen Fusarium oxysporum carries a unique set of lineage-specific chromosomes. Communications Biology. 2020;3: 1–12.

11. Genome-wide analysis of Fusarium verticillioides reveals inter-kingdom contribution of horizontal gene transfer to the expansion of metabolism. Fungal Genet Biol. 2019;128: 60–73.

12. Sieber CMK, Lee W, Wong P, Münsterkötter M, Mewes H-W, Schmeitzl C, et al. The Fusarium graminearum Genome Reveals More Secondary Metabolite Gene Clusters and Hints of Horizontal Gene Transfer. PLoS One. 2014;9: e110311.

13. Guarro J. Fusariosis, a complex infection caused by a high diversity of fungal species refractory to treatment. Eur J Clin Microbiol Infect Dis. 2013;32: 1491–1500.

14. van Diepeningen AD, Al-Hatmi AMS, Brankovics B, de Hoog GS. Taxonomy and Clinical Spectra of Fusarium Species: Where Do We Stand in 2014? Current Clinical Microbiology Reports. 2014;1: 10–18.

15. van Diepeningen AD, Feng P, Ahmed S, Sudhadham M, Bunyaratavej S, de Hoog GS. Spectrum of Fusarium infections in tropical dermatology evidenced by multilocus sequencing typing diagnostics. Mycoses. 2015;58: 48–57.

16. Short DPG, O'Donnell K, Zhang N, Juba JH, Geiser DM. Widespread occurrence of diverse human pathogenic types of the fungus Fusarium detected in plumbing drains. J Clin Microbiol. 2011;49: 4264–4272.

17. Garnica M, Nucci M. Epidemiology of Fusariosis. Curr Fungal Infect Rep. 2013;7: 301–305.

18. Hoenigl M, Jenks JD, Egger M, Nucci M, Thompson GR. Treatment of Fusarium Infection of the Central Nervous System: A Review of Past Cases to Guide Therapy for the Ongoing 2023 Outbreak in the United States and Mexico. Mycopathologia. 2023;188: 973–981.

19. Strong N, Meeks G, Sheth SA, McCullough L, Villalba JA, Tan C, et al. Neurovascular Complications of Iatrogenic Meningitis. N Engl J Med. 2024;390: 522–529.

20. García-Rodríguez G, Duque-Molina C, Kondo-Padilla I, Zaragoza-Jiménez CA, González-Cortés VB, Flores-Antonio R, et al. Outbreak of Meningitis in Immunocompetent Persons Associated With Neuraxial Blockade in Durango, Mexico, 2022-2023. Open Forum Infect Dis. 2024;11: ofad690.

21. Gibert S, Edel-Hermann V, Gautheron E, Gautheron N, Sol J-M, Capelle G, et al. First Report of Fusarium avenaceum, Fusarium oxysporum, Fusarium redolens, and Fusarium solani Causing Root Rot in Pea in France. Plant Dis. 2022 [cited 5 Aug 2024]. doi:10.1094/PDIS-04-21-0833-PDN

22. Pérez-Hernández A, Porcel-Rodríguez E, Gómez-Vázquez J. Survival of Fusarium solani f. sp. cucurbitae and Fungicide Application, Soil Solarization, and Biosolarization for Control of Crown and Foot Rot of Zucchini Squash. Plant Dis. 2017 [cited 5 Aug 2024]. doi:10.1094/PDIS-06-16-0883-RE

23. Gómez J, Serrano Y, Pérez A, Porcel E, Gómez R, Aguilar MI. Fusarium solani f. sp. cucurbitae, affecting melon in Almería Province, Spain. Australas Plant Dis Notes. 2014;9: 1–3.

24. Wang R-Y, Gao B, Li X-H, Ma J, Chen S-L. First Report of Fusarium solani Causing Fusarium Root Rot and Stem Canker on Storage Roots of Sweet Potato in China. 2013 [cited 5 Aug 2024]. doi:10.1094/PDIS-06-13-0651-PDN

25. Short DPG, O'Donnell K, Geiser DM. Clonality, recombination, and hybridization in the plumbing-inhabiting human pathogen Fusarium keratoplasticuminferred from multilocus sequence typing. BMC Evol Biol. 2014;14: 1–14.

26. Vega-Gutiérrez TA, López-Orona CA, Molina-Cárdenas L, López-Urquídez GA, Payán-Arzapalo MA, Tirado-Ramírez MA. First Report of Fusarium keratoplasticum Causing Strawberry Root Rot in Sinaloa, Mexico. Plant Dis. 2023 [cited 17 May 2024]. doi:10.1094/PDIS-10-22-2441-PDN

784 27. Lu L-Y, Ou J-H, Hui RC-Y, Chuang Y-H, Fan Y-C, Sun P-L. High Diversity of Fusarium
785 Species in Onychomycosis: Clinical Presentations, Molecular Identification, and Antifungal
786 Susceptibility. Journal of Fungi. 2023;9: 534.

787 28. Walther G, Stasch S, Kaerger K, Hamprecht A, Roth M, Cornely OA, et al. Fusarium
788 Keratitis in Germany. J Clin Microbiol. 2017 [cited 25 Jun 2024]. doi:10.1128/jcm.00649-17

789 29. Peinado-Acevedo JS, Ramírez-Sánchez IC. Endocarditis by Fusarium keratoplasticum.
790 Mycopathologia. 2020;186: 131–133.

791 30. Phylogenetic relationships among members of the Fusarium solani species complex in
792 human infections and the descriptions of F. keratoplasticum sp. nov. and F. petroliphilum
793 stat. nov. Fungal Genet Biol. 2013;53: 59–70.

794 31. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen L-L, et al. Long non-
795 coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell
796 Biol. 2023;24: 430–447.

797 32. Kim W, Miguel-Rojas C, Wang J, Townsend JP, Trail F. Developmental Dynamics of Long
798 Noncoding RNA Expression during Sexual Fruiting Body Formation in Fusarium
799 graminearum. MBio. 2018;9. doi:10.1128/mBio.01292-18

800 33. Parra-Rivero O, Pardo-Medina J, Gutiérrez G, Limón MC, Avalos J. A novel lncRNA as a
801 positive regulator of carotenoid biosynthesis in Fusarium. Sci Rep. 2020;10: 1–14.

802 34. Pegueroles C, Iraola-Guzmán S, Chorostecki U, Ksiezopolska E, Saus E, Gabaldón T.
803 Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species
804 of the genus Caenorhabditis. RNA Biol. 2019;16: 320–329.

805 35. Kingsbury JM, Yang Z, Ganous TM, Cox GM, McCusker JH. Cryptococcus neoformans
806 Ilv2p confers resistance to sulfometuron methyl and is required for survival at 37 °C and in
807 vivo. Microbiology. 2004;150: 1547–1558.

808 36. Kingsbury JM, McCusker JH. Cytocidal amino acid starvation of Saccharomyces cerevisiae
809 and Candida albicans acetolactate synthase (ilv2Δ) mutants is influenced by the carbon
810 source and rapamycin. Microbiology. 2010;156: 929–939.

811 37. Oh Y, Franck WL, Han S-O, Shows A, Gokce E, Muddiman DC, et al. Polyubiquitin Is
812 Required for Growth, Development and Pathogenicity in the Rice Blast Fungus
813 Magnaporthe oryzae. PLoS One. 2012;7. doi:10.1371/journal.pone.0042868

814 38. Rodier M-H, El Moudni B, Kauffmann-Lacroix C, Daniault G, Jacquemin J-L. A Candida
815 albicans metallopeptidase degrades constitutive proteins of extracellular matrix. FEMS
816 Microbiol Lett. 1999;177: 205–210.

817 39. Silva BA, Souza-Gonçalves AL, Pinto MR, Barreto-Bergter E, Santos ALS.
818 Metallopeptidase inhibitors arrest vital biological processes in the fungal pathogen
819 Scedosporium apiospermum. Mycoses. 2011;54: 105–112.

820 40. Park J, Han JW, Lee N, Kim S, Choi S, Lee H-H, et al. Sulfur metabolism-mediated fungal
821 glutathione biosynthesis is essential for oxidative stress resistance and pathogenicity in the
822 plant pathogenic fungus Fusarium graminearum. MBio. 2024 [cited 20 Aug 2024].

823 doi:10.1128/mbio.02401-23

824 41. Oxidative responses and fungal infection biology. Semin Cell Dev Biol. 2019;89: 34–46.

825 42. Arana DM, Prieto D, Román E, Nombela C, Alonso-Monge R, Pla J. The role of the cell wall
826 in fungal pathogenesis. Microb Biotechnol. 2009;2: 308–320.

827 43. Garcia-Rubio R, de Oliveira HC, Rivera J, Trevijano-Contador N. The Fungal Cell Wall: , ,
828 and Species. Front Microbiol. 2019;10: 2993.

829 44. Identification of gene clusters associated with fusaric acid, fusarin, and perithecial pigment
830 production in Fusarium verticillioides. Fungal Genet Biol. 2012;49: 521–532.

831 45. Jihua Wei BW. Chemistry and bioactivities of secondary metabolites from the genus
832 Fusarium. Fitoterapia. 2020;146: 104638.

833 46. Dawson MJ, Farthing JE, Marshall PS, Middleton RF, O'neill MJ, Shuttleworth A, et al. THE
834 SQUALESTATINS, NOVEL INHIBITORS OF SQUALENE SYNTHASE PRODUCED BY A
835 SPECIES OF PHOMA I. TAXONOMY, FERMENTATION, ISOLATION, PHYSICO-
836 CHEMICAL PROPERTIES AND BIOLOGICAL ACTIVITY. J Antibiot . 1992;45: 639–647.

837 47. Kurobane I, Vining LC, Gavin Mcinnes A, Gerber NN. METABOLITES OF FUSARIUM
838 SOLANI RELATED TO DIHYDROFUSARUBIN. J Antibiot . 1980;33: 1376–1379.

839 48. Kundu A, Mandal A, Saha S, Prabhakaran P, Walia S. Fungicidal activity and molecular
840 modeling of fusarubin analogues from Fusarium oxysporum. Toxicol Environ Chem. 2020
841 [cited 31 May 2024]. doi:10.1080/02772248.2020.1770253

842 49. Costa JH, Wassano CI, Angolini CFF, Scherlach K, Hertweck C, Pacheco Fill T. Antifungal
843 potential of secondary metabolites involved in the interaction between citrus pathogens. Sci
844 Rep. 2019;9: 1–11.

845 50. Kato S, Motoyama T, Futamura Y, Uramoto M, Nogawa T, Hayashi T, et al. Biosynthetic
846 gene cluster identification and biological activity of lucilactaene from Fusarium sp. RK97-94.
847 Biosci Biotechnol Biochem. 2020;84: 1303–1307.

848 51. Suga H, Arai M, Fukasawa E, Motohashi K, Nakagawa H, Tateishi H, et al. Genetic
849 Differentiation Associated with Fumonisin and Gibberellin Production in Japanese Fusarium
850 fujikuroi. Appl Environ Microbiol. 2018 [cited 19 Jul 2024]. doi:10.1128/AEM.02414-18

851 52. Hossain MT, Khan A, Chung EJ, Rashid MH-O, Chung YR. Biological Control of Rice
852 Bakanae by an Endophytic Bacillus oryzicola YC7007. Plant Pathol J. 2016;32: 228–241.

853 53. Fayyaz A, Robinson G, Chang PL, Bekele D, Yimer S, Carrasquilla-Garcia N, et al. Hiding
854 in plain sight: Genome-wide recombination and a dynamic accessory genome drive
855 diversity in Fusarium oxysporum f.sp. ciceris. Proceedings of the National Academy of
856 Sciences. 2023;120: e2220570120.

857 54. van Westerhoven AC, Aguilera-Galvez C, Nakasato-Tagami G, Shi-Kunne X, de la Parte
858 EM, Chavarro-Carrero E, et al. Segmental duplications drive the evolution of accessory
859 regions in a major crop pathogen. New Phytol. 2024;242: 610–625.

860 55. Barber AE, Sae-Ong T, Kang K, Seelbinder B, Li J, Walther G, et al. Aspergillus fumigatus

861 pan-genome analysis identifies genetic variants associated with human infection. Nature
862 Microbiology. 2021;6: 1526–1536.

863 56. McCarthy CGP, Fitzpatrick DA. Pan-genome analyses of model fungal species. Microbial
864 Genomics. 2019;5: e000243.

865 57. Gangurde SS, Korani W, Bajaj P, Wang H, Fountain JC, Agarwal G, et al. Aspergillus
866 flavus pangenome (AflaPan) uncovers novel aflatoxin and secondary metabolite associated
867 gene clusters. BMC Plant Biol. 2024;24: 1–17.

868 58. Anne Hatmaker E, Barber AE, Drott MT, Sauters TJC, Alastruey-Izquierdo A, Garcia-
869 Hermoso D, et al. Pathogenicity is associated with population structure in a fungal
870 pathogen of humans. bioRxiv. 2024. p. 2024.07.05.602241.
871 doi:10.1101/2024.07.05.602241

872 59. Dewar AE, Hao C, Belcher LJ, Ghoul M, West SA. Bacterial lifestyle shapes pangenomes.
873 PNAS. 2024;121. doi:10.1073/pnas.2320170121

874 60. Perrier M, Barber AE. Unraveling the genomic diversity and virulence of human fungal
875 pathogens through pangenomics. PLoS Pathog. 2024;20: e1012313.

876 61. Möller M, Stukenbrock EH. Evolution and genome architecture in fungal plant pathogens.
877 Nat Rev Microbiol. 2017;15: 756–771.

878 62. Fokkens L, Guo L, Dora S, Wang B, Ye K, Sánchez-Rodríguez C, et al. A Chromosome-
879 Scale Genome Assembly for the Fusarium oxysporum Strain Fo5176 To Establish a Model
880 Arabidopsis-Fungal Pathosystem. G3 Genes|Genomes|Genetics. 2020;10: 3549–3555.

881 63. McCourt JA, Duggleby RG. Acetohydroxyacid synthase and its role in the biosynthetic
882 pathway for branched-chain amino acids. Amino Acids. 2006;31: 173–210.

883 64. Steyer JT, Todd RB. Branched-chain amino acid biosynthesis in fungi. Essays in
884 Biochemsitry. 2023;67: 865–876.

885 65. Richie DL, Thompson KV, Studer C, Prindle VC, Aust T, Riedl R, et al. Identification and
886 Evaluation of Novel Acetolactate Synthase Inhibitors as Antifungal Agents. Antimicrob
887 Agents Chemother. 2013;57: 2272.

888 66. Pokhrel A, Coleman JJ. Inventory of the Secondary Metabolite Biosynthetic Potential of
889 Members within the Terminal Clade of the Species Complex. J Fungi (Basel). 2023;9.
890 doi:10.3390/jof9080799

891 67. Lim FY, Won TH, Raffa N, Baccile JA, Wisecaver J, Rokas A, et al. Fungal Isocyanide
892 Synthases and Xanthocillin Biosynthesis in Aspergillus fumigatus. MBio. 2018 [cited 12 Aug
893 2024]. doi:10.1128/mbio.00785-18

894 68. Harris NC, Sato M, Herman NA, Twigg F, Cai W, Liu J, et al. Biosynthesis of isonitrile
895 lipopeptides by conserved nonribosomal peptide synthetase gene clusters in
896 Actinobacteria. Proceedings of the National Academy of Sciences. 2017;114: 7025–7030.

897 69. Won TH, Bok JW, Nadig N, Venkatesh N, Nickles G, Greco C, et al. Copper starvation
898 induces antimicrobial isocyanide integrated into two distinct biosynthetic pathways in fungi.
899 Nat Commun. 2022;13: 1–14.

900 70. Hovhannisyan H, Gabaldón T. The long non-coding RNA landscape of Candida yeast
901 pathogens. Nat Commun. 2021;12. doi:10.1038/s41467-021-27635-4

902 71. Kalem MC, Panepinto JC. Long Non-Coding RNAs in Cryptococcus neoformans: Insights
903 Into Fungal Pathogenesis. Front Cell Infect Microbiol. 2022;12: 858317.

904 72. Glad HM, Tralamazza SM, Croll D. The expression landscape and pangenome of long non-
905 coding RNA in the fungal wheat pathogen Zymoseptoria tritici. Microbial Genomics. 2023;9:
906 001136.

907 73. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie
908 enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol.
909 2015;33: 290–295.

910 74. Sims D, Ilott NE, Sansom SN, Sudbery IM, Johnson JS, Fawcett KA, et al. CGAT:
911 computational genomics analysis toolkit. Bioinformatics. 2014;30: 1290–1291.

912 75. Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, et al. CPC2: a fast and accurate
913 coding potential calculator based on sequence intrinsic features. Nucleic Acids Res.
914 2017;45: W12–W16.

915 76. Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, Leeb T, et al. FEELnc: a tool for long
916 non-coding RNA annotation and its application to the dog transcriptome. Nucleic Acids Res.
917 2017;45: e57.

918 77. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for
919 assigning sequence reads to genomic features. Bioinformatics. 2014;30: 923–930.

920 78. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast
921 and versatile genome alignment system. PLoS Comput Biol. 2018;14: e1005944.