


RESEARCH ARTICLE

Open Access



Comparative genomic and transcriptomic analyses of trans-kingdom pathogen *Fusarium solani* species complex reveal degrees of compartmentalization

Daphne Z. Hoh^{1,2,3}, Hsin-Han Lee¹, Naohisa Wada¹, Wei-An Liu¹, Min R. Lu¹, Cheng-Kuo Lai^{1,4}, Huei-Mien Ke¹, Pei-Feng Sun^{1,2,3}, Sen-Lin Tang^{1,2}, Wen-Hsin Chung⁵, Ying-Lien Chen⁶, Chia-Lin Chung⁶ and Isheng Jason Tsai^{1,2,4*} 

Abstract

Background: The *Fusarium solani* species complex (FSSC) comprises fungal pathogens responsible for mortality in a diverse range of animals and plants, but their genome diversity and transcriptome responses in animal pathogenicity remain to be elucidated. We sequenced, assembled and annotated six chromosome-level FSSC clade 3 genomes of aquatic animal and plant host origins. We established a pathosystem and investigated the expression data of *F. falciforme* and *F. keratoplasticum* in Chinese softshell turtle (*Pelodiscus sinensis*) host.

Results: Comparative analyses between the FSSC genomes revealed a spectrum of conservation patterns in chromosomes categorised into three compartments: core, fast-core (FC), and lineage-specific (LS). LS chromosomes contribute to variations in genomes size, with up to 42.2% of variations between *F. vanettenii* strains. Each chromosome compartment varied in structural architectures, with FC and LS chromosomes contain higher proportions of repetitive elements with genes enriched in functions related to pathogenicity and niche expansion. We identified differences in both selection in the coding sequences and DNA methylation levels between genome features and chromosome compartments which suggest a multi-speed evolution that can be traced back to the last common ancestor of *Fusarium*. We further demonstrated that *F. falciforme* and *F. keratoplasticum* are opportunistic pathogens by inoculating *P. sinensis* eggs and identified differentially expressed genes also associated with plant pathogenicity. These included the most upregulated genes encoding the CFEM (Common in Fungal Extracellular Membrane) domain.

Conclusions: The high-quality genome assemblies provided new insights into the evolution of FSSC chromosomes, which also serve as a resource for studies of fungal genome evolution and pathogenesis. This study also establishes an animal model for fungal pathogens of trans-kingdom hosts.

Keywords: Genome compartments, *Fusarium solani* species complex, Chromosome evolution, Opportunistic pathogen, Animal pathogenicity, Turtle

Background

The filamentous fungi in the genus *Fusarium* are amongst the most virulent pathogens affecting multi-kingdom hosts [1, 2], but can also exist as saprophytes. The genus, commonly identified as plant pathogens, has caused

*Correspondence: ijtsai@gate.sinica.edu.tw

¹ Biodiversity Research Center, Academia Sinica, 115 Nangang, Taipei, Taiwan
Full list of author information is available at the end of the article



devastating losses in the global agricultural industry [3] and is recognised as one of the most prevalent clinical pathogens causing superficial and invasive disease in immunocompromised humans [4, 5]. In recent decades, an increasing number of fusariosis cases associated with various types of animals have been reported worldwide [6–10], but research on the virulence of pathogens causing fusariosis beyond plant hosts remains limited [11, 12]. The disease is now considered to be a serious emerging fungal threat potentially inducing host population loss and extinction [2, 13].

The most prevalent *Fusarium* pathogens associated with veterinary infection are from the species-rich clade three of *Fusarium solani* species complex (FSSC) [1, 2, 14, 15]. Most species in this clade are ubiquitous in the natural environment [1], except *F. keratoplasticum*, which is mostly isolated from plumbing systems [16]. Several FSSC species were reported to cause disease in aquatic animals such as grey seals, shrimps, dolphins, and sharks [2]. Two species from the FSSC clade three—*F. falciforme* and *F. keratoplasticum*—were the predominant species occurring in diseased sea turtle nests worldwide [17], and the latter was also recently reported to infect *Podocnemis unifilis*, an endangered freshwater turtle species [18]. Koch's postulates were fulfilled for *F. keratoplasticum*, which can cause disease and high mortality rate (83.3%) in sea turtle *Caretta caretta* eggs [19]. The disease is now termed sea turtle egg fusariosis [20] and is responsible for low hatching success of eggs from both natural nests [17, 21] and man-made hatcheries [22, 23].

One particular genome characteristic in many pathogenic fungi are that their chromosomes can be differentiated into two compartments: the core chromosome (CC), which contain essential genes required for survival and reproduction, and lineage-specific chromosome (LSC;

also known as accessory chromosome), which is repeat-rich and enriched with genes mostly associated with niche adaptation and pathogenicity [24–26]. LSC can be dispensable and do not affect fungal growth in several *Fusarium* species such as *F. oxysporum* f. sp. *lycopercisi* [24] and *F. vanettenii* (previously *Nectria haematococca*) [27]. The LSC of *F. vanettenii* harbours pea-specific pathogenic genes—the PEP cluster [27–29], and the deletion of the entire LSC reduced the pathogen's virulence towards pea plants [11, 27].

In this study, we produced six high-quality genome assemblies for species in the FSSC clade 3 isolated from hosts belonging to different kingdoms. Comparative analyses of six FSSC and three other *Fusarium* genomes allowed us to divide the chromosomes into multiple compartments based on genome alignments and clustering of single-copy orthologs. These compartments harboured distinct genome features as a result of different evolutionary dynamics. We show that *F. falciforme* and *F. keratoplasticum* that are responsible for mass mortalities in sea turtles can penetrate Chinese softshell turtle eggshells and colonise egg inclusions in the laboratory setting. By employing a dual RNA-seq approach, gene expression data provided a detailed description of the two species and the host during infection. Together, the availability of these genomes and the transcriptome data underpin an initial effort to study animal parasitism in these trans-kingdom pathogens.

Results

Genome characteristics of six sequenced FSSC isolates

We sequenced six genomes of five species within clade three of FSSC (Table 1). This included *F. falciforme* (Fu3), *F. keratoplasticum* (Fu6 and LHS11), and *Fusarium* sp. FSSC12 (LHS14) isolated from various aquatic

Table 1 Genome statistics of six sequenced isolates of *Fusarium solani* species complex

Species	<i>F. falciforme</i>	<i>F. keratoplasticum</i>	<i>F. keratoplasticum</i>	<i>Fusarium</i> sp. FSSC12	<i>Fusarium</i> sp.	<i>F. vanettenii</i>
Strain ID	Fu3	Fu6	LHS11	LHS14	Ph1	Fs6
Estimated haploid genome size (Mb)	54.5	51.6	56.2	59.6	50.9	72.9
Assembly size (Mb)	53.6	49.5	53.5	56.3	51.5	72.9
Number of contig	14	25	26	35	24	40
Largest contig (Mb)	6.8	6.4	6.6	6.1	5.4	6.7
N50 Mb (L50)	3.8 (6)	4.2 (5)	3.4 (6)	3.8 (6)	3.2 (6)	3.3 (9)
N90 Mb (L90)	2.7 (12)	2.2 (11)	1.5 (13)	0.9 (14)	1.1 (17)	1.3 (23)
GC %	49.4	51.3	51.4	48.6	51.1	49.4
Number of protein coding genes	15137	14935	15051	15937	15498	18862
Protein BUSCO % (fungi)	99	98.4	98.4	98.3	100	98.6
Protein BUSCO % (ascomycota)	98.8	98.8	98.1	98.6	99.8	98.7

Genome size was estimated via GenomeScope 2.0

animal hosts. Two species of plant host origins—*Fusarium* sp. (Ph1) from orchid and *F. vanettenii* (Fs6) from pea—were also chosen for comparative purposes. The initial assemblies were produced from averaging 121X of Oxford Nanopore reads (sequence N50 = 14–23kb) using the Flye assembler [30] and polished by Illumina reads (Additional file 1: Table S1). The final assemblies were in 14 to 40 contigs with N50 3.2–4.2Mb (Table 1) averaging 56 Mb, which is more contiguous and larger than other representative *Fusarium* genomes (ranging 12–4197 contigs averaging 47Mb; Additional file 1: Table S2). Interestingly, *F. vanettenii* Fs6 has an assembly of 72.9Mb, the largest *Fusarium* genome reported to date and 42.2% larger than the published *F. vanettenii* MPVI 77-13-4 (abbreviated as FVANE) genome of 51.2Mb [31], suggesting high intraspecies variation. The *F. falciforme* Fu3 reference was amongst the most complete genome, consisting of 14 contigs with six telomere-to-telomere gapless chromosomes (Additional file 1: Table S3) and was on average 17 times more contiguous than the published FSSC genomes (N90: 2.7Mb vs. 0.4kb–2.6Mb, respectively; last accessed date 18th January 2022; Additional file 1: Table S4).

We predicted 14,927 to 18,862 protein-encoding genes from these assemblies using the MAKER2 [32] pipeline based on evidence using proteomes from closely related fungi and transcriptome reads generated from mycelium. Benchmarking Universal Single-Copy Orthologs (BUSCO) [33] analysis indicated that the proteomes were 98.3 to 100% complete (Table 1) and 70–77.1% of these genes were assigned to a protein family or domain (Pfam) via eggNOG-mapper [34], indicating high completeness in these assemblies. The orthology of the six sequenced assemblies, 17 other *Fusarium* genomes, and an outgroup species *Beauveria bassiana* (Additional file 1: Table S2) was inferred using OrthoFinder [35]. A total of 24,203 orthogroups (OGs) were inferred, of which 16,154 (66.7%) and 235 (1%) OGs were *Fusarium* and FSSC-specific, respectively. All six genomes contained a predicted average of 420 (2.7% of total gene) and 487 genes, encoding effectors and carbohydrate-active enzymes (CAZymes), respectively (Additional file 1: Table S5 and S6). In addition, an average of 44 secondary metabolite biosynthetic gene clusters (SMBGCs) were detected in the six genomes, some of which, including a fusarin gene cluster, are associated with plant pathogenicity (Additional file 1: Table S7 and S8) [36]. The numbers of these gene families are comparable to other *Fusarium* genomes [36, 37]. The repeat proportion of FSSC genomes typically range 4.2–9.4%, except for *F. vanettenii* Fs6, which contained 18.8% repeats (Additional file 2: Fig. S1a). DNA transposons constituted the largest proportion

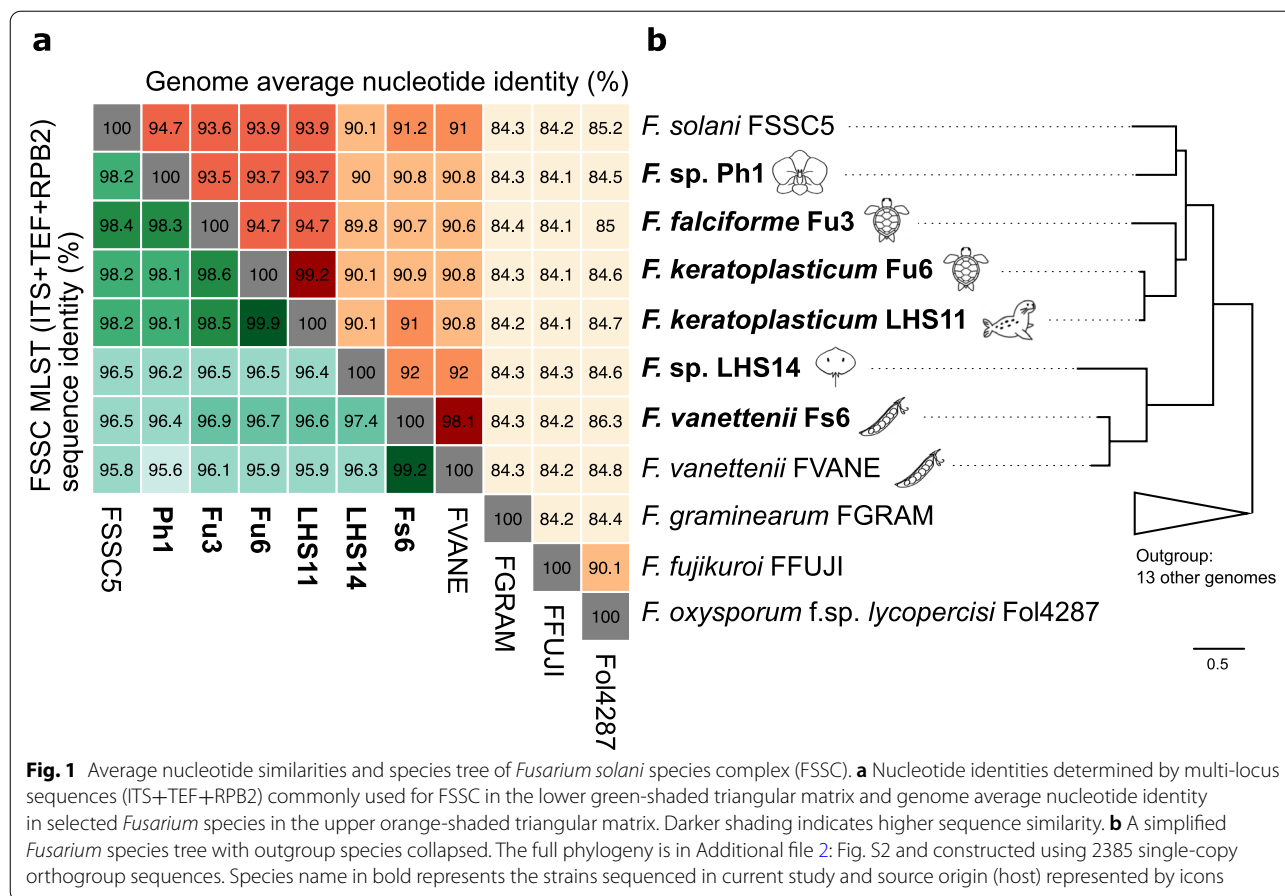
(1.4 to 6.4%) of FSSC genomes (Additional file 2: Fig. S1b), followed by LTR retrotransposons (0.9 to 5.8%).

To determine the phylogenetic position of our isolates, we compared the sequence identities and constructed a maximum likelihood phylogeny based on 40 FSSC species using the commonly employed multi-locus sequence typing (MLST) targeting the internal transcribed spacer (ITS) region, translation elongation factor (TEF), and RNA polymerase II (RPB2) (Fig. 1a; Additional file 2: Fig. S2 and Additional file 1: Table S9). With the new genomes, we computed the genome average nucleotide identity (ANI) and constructed a species phylogeny using 2385 single-copy orthologs, which recapitulated the relationship of the MLST phylogeny, and found that the species relationships were not grouped by animal or plant hosts (Fig. 1b; Additional file 2: Fig. S3 and Additional file 1: Table S2). Ph1 was most closely related to *F. solani* haplotype FSSC5 (MLST and genome ANI: 98.2 and 94.7%). *Fusarium* sp. LHS14 was designated as haplotype FSSC12 and the first genome assembly of this species (Additional file 2: Fig. S3).

Evolutionary dynamics of FSSC chromosomes

To investigate whether lineage-specific chromosomes (LSCs) were present in each *Fusarium* species, we assessed the extent of chromosome linkage between species via pairwise single-copy orthologs, proportions of repetitive elements and FSSC-specific genes in each chromosome. In addition to assigning the core and LSC as in previous *Fusarium* studies [24, 31], FSSC chromosomes were categorised into two compartments, revealing a spectrum of conservation (Fig. 2a; Additional file 2: Fig. S4a). The majority of chromosomes were part of the linkage group that was shared by all six species; we designated these the core chromosomes (CCs; Additional file 2: Fig. S4a). On the other end of the conservation spectrum were LSCs that contained single-copy orthologues shared by mostly two out of six genomes and a lower proportion of shared genes. Further examination of linkage between FSSC chromosome with three other non-FSSC *Fusarium* species revealed an additional compartment within the FSSC, which had a lower proportion of genes with orthology detected beyond FSSC (Fig. 2b; Additional file 2: Fig. S4b). This additional compartment comprised mainly three linkage groups corresponding to chromosomes 7, 11, and 12 of *F. falciforme* Fu3 genome; these were designated as fast-core chromosomes (FCCs).

Synteny analysis using *F. falciforme* Fu3 as reference revealed that the CCs and FCCs were more syntenic than the LSCs between the FSSC species (Additional file 2: Fig. S5). The CCs harboured significantly fewer repeats (Wilcoxon rank sum test, $p < 0.01$; Fig. 2c) and lower proportions of FSSC-specific genes than FCCs and



LSCs in each FSSC genome (Wilcoxon rank sum test, $p < 0.0001$; Fig. 2d). We designated a total of 9 CCs, 3 FCCs, and 2–11 LSCs in each FSSC genome. The CCs and FCCs constituted the majority (67.7–94.7%) of genomes, while the number of LS chromosomes varied and were responsible for the aforementioned genome differences between the two *F. vanettenii* strains (72.9Mb in Fs6 versus 51.2Mb in FVANE). These LS chromosomes exhibited a lack of synteny (Fig. 2e; Additional file 2: Fig. S6), suggesting that these chromosomes were differentially maintained or distributed through horizontal chromosome transfer [24]. Additionally, the *F. vanettenii* CCs comprise 14,513 and 14,961 genes in FVANE and Fs6, respectively, but the latter contained more genes (3395 vs. 1195) located on these additional LSCs. The gapless FSSC LS chromosomes averaged 1.2Mb, which is a similar mark up as the *F. oxysporum* genome [24]. However, telomere-to-telomere LSC length were observed in FSSCs, ranging from 0.8Mb in chromosome 13 of *Fusarium* sp. Ph1 to 3.3Mb in chromosome 14 of *F. vanettenii* Fs6, demonstrating the characteristic dynamics of non-CCs in *Fusarium* genomes [38, 39]. The combined mean size of FCCs was approximately 2.8Mb, with FSSC-specific genes ranging

from 26.8 to 35.4% of the total gene content (vs. 7.7 to 22.1% in CCs). A lower number of FSSC-specific genes were detected in LSCs compared to CCs (averaging 59 vs. 272 genes), but the former constitutes at least one-fifth of the total gene content (Additional file 1: Table S10). The FCCs appeared as an intermediate category between the CCs and LSCs based on genome features, similar to *F. oxysporum* [38]. In terms of gene locality, sub-telomeric bias was detected for these FSSC-specific genes, a similar trend as previously reported in *F. graminearum* [40] and *Aspergillus* genomes [41], but not in FCCs, where the genes were distributed across the entire chromosome (Additional file 2: Fig. S7). Gene Ontology (GO) enrichment analysis in each chromosome type revealed FCCs and LSCs harbour genes that are feasibly linked to pathogenicity and expansion of new niches such as environment or host, compared to CCs which mainly harbour genes for essential biological functions such as growth and development (Additional file 3). FCCs had the highest mean proportion and number of effectors, CAZymes and SMBGCs amongst the chromosome types (Additional file 1: Table S11 and S12), consistent with observations in *F. oxysporum* [42]. Together these results

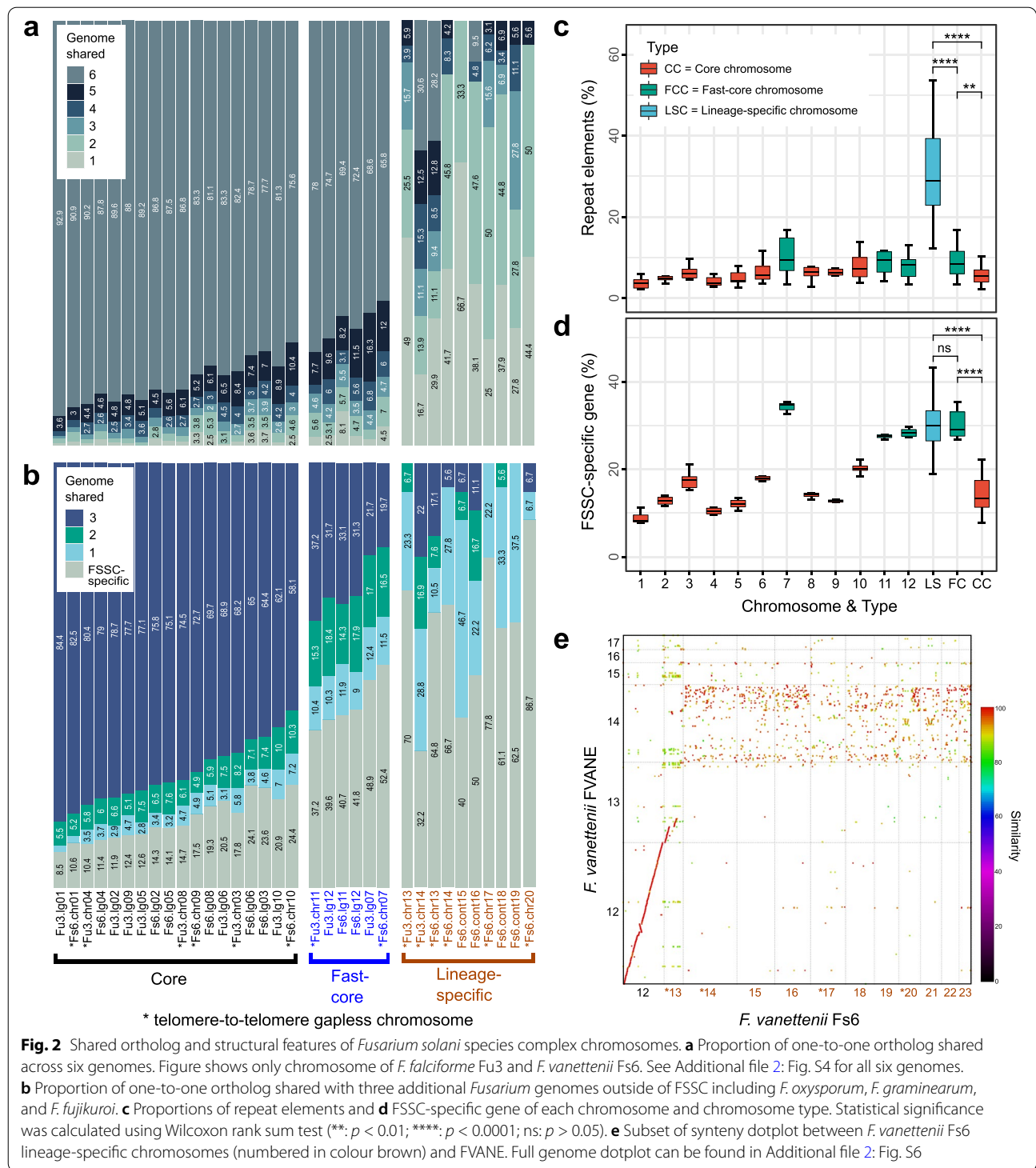


Fig. 2 Shared ortholog and structural features of *Fusarium solani* species complex chromosomes. **a** Proportion of one-to-one ortholog shared across six genomes. Figure shows only chromosome of *F. falciforme* Fu3 and *F. vanettenii* Fs6. See Additional file 2: Fig. S4 for all six genomes. **b** Proportion of one-to-one ortholog shared with three additional *Fusarium* genomes outside of FSSC including *F. oxysporum*, *F. graminearum*, and *F. fujikuroi*. **c** Proportions of repeat elements and **d** FSSC-specific gene of each chromosome and chromosome type. Statistical significance was calculated using Wilcoxon rank sum test (**: $p < 0.01$; ****: $p < 0.0001$; ns: $p > 0.05$). **e** Subset of synteny dotplot between *F. vanettenii* Fs6 lineage-specific chromosomes (numbered in colour brown) and FVANE. Full genome dotplot can be found in Additional file 2: Fig. S6

suggest that, in addition to LSCs, genes in FCCs also play an important role in pathogenicity processes such as host colonisation and infection.

To investigate the evolutionary differences of each chromosome type, we estimated the ratio of non-synonymous

substitutions to synonymous substitutions per gene (d_N/d_S) across single-copy orthologs between *F. falciforme* Fu3 and *F. keratoplasticum* Fu6. Overall, genes located in LSCs had higher d_N/d_S than FCCs, followed by CCs (median $\omega = 0.24, 0.10, \text{ and } 0.06$, respectively, Wilcoxon

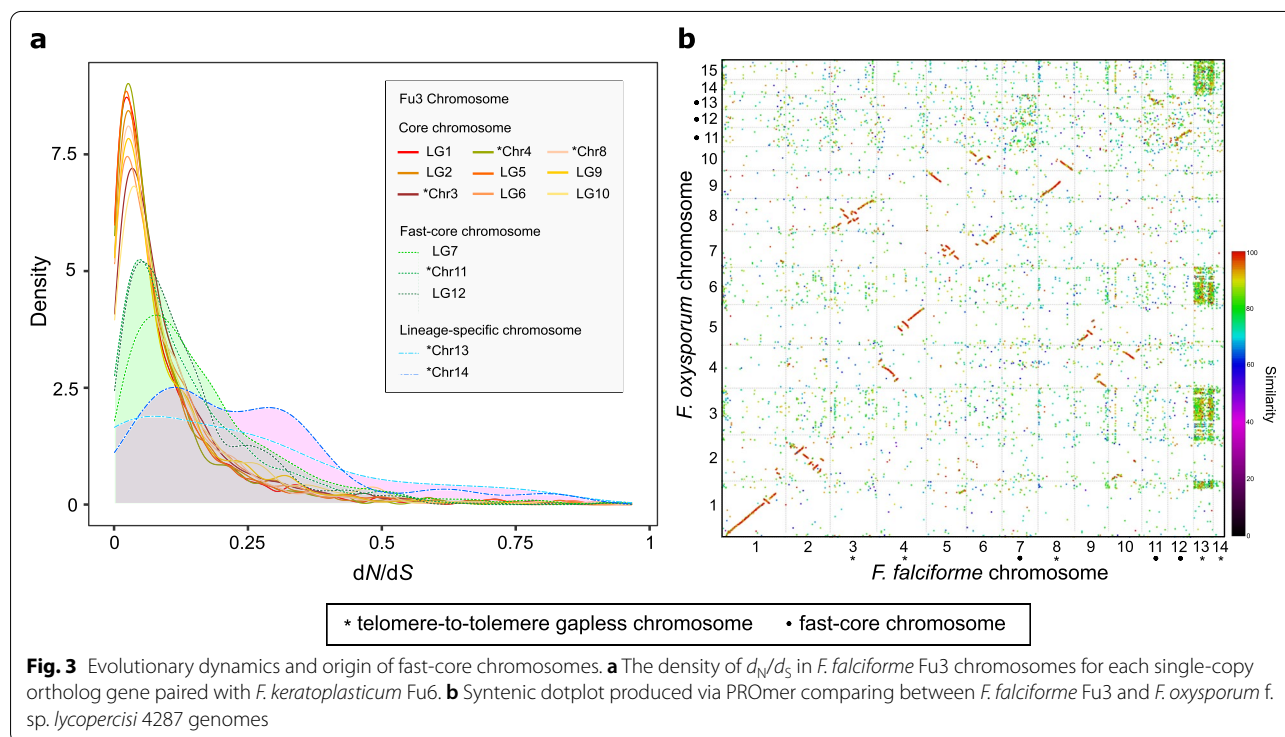
rank sum test, $p < 0.05$ in all comparison pairs; Fig. 3a), indicating different levels of purifying selections amongst chromosome types. In addition, genes that were translocated to other linkage groups were also comprised of higher d_N/d_S (median $\omega = 0.09, 0.08,$ and $0.05,$ in LSCs, FCCs, and CCs, respectively; Wilcoxon rank sum test, $p < 0.0001$ in all comparison pairs; Additional file 2: Fig. S8), suggesting differential selection after rearrangement, such as relaxed purifying selection.

Several distinctive differences in chromosome 7 of FSSC genomes were observed when compared to the other FCCs. For instance, the former was mostly located at the end of FCCs in the conservation spectrum (Additional file 2: Fig. S4); it had the highest proportions of repeat and FSSC-specific genes (Fig. 2c, d) and higher d_N/d_S ratio than other FCCs and CCs (Fig. 3a). Furthermore, frequent rearrangements were observed in chromosome 7 across FSSC genomes (Additional file 2: Fig. S5). Interestingly, intraspecies chromosomal structural variations were observed between *F. keratoplasticum* Fu6 and LHS11 in which translocations of *F. falciforme* Fu3 chromosome 7 were observed in chromosomes 3, 13, and 14 of *F. keratoplasticum* Fu6 but dissimilar in LHS11 (Additional file 2: Fig. S5). Hence, the shorter length of chromosome 7 of Fu6 suggested that accessory regions were lost and translocated to other chromosomes in the genome. Enriched clustering of single-copy orthologs and synteny were found between FCCs of the

F. falciforme Fu3 and *F. oxysporum* f. sp. *lycopercisi* strain 4287 genomes (Fig. 3b; Additional file 2: Fig. S9), but not other *Fusarium* species (Additional file 2: Fig. S10), suggesting the FCCs in the FSSCs were present in the last common ancestor of *Fusarium* species and were differentially maintained.

Interplay between DNA methylation and repeat in different compartments

Examination of 5-methylcytosine (5mC) in the DNA of four of our FSSC genomes revealed that methylation levels were different between species ranging from 3 to 6.6% (Fig. 4a). The distinct differences were not associated to phylogenetic positions, but higher methylation abundance was observed in the strains isolated from animals (Fig. 1b). Significantly higher DNA methylation levels were observed in FCCs and LSCs than in CCs (Wilcoxon rank sum test, $p < 0.0001$; Fig. 4a; Additional file 2: Fig. S11). Consistent with other fungi [43, 44], the 5mC methylation level in the FSSCs were typically higher in the repeat regions than in coding regions (Fig. 4b; Additional file 2: Fig. S12a to S14a). As expected, we found methylation levels along the genome were negatively and positively correlated with the presence of coding sequence and repeat region, respectively (Kendall's Tau correlation, $p < 0.0001$ in all four species; Fig. 4c; Additional file 2: Fig. S12b to S14b). Strikingly, we found uniformly low



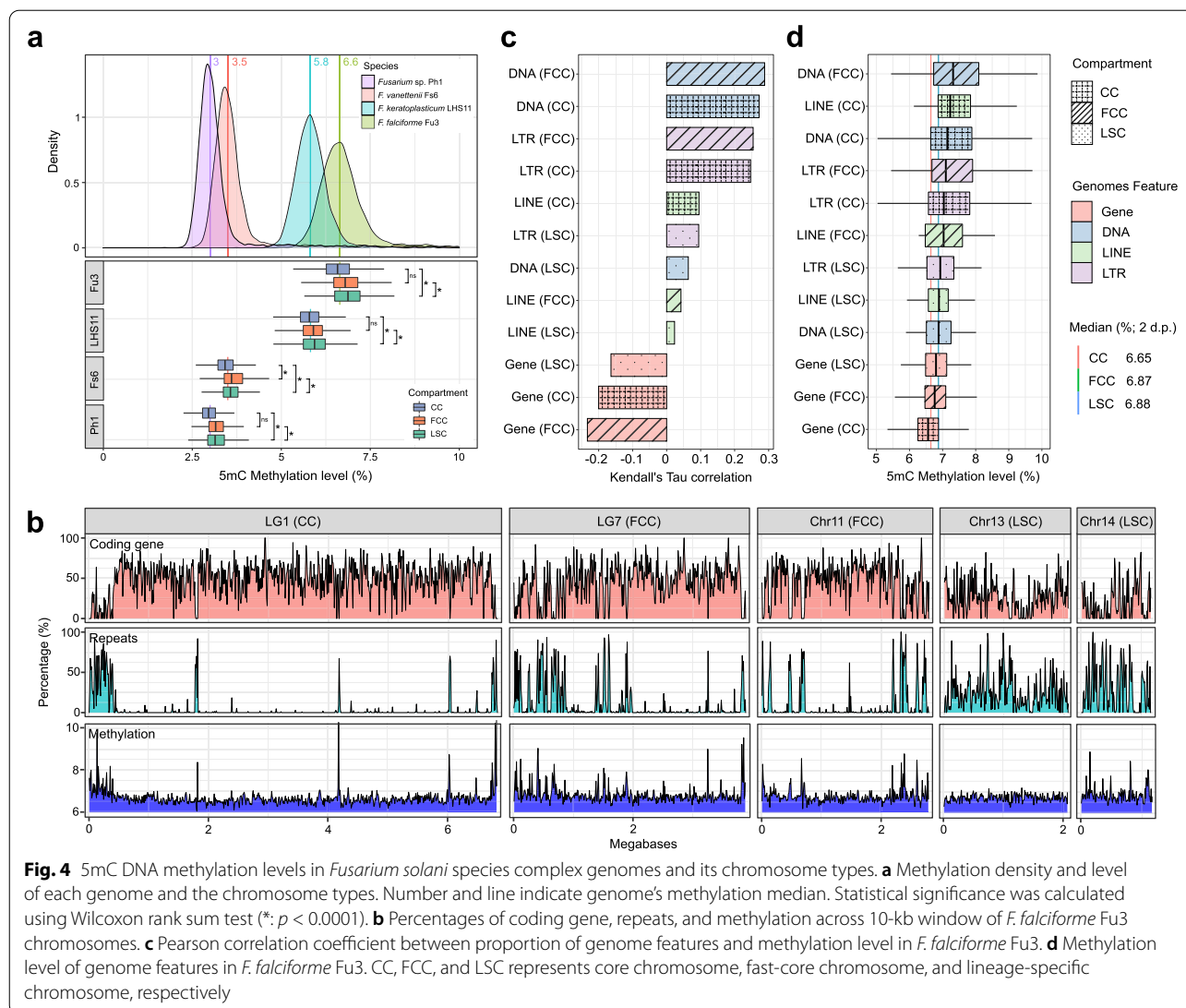
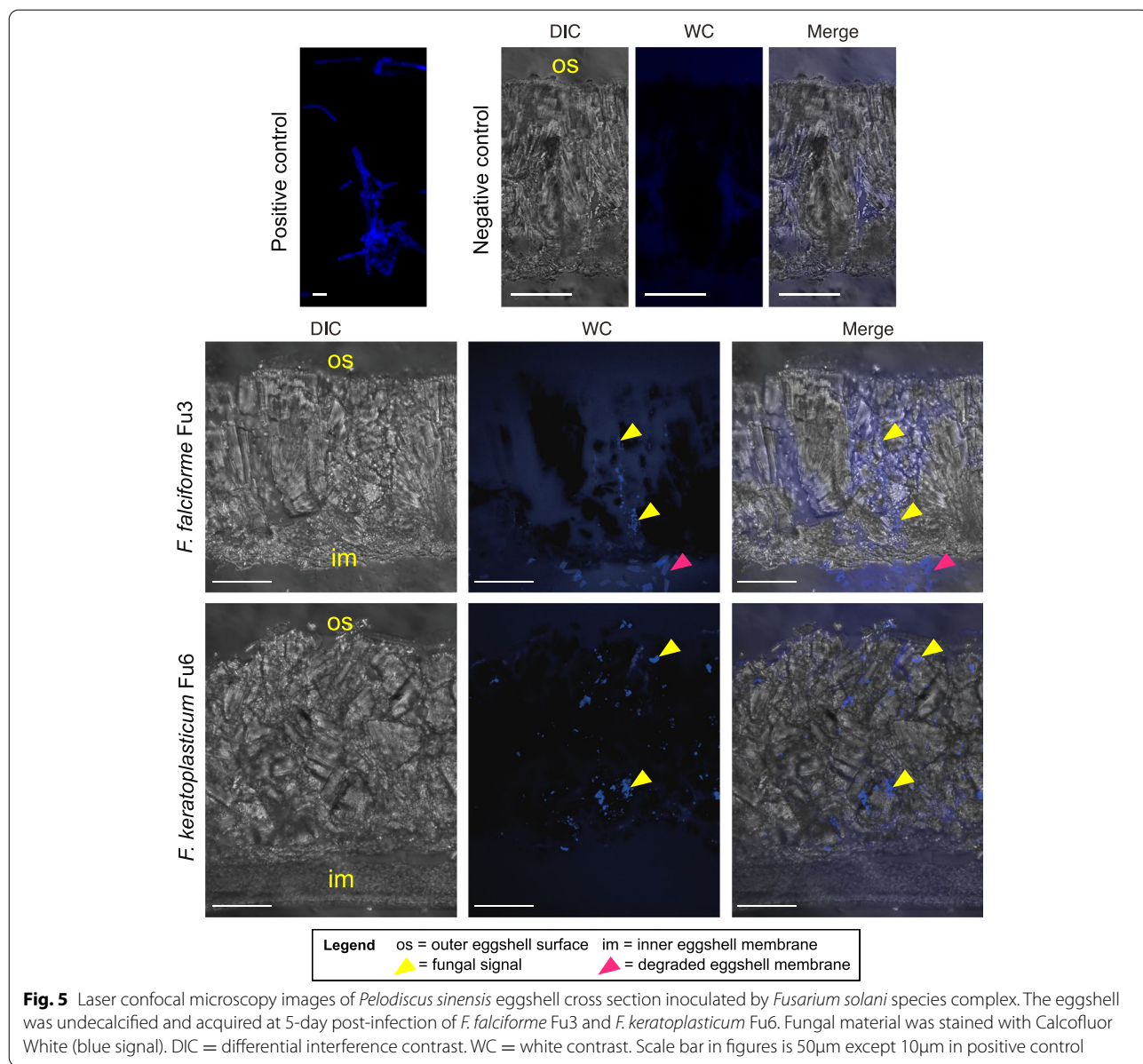


Fig. 4 5mC DNA methylation levels in *Fusarium solani* species complex genomes and its chromosome types. **a** Methylation density and level of each genome and the chromosome types. Number and line indicate genome's methylation median. Statistical significance was calculated using Wilcoxon rank sum test (*: $p < 0.0001$). **b** Percentages of coding gene, repeats, and methylation across 10-kb window of *F. falciforme* Fu3 chromosomes. **c** Pearson correlation coefficient between proportion of genome features and methylation level in *F. falciforme* Fu3. **d** Methylation level of genome features in *F. falciforme* Fu3. CC, FCC, and LSC represents core chromosome, fast-core chromosome, and lineage-specific chromosome, respectively

methylation levels across LSCs (example with chr. 13 and 14 in Fig. 4b; Additional file 2: Fig. S12a and S13a). In contrast to CCs and FCCs, methylation levels were similar between the repetitive elements and coding regions in LSCs (Tukey HSD test; Fig. 4d; Additional file 1: Table S13; Additional file 2: Fig. S12c to S14c), which were also consistent with the lower correlation between methylation level and genome features (Fig. 4c; Additional file 2: Fig. S12b to S14b). Together, the overall different methylation levels between the CCs and FCCs were due to the different markup of coding and repeat types (Fig. 2c; Additional file 2: Fig. S15). Despite higher repeat content in the LSCs, DNA methylation levels were uniform and overall similar to the FCCs.

***Fusarium falciforme* and *F. keratoplasticum* are opportunistic pathogens of turtle eggs**

F. falciforme Fu3 and *F. keratoplasticum* Fu6 are frequently isolated from sea turtle eggs [17, 22], and we seek to understand how these two species colonise and invade egg hosts by inoculating Chinese softshell turtle (*Pelodiscus sinensis*) eggs. These two *Fusarium spp.* did not actively seek hosts because attraction assays revealed no significant difference in hyphal growth rates with or without the presence of turtle eggs in each *Fusarium* species (Additional file 2: Fig. S16; Wilcoxon test, $p = 0.67$ in *F. falciforme* and $p = 0.86$ in *F. keratoplasticum*) or between species (Wilcoxon test, $p = 0.86$ in control and $p = 0.86$ in treatment). After 5 days of inoculation, hyphal growth in both species was



observed on the eggshell surface (Fig. 5) with some occasionally growing into the cavity-like structures (Additional file 2: Fig. S17). Cryosectioning and histological observations of undecalcified eggshell cross sections (see Additional file 3 for sample preparations) [45] revealed the presence of both *Fusarium* species on the outer, within the calcareous, and inner layers of the eggshells, confirming that hyphae vertically penetrated the eggshell. Degradations were sometimes observed on the eggshell membrane (Fig. 5). We further examined the symptoms of *F. falciforme* and *F. keratoplasticum* colonisation on eggs at 3 and 4 days post-inoculation

(dpi) (Additional file 2: Fig. S18). Mycelial mass was observed growing on the membrane (Additional file 2: Fig. S18a) and embryo (Additional file 2: Fig. S18b and c). Some inoculated eggs exhibited reduced branching points and disrupted blood capillaries on the microvascular system despite the embryo still being alive during the examination (Additional file 2: Fig. S18c). Together, these observations suggested that these two FSSC species are opportunistic animal pathogens and, upon contact, their hyphae can penetrate eggshells via natural openings and subsequently lyse and colonise egg inclusions.

Transcriptome profiles of FSSC pathogens during egg infection

To better understand how FSSC pathogens infect aquatic animals, we inoculated *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 on *Pelodiscus sinensis* eggs and compared transcriptome-wide gene expression data of both species. The inoculated samples exhibited blotch and membrane phenotypes (Additional file 2: Fig. S18) and were grouped together and distinct from the control samples in the principal component analyses (Additional file 2: Fig. S19 and S20). Globally, both species adopted similar colonisation and infection strategy at a transcriptome level while contacting eggs (Fig. 6a; Additional file 2: Fig. S21; Additional file 3) and we identified a total of 4111

(1823 up- and 2362 downregulated) and 4185 (2241 up- and 1870 downregulated) genes that were differentially expressed (DE; adjusted *p* value < 0.05) in *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 (Fig. 6b), respectively. When considering these genes regardless of expression level, each pathogen underwent differential responses during egg infection—GO enrichment revealed that *F. keratoplasticum* Fu6 exhibited reactions related to pathogenicity such as responses to host, stimulus, and toxic substances, cell adhesion, and regulation of immune system whereas *F. falciforme* Fu3 was mostly ribosome-associated processes such as biogenesis, maturation, and transport (Additional file 1: Table S14). We speculated a larger proportion of genes to be differentially expressed

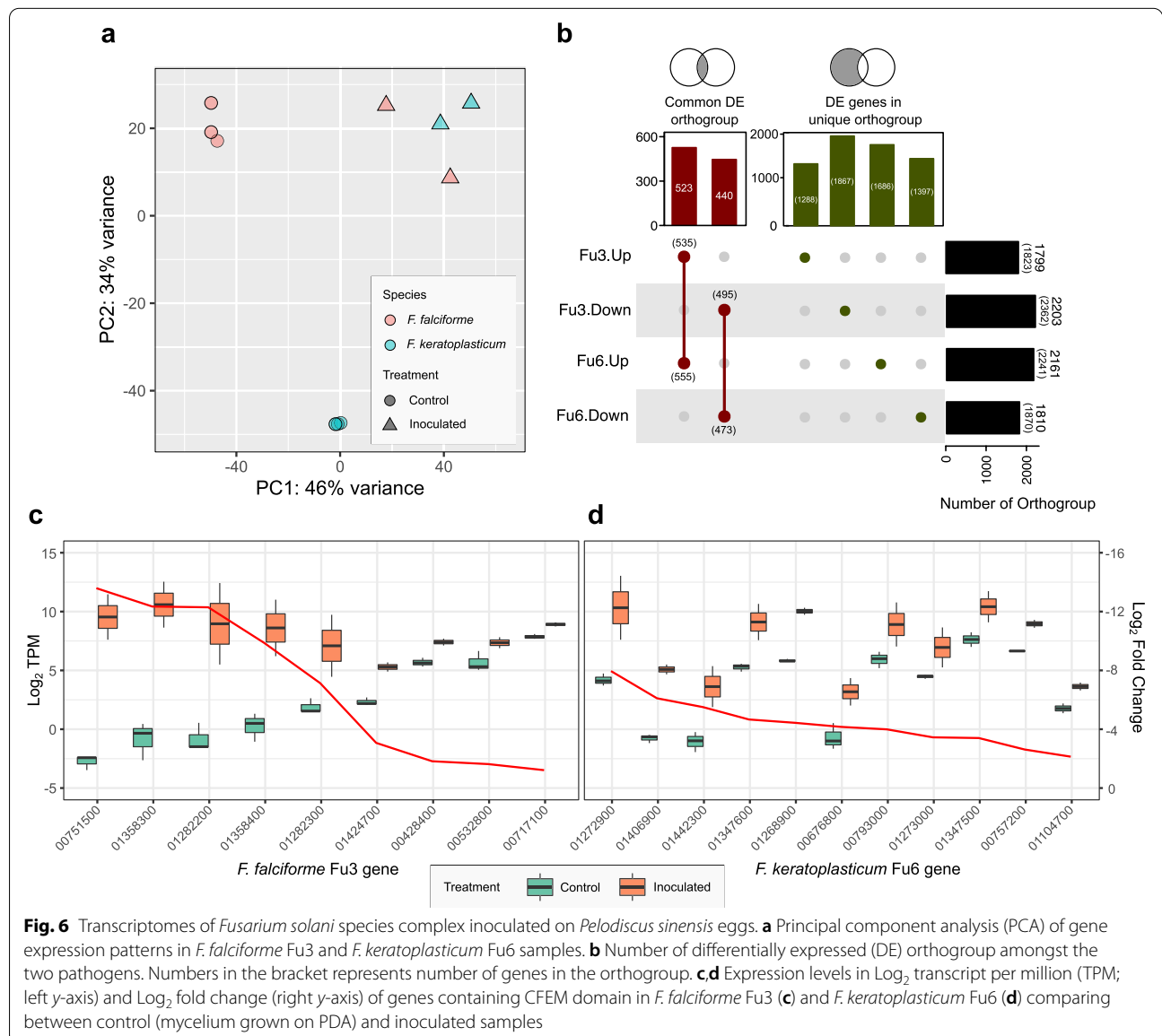


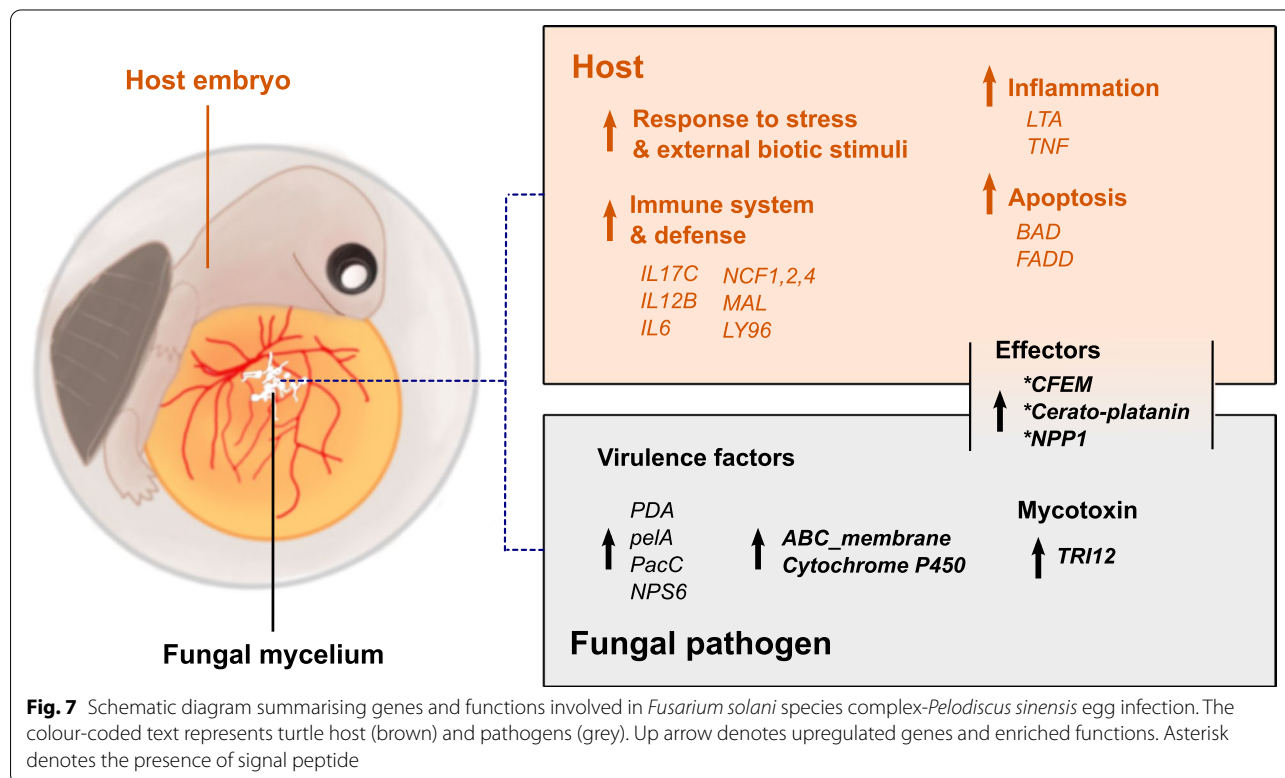
Fig. 6 Transcriptomes of *Fusarium solani* species complex inoculated on *Pelodiscus sinensis* eggs. **a** Principal component analysis (PCA) of gene expression patterns in *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 samples. **b** Number of differentially expressed (DE) orthogroup amongst the two pathogens. Numbers in the bracket represents number of genes in the orthogroup. **c,d** Expression levels in Log₂ transcript per million (TPM; left y-axis) and Log₂ fold change (right y-axis) of genes containing CFEM domain in *F. falciforme* Fu3 (**c**) and *F. keratoplasticum* Fu6 (**d**) comparing between control (mycelium grown on PDA) and inoculated samples

in FCCs and LSCs but observed a reverse pattern where the majority of such genes were located in CCs (Additional file 2: Fig. S22; chi-squared test: Fu3: $\chi^2_1 = 10.7$, $p = 0.005$; Fu6: $\chi^2_1 = 205.84$, $p < 0.001$). For instance, only 114 and no genes were DE in LSCs of *F. falciforme* Fu3 and *F. keratoplasticum* Fu6, respectively, indicating genes in FCCs and LSCs were not necessarily enriched during animal pathogenesis.

Several genes that have been previously identified to be involved in *Fusarium*-plant infection were upregulated in the egg-inoculated samples (Additional file 1: Table S15 and S16). The majority of these genes were predicted as effectors or contained a signal peptide. Of particular interest, genes annotated to contain a CFEM (Common in Fungal Extracellular Membrane) domain, a cysteine-rich protein domain found in diverse phytopathogenic fungal species, comprised some of the most DE genes upregulated in both *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 treatments (11 and 15 genes, respectively; Fig. 5c, d). Other examples included the effectors necrosis-inducing proteins (NPP1) and cerato-platanin (CP), which were required for virulence of *F. oxysporum* [46, 47]; ABC membrane and transporter, cytochrome P450 or termed pisatin demethylase (PDA), pectate lyases, whose deletion or inactivation reduced the virulence of *F. vanettenii* FSSC11 on pea [27, 48, 49]. The results suggested that a similar repertoire of genes was utilised

during infection regardless of host types by *F. falciforme* and *F. keratoplasticum*.

In addition, a total of 535 and 555 genes from *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 treatment, respectively belonging 523 OGs were co-upregulated (Fig. 6b). GO analysis and functional annotations of these genes revealed that both pathogens interacted with the host by positive regulations of their immune system processes and defence responses (Additional file 1: Table S17). These genes include a TRI12 encoding a major facilitator superfamily protein (MFS_1) involved in the export of mycotoxin trichothecene [50], a nonribosomal peptide synthetase (NPS6), a fungal effector involved in secondary metabolite biosynthesis producing AM-toxin, and PacC, a transcription factor-dependent of pH during pathogenesis (Fig. 7; Additional file 1: Table S18 and S19) [51, 52]. In contrast, 440 OGs downregulated in pathogens during egg infection were involved in transmembrane transports of metal ions, spore development, and growth (Additional file 1: Table S20). Furthermore, GO-enriched biological processes of species-specific upregulated genes were similar to GO enrichment of all upregulated DE genes in each pathogen species. Upon examination of these upregulated species-specific genes, it was found that 25.7 and 50.4% of the protein domains in genes of *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 respectively, were present in the co-upregulated DE genes



of both species (Additional file 1: Table S21 and S22), suggesting similar infective mechanism were adopted by both pathogen species in animal pathogenicity despite diverged gene sequences.

Gene expression profile of FSSC-infected animal host

Combined with transcriptome data of various development stages in the *P. sinensis* embryos [53], the presence (75.0–81.5%) of *F. falciforme*- and *F. keratoplasticum*-inoculated turtle RNA-seq reads in samples allowed us to determine host responses to the two pathogens (Additional file 1: Table S23). A large difference can be seen between the gene expression pattern of the host inoculated by FSSC and the natural developing host embryo on similar egg incubation days (PC1: 95% variance; Additional file 2: Fig. S23a), signifying host responded distinctly to FSSC infection. Besides, gene expression in host inoculated with either *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 were highly correlated ($R^2 = 0.98$, $p < 0.0001$; Additional file 2: Fig. S23b), suggesting the turtle host responded similarly to FSSC pathogens. In addition to genes that were related to embryo development, other upregulated genes were enriched in biological processes involving host immunity, response, and defence to another organism (Additional file 1: Table S24). Specifically, positive regulations of the immune response towards stress and external biotic stimulus were detected. These regulations included leukocyte activation, cytokine production, apoptotic process, and defence response (Fig. 7).

Discussion

Understanding the genetic diversity within FSSC and their ability to infect cross-kingdoms is of fundamental evolutionary interest and essential for the management of emerging wildlife diseases. Here, we produced highly contiguous assemblies for five FSSC species, established the first *Fusarium*-aquatic animal infection model which utilised high-throughput sequencing technology on Chinese softshell turtle (*Pelodiscus sinensis*) egg, and examined gene expression patterns of *F. falciforme* and *F. keratoplasticum* during infection on the animal host. We uncovered diverse evolutionary dynamics of FSSC chromosomes, and these variations were not associated with egg infection. The availability of these genomes serves as a useful reference for information regarding *Fusarium* evolution and FSSC opportunistic infection on animal host.

Comparative analyses of the six chromosome-level FSSC assemblies allowed the distinction of multiple chromosome compartments based on various structural features. Regions of CCs were highly conserved amongst the FSSC genomes and large synteny were also

detected in other non-FSSC species such as *F. oxysporum*, *F. graminearum*, and *F. fujikuroi*, indicating these are the CCs across *Fusarium* genera. When Fokkens et al. [38] compared 59 *F. oxysporum* strains, they coined the term ‘fast-core chromosome’ (FCC), which referred to the three smallest core chromosomes (chr. 11, 12, and 13) that contained features that were intermediate between core and lineage-specific chromosomes. Of these three, two chromosomes shared ancestries with the FCCs identified in our study between FSSCs, suggesting that these chromosomes were already evolving differently compared to CCs since the last common ancestor of *Fusarium* species. The FCCs were gene-rich, enriched in the number of effectors, CAZYmes and SMBGCs, and reduced in repeats compared to LSCs, but had a higher SNP density [38] and were less conserved at strain- and species-level comparisons than CCs. We hereby confirm and verify multi-compartmentalization in FSSC genomes includes fast-core chromosomes in addition to the more commonly discussed ‘two-speed genome’ comprising core and lineage-specific chromosomes in fungal pathogens [25, 26, 54].

Although the repeat portion of the fungal genome is typically accompanied by a higher abundance of methylated DNA bases [43], we found methylation was uniformly reduced along repetitive content of LSC of FSSC genomes. Enriched H3K27me3 and an increased methylation level were independently found in the FCCs of *F. oxysporum* [38] and FSSC, respectively. In *Arabidopsis thaliana*, H3K27me3 was exclusively associated with DNA methylation at transposable elements [55]. We speculate enrichment of both marks may be the hallmarks of all *Fusarium* FCCs, and the interplay of DNA methylation and H3K27me3 may mediate transcriptional responses in the enriched gene families involved in the pathogenicity processes [56]. Chromosomes harbouring similar features have been reported in other *Fusarium* species [38, 54], and additional sequencing may identify and elucidate the molecular mechanisms of FCCs as they displayed differential evolutionary dynamics in each species. DNA methylation levels of FSSC genomes were surprisingly higher than most previously determined ascomycetes in Bewick et al. [43], which coincides with the only exception of high methylation, *Pseudogymnoascus destructans*, which is also an animal pathogen. Further comparisons are needed to elucidate the potential association between DNA methylation and animal virulence in fungal pathogens.

Some FSSC species are model systems for cell biology [57], biocatalytic applications [58], and the most extensively studied plant pathosystem that involves *F. vanettenii* on pea [59]. Our attraction assays indicated no signs of egg attraction suggesting chance-encounter

of FSSC pathogens on eggs, unlike some pathogens that seek hosts actively in the environment [60, 61]. During plant infection, hyphae of fungal pathogens penetrate host tissue through natural openings such as stomata or lenticels [62]. We observed the initial stage of disease development which include spore germination, hyphal extension, and colonisation of *F. falciforme* and *F. keratoplasticum* on turtle eggs, both on the eggshell and internal embryo. While eggshell serves as a protection for the developing embryos, we observed both FSSC pathogens were capable of invading egg inclusions through the eggshell and caused tissue degradation of egg membrane. Previous study had also shown calcium depletion of sea turtle eggshell post-infection by *F. solani* [63], suggesting lytic activity of FSSC pathogens. Nevertheless, the eggshell of *P. sinensis* is thicker than sea turtle eggshell with an expanded calcareous layer [64], emphasising that egg penetration may have played a more primary role in establishing infection at least for the *P. sinensis* host.

Conserved pathogenicity traits such as effectors in pathogen of different strains might play a role in infecting the same host [65]. In our egg infection experiment, only a few differences were found between gene expression and overall regulated biological processes of both FSSC pathogens. Interestingly, we detected that plant pathogenicity-related genes were also upregulated in animal infection. These include carbohydrate-active enzymes such as cellulase which functions to degrade plant cell wall materials, and pathogenicity genes containing CFEM domain, which were few of the top most expressed genes in both pathogens. CFEM is found in diverse phytopathogenic fungal species with various virulent roles such as appressoria formation in *Pyricularia oryzae* (syn. *Magnaporthe oryzae*) [66] and root colonisation in *F. oxysporum* [67]. The most possible explanation for its role in animal infection is biofilm formation, as seen in *Candida albicans* [68] and white blotches surrounding the embryo. Nevertheless, the role of CFEM domain and plant virulence genes in animal infection remained to be elucidated. This combined evidence suggested that FSSC species may adopt a similar repertoire of genes in establishing infection across kingdoms. Intriguingly, genes involved in pathogenesis are usually enriched and highly expressed during the infection process in the FCCs or LSCs of *Fusarium* species [11, 38, 42, 69, 70] but we found no such association in *F. falciforme* and *F. keratoplasticum* during the egg infection. The differences suggest that the evolution of FSSC genomes may be shaped by lifestyles other than animal parasitism and that the differentially expressed genes during the infection process thus represented a more generic infection process.

Conclusions

In this study, we identified three compartments in the genomes of the *Fusarium solani* species complex (FSSC) that appear to be the hallmarks of *Fusarium* chromosome evolution. Moreover, the reference quality FSSC genomes offer prospects for future research in pathogenesis regardless of the host species. We present a landscape of transcriptome profiling of FSSC genes during egg infection, some of which were also highly expressed when infecting plant hosts. The combined results provided new insights into the genomic characteristics of animal-infecting FSSC species and their disease development, particularly on turtle eggs. This research represents the beginning of critical steps to understanding FSSC infection on turtles towards data-informed decision making and management of disease epidemics to reduce disease occurrences in the wild and man-managed settings.

Methods

Nucleic acid isolation, genome and transcriptome sequencing

Six isolates from FSSC clade 3 were chosen for genome sequencing and comparative analyses (see Additional file 3 for fungal culturing conditions and species identification of isolates) [14, 22, 71]. Fresh mycelium from the cultures was scraped off from the culture media for gDNA and RNA isolation. Genomic DNA was extracted following protocols designed for high-molecular-weight gDNA sequencing [72] and size selection and purification of isolated gDNA was performed following [73]. The integrity of gDNA was evaluated using Fragment Analyzer 5200 (Advanced Analytic Technologies, Inc., Ankeny, USA) and the fragment lengths were determined using PROsize 2 software (Advanced Analytic Technologies, Inc., Ankeny, USA). Genomes were sequenced using both Illumina and Oxford Nanopore platforms. Detailed information such as sequencing platforms, library kits used, and sequence accession number for each sample can be found in Additional file 1: Table S25. The summary of DNA sequencing data is shown in Additional file 1: Table S1. For gene model prediction and annotation, RNA of the six isolates were extracted following the TRIzol reagent protocol (Thermo Fisher Scientific, Waltham, MA, USA, Cat. #15596026). The integrity of the isolated RNA was checked using 1.5% agarose gel, and quantity was determined using Invitrogen Qubit[®] 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) before library preparation and sequencing (Additional file 1: Table S25).

Genome assembly and annotations

Guppy (v3.2.4 and v4.0.11; Oxford Nanopore Technology) or albacore (for 1D² reads; v2.2.7; Oxford Nanopore Technology) were used to perform basecalling of Nanopore raw signals (Additional file 1: Table S1) and assembled using Flye (v2.5) [30] and further polish and correct the assembled sequences with consensus mapped reads from Illumina using Pilon (v1.22) [74]. Haplotypes were collapsed with HaploMerger2 (ver. 20180603) [75]. Genome haploid length was estimated via GenomeScope 2.0 [76]. Annotation of repeat elements of the genomes was performed following Berriman et al. [77] except described otherwise in the following: a consensus repeat library was created using repeat elements identified via RepeatModeler (v.open-1.0.8) and TransposonPSI (release 08222010; <http://transposonpsi.sourceforge.net/>) and merged using usearch (v8.1.1861) [78]. RepeatMasker (v.open-4.0.7; option -s; <https://www.repeatmasker.org>) was used to mask the predicted repeat regions in each genome. Telomeric repeats of each scaffold were determined using Tandem Repeat Finder (v4.09; default parameters) [79]. Enriched hexamers were identified (TTAGGG) manually using Python script on the terminal 5 kb regions of each scaffold to confirm the presence of telomeres. Genes of the assemblies were predicted using AUGUSTUS (v3.3.3) [80] and trained with BRAKER2 (option fungi and softmasked; v2.1.4) [81]. MAKER2 (v2.31.9) [32] was then used to combine evidence from the assembled transcripts, reference proteomes, and transcript evidence obtained from RNA-seq of the mycelium to produce a final gene annotation set. Completeness of each genome was assessed by BUSCO (v5.2.2; Table 1) [33]. Functional annotations of the amino acid sequences were carried out using eggNOG-mapper v2 (<http://eggno-mapper.embl.de>; default parameters) [34] on eggNOG v5 database [82]. Protein domains and families were identified using pfam_scan.pl (v1.6; <http://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/>) on Pfam database (release 34.0) [83]. Additional annotations were carried out as follows: carbohydrate-active enzymes were determined using dbCAN2 (v2.0.11) [84]; secondary metabolite detection via antiSMASH (v6.0) [85]; and fungal effectors were predicted using EffectorP (v3.0) [86] on amino acid sequences which passed the signal peptide prediction via signalP (v6.0) [87].

Comparative genomic analyses

Orthology was assigned using OrthoFinder (v2.3.8) [35] with proteomes of six species from this study, 17

other *Fusarium* published assemblies from and an out-group species *Beauveria bassiana* (Additional file 1: Table S2). A total of 2385 single-copy orthogroups (OG) were determined from proteomes of 24 species and were aligned using MAFFT (v7.487; option -max-iterate 1000) [88]. The alignments from each of the OG were sent to RAxML-NG for maximum likelihood phylogeny inference (v0.9.0; option --model LG+I+F+G4, --bs-trees 100) [89]. The maximum likelihood trees and bootstrap-supported replicates generated from the previous step were combined for the construction of consensus species tree using ASTRAL-III (v5.7.7; option -r 100) [90]. Lastly, a maximum likelihood phylogeny from concatenated amino acid alignments of single-copy OG was built using RAxML-NG (v0.9.0; option --model LG+I+F+G4, --bs-trees 100) [89] with 100 bootstrap replicates. For d_N/d_S analysis, codon alignments of one-to-one orthologs between *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 were produced using TranslatorX (v1.1) [91]. Ratio of d_N/d_S were calculated for each ortholog alignment using PAML² of CODEML programme (v4.9e; option runmode=-2, seqtype=1, CondonFreq=3, and fix_omega=0) [92].

Core, fast-core, and lineage-specific chromosomes assignment

Synteny analysis between FSSC species (i.e. *F. falciforme* Fu3 vs. *F. vanettenii* 77-13-4) and between FSSC and non-FSSC species were carried out using MUMmer4 [93] and PROmer (v3.23) [94], respectively (see Additional file 3). From the synteny analysis, a total of 12 conserved linkage groups were assigned unambiguously in the six FSSC genomes, and these were defined as core chromosomes. Lineage-specific chromosomes were assigned based on lack of synteny regions linked via one-to-one gene with Fu3 chromosomes (Additional file 2: Fig. S5). Contigs of less than 450kb were not included in this assignment due to an unclassifiable region via single-copy orthologues or genome alignment via nucmer [93]. The excluded contigs comprised 0.3–3.5% of assemblies and ranged from 11 to 404 genes. The designated lineage-specific chromosome includes chromosomes 13 and 14 of Fu3, chromosomes 13, 14, and 15 of Fu6 and Ph1, chromosomes 13, 14, 15, and 16 of LHS11, chromosomes 13 to 19 of LHS14, and chromosomes 13, 14 to 23 of Fs6. From the core chromosomes, fast-core chromosomes were designated based on higher proportion of FSSC-specific gene (Fig. 2d). FSSC-specific gene were defined as genes with no orthologs identified in *F. fujikuroi* IM158289, *F. graminearum* PH-1, and *F. oxysporum* f. sp. *lycopercisi* 4287 genomes [24, 40, 95]. The designated fast-core chromosomes include chromosomes 7, 11, and 12 of all six FSSC assemblies.

Methylation analysis

To examine 5-methylcytosine (5mC) in DNA sequences, Nanopore raw FAST5 files of *F. falciforme* Fu3, *F. keratoplasticum* LHS11, *Fusarium* sp. Ph1, and *F. vanettenii* Fs6 were used to run Megalodon (v2.3.3; Oxford Nanopore Technologies) and Guppy (v5.0.11; Oxford Nanopore Technologies) using the default parameters. To determine the methylation level of each CpG site, we calculated the ratio of methylated reads including both strands. Calculation of methylation levels was carried out using a 10-kb window.

Animal inoculation experiments

We incubated softshelled turtle (*Pelodiscus sinensis*) eggs up to day 30, at which point the embryo was estimated to have developed to either stage TK21 to 22 [96]. Dead eggs were removed, and alive eggs were kept for experiments which included pathogen inoculation, host attraction assay, and eggshell observations. To determine if the pathogens *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 can be attracted to the egg host, an attraction assay experiment was performed and hyphal growth towards the host was measured. Eggshell fragments inoculated with either *F. falciforme* Fu3 or *F. keratoplasticum* Fu6 pathogens were collected for histological observations using scanning electron microscopy (SEM) and laser scanning confocal microscopy (LSCM) for invasion route. For transcriptome sequencing of FSSC infection on *P. sinensis* eggs, we injected 10^7 spores/mL suspension into the eggs and incubated it for another 3 to 4 days before RNA isolation. Detailed methods can be found in Additional file 3.

Total RNA isolation, library preparation, and sequencing

RNA isolation of the animal inoculation experiment was carried out according to the TRIzol reagent protocol. At the cell lysis step, each sample in 2mL screw-cap tube was mixed with six to eight 0.8-mm stainless steel beads, flash-freezing the tubes in liquid nitrogen, and homogenised using PowerLyzer24 homogeniser (MoBio Laboratories, Carlsbad, USA, Cat. #13155) set to $3000 \times g$ for 20 s and repeated at least twice to ensure the sample was homogenised. Isolated RNA was quantified using Invitrogen Qubit[®] 4 Fluorometer (Thermo Fisher Scientific, Waltham, USA) and checked for integrity in 1.5% agarose gel. Samples with adequate concentration and show no degradation was chosen for RNA sequencing. In total, 20 samples were sequenced, which include three *F. falciforme* Fu3 and three *F. keratoplasticum* Fu6 positive controls (mycelium from PDA media), seven *F. falciforme* Fu3 and seven *F. keratoplasticum* Fu6 inoculated samples (Additional file 1: Table S26). The inoculated samples were made up of four blotch and three membrane samples. Paired-end library was prepared for the RNA samples using NED Next[®] Ultra[™] RNA Library Prep Kit and

sequenced on Illumina NovaSeq 6000 instrument with 150 bp paired-end reads.

Analysis of RNA-seq reads

Raw RNA reads were trimmed to remove adaptor sequences and poor-quality reads using fastp (option -l 30; v0.20.1) [97]. Trimmed reads of each sample were mapped to their respective genome (*F. falciforme* Fu3 or *F. keratoplasticum* Fu6) according to the inoculation treatment and the host *P. sinensis* genome (GCF_000230535.1_PelSin_1.0) [53] using STAR (v2.7.7a) [98]. To ensure RNA reads which mapped on the *Fusarium* genomes were not from the host, reads mapped onto both *Fusarium* and host genome and had lower CIGAR scores in *Fusarium* were excluded from further analyses. Raw read count of each gene was calculated using featureCounts (-p -s 2; v2.0.1) [99]. We first examined the inoculated samples from the blotch and membrane to see if they had a dissimilar gene expression pattern through principal component analysis (PCA) and found that 43% of the variances could be explained between these two sample types in either treatment (Additional file 2: Fig. S19). Since our aim is to determine the general pattern of FSSC and animal pathogenesis, we grouped the blotch and membrane samples as 'inoculated samples' for further analyses. Differentially expressed genes (DEGs) between inoculated and control samples in *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 were inferred using DESeq2 (*padj* < 0.05; v1.24) [100]. Functional enrichment of the DEGs was identified using the 'topGO' (v2.36.0) [101] package. We used the same pipeline as described for the analysis of the host's RNA reads to determine the DEGs of the inoculated host using the *P. sinensis* genome [53] as reference. For comparison, the RNA-seq dataset of *P. sinensis* embryos from different embryo developmental stages (TK19 and TK23 defined as in [96] generated by [53]) was chosen as the control because our samples were collected between TK21 to 22. Of those, 12,419 genes were differentially expressed (adjusted *p* value < 0.05), with 5815 up- and 6604 downregulated genes during the infection experiment.

Statistical analyses

We used R package 'ggpubr' [102] to perform Wilcoxon rank sum test for mean comparisons of groups in analyses which include dN/dS, proportions of repeat content and FSSC-specific genes, methylation levels, and host attraction assay. For correlation analyses, we used R package 'stats' [103] Kendall's Tau method to compare methylation levels amongst the genome features and used Pearson method to compare orthologous gene expression between *F. falciforme* Fu3 and *F. keratoplasticum* Fu6. Chi-squared test of R package 'stats' [103] was used to compare proportion of differentially expressed genes in different chromosome types.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-022-01436-7>.

Additional file 1: Table S1. Read summary of genome assemblies from each sequencing platform. **Table S2.** *Fusarium* genomes and an outgroup genome used in the phylogenomic analysis and comparative studies. **Table S3.** Scaffold assembly completeness. **Table S4.** FSSC genome assemblies used to compare N90. **Table S5.** Number of genes with predicted signal peptide and effector using SignalP and EffectorP, respectively. **Table S6.** Number of genes with predicted carbohydrate-active enzymes (CAZymes) and its categories using dbCAN2. **Table S7.** Number of secondary metabolites biosynthetic gene cluster detected via antiSMASH. **Table S8.** Type and presence of secondary metabolite products in each FSSC genomes predicted via antiSMASH. **Table S9.** *Fusarium solani* species complex (FSSC) sequences used in multi-locus sequence typing (MLST) phylogenetic tree in Fig. S2. **Table S10.** Structural features of FSSC genome assemblies. **Table S11.** Gene number and proportion of effector, carbohydrate active enzymes and secondary metabolite gene cluster in each FSSC chromosome. **Table S12.** Average number of genes predicted to encode for effector, carbohydrate-active enzymes (CAZymes) and secondary metabolite biosynthetic gene clusters (SMBGCs) of FSSC genomes by chromosome types. **Table S13.** Multiple mean comparisons of methylation levels among the chromosome type and genome features in *F. falciforme* Fu3. **Table S14.** Top 50 enriched biological processes of FSSC pathogen *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 during *P. sinensis* host infection. **Table S15.** Subset of differentially expressed and upregulated genes of *F. falciforme* Fu3 during *P. sinensis* host infection previously known to associate with plant host infection. **Table S16.** Subset of differentially expressed and upregulated genes of *F. keratoplasticum* Fu6 during *P. sinensis* host infection previously known to associate with plant host infection. **Table S17.** Selected enriched biological processes of co-upregulated genes of *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 during *P. sinensis* host infection. **Table S18.** Top 50 differentially expressed genes of *F. falciforme* Fu3 which were co-upregulated with *F. keratoplasticum* Fu6 (Table S19) during *P. sinensis* host infection. **Table S19.** Top 50 differentially expressed genes of *F. keratoplasticum* Fu6 which were co-upregulated with *F. falciforme* Fu3 (Table S18) during *P. sinensis* host infection. **Table S20.** Selected enriched biological processes of co-downregulated genes of *F. falciforme* Fu3 and *F. keratoplasticum* Fu6 during *P. sinensis* host infection. **Table S21.** Top 50 differentially expressed genes of *F. falciforme* Fu3 which were unique to *F. keratoplasticum* Fu6 (by orthogroup) during *P. sinensis* host infection. **Table S22.** Top 50 differentially expressed genes of *F. keratoplasticum* Fu6 which were unique to *F. falciforme* Fu3 (by orthogroup) during *P. sinensis* host infection. **Table S23.** RNA-seq and read statistics of FSSC-animal infection experiment. **Table S24.** Upregulated biological processes of *P. sinensis* host infected by pathogen *F. falciforme* and *F. keratoplasticum*. **Table S25.** Sequencing information of gDNA and RNA for genome assembly and gene annotation. **Table S26.** All sequenced samples and metadata for inoculation experiment.

Additional file 2: Fig. S1. Genome features in FSSC assemblies. **Fig. S2.** Multi-locus sequence typing (MLST) phylogeny tree of FSSC. **Fig. S3.** Genome phylogeny of *Fusarium* species. **Fig. S4.** Orthologue sharing amongst *Fusarium* chromosome. **Fig. S5.** FSSC genome synteny. **Fig. S6.** Synteny between *F. vanettenii* Fs6 and FVANE. **Fig. S7.** Location of FSSC-specific genes across genomes. **Fig. S8.** Selection level in FSSC chromosomes. **Fig. S9.** Orthogroup number comparison between *F. falciforme* Fu3 and *F. oxysporum* f. sp. *lycopercisi* 4287. **Fig. S10.** Syntenic dotplot between FSSC and non-FSSC species. **Fig. S11.** Methylation level of each chromosome in FSSC. **Fig. S12.** Methylation analysis of *F. keratoplasticum* LHS11. **Fig. S13.** Methylation analysis of *Fusarium* sp. Ph1. **Fig. S14.** Methylation analysis of *F. vanettenii* Fs6. **Fig. S15.** Gene density in 10 kb window per chromosome or chromosome types. **Fig. S16.** Host attraction assay. **Fig. S17.** Scanning electron microscopy images. **Fig. S18.** Chinese soft-shelled turtle *P. sinensis* egg inoculated with conidia of *F. falciforme* Fu3 or *F. keratoplasticum* Fu6. **Fig. S19.** Principal component analyses of pathogens' gene expression pattern of inoculated samples. **Fig. S20.** Principal component analyses of pathogens' gene expression pattern of all samples. **Fig. S21.** Correlation plots of log₂ normalized transcript per

million (TPM) of one-to-one orthologous gene of *F. falciforme* Fu3 and *F. keratoplasticum* Fu6. **Fig. S22.** Distribution of all differentially expressed genes of FSSC pathogens during egg inoculation experiment. **Fig. S23.** Gene expression pattern of the animal host *P. sinensis*.

Additional file 3. Extended Results and Methods.

Acknowledgements

We thank the Department of Fisheries Malaysia, World Wide Fund for Nature Malaysia for sampling permission in Padang Kemunting Turtle Hatchery, Melaka, and Maximus Cheing YC for collecting Fu3 and Fu6 strains. We are grateful to Dr Stéphane Hacquard, Dr Francis Martin, and the '1000 Fungal Genomes-Deep Sequencing of Ecologically-relevant Dikarya' project for access to unpublished genome data. The genome sequence data were produced by the US Department of Energy Joint Genome Institute in collaboration with the user community. We thank the National Center for High-performance Computing (NCHC) for providing computational and storage resources. We thank High Throughput Genomics Core, Biodiversity Research Center Academia Sinica for the sequencing service; Division of Core Facilities Imaging, Institute of Cellular and Organismic Biology, Academia Sinica for SEM sample preparation service; Dr. Po-Yang Chen and Division of Electron Microscope, Cell Biology Core Lab, Institute of Plant and Microbial Biology, Academia Sinica for assistance in using SEM; Dr. Yoko Nozawa for providing histological materials and Neuro-imaging Core, Neuroscience Program of Academia Sinica for LSM imaging service.

Authors' contributions

IJT conceived and led the study; DZH conducted the experiments; DZH, WAL, HMK, and PFS isolated samples' nucleic acids; DZH, NW, and SLT conducted the laser confocal microscopy experiments; DZH and IJT performed assemblies and annotations of FSSC genomes; DZH and IJT carried out comparative genomic analyses; DZH, HHL, and IJT carried out the methylation analyses; CKL carried out the d_n/d_s analyses; DZH designed and conducted the inoculation experiments; DZH and MRL analysed the RNA-seq data; WHC and YLC provided FSSC strains. DZH and IJT wrote the manuscript with input from CLC and others. All authors read and approved the final manuscript.

Funding

This work was supported by Academia Sinica grant (AS-CDA-107-L01) and National Science and Technology Council (111-2628-B-001-021-) to IJT. DZH and PFS are supported by the doctorate fellowship of the Taiwan International Graduate Program, Academia Sinica of Taiwan.

Availability of data and materials

All sequences generated from this study were deposited on NCBI under BioProject PRJNA782245 [104] and accession number of gene sequences from MLST of isolates can be found in Additional file 1: Table S9.

Declarations

Ethics approval and consent to participate

Animal experiments were evaluated and approved by the Institutional Animal Care and Use Committee (IACUC) of Academia Sinica, Taiwan, under protocol ID 18-05-1214. Any alive embryos remaining after all experiments were humanely euthanised following the standardised procedures approved by IACUC.

Consent for publication

Not applicable.

Competing interests

Not applicable.

Author details

¹Biodiversity Research Center, Academia Sinica, 115 Nangang, Taipei, Taiwan. ²Biodiversity Program, Taiwan International Graduate Program, Academia Sinica and National Taiwan Normal University, Taipei, Taiwan. ³Department of Life Science, National Taiwan Normal University, 116 Wenshan, Taipei, Taiwan. ⁴Genome and Systems Biology Degree Program, National Taiwan

University and Academia Sinica, Taipei, Taiwan. ⁵Department of Plant Pathology, National Chung Hsing University, Taichung, Taiwan. ⁶Department of Plant Pathology and Microbiology, National Taiwan University, Taipei 10617, Taiwan.

Received: 30 March 2022 Accepted: 13 October 2022
Published online: 20 October 2022

References

- 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(6):2186–90.
- O'Donnell K, Sutton DA, Wiederhold N, Robert VARG, Crous PW, Geiser DM. Veterinary Fusarioses within the United States. *J Clin Microbiol.* 2016;54(11):2813–9.
- Dean R, Van Kan Ja L, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 2012;13(4):414–30.
- Al-Hatmi AMS, Meis JF, de Hoog GS. *Fusarium*: molecular diversity and intrinsic drug resistance. *Plos Pathog.* 2016;12(4):e1005464.
- Walther G, Stasch S, Kaerger K, Hamprecht A, Roth M, Cornely OA, et al. *Fusarium Keratitis* in Germany. *J Clin Microbiol.* 2017;55(10):2983–95.
- Cabañas FJ, Alonso JM, Castellá G, Alegre F, Domingo M, Pont S. Cutaneous hyalohyphomycosis caused by *Fusarium solani* in a loggerhead sea turtle (*Caretta caretta* L.). *J Clin Microbiol.* 1997;35(12):3343–5.
- Kano R, Okayama T, Hamamoto M, Nagata T, Ohno K, Tsujimoto H, et al. Isolation of *Fusarium solani* from a dog: identification by molecular analysis. *Med Mycol.* 2002;40:435–7.
- Fernando N, Hui SW, Tsang CC, Leung SY, Ngan AHY, Leung RWW, et al. Fatal *Fusarium solani* species complex infections in elasmobranchs: the first case report for black spotted stingray (*Taeniura melanopsila*) and a literature review. *Mycoses.* 2015;58(7):422–31.
- Brown T, Millar Z, Evans D, Pham PH, LePage V, Lumsden JS. *Fusarium solani* haplotype 12-b and aortic and branchial arteritis in *Hippocampus erectus* Perry. *J Fish Dis.* 2020;43(2):301–4.
- Hsu LH, Su CY, Sun PL, Chen YL. *Fusarium solani* species complex infection in elasmobranchs: a case report for rough-tail stingray with valid antifungal therapy. *Med Mycol Case Rep.* 2021;32:34–8.
- Coleman JJ. The *Fusarium solani* species complex: ubiquitous pathogens of agricultural importance. *Mol Plant Pathol.* 2016;17(2):146–58.
- Rep M, Kistler HC. The genomic organization of plant pathogenicity in *Fusarium* species. *Curr Opin Plant Biol.* 2010;13(4):420–6.
- Fisher MC, Daniel A H, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, et al. Emerging fungal threats to animal, plant and ecosystem health. *Nature.* 2012;484(7393):186–94.
- O'Donnell K, Sutton DA, Fothergill A, McCarthy D, Rinaldi MG, Brandt ME, et al. Molecular phylogenetic diversity, multilocus haplotype nomenclature, and in vitro antifungal resistance within the *Fusarium solani* species complex. *J Clin Microbiol.* 2008;46(8):2477–90.
- Schroers HJ, Samuels GJ, Zhang N, Short DPG, Juba J, Geiser DM. Epitypification of *Fusiformium (Fusarium) solani* and its assignment to a common phylogenetic species in the *Fusarium solani* species complex. *Mycologia.* 2016;108(4):806–19.
- Short DP, O'Donnell K, Geiser DM. Clonality, recombination, and hybridization in the plumbing-inhabiting human pathogen *Fusarium keratoplaticum* inferred from multilocus sequence typing. *BMC Evol Biol.* 2014;14(1):91.
- Sarmiento-Ramírez JM, Abella-Pérez E, Phillott AD, Sim J, van West P, Martín MP, et al. Global distribution of two fungal pathogens threatening endangered sea turtles. *Plos One.* 2014;9(1):e85853.
- García-Martín JM, Sarmiento-Ramírez JM, Diéguez-Urbeondo J. Beyond Sea Turtles: *Fusarium keratoplaticum* in eggshells of *Podocnemis unifilis*, a threatened Amazonian freshwater turtle. *J Fungi.* 2021;7(9):742.
- Sarmiento-Ramírez JM, Abella E, Martín MP, Tellería MT, López-Jurado LF, Marco A, et al. *Fusarium solani* is responsible for mass mortalities in nests of loggerhead sea turtle, *Caretta caretta*, in Boavista, Cape Verde: *Fusarium solani* in sea turtle. *FEMS Microbiol Lett.* 2010;312(2):192–200.
- Smyth CW, Sarmiento-Ramírez JM, Short DPG, Diéguez-Urbeondo J, O'Donnell K, Geiser DM. Unraveling the ecology and epidemiology of an emerging fungal disease, sea turtle egg fusariosis (STEF). *Plos Pathog.* 2019;15(5):e1007682.
- Phillott AD, Parmenter CJ, Limpus CJ. Occurrence of mycobiota in eastern Australian sea turtle nests. *Memoir Queensl Mus.* 2004;49:701–3.
- Hoh DZ, Lin YF, Liu WA, Sidique SNM, Tsai IJ. Nest microbiota and pathogen abundance in sea turtle hatcheries. *Fungal Ecol.* 2020;47:100964.
- Sidique SNM, Azuddin NF, Joseph J. First report of *Fusarium* species at nesting sites of endangered sea turtle in Terengganu and Melaka, Malaysia. *Malays Appl Biol.* 2017;46(3):195–205.
- Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature.* 2010;464(7287):367–73.
- Bertazzoni S, Williams AH, Jones DA, Syme RA, Tan KC, Hane JK. Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic fungi. *Mol Plant Microbe Interact.* 2018;31(8):779–88.
- Möller M, Stukenbrock EH. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol.* 2017;15(12):756–71.
- Wasmann CC, VanEtten HD. Transformation-mediated chromosome loss and disruption of a gene for pisinin demethylase decrease the virulence of *Nectria haematococca* on pea. *Mol Plant Microbe Interact.* 1996;9(9):793–803.
- Han Y, Liu X, Benny U, Kistler HC, VanEtten HD. Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen *Nectria haematococca*. *Plant J.* 2001;25(3):305–14.
- Temporini E, VanEtten H. Distribution of the pea pathogenicity (PEP) genes in the fungus *Nectria haematococca* mating population VI. *Curr Genet.* 2002;41(2):107–14.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37(5):540–6.
- Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, et al. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *Plos Genet.* 2009;5(8):e1000618.
- Holt C, Yandell M. MAKER2: an annotation pipeline and genome-data-base management tool for second-generation genome projects. *BMC Bioinformatics.* 2011;12(1):491.
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 2021;38(10):4647–54.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38(12):5825–9.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238.
- Ma LJ, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, Trail F, et al. *Fusarium* Pathogenomics. *Annu Rev Microbiol.* 2013;67(1):399–416.
- Niehaus EM, Münsterkötter M, Proctor RH, Brown DW, Sharon A, Idan Y, et al. Comparative “omics” of the *Fusarium fujikuroi* species complex highlights differences in genetic potential and metabolite synthesis. *Genome Biol Evol.* 2016;8(11):3574–99.
- Fokkens L, Shahi S, Connolly LR, Stam R, Schmidt SM, Smith KM, et al. The multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events. *BioRxiv.* 2018:465070 Available from: <https://www.biorxiv.org/content/10.1101/465070v1>.
- Zhang Y, Yang H, Turra D, Zhou S, Ayhan DH, Delulio GA, et al. The genome of opportunistic fungal pathogen *Fusarium oxysporum* carries a unique set of lineage-specific chromosomes. *Commun Biol.* 2020;3(1):50.
- Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science.* 2007;317(5843):1400–2.
- Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, et al. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *Plos Genet.* 2008;4(4):e1000046.

42. Armitage AD, Taylor A, Sobczyk MK, Baxter L, Greenfield BPJ, Bates HJ, et al. Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepaee*. *Sci Rep*. 2018;8(1):13530.
43. Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Grigoriev IV, James TY, et al. Diversity of cytosine methylation across the fungal tree of life. *Nat Ecol Evol*. 2019;3(3):479–90.
44. Ke HM, Lee HH, Lin CY, Liu YC, Lu MR, Hsieh JWA, et al. *Mycena* genomes resolve the evolution of fungal bioluminescence. *Proc Natl Acad Sci*. 2020;117(49):31267–77.
45. Wada N, Kawamoto T, Sato Y, Mano N. A novel application of a cryosectioning technique to undecalcified coral specimens. *Mar Biol*. 2016;163:117.
46. Gijzen M, Nürnberg T. Nep1-like proteins from plant pathogens: recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry*. 2006;67(16):1800–7.
47. Liu S, Wu B, Yang J, Bi F, Dong T, Yang Q, et al. A cerato-platanin family protein FocCP1 is essential for the penetration and virulence of *Fusarium oxysporum* f. sp. *cupense* Tropical Race 4. *Int J Mol Sci*. 2019;20(15):3785.
48. Rogers LM, Kim YK, Guo W, Gonzalez-Candelas L, Li D, Kolattukudy PE. Requirement for either a host- or pectin-induced pectate lyase for infection of *Pisum sativum* by *Nectria haematococca*. *Proc Natl Acad Sci*. 2000;97(17):9813–8.
49. Coleman JJ, White GJ, Rodriguez-Carres M, VanEtten HD. An ABC Transporter and a Cytochrome P450 of *Nectria haematococca* MPVI Are Virulence Factors on Pea and Are the Major Tolerance Mechanisms to the Phytoalexin Pisatin. *Mol Plant Microbe Interact*. 2011;24(3):368–76.
50. Alexander NJ, McCormick SP, Hohn TM. TRI12, a trichothecene efflux pump from *Fusarium sporotrichioides*: gene isolation and expression in yeast. *Mol Gen Genet MGG*. 1999;261(6):977–84.
51. Caracul Z, Roncero MIG, Espeso EA, González-Verdejo CI, García-Mac-eira FI, Di Pietro A. The pH signalling transcription factor PacC controls virulence in the plant pathogen *Fusarium oxysporum*: PacC controls virulence in *Fusarium*. *Mol Microbiol*. 2003;48(3):765–79.
52. Lee BN, Kroken S, Chou DYT, Robbertse B, Yoder OC, Turgeon BG. Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress. *Eukaryot Cell*. 2005;4(3):545–55.
53. Wang Z, Pascual-Anaya J, Zadissa A, Li W, Niimura Y, Huang Z, et al. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet*. 2013;45(6):701–6.
54. Wang Q, Jiang C, Wang C, Chen C, Xu JR, Liu H. Characterization of the two-speed subgenomes of *Fusarium graminearum* reveals the fast-speed subgenome specialized for adaption and infection. *Front Plant Sci*. 2017;8:140.
55. Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G, Schubert D, et al. Loss of the DNA methyltransferase MET1 induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in *Arabidopsis thaliana*. *Plos Genet*. 2012;8(11):e1003062.
56. Reyes-Dominguez Y, Boedi S, Sulyok M, Wiesenberger G, Stoppacher N, Krška R, et al. Heterochromatin influences the secondary metabolite profile in the plant pathogen *Fusarium graminearum*. *Fungal Genet Biol*. 2012;49(1):39–47.
57. Aist JR, Bayles CJ. Ultrastructural basis of mitosis in the fungus *Nectria haematococca* (sexual stage of *Fusarium solani*). *Protoplasma*. 1991;161:111–22.
58. Winkler M, Kaplan O, Vejvoda V, Klempier N, Martinková L. Biocatalytic application of nitrilases from *Fusarium solani* O1 and *Aspergillus niger* K10. *J Mol Catal B Enzym*. 2009;59(4):243–7.
59. Hadwiger LA. Pea–*Fusarium solani* interactions contributions of a system toward understanding disease resistance. *Phytopathology*. 2008;98(4):372–9.
60. Ruiz F, Castelletto ML, Gang SS, Hallem EA. Experience-dependent olfactory behaviors of the parasitic nematode *Heligmosomoides polygyrus*. *Plos Pathog*. 2017;13(11):e1006709.
61. Van Rooij P, Martel A, Haesebrouck F, Pasmans F. Amphibian chytridiomycosis: a review with focus on fungus-host interactions. *Vet Res*. 2015;46(1):137.
62. Ikeda K, Park P, Nakayashiki H. Cell biology in phytopathogenic fungi during host infection: commonalities and differences. *J Gen Plant Pathol*. 2019;85(3):163–73.
63. Phillott AD, Parmenter CJ, McKillup SC. Calcium depletion of eggshell after fungal invasion of sea turtle eggs. *Chelonian Conserv Biol*. 2006;5(1):146.
64. Kusuda S, Yasukawa Y, Shibata H, Saito T, Doi O, Ohya Y, et al. Diversity in the matrix structure of eggshells in the testudines (Reptilia). *Zoolog Sci*. 2013;30(5):366–74.
65. Williams AH, Sharma M, Thatcher LF, Azam S, Hane JK, Sperschneider J, et al. Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum* formae speciales facilitates identification of candidate effectors. *BMC Genomics*. 2016;17(1):191.
66. Kou Y, Tan YH, Ramanujam R, Naqvi NI. Structure–function analyses of the Pth11 receptor reveal an important role for CFEM motif and redox regulation in rice blast. *New Phytol*. 2017;214(1):330–42.
67. Ling J, Zeng F, Cao Y, Zhang J, Chen G, Mao Z, et al. Identification of a class of CFEM proteins containing a new conserved motif in *Fusarium oxysporum*. *Physiol Mol Plant Pathol*. 2015;89:41–8.
68. Pérez A, Pedrós B, Murgui A, Casanova M, López-Ribot JL, Martínez JP. Biofilm formation by *Candida albicans* mutants for genes coding fungal proteins exhibiting the eight-cysteine-containing CFEM domain. *FEMS Yeast Res*. 2006;6(7):1074–84.
69. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev*. 2015;35:57–65.
70. Yang H, Yu H, Ma LJ. Accessory Chromosomes in *Fusarium oxysporum*. *Phytopathology*. 2020;110(9):1488–96.
71. Liu YJ, Whelen S, Hall BD. Phylogenetic relationships among ascomycetes: evidence from an RNA polymerase II subunit. *Mol Biol Evol*. 1999;16(12):1799–808.
72. Mayjonade B, Gouzy J, Donnadieu C, Pouilly N, Marande W, Callot C, et al. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques*. 2016;61(4):203–5.
73. Chouikh Y, Volovitch M, Yot P. A simple and fast electrophoretic method for elution of nucleic acids from gels. *Mol Biol Rep*. 1979;5(4):237–9.
74. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *Plos One*. 2014;9(11):e112963.
75. Huang S, Kang M, Xu A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*. 2017;33(16):2577–9.
76. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11(1):1432.
77. Berriman M, Coghlan A, Tsai IJ. Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. *Protoc Exch*. 2018; Available from: <http://www.nature.com/protocolexchange/protocols/6761>.
78. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460–1.
79. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80.
80. Hoff KJ, Stanke M. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc Bioinforma*. 2018;65(1):e57.
81. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma*. 2021;3(1):1–11.
82. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309–14.
83. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res*. 2021;49(D1):D412–9.
84. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46(W1):W95–101.

85. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 2021;49(W1):W29–35.
86. Sperschneider J, Dodds PN. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol Plant Microbe Interact.* 2022;35(2):146–56.
87. Teufel F, Armenteros JJA, Johansen AR, Gíslason MH, Piñi SI, Tsirigos KD, et al. SignalP 6.0 achieves signal peptide prediction across all types using protein language models. *BioRxiv.* 2021; Available from: <https://www.biorxiv.org/content/10.1101/2021.06.09.447770v1>.
88. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
89. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* 2019;35(21):4453–5.
90. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 2018;19(6):153.
91. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 2010;38:W7–13.
92. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91.
93. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *Darling AE, editor. plos Comput Biol.* 2018;14(1):e1005944.
94. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5(2):R12.
95. Wiemann P, Sieber CMK, von Bargen KW, Studt L, Niehaus EM, Espino JJ, et al. Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *Plos Pathog.* 2013;9(6):e1003475.
96. Tokita M, Kuratani S. Normal embryonic stages of the chinese softshelled turtle *Pelodiscus sinensis* (Trionychidae). *Zoolog Sci.* 2001;18(5):705–15.
97. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics.* 2018;34(17):i884–90.
98. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
99. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923–30.
100. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
101. Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO. *Bioconductor.* 2019;27:1–26.
102. Kassambara A, Kassambara M. Package ‘ggpubr’. R package version, vol. 0; 2020. p. 1.
103. R Core Team. R: a language and environment for statistical computing. 2022. Available from: <https://www.R-project.org/>
104. Academia Sinica. *Fusarium solani* species complex Clade 3, Genomic basis of animal pathogenicity in *Fusarium solani* species complex: NCBI:BioProject PRJNA782245; 2022. Available from: <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA782245>

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

