


# Lifestyle Transitions in Fusarioid Fungi are Frequent and Lack Clear Genomic Signatures

Rowena Hill <sup>\*,1,2</sup> Richard J.A. Buggs,<sup>1,2</sup> Dang Toan Vu,<sup>3</sup> and Ester Gaya<sup>1</sup>

<sup>1</sup>Comparative Fungal Biology, Royal Botanic Gardens Kew, Jodrell Laboratory, Richmond, United Kingdom

<sup>2</sup>School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK

<sup>3</sup>Research Planning and International Cooperation Department, Plant Resources Center, Hanoi, Vietnam

\*Corresponding author: E-mails: R.Hill@kew.org; rowena.c.hill@gmail.com.

Associate editor: Crystal Hepp

## Abstract

The fungal genus *Fusarium* (Ascomycota) includes well-known plant pathogens that are implicated in diseases worldwide, and many of which have been genome sequenced. The genus also encompasses other diverse lifestyles, including species found ubiquitously as asymptomatic-plant inhabitants (endophytes). Here, we produced structurally annotated genome assemblies for five endophytic *Fusarium* strains, including the first whole-genome data for *Fusarium chuoi*. Phylogenomic reconstruction of *Fusarium* and closely related genera revealed multiple and frequent lifestyle transitions, the major exception being a monophyletic clade of mutualist insect symbionts. Differential codon usage bias and increased codon optimisation separated *Fusarium sensu stricto* from allied genera. We performed computational prediction of candidate secreted effector proteins (CSEPs) and carbohydrate-active enzymes (CAZymes)—both likely to be involved in the host–fungal interaction—and sought evidence that their frequencies could predict lifestyle. However, phylogenetic distance described gene variance better than lifestyle did. There was no significant difference in CSEP, CAZyme, or gene repertoires between phytopathogenic and endophytic strains, although we did find some evidence that gene copy number variation may be contributing to pathogenicity. Large numbers of accessory CSEPs (i.e., present in more than one taxon but not all) and a comparatively low number of strain-specific CSEPs suggested there is a limited specialisation among plant associated *Fusarium* species. We also found half of the core genes to be under positive selection and identified specific CSEPs and CAZymes predicted to be positively selected on certain lineages. Our results depict fusarioid fungi as prolific generalists and highlight the difficulty in predicting pathogenic potential in the group.

**Key words:** *Fusarium*, fungal endophytes, plant pathogens, lifestyle evolution, effectors, CAZymes.

## Introduction

*Fusarium* (Hypocreales, Ascomycota) is a globally distributed genus of approximately 230 species (<https://www.fusarium.org/>), many of which are implicated in devastating fungal diseases of plants. For instance, throughout the first half of the 20th century, *Fusarium* wilt of banana single-handedly wiped out the main globally traded banana cultivar—equivalent to losses of at least US\$2.3 billion in 2000 (Ploetz 2005). A new *Fusarium* epidemic is now affecting the current dominant banana cultivar (Ordóñez et al. 2015). Moreover, on the much-cited list of the top 10 fungal plant pathogens by Dean et al. (2012), two spots belong to *Fusarium* species. Beyond plant pathogenicity, however, many species are also reported to exhibit an array of other fungal lifestyles (see [supplementary table S1, Supplementary Material](#) online), and *Fusarium* strains are also frequently isolated from inside healthy plant tissues (e.g., Parsa et al. 2016; Zakaria and Aziz 2018; Rashmi et al. 2019). Fungal inhabitants of plant tissues which cause no symptoms of disease are

known as fungal endophytes, hyperdiverse microfungi that are omnipresent in plant microbiomes (Rodríguez et al. 2009; Hardoim et al. 2015).

There is no single role that endophytes play in the plant host, as the endophytic lifestyle represents a functional range between pathogenicity and mutualism, which has been dubbed the “endophytic continuum” (Schulz and Boyle 2005). The outcome of endophyte colonisation can be highly dependent on the context of the plant–fungal interaction, such as the status of the plant immune system and nutrient conditions (Junker et al. 2012; Lahrman et al. 2015; Hacquard et al. 2016; Hiruma et al. 2016), as well as the presence of other endophytes within the microbiome (Redman et al. 2001; Durán et al. 2018; Mesny et al. 2021; Wolinska et al. 2021) and even light conditions (Álvarez-Loayza et al. 2011). The transient status of endophytism for many taxa is evident from observations of endophytes becoming decayers (saprotrophs) or pathogens following some change in host or abiotic conditions (Slippers and Wingfield 2007; Arnold et al. 2009;

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Promptuttha et al. 2010; Swett and Gordon 2015; Nelson et al. 2020). In some cases, however, an evolutionary transition from pathogenicity to endophytism may represent a permanent switch to obligate commensalism or mutualism (Gazis et al. 2016), and it has also been hypothesised that endophytism may have been an ancestral “waiting room” for the evolution of mycorrhizal symbiosis (Selosse et al. 2018).

The need to categorise pathogenic potential of *Fusarium* taxa is obvious considering the ubiquity of *Fusarium* endophytes in our crops (e.g., Rubini et al., 2005; Leoni et al., 2013; Sandoval-Denis et al., 2018) and the ramifications of pathogenic *Fusarium* strains for food security (e.g., Menzies et al. 1990; Kokkonen et al. 2010; Okello et al. 2020). In >200 years since *Fusarium* was first described, the generic concept has been the source of lively debate (Summerell 2019). In recent years, many *Fusarium* species complexes have been reclassified into distinct “fusarioid” genera based on phenotypic and phylogenetic evidence—such as *Albonectria*, *Bisifusarium*, *Cyanonectria*, *Geejayessia*, *Neocosmospora*, and *Rectifusarium* (Schroers et al. 2011; Lombard et al. 2015; Sandoval-Denis et al. 2019)—resulting in a narrower definition of the genus, *Fusarium sensu stricto*. This has been opposed in some quarters, with the argument that retaining a broader definition of the genus (*Fusarium sensu lato*) is desirable to facilitate communication between scientists and practitioners dealing with agriculturally and clinically important species that have historically been classified under *Fusarium* (Geiser et al. 2021; O’Donnell et al. 2020). Crous, Lombard, et al. (2021) countered that, in light of ever-increasing species discovery and recognised chemical and morphological differences between these clades, reclassification of certain species complexes into different genera is both biologically and practically meaningful. However, both sides of the debate note that ecology is similar among many of these taxa, and so questions regarding lifestyle warrant a perspective that includes allied fusarioid genera.

An evolutionary genomics approach using genomes from diverse lifestyles of fusarioid fungi could address this issue of detecting where strains fall on the pathogenic-mutualistic spectrum. A phylogenomic framework could not only shed light on the timing and frequency of lifestyle transitions in the group, but also inform to what extent genetic content is shared between taxa due to ancestry versus lifestyle. In addition to comparing gene repertoires, detecting signatures of selection may also help to uncover the genetic basis of recently evolved traits. Methods based on the ratio of nonsynonymous to synonymous substitutions (dN/dS) and the phenomenon of codon usage bias—where certain codons appear more frequently than others despite encoding the same amino acid—can be used to investigate the extent of selection acting on gene content.

One genetic feature that can be particularly illuminating to compare between lifestyles is genes that encode effector proteins. Fungal effectors (known as candidate-secreted effector proteins [CSEPs] when computationally predicted)

are small secreted proteins produced by fungi which mediate the plant–fungal interaction. While best-studied in the context of pathogenicity (Stergiopoulos and de Wit 2009; De Jonge et al. 2011), we now know that effectors are also essential for mutualistic or commensal fungi to form associations with plant hosts by evading the host immune response (Rafiqi et al. 2012; Plett and Martin 2015; Lo Presti et al. 2015). Effector repertoires have been shown to differentiate host-specific strains (*forma specialis*) in the *Fusarium oxysporum* species complex (FOSC) (van Dam et al. 2016), and could potentially further distinguish pathogenic and endophytic FOSC strains (Czislowski et al. 2021). Another frequently studied group of proteins involved in the plant–fungal interaction are carbohydrate-active enzymes (CAZymes), many of which act as plant cell wall–degrading enzymes (PCWDEs) (Kubicek et al. 2014). CAZymes are often referred to as saprotrophic features (Lebreton et al. 2021), but are also abundant in plant pathogens and endophytes (e.g., Zhao et al. 2013; Knapp et al. 2018; Mesny et al. 2021), and, although present in lower numbers in mycorrhizal fungi (Kohler et al. 2015; Peter et al. 2016; Miyauchi et al. 2020), certain CAZymes play key roles in the establishment and maintenance of the symbiosis (Veneault-Fourrey et al. 2014; Doré et al. 2017; Marqués-Gálvez et al. 2021). Comparing CSEP and CAZyme repertoires is therefore highly relevant to exploring genetic differences in plant associated lifestyles of fusarioid fungi.

Here, we performed whole genome sequencing, assembly, and structural annotation of five novel endophytic *Fusarium* strains (supplementary table S2, Supplementary Material online), including the first whole genome sequencing data and annotated assembly for the recently described species, *Fusarium chuoi* (Crous, Osieck, et al. 2021). Using predicted genes from these and other publicly available fusarioid strains, we produced a genome-scale phylogeny of *Fusarium* and allied genera with time calibration and compared CSEP and CAZyme content to answer the following questions: 1) How are lifestyles distributed across the phylogeny? 2) Can we distinguish plant pathogens and endophytes from genome sequences alone? and 3) How is selection acting on different lifestyles?

## Results

### Both Single- and Multi-Copy Genes Inferred the Same Backbone for *Fusarium s. str.*

To infer the genome-scale phylogeny of *Fusarium*, we used both concatenation and coalescent-based approaches, using single-copy genes with and without multi-copy genes also included (see our bioinformatics pipeline in supplementary fig. S1, Supplementary Material online). Including multi-copy genes had a greater impact on topology than tree building approach (i.e., concatenation versus coalescent) (fig. 1A). This was seen chiefly from a change in divergence order of allied genera—*Neocosmospora* (= *Fusarium solani* species

concept, “FSSC”), *Geejayessia* (= *Fusarium staphyleae* species concept, “FSTSC”), and *Albonectria* (= *Fusarium decemcellulare* species concept, “FDESC”)—when including multi-copy genes (fig. 1B). All methods, however, produced the same divergence order for *Fusarium s. str.* species concepts. Disregarding differences in the naming of species, our estimations of *Fusarium s. str.* from 1,060 loci were in broad agreement with the most recent phylogenetic analyses by Crous, Lombard, et al. (2021) and Geiser et al. (2021).

We additionally compared the impact of alignment trimming tools—trimAl (Capella-Gutiérrez et al. 2009) versus BMGE (Crisuolo and Grimaldo 2010)—on species tree topology. The RAxML-NG (Kozlov et al. 2019) species tree was identical for both trimming tools, but trimming tool impacted topology for IQ-TREE (Minh et al. 2020) and ASTRAL-III (Zhang, Rabiee, et al. 2018), with discordance in the ambrosia clade of *Neocosmospora* (supplementary fig. S2, Supplementary Material online). The gene trees trimmed with trimAl were selected for downstream analyses based on its reported accuracy relative to BMGE in the literature (Tan et al. 2015; Steenwyk et al. 2020). The RAxML-NG species tree was selected for downstream analyses as its topology was identical for both trimming tools while having branch length units as substitutions per site as opposed to coalescent units.

### Dated Genome-Scale Phylogeny of *Fusarium* and Allied Genera

For divergence time estimation of the RAxML-NG species tree, we used both the independent-rates (IR) and auto-correlated-rates (AR) relaxed clock models, implemented in MCMCTree (Yang and Rannala 2006). Testing best-fit of clock models in MCMCTree (see dos Reis et al., 2018) is not possible using amino acid data, and so our assessment of divergence time estimation from the two clock models was restricted to comparisons against previous studies. The IR model generally shifted nodes towards more recent divergence times in comparison to the AR model (supplementary fig. S3, Supplementary Material online). The crown age of *Fusarium s. lat.* was estimated to fall in the late Cretaceous by both the IR (71 Ma) and AR (84 Ma) models, although the latter was closer to the estimate by O’Donnell et al. (2013) (83 Ma). The crown age of *Fusarium s. str.* estimated in the Eocene (49 Ma) by the same study was much closer to our result from the IR model (51 Ma) compared with the AR model (69 Ma, late Cretaceous). The middle Miocene crown age of the ambrosia clade in *Neocosmospora* from previous estimates by Kasson et al. (2013) (13 Ma) and O’Donnell et al. (2015) (9 Ma) were also in closer agreement with the IR model (7 Ma) compared with the AR model (25 Ma). The crown age of Xyleborini beetle hosts estimated by Jordal and Cognato (2012) (21 Ma) corresponded more closely with the IR estimate of the divergence of the ambrosia clade from non-insect mutualists (15 Ma) compared with the AR estimate (41 Ma). The dating of the

diversification of various *formae speciales* in the FOSC by our IR model was also a better fit with their crop hosts having been domesticated within the last ~10,000 years (Meyer et al. 2012).

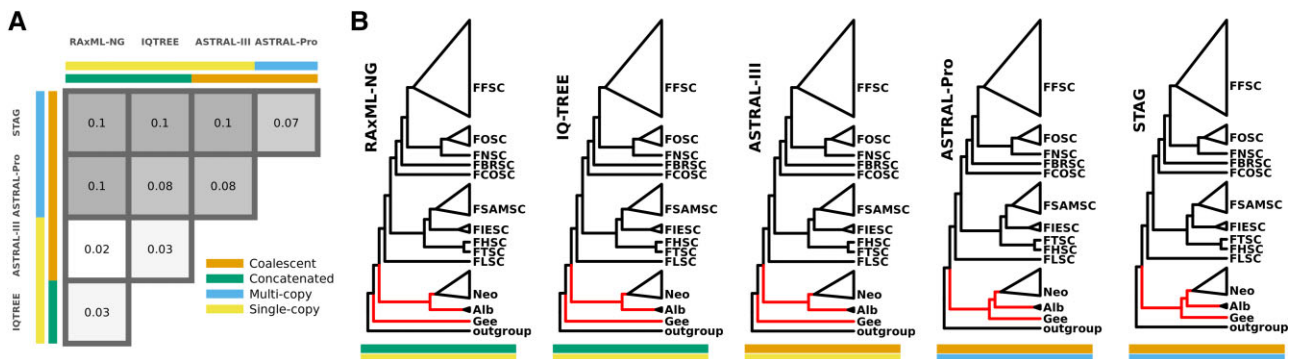
### Gene, CSEP, and CAZyme Repertoires were Broadly Shared Across Lifestyles, But Plant Pathogens Included Copy Number Outliers

There was no significant difference in number of genes, CSEPs or CAZymes across lifestyles (supplementary table S3, Supplementary Material online). Most genes, CSEPs and CAZymes were either core (present in all fusarioid taxa) or accessory (present in more than one taxon but not all), with very few being strain-specific, indeed strain-specific CAZymes being almost non-existent (fig. 2A). The number of strain-specific genes or CSEPs was not significantly different across lifestyles (supplementary fig. S4A, supplementary table S3, Supplementary Material online). Global pairwise permutational analysis of variance (PERMANOVA) showed that gene, CSEP, and CAZyme content were better described by phylogenetic relatedness (35–42% variance) than lifestyle (9% variance) (fig. 2B, supplementary table S4, Supplementary Material online). Nonetheless, pairwise PERMANOVA identified the insect mutualist lifestyle as the most genetically distinct, with insect mutualist taxa having significantly different gene, CSEP, and CAZyme repertoires compared with all other lifestyles other than mycoparasite. While most other lifestyles were genetically similar, endophytes and saprotrophs were also found to be significantly different in terms of CSEPs. In a similar pattern to the number of strain-specific genes, mean gene, CSEP, and CAZyme copy number were not found to be significantly different between lifestyles (supplementary fig. S4B, supplementary table S3, Supplementary Material online), but there were extreme outliers in copy number amongst plant pathogens (fig. 2C). The greatest copy number outlier by a considerable margin was predicted to be both a CSEP and CAZyme belonging to *F. oxysporum* f. sp. *conglutinans*, annotated as a glycosyltransferase in the GT4 family:  $\alpha,\alpha$ -trehalose phosphorylase (configuration-retaining) (EC 2.4.1.231).

### Almost Half of Core Single-Copy Genes were Under Positive Selection

While gene, CSEP, and CAZyme repertoires may have been broadly shared, we were interested in whether genes were evolving in a lifestyle-directed manner. Of the 1,054 core single-copy genes used in the selection analyses, 469 (44%) were found to be under episodic positive selection by both BUSTED (Murrell et al. 2015) and aBSREL (Smith et al. 2015) (fig. 3A). This included 11 of 31 (35%) core CSEPs and 6 of 11 (55%) core CAZymes. The branch at the root of the more conservative generic concept, *Fusarium s. str.*, was a particular “hotspot” of positive selection, with 52 core single-copy genes positively selected according to BUSTED and aBSREL (supplementary fig. S5,





**Fig. 1.** (A) Pairwise comparison of normalised Robinson–Foulds distances between topologies from all species tree estimation methods, with grid cells coloured from most similar topology (lighter) to most dissimilar (darker). (B) Summary of species trees with red branches indicating topological discordance between methods. Labels indicate species complex (SC) or allied genus (FFSC = *Fusarium fujikuroi* species complex, FOOSC = *Fusarium oxysporum* species complex, FNSC = *Fusarium nisikadoi* species complex, FBRSC = *Fusarium burgessii* species complex, FCOSC = *Fusarium concolor* species complex, FSAMSC = *Fusarium sambucinum* species complex, FIESC = *Fusarium incarnatum-equiseti* species complex, FHSC = *Fusarium heterosporum* species complex, FTSC = *Fusarium tricinctum* species complex, FLSC = *Fusarium lateritium* species complex, Neo = *Neocosmospora*, Alb = *Albonectria*, Gee = *Geejayessia*).

Supplementary Material online). A few external branches also had a notably high number of positively selected core genes: insect mutualist *N. oligoseptata*; saprotrophic *F. culmorum* in the *F. sambucinum* species complex (FSAMSC); and plant pathogenic *F. oxysporum* f. sp. *lycopersici* in the FOOSC. There was no significant difference in the number of positively selected genes on external branches between lifestyles according to analysis of variance (ANOVA,  $P=0.7$ ; supplementary table S3, Supplementary Material online).

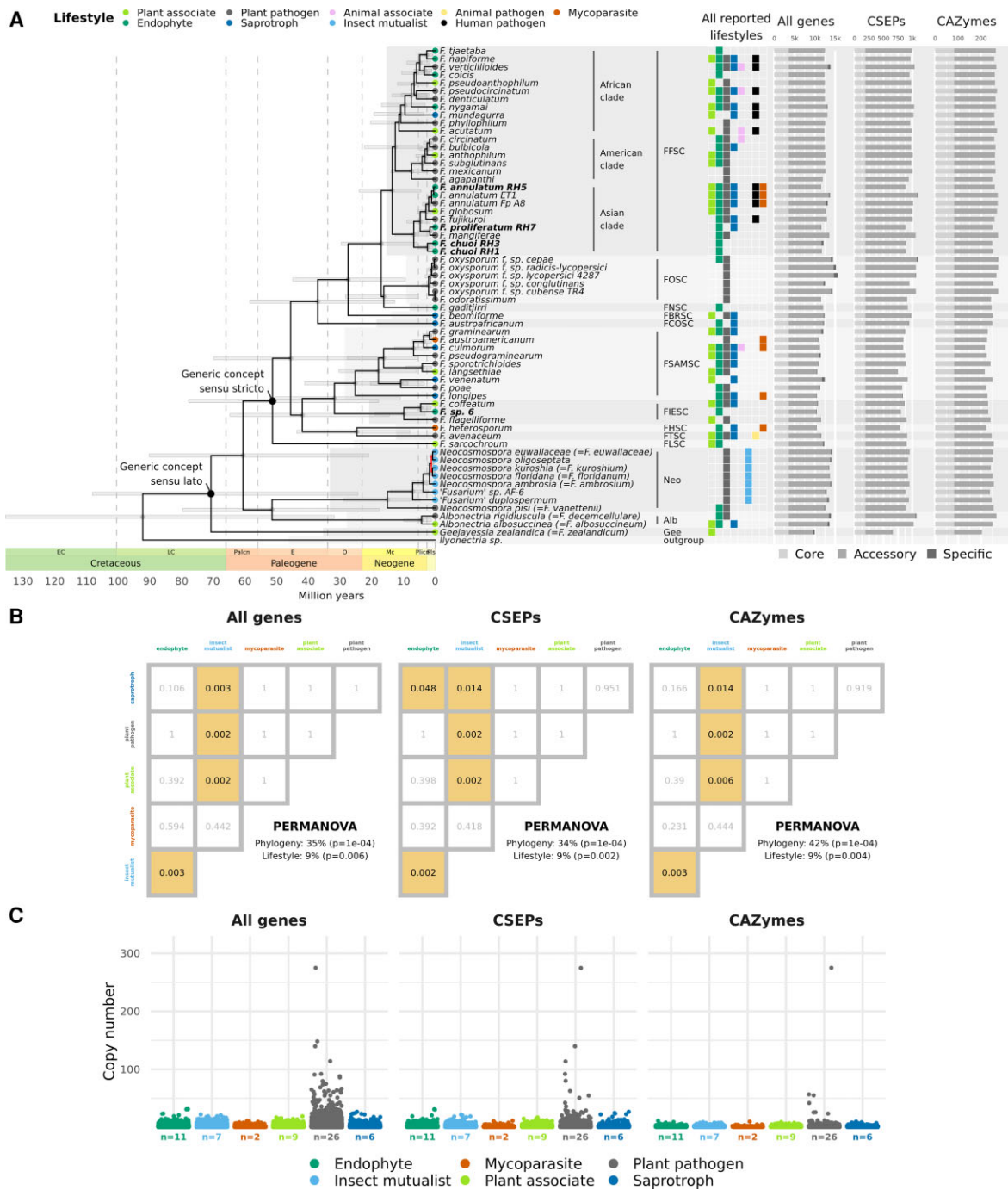
Although a minority of all CSEPs (11%) could be assigned known gene names using the PHI-base database (Urban et al. 2020), two core CSEPs with signatures of selection could be classified as known genes: 5680 as FGSG\_00806 and 6786 as FgPR-IL-2 (fig. 3A). Based on PHI-base records of gene knockouts in *F. graminearum* inoculated on wheat, both FGSG\_00806 and FgPR-IL-2 had the mutant phenotype of unaffected pathogenicity (supplementary fig. S6, Supplementary Material online). Of the six core CAZymes which had undergone positive selection, four are known to act on plant cell wall substrates (supplementary fig. S7, Supplementary Material online): glycoside hydrolase GH35 ( $\beta$ -galactosidase) on hemicellulose and pectin and GH51 (non-reducing end  $\alpha$ -L-arabinofuranosidase) on cellulose, hemicellulose and pectin; carbohydrate esterase CE12 (rhamnogalacturonan acetyltransferase) on pectin; and an enzyme of auxiliary activities AA3\_2 (5'-oxooverantoin cyclase) on lignin.

Most CSEPs and CAZymes reported as positively selected by both BUSTED and aBSREL were also found to contain sites with a higher relative selective pressure in certain lifestyles by Contrast-FEL (Kosakovsky Pond et al. 2021) (fig. 3B). In most cases only one site per gene was found to have a difference in relative selective pressure. The insect mutualist lifestyle had significantly more sites per gene under higher selective pressure compared with most other lifestyles (fig. 3B). We should emphasise that Contrast-FEL does not inform whether positive or negative

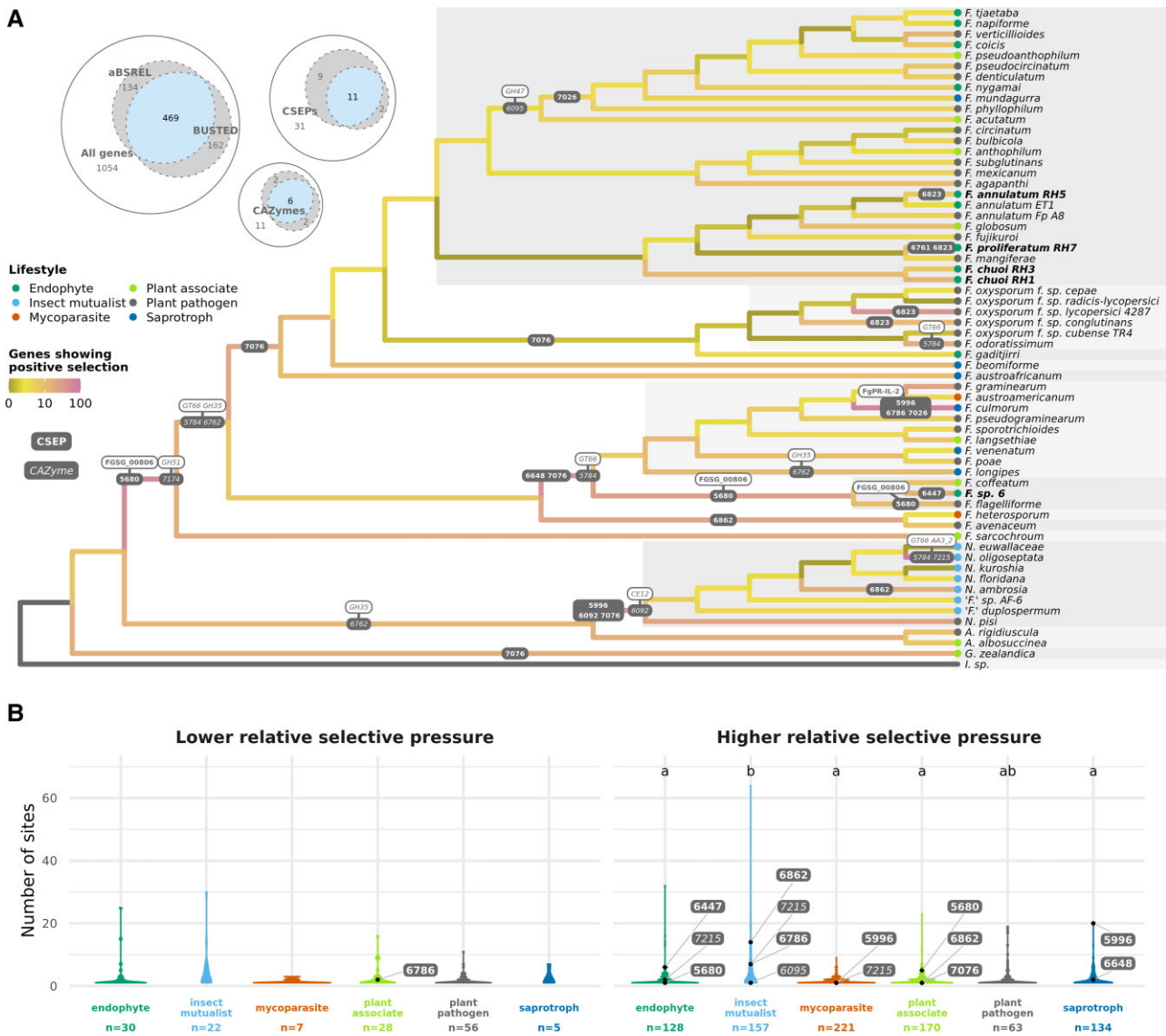
selection is occurring on a branch set, only that there is a relative increase or decrease in dN/dS, and thus higher or lower selective pressure, compared with other branches. We reasoned that if a CSEP or CAZyme with higher relative selective pressure for a lifestyle was also found to be positively selected on an external lineage of that lifestyle, then it could suggest that the selective pressure is imposed by lifestyle. This was the case for 4 of the 9 core CSEPs and 1 of the 3 core CAZymes identified as positively selected on external lineages: CSEPs 6447 (*F. sp. 6*, endophyte); 5996 (*F. culmorum*, saprotroph); 6862 (*N. ambrosia*, insect mutualist); and 7076 (*Geejayessia zealandica*, plant associate); and CAZyme 7215 of lignin degrading subfamily AA3\_2 (*N. oligoseptata*, insect mutualist).

### Codon Optimization was Higher in *Fusarium s. str.*

As dN/dS methods are biased by the erroneous assumption that all synonymous substitutions are neutral (Hershberg and Petrov 2008; Rahman et al. 2021), we also explored whether translational selection (i.e., bias towards certain codons in more highly expressed genes) may be acting on synonymous substitutions by assessing the extent of codon optimisation (S) across fusarioid taxa (dos Reis et al. 2004). Codon optimisation of 1,054 core single-copy genes was generally high for all taxa (between 0.4 and 0.6, on a scale from  $-1$  to  $1$ ), but it was significantly lower in insect mutualists compared with endophytes, plant pathogens and saprotrophs (fig. 4A, supplementary table S5, Supplementary Material online). S values were found to be significantly higher in CSEPs and CAZymes than other core single-copy genes for all lifestyles (excluding mycoparasite, which could not be tested due to small sample size); furthermore, codon optimisation of CAZymes was also significantly higher than CSEPs for insect mutualists and plant pathogens (fig. 4B, supplementary table S5, Supplementary Material online). CSEPs and CAZymes



**FIG. 2.** (A) Genome-scale phylogeny of fusarioid taxa produced by RAXML-NG from 1,060 core single-copy genes. All branches were significantly supported ( $\geq 70$  FBP), except those in red. A time scale for node ages estimated by the IR relaxed clock model is shown below the phylogeny, with highest posterior density 95% confidence intervals shown as bars on nodes. For the AR model results and the exact ages and confidence intervals estimated for every node, see [supplementary figure S3, Supplementary Material](#) online. Clades corresponding to species complexes (SC) and allied genera are highlighted with alternating boxes and annotated to the right of taxon names (FFSC = *Fusarium fujikuroi* species complex, FOSC = *Fusarium oxysporum* species complex, FNESC = *Fusarium nisikadoi* species complex, FBRSC = *Fusarium burgessii* species complex, FCOSC = *Fusarium concolor* species complex, FSAMSC = *Fusarium sambucinum* species complex, FIESC = *Fusarium incarnatum-equiseti* species complex, FHSC = *Fusarium heterosporum* species complex, FTSC = *Fusarium tricinctum* species complex, FLSC = *Fusarium lateritium* species complex, Neo = *Neocosmospora*, Alb = *Albonectria*, Gee = *Geejayessia*). Lifestyles of the strains used in this study are indicated by coloured circles on tips, with other lifestyles reported from the literature summarised in the central grid (see [supplementary table S1, Supplementary Material](#) online for references). Bar graphs on the right indicate the number of genes, CSEPs and CAZymes for each taxon, with lightest to darkest colour indicating whether genes are core, accessory, or strain-specific. (B) Matrix of P-values showing whether gene, CSEP, and CAZyme content were significantly different between lifestyles according to pairwise PERMANOVA. Coloured boxes indicate significant P values ( $< 0.05$ ). Global PERMANOVA results are reported in the bottom right of plots (see also [supplementary table S4, Supplementary Material](#) online). (C) Scatterplot showing variation in gene copy number across all genes, CSEPs and CAZymes for different lifestyles. Points are jittered to reduce overlap. Sample size (the number of strains) is reported under x-axis labels.



**Fig. 3.** Results of dN/dS analyses on 1,054 core single-copy genes. (A) The Euler diagrams show the number of genes, CSEPs and CAZymes found to be under positive selection by both aBSREL and BUSTED. For the 469 cases where there was consensus between the two methods, the number of positively selected genes for each lineage according to aBSREL are shown by coloured branches on the species tree. The colour scale was pseudo log transformed for easier visualisation. For the exact number of positively selected genes on every branch, see [supplementary figure S5, Supplementary Material](#) online. Branches on which CSEPs (bold) and CAZymes (italic) were positively selected are labelled with the gene ID(s) and, where possible, more detailed functional annotation is also indicated in white labels. Lifestyles of strains are indicated by coloured circles on tips. (B) Violin plot showing, for genes with at least 1 site with different selective pressure, the number of sites per genes for each lifestyle with lower (left) or higher (right) selective pressure relative to all other lifestyles according to Contrast-FEL. CSEPs (bold) and CAZymes (italic) that were also reported to be positively selected by BUSTED and aBSREL are indicated by points and labelled with the gene ID. Lifestyles with significant difference of means as calculated by the Games Howell test are shown by letters to the top of the plots (see [supplementary tables S3 and S5, Supplementary Material](#) online for full statistical test results). Sample size (the number of genes) is reported under x-axis labels.

also encompassed greater extremes of codon optimisation than other core genes.

As high levels of codon optimisation has been linked to host generalism in fungi (Badet et al. 2017) and codon usage bias to wide habitat range in prokaryotes (Botzman and Margalit 2011), we speculated that higher codon optimisation may be associated with lifestyle generalism, that is, taxa being capable of exhibiting multiple lifestyles. When no data correction was performed, there was a medium strength positive correlation between the

number of reported lifestyles or “lifestyle range” and S values (Pearson’s  $r = 0.3$ ,  $P = 0.01$ ), but the statistical significance of this correlation did not hold when accounting for phylogenetic relationships with phylogenetic generalised least squares (PGLS) analysis ( $P = 0.06$ ) ([supplementary fig. S8, Supplementary Material](#) online).

There was significantly higher codon optimisation in species complexes belonging to *Fusarium* s. str. compared with allied genera ( $t$ -test,  $P = 6e-11$ ; [fig. 5A](#) inset). Codon optimisation for CSEPs was shown to be strongly



correlated with phylogeny as shown by the fit of  $S$  values to a principal component analysis (PCA) of phylogenetic distances (fig. 5A). This was not the case for CAZymes, however, for which the fit of codon optimisation to the PCA was not significant ( $P=0.2$ ). Hierarchical clustering of taxa by normalised relative synonymous codon usage (RSCU) was also reasonably concordant with the species tree, with a Robinson–Foulds distance of 0.4 ( $P=0$ ; fig. 5B), indicating that codon usage bias, for CSEPs if not CAZymes, is likely to be influenced by shared ancestry more than lifestyle.

## Discussion

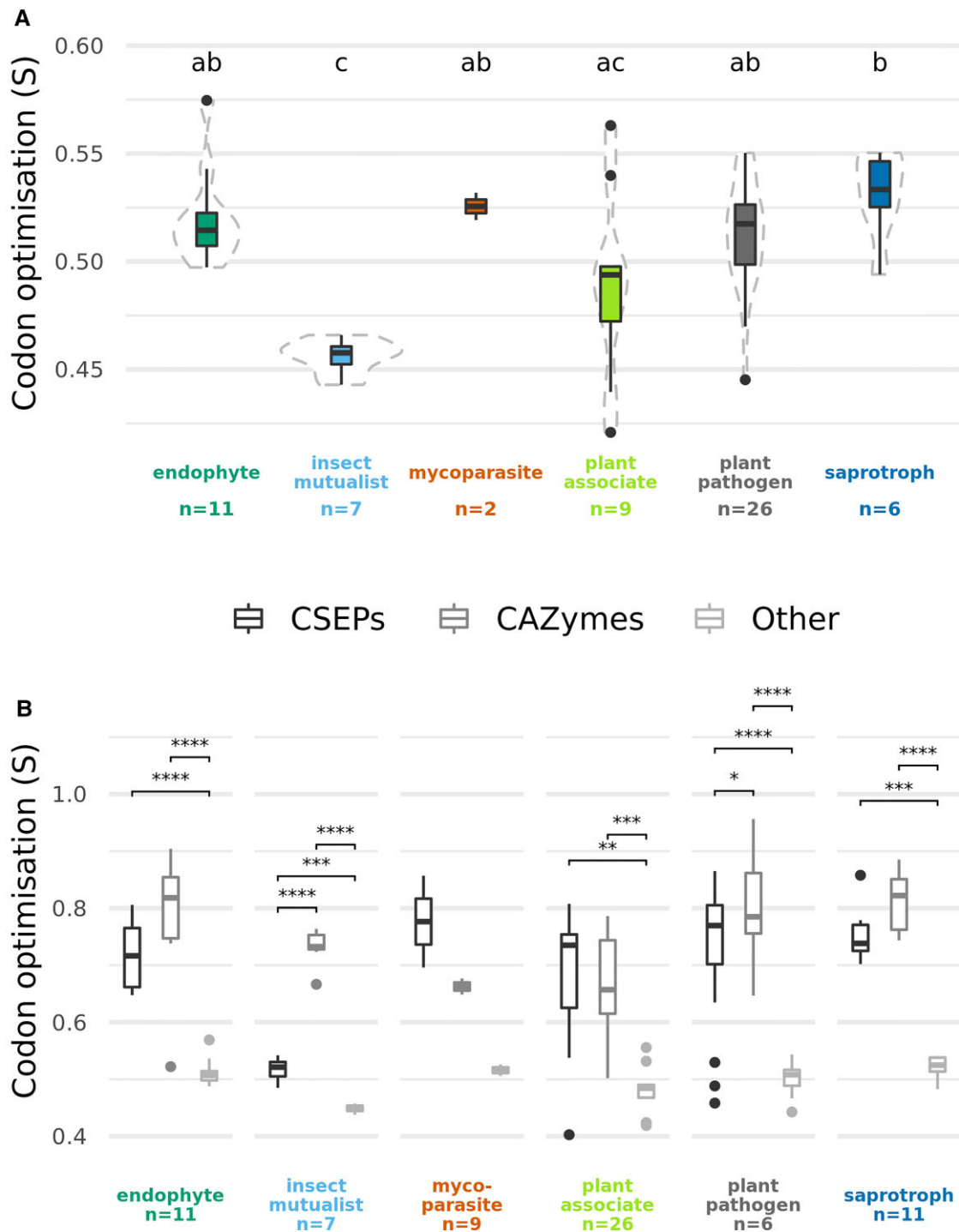
In this study, we inferred a phylogeny of *Fusarium* and allied genera using the greatest number of loci to date, with almost all branches significantly supported (fig. 2A). This adds to numerous recent efforts to produce high quality fungal phylogenies from genome-scale data (e.g., Spatafora et al. 2016; Steenwyk et al. 2019; Varga et al. 2019; Li et al. 2021). Trimming method and inclusion/exclusion of multi-copy genes had some impact on species tree topology (fig. 1, supplementary fig. S2, Supplementary Material online), but the *Fusarium s. str.* backbone was consistent across all approaches and in general agreement with the most recently published phylogeny of the group (Crous, Osieck, et al. 2021). Discordance was concentrated in the ambrosia clade in *Neocosmospora*, perhaps due to the occurrence of interspecific hybridization in this lineage (Kasson et al. 2013) or horizontal gene transfer via the exchange of strains by beetles (Hulcr and Cognato 2010). The objectives of this study were not concerned with the taxonomic debates surrounding the *Fusarium* generic concept, but our results did show that the divergence between *Fusarium s. str.* and other fusarioid taxa was associated with positive selection on a considerable number of core genes (fig. 3A); an upwards shift in translational selection (fig. 5A); and distinct patterns in codon usage bias (fig. 5B). While these results obviously do not directly contribute to characterisation of the taxa involved, they might be seen as a symptom of a “larger and more abrupt” divergence than that between species within the same genus (Booth 1978), contrary to *Fusarium s. lat.* (O’Donnell et al. 2020; Geiser et al. 2021).

We generally found the IR molecular clock model to produce dating estimates that were more concordant with estimates from other studies assessing divergence times of fusarioid fungi (e.g., Kasson et al., 2013; O’Donnell et al., 2013, 2015), which was largely to be expected considering that these studies also used IR models (but different secondary calibrations). The IR model estimated the divergence of obligate insect mutualists to correspond more closely to the crown age of their insect hosts, as estimated with insect fossil calibrations (Jordal and Cognato 2012). By contrast, the AR model appeared to produce less congruent ages for recently diverged lineages, such as the highly specialised FOSC strains

diverging before their host plants are likely to have existed. AR models have generally been thought appropriate for plants and animals considering the correlation between substitution rate and life-history traits (Lartillot et al. 2016), and it has furthermore been suggested that AR is the norm across all kingdoms of life (Tao et al. 2019). On the other hand, Taylor and Berbee (2006) found no lineage-specific correlation of substitution rates across the kingdom Fungi. Similarly, Linder et al. (2011) did not find strong evidence for rate autocorrelation across plant and simian datasets, instead finding the IR model to have more explanatory power. The AR model is not immune to bias (Lartillot and Delsuc 2012), and has been shown to produce older estimates for simulated datasets across dating tools, including MCMCTree (Miura et al. 2020). The presence of short-term rate fluctuations in mammals suggest that mixed relaxed clock models accounting for both autocorrelation and jumps in rate variation are needed (Ho 2009; Lartillot et al. 2016).

Sources of error in divergence time estimation are manifold, as evidenced by the large confidence intervals in our analysis (supplementary fig. S3, Supplementary Material online). Beyond the difficulty surrounding choice and implementation of molecular clock models, a major source of error is the use of secondary calibrations—a necessity due to the general lack of fungal fossil data (Beimforde et al. 2014)—which can impact the precision and accuracy of divergence time estimates (Shaul and Graur 2002; Graur and Martin 2004; Sauquet et al. 2012; Schenk 2016). For this reason, we incorporated the error from node ages estimated using primary fossil calibrations (Lutzoni et al. 2018) using confidence intervals to provide upper and lower bounds, as recommended when using secondary calibrations (Graur and Martin 2004; Forest 2009; Hipsley and Müller 2014). An alternative approach is to expand taxon sampling until fossil data can be incorporated, although secondary calibrations have been shown to produce divergence time estimates with similar accuracy to those from distant primary calibrations, albeit with lower precision (Powell et al. 2020). Our motivation for divergence time estimation was not to test specific time-dependent hypotheses, but rather to calibrate branch lengths for more realistic measures of phylogenetic distance in subsequent comparative analyses. As with any divergence time analysis, major uncertainties are still associated with the divergence times of fusarioid fungi.

All taxa had a similar number of genes, CSEPs and CAZymes, very few of which were strain-specific (fig. 2A). It has previously been suggested that the number of species-specific secreted proteins (and by extension, we assume, effectors) is generally higher in fungal lifestyles which associate with plants without killing or decaying them, such as mutualistic symbionts and biotrophic pathogens, compared with saprotrophs and necrotrophic pathogens (Kim et al. 2016), the reasoning being that the former have to negotiate the plant–fungal interaction for an extended period. In the genus *Colletotrichum*, however, a reduction in the number of species-specific CSEPs was

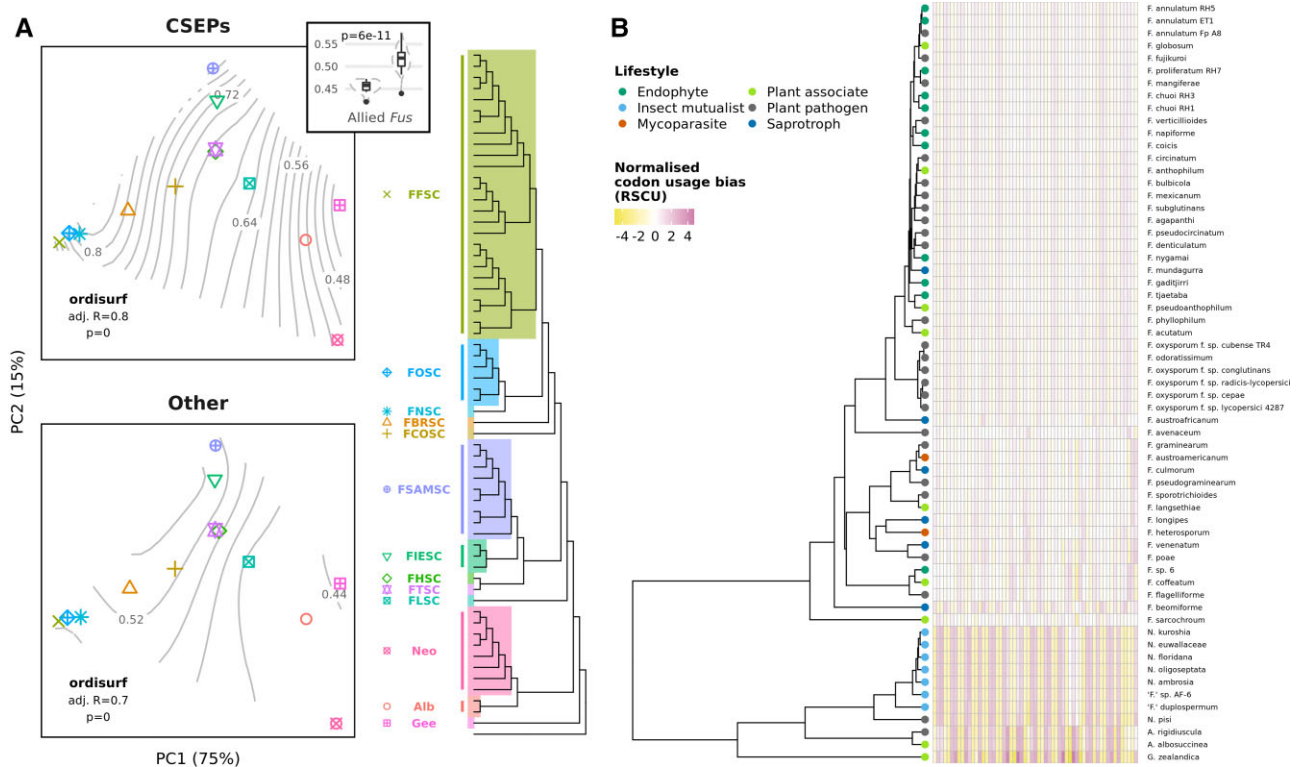


**FIG. 4.** Boxplots showing codon optimisation ( $S$ ) of core single-copy genes across lifestyles. Sample size (the number of strains) is reported under x-axis labels. (A) Difference in overall  $S$  values between lifestyles, with significant difference of means as calculated by TukeyHSD shown by letters at the top of the plot (see [supplementary tables S4 and S6, Supplementary Material](#) online for full statistical test results). (B) Difference in  $S$  values between CSEPs, CAZymes, and other genes for each lifestyle, with significant difference of means between the gene type as calculated by the Games Howell test shown by bars across significantly different categories ( $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$ ; see [supplementary tables S3 and S5, Supplementary Material](#) online for full statistical test results).

observed alongside the transition from phytopathogenicity to beneficial endophytism (Hacquard et al. 2016), showing that CSEPs and their impact on the plant–fungal interaction can be highly lineage-specific. We saw no significant difference in the number of strain-specific CSEPs

(or genes) between any lifestyles ([supplementary fig. S4, Supplementary Material](#) online). This, combined with the fact that plant pathogens are often also reported as endophytes and vice versa ([fig. 2A](#)), and that plant pathogen and endophyte strains were not significantly different in





**FIG. 5.** (A) PCA of phylogenetic distances between taxa, with points representing centroids for species complexes/allied genera, differentiated by shape and colour, as indicated by the tree legend. The percentage of variance explained by each principal component is shown on axis labels. Contours indicate the fit of codon optimisation ( $S$  values), of both core CSEPs and other core genes, to the ordination; the fit of CAZyme codon optimisation is not shown as it was not significant ( $P = 0.2$ ). The inset boxplot shows the significant difference ( $t$ -test,  $P = 6e-11$ ) in overall  $S$  values between *Fusarium* and allied genera. (B) Hierarchical clustering of taxa according to normalised codon usage bias (RSCU). Heatmap columns represent codons (excluding Trp, Met, and stop codons) with cells coloured by normalised RSCU, where positive values represent higher than expected codon usage and negative values represent lower than expected codon usage.

terms of gene and CSEP content (fig. 2B), suggests that fusarioid taxa have a shared genetic capacity for phytopathogenicity and/or endophytism. Having a high proportion of species-specific CSEPs has also been associated with the connected factor of host specialisation (Spanu et al. 2010), which, considering we report very low numbers of strain-specific genes, may also explain the status of many *Fusarium* taxa as host generalists. Our results were also similar to those comparing pathogenic and non-pathogenic taxa in another genus of broad generalists, *Aspergillus* (Mead et al. 2021).

We did not identify common genetic signatures for the endophytic lifestyle in terms of gene, CSEP or CAZyme content, reinforcing the current understanding that there is no universal “toolkit” associated with the endophytic lifestyle (Hacquard et al. 2016; Knapp et al. 2018). This contrasts with other well-defined lifestyles such as that of mycorrhizal fungi, for which specific genetic features have been associated with lifestyle in both ascomycetes and basidiomycetes (Martin et al. 2010; Delaux et al. 2013; Kohler et al. 2015; Peter et al. 2016; Miyauchi et al. 2020; Rich et al. 2021). One observed hallmark of the transition to mycorrhizal symbiosis is the loss of genes encoding PCWDEs (Kohler et al. 2015; Peter et al. 2016; Miyauchi et al. 2020), but, as we found here, these are retained in

various endophytic taxa (Zuccaro et al. 2014; Lahrman et al. 2015; Hacquard et al. 2016; Franco et al. 2022; Mesny et al. 2021). As PCWDEs have often been treated predominantly as features of saprotrophy, this has fed into the hypothesis that many endophytes are latent saprotrophs, but in a broad comparison of CAZymes across the Dikarya, Zhao et al. (2013) demonstrated that plant pathogens have on average more CAZymes belonging to typical PCDWE families than saprotrophs. As there was no significant difference in total number or repertoire of CAZymes between plant pathogens, endophytes and saprotrophs, it indicates that fusarioid fungi retain the same machinery for plant cell wall degradation and/or remodelling, regardless of lifestyle. We did, however, find a significant difference in CSEP content between saprotrophs and endophytes (fig. 2B), which could suggest that fusarioid endophytes are more likely to be latent pathogens than saprotrophs.

The major exception to the apparent lifestyle flexibility among fusarioid fungi is the insect mutualist lifestyle, which formed a monophyletic group (the ambrosia clade) in *Neocosmospora* (fig. 2A). The insect mutualist lifestyle was also the most distinct in terms of gene and CSEP content, being significantly different from all other lifestyles apart from the mycoparasitic lifestyle (fig. 2B), but the

very small sample size for the latter will have impacted the test's power in that case (Alekseyenko 2016). The transition to symbiotic mutualism in *Neocosmospora* was not associated with a reduction in total number of genes, CSEPs or CAZymes, in agreement with results from other ectosymbiotic insect mutualists (Biedermann and Vega 2020). As the representative strains used in this study are all known to cause disease on the trees they colonise with their beetle partner (Freeman et al. 2013; O'Donnell et al. 2016; Na et al. 2018; Aoki et al. 2019), it follows that they would have retained many of the genetic mechanisms from their (presumably) plant associated ancestors. Some strains have been found to cause disease *in vitro* in the absence of their beetle partners (e.g., Eskalen et al., 2012; Na et al., 2018), however, to our knowledge, fusarioid ambrosia fungi have never been reported as free-living in the wild.

Although we did not identify significant differences in the genetic repertoires between fusarioid endophytes and plant pathogens, we did find some evidence that copy number variation—genes or regions that are either duplicated or deleted in reference to other taxa—may be contributing to lifestyle. There was no significant difference in mean gene copy number between lifestyles, but plant pathogens included extreme outliers in gene copy number compared with other lifestyles (fig. 2C). Extensive gene duplication has been suggested as a key strategy for pathogenicity in basidiomycete rusts (Pendleton et al. 2014), and copy number of the pectin degrading CAZyme subfamily PL1\_7 across 41 root-colonising fungi was shown to correlate with pathogenicity in *Arabidopsis* (Mesny et al. 2021). Gene duplication is regarded as the primary resource for the evolution of functional novelties, and the persistence of gene duplicates is indicative of neofunctionalisation and/or subfunctionalisation, as a functionally redundant gene copy will be rapidly lost due to the absence of selective pressure to retain it (Lynch and Conery 2000; He and Zhang 2005). The most common functional innovations of gene copies in fungi are regulatory changes (Wapinski et al. 2007). Indeed, copy number variation is known to be correlated with differential gene expression (Stranger et al. 2007; Steenwyk and Rokas 2018; Shao et al. 2019), and has been shown to contribute to phenotypic or pathological differences in fungi (Steenwyk et al. 2016; Zhao and Gibbons 2018).

This aligns with mounting evidence that a major factor impacting lifestyle of closely related phytopathogens and endophytes is not gene repertoire itself, but expression profiles. Returning to *Colletotrichum*, Hacquard et al. (2016) found that a pathogenic taxon had a different pattern of gene expression during host colonisation, including upregulation of CSEPs, compared with a closely related and genetically similar beneficial endophyte. The authors noted that this also makes the beneficial endophyte genetically capable of reverting to pathogenicity (and, presumably, the closely related pathogens capable of inhabiting plants as endophytes). The aforementioned CAZyme subfamily PL1\_7, which we found between 2 and 4 copies of in all fusarioid taxa (supplementary fig. S7, Supplementary

Material online), was also more highly expressed in the pathogenic *Colletotrichum* taxon. The importance of expression has already been seen in *Fusarium*, where expression of secondary metabolites differed between endophytic and pathogenic strains of the same species, *F. annulatum* (as *F. proliferatum*, FFSC), despite generally sharing secondary metabolite gene clusters (Niehaus et al. 2016). Generating *in planta* expression profiles for both pathogenic and non-pathogenic strains across the group could reveal whether there is convergence in expression patterns for certain lifestyles.

Regulation of certain genes located on accessory chromosomes has also been seen to direct plant infection phenotypes in an endophytic versus pathogenic FOSC strain (Guo et al. 2021). Accessory chromosomes—chromosomes that are not essential for survival, but potentially confer functional advantages (Bertazzoni et al. 2018)—are likely another important factor impacting lifestyle in *Fusarium*. The first acc. chromosomes in fungi were discovered in the fusarioid species *Neocosmospora haematococca* (as *Nectria haematococca*) (Coleman et al. 2009), with further reports in at least nine other fusarioid strains (Bertazzoni et al. 2018). They have mostly been studied in the FOSC, in which horizontal transfer of acc. chromosomes can confer pathogenicity (Ma et al. 2010; Li et al. 2020). Not only are acc. chromosomes deemed to be a key innovation for rapid adaptation by plant pathogens (Croll and McDonald 2012) they have also been implicated in adaptation of FOSC strains to human pathogenicity (Zhang, Yang, et al. 2020). Exploring the extent of acc. chromosomes broadly across fusarioid fungi, as well as phenomena impacting genomic architecture such as transposable elements (Muszewska et al. 2019), may shed light on the mechanisms underlying lifestyle flexibility in the group (Ma et al. 2013).

As effectors are highly diverged and often lineage-specific, if not strain-specific, only a small proportion of the CSEPs predicted here could be matched to experimentally verified genes from PHI-base. Of these, the majority were genes known to impact virulence to some degree or not at all in the hosts they have been tested on (supplementary fig. S6, Supplementary Material online), although the knockout mutant phenotype for a certain gene will not necessarily be the same for different fungal strains or on different hosts. PHI-base is also explicitly dedicated to pathogen-host genes, and similar high quality, curated resources are needed for genes involved in non-pathogenic fungal–host interactions. Nonetheless, our results give us a broad perspective on CSEP distributions across fusarioid fungi. Some CSEPs exhibited phylogenetic patterns (such as lower copy number in *Fusarium s. lat.* compared with *Fusarium s. str.* for MoCDIP4, which was first discovered in *Magnaporthe oryzae* (Chen et al. 2013) and since reported in *F. oxysporum* f. sp. *pisi* (Achari et al. 2021)), but most had scattered distributions across the group (supplementary fig. S6, Supplementary Material online), which may be the result of frequent horizontal gene transfer (e.g., van Dam and Rep 2017; Peck et al. 2021).

A slightly lower proportion of core CSEPs were found to be positively selected than non-CSEPs according to dN/dS calculations (fig. 3A). This may be seen as surprising, as effectors that promote virulence are assumed to be under strong selective pressure during the evolutionary arms race between fungus and host (De Jonge et al. 2011; Lo Presti et al. 2015). For instance, CSEPs have been found to more frequently be under positive selection compared with non-CSEPs in phytopathogenic *Microbotryum* species (Beckerson et al. 2019). High rates of selection on CSEPs are not only a hallmark of pathogenicity, however, as these have also been observed for obligate, host-specific *Epichloë* endophytes (Schirrmann et al. 2018); the arbuscular mycorrhizal fungus *Rhizophagus irregularis* (Schmitz et al. 2019); and the saprotroph *Verticillium tricorpus* (Seidl et al. 2015), emphasizing the broader roles played by effectors in host–fungal interactions. Our results could be explained by the fact that we focused on core genes, and so the CSEPs in question are presumably contributing to integral host–fungal interactions that would be under similar selective pressure as other core functions, rather than specialised CSEPs more likely to be under strong selective pressure from the host. We should also note that detection of positive selection with dN/dS methods is biased against shorter genes (Derbyshire et al. 2021), which CSEPs by definition are, and so this may have impacted our results.

We identified five cases where positive selection of core CSEPs and CAZymes may be connected to lifestyle by comparing aBSREL analysis of positive selection on external branches to Contrast-FEL analysis of relative selection pressures between lifestyles. Interestingly, there were no core CSEPs with higher selective pressure in plant pathogens relative to other lifestyles, which could be interpreted as evidence that the ancestral state of the group is phytopathogenic rather than endophytic, but the unbalanced sample sizes for the different lifestyles will have influenced the Contrast-FEL results. Once again, the insect mutualist lifestyle was shown to be distinct, with a greater number of sites per gene undergoing higher selective pressure relative to other lifestyles (fig. 3B). This may be associated with the fact that these ambrosia taxa have evolved via insect farming, in what could be interpreted as some level of “artificial selection” (Mueller et al. 2005). We were only able to tentatively link the positive selection of one core CAZyme to lifestyle: 5′-oxoaverantin cyclase in the AA3\_2 subfamily, which was positively selected for in the insect mutualist *N. oligoseptata* (fig. 3A). Other members of the same subfamily are implicated in lignin degradation (Levasseur et al. 2013; Miyauchi et al. 2017), but 5′-oxoaverantin cyclase was first identified as an intermediate in aflatoxin biosynthesis in *Aspergillus parasiticus* (Sakuno et al. 2003). Another insect–fungus mutualism between the navel orangeworm and *A. flavus* has shown that aflatoxin tolerance is a key adaptation of the insect to its fungal diet (Niu et al. 2009; Ampt et al. 2016), and as fusarioid fungi are known to produce an array of mycotoxins (Desjardins and Proctor 2007), it would be interesting to determine

whether there is a similar dynamic in the evolution of the ambrosia mutualism.

Conventional dN/dS methods to detect selection such as aBSREL and BUSTED make the assumption that synonymous substitutions are always selectively neutral, but we now know that selection does occur on synonymous mutations (Ohta 1996; Chen et al. 2004; Hershberg and Petrov 2008). Subsequently dN/dS methods have been shown to overestimate the frequency of positive selection and underestimate the strength of negative selection in bacteria, even when selection on synonymous sites is weak (Rahman et al. 2021). Furthermore, using dN/dS > 1 as a signifier of positive selection has been declared arbitrary (Tamuri and dos Reis 2021). As flexible dN/dS methods accounting for selection on synonymous substitutions have yet to be integrated into the widely used tools for detecting positive selection, this remains a caveat of our dN/dS analyses. Additionally, even a low incidence of sequence inaccuracies can result in false-positive signals of selection (Mallick et al. 2009), so ideally candidate genes should be resequenced to detect errors and confirm whether sites are truly under selection. A further limitation of the selection analyses is that they were restricted to core genes due to the requirement of a robust species tree to estimate dN/dS across lineages, which necessarily excludes a large proportion of the gene content (Derbyshire et al. 2021). Further exploration of selection dynamics in the extensive accessory content would undoubtedly shed more light on the evolution of the group.

When exploring the issue of selection on synonymous substitutions, we showed that codon optimisation of the core single-copy genes—that is, the extent of translational selection on codon usage—was higher in CSEPs and CAZymes than other genes (fig. 4B), as was previously found in the *F. oxysporum* f. sp. *cepae* pangenome (Armitage et al. 2018). Insect mutualists had a much larger difference in codon optimisation between CSEPs and CAZymes (fig. 4B). One possible explanation for this result is that these taxa may have less translational selective pressure on CSEPs that are required for plant invasion—being farmed by insects which excavate and weaken the plant hosts—but retain higher translational selective pressure on CAZymes that are required for assimilation of nutrients, which ultimately maintains the insect–fungus mutualism. Following this broad perspective on codon optimisation, further functional annotation could allow the use of a “reverse ecology framework” to explore whether genes with the highest codon optimisation correspond with lifestyle (LaBella et al. 2021).

We also found that correlation between lifestyle range and codon optimisation was not significant after correcting for phylogenetic relationships (supplementary fig. S8, Supplementary Material online), contrary to expectation from previous studies (Botzman and Margalit 2011; Badet et al. 2017). Our approach to assess lifestyle range was limited by the availability of published reports of fusarioid taxa, and so we will undoubtedly have underestimated the number of lifestyles exhibited by some species.



Furthermore, fusarioid species are often hard to distinguish, and lifestyle reports may therefore be misattributed. To mitigate against this issue, we only included studies that used appropriate genetic markers to distinguish taxa—not, for instance, solely using ITS (Geiser et al. 2004)—and crosschecked phylogenetic analyses for misclassifications. Despite this, we may have inadvertently included lifestyle reports for species that were incorrectly classified in the original study. A comprehensive meta-analysis is needed to better understand the extent of lifestyle and host range for fusarioid taxa.

A major caveat of our comparative analyses is that we were forced to attribute a single lifestyle to the strains being used, despite the current understanding, which our own results support, that these lifestyles are not necessarily mutually exclusive (Selosse et al. 2018). Furthermore, treating lifestyles as categorical traits does not accurately reflect the range of outcomes we know can exist within even one lifestyle, such as different pathogenic strains within the same species varying in “aggressiveness” (e.g., Holtz et al. 2011; Chen et al. 2014; Šišić et al. 2018). These both remain central issues with current approaches to fungal lifestyle comparison at large (e.g., Knapp et al. 2018; Miyauchi et al. 2020; Mesny et al. 2021; Franco et al. 2022). New methods that can effectively incorporate multiple lifestyle hypotheses, or treat lifestyles as points on a continuous spectrum, are sorely needed to encapsulate the nuance of these highly context-dependent interactions.

## Conclusions

We found an apparent shared genetic capacity for phytopathogenicity and endophytism in *Fusarium*, which suggests that, while strains may be reported as plant pathogens or endophytes, their lifestyle is potentially transient. Were fusarioid taxa to make the transition to obligate, mutualistic endophytism, we might expect to see genetic hallmarks more akin to those seen in the transition to obligate symbiosis in mycorrhizal lifestyles (e.g., Delaux et al., 2013). Despite multiple reports of certain endophytic *Fusarium* strains being beneficial to certain plant hosts (e.g., Kavroulakis et al., 2007; Mendoza and Sikora, 2009; Bilal et al., 2018), large uncertainties remain as to the stability of these interactions. Our results depict fusarioid fungi as prolific generalists and highlight the difficulty in predicting pathogenic potential in the group. Considering the importance of plant immune response, biotic and abiotic conditions to the plant–fungal interaction, such endophytes may not be the “silver bullet” for biocontrol that they are sometimes touted to be.

## Materials and Methods

### Genome Sequencing, Assembly, and Structural Annotation

We selected five endophytic *Fusarium* strains for whole-genome sequence which were representatives of species

hypotheses that had previously been isolated and clustered into 99% similarity operational taxonomic units by Hill et al. (2021), with taxonomic identification confirmed where possible via morphological assessment by the Westerdijk Institute (supplementary table S2, Supplementary Material online). For DNA extractions, a fragment of mycelium from axenic cultures was transferred to 500 ml of 2% malt extract nutrient broth using a sterile needle and grown at 25 °C in ambient light conditions on an orbital shaker at 120 rpm for ~1 week. Mycelia were collected via vacuum filtration and frozen at –80 °C before being pulverised with two sterile stainless-steel beads in a 2 ml Eppendorf using a Mixer Mill MM 400 (Retsch, Germany).

DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, CA, USA) according to the manufacturer’s protocol and eluted in 70 µl of TE buffer. Sufficient DNA concentration (more than 20 ng/µl) was confirmed with a Quantus™ Fluorometer (Promega, WI, USA) and purity (260/280 absorbance ratio of approximately 1.8) confirmed with a NanoDrop spectrophotometer (Thermo Fisher Scientific, MA, USA). DNA extractions were sent to Macrogen (Macrogen Inc., South Korea) for library preparation and sequencing: library preparation was performed using the TruSeq DNA PCR-free Sample Preparation Kit with a 550 bp insert size and 151 bp paired-end reads were sequenced using the NovaSeq 6000 platform (Illumina, San Diego, CA, USA).

Our assembly and annotation methodology are described in full detail in Supplementary Material. In brief, reads were assembled with ABySS v2.0.2 (Simpson et al. 2009) and polished with Pilon v1.23 (Walker et al. 2014). Contiguity was assessed using QUAST v5.0.2 (Gurevich et al. 2013); completeness as measured by gene sets was assessed using BUSCO v3.0.1 (Simão et al. 2015); and absence of contamination was checked using BlobTools v1.1 (Laetsch and Blaxter 2017). Assemblies were annotated following the MAKER pipeline (Cantarel et al. 2008). See supplementary table S6, Supplementary Material online for a summary of assembly quality statistics.

### Phylogenomic Analyses

Predicted genes from 57 additional publicly available strains of *Fusarium* and allied genera were downloaded from NCBI (supplementary table S1, Supplementary Material online) and orthogroups (referred to here as genes) were inferred from amino acid sequences of the total 62 strains using OrthoFinder v2.4.0 (Emms and Kelly 2019). We aligned 1,060 core (i.e., shared between all fusarioid taxa including the outgroup) single-copy genes using MAFFT v7.310 with default settings (Katoh and Standley 2013) and removed ambiguously aligned regions using both BMGE v1.12 (Criscuolo and Gribaldo 2010) and trimAl v1.4.rev15 with the gappyout option (Capella-Gutiérrez et al. 2009) to compare the impact of trimming tools on the resulting species trees.

For a concatenation-based approach, core single-copy gene alignments were concatenated with AMAS v0.98



(Borowiec 2016). We compared two tools for maximum likelihood (ML) species tree estimation: IQ-TREE v2.1.2 (Minh et al. 2020) and RAXML-NG v1.0.1 (Kozlov et al. 2019), with the concatenated alignment partitioned by gene in both cases. For IQ-TREE, the best-fit amino acid substitution model for each partitioned gene was selected by the inbuilt tool ModelFinder (Kalyaanamoorthy et al. 2017) using Bayesian information criterion (BIC) values, and branch support was computed via 1,000 ultrafast bootstrap replicates (UFBoot) (Hoang et al. 2018). For RAXML-NG, ModelTest-NG v0.1.6 (Darriba et al. 2020) was used to select substitution models for each gene using Akaike information criterion (AIC) values, and branch support was computed via 100 Felsenstein's bootstrap replicates (FBPs). FBP convergence was confirmed with the `-bsconverge` option using the default 3% cutoff for weighted Robinson–Foulds distances (Pattengale et al. 2010).

For a coalescent-based approach, ML gene trees were inferred from each core single-copy gene alignment with RAXML-NG using the best-fit model selected by ModelTest-NG during the concatenated analysis. The resulting ML gene trees were used for coalescent-based species tree reconstruction using ASTRAL-III v5.7.3 (Zhang, Yohe, et al. 2018) with local posterior probability (LPP) branch support estimation (Sayyari and Mirarab 2016). ASTRAL-Pro v1.2 (Zhang, Scornavacca, et al. 2020) was additionally run with LPP support estimation on the 20,343 gene trees produced by OrthoFinder, which represented both single- and multi-copy “total” genes. OrthoFinder itself also produces a coalescent-based species tree topology by default using STAG (Emms and Kelly 2018), which used 3,449 core single- and multi-copy genes. All species tree topologies were compared by computing the normalised Robinson–Foulds metric using the `RF.dist` function from the `phangorn` v2.7.0 package (Schliep et al. 2017) in R v4.0.4 (R Core Team 2020).

### Molecular Clock Analyses

The species tree topology inferred by RAXML-NG was used to perform molecular clock analyses with MCMCTree (Yang and Rannala 2006) in PAML v4.9 (Yang 2007) using the top 10 “clock-like” core single-copy genes, as inferred by SortaDate based on root-to-tip variance (Smith et al. 2018). Divergence times were estimated using the approximate likelihood method (dos Reis and Yang 2011) with the WAG amino acid substitution model (Whelan and Goldman 2001). Due to the sparse fossil record for the fungi at large, a previous fossil-calibrated study of the kingdom including *Fusarium* species was used to inform secondary calibrations of the tree root at 0.9–1.35 (1 time unit being 100 My) and the node representing the split between *F. graminearum* and “*F.*” *solani* at 0.5–0.9 (Lutzoni et al. 2018). For details of MCMCTree priors and run settings, see [Supplementary Material](#).

### Computational Prediction of CSEPs and CAZymes

CSEPs were identified from predicted genes using a framework inspired by Beckerson et al. (2019) and summarised

in [supplementary figure 9A, Supplementary Material](#) online including the following: 1) signal peptide detection; 2) filtering for contradictory cellular localisation signals; and 3) cross-checking against machine learning-based effector prediction using EffectorP 3.0 (Sperschneider and Dodds 2021). A custom bash script, `CSEPFILTER`, was written to perform the filtering of gene sets at each stage. To match CSEPs to experimentally verified genes, sequences were searched against the PHI-base database (downloaded 09/02/2022; Urban et al. 2020) using BLAST 2.7.1+ (Camacho et al. 2009). CAZymes were identified from predicted genes using `run_dbCAN` v3.0.2 ([https://github.com/linnabrown/run\\_dbcan](https://github.com/linnabrown/run_dbcan)) from the dbCAN2 CAZyme annotation server (Zhang, Yohe, et al. 2018) and assigned names using the ExplorEnz website (McDonald et al. 2009). For full details on both CSEP and CAZyme prediction, see [Supplementary Material](#).

CSEPs and CAZymes were matched to gene orthogroups with a custom R script, `orthogroup_parser.r`, where a gene was defined as a CSEP/CAZyme if it was predicted to be so in at least one taxon. We checked that genome assembly quality did not significantly influence the number of predicted genes, CSEPs or CAZymes by confirming that there was no correlation between assembly N50 (extracted from NCBI metadata for assemblies produced outside this study) and number of genes/CSEPs/CAZymes using the `cor.test` function in R.

### Comparative Genomics of Lifestyle

Lifestyles of all the strains used in this study were inferred from the host/substrate and other relevant data (such as pathogenicity tests) sourced from the literature, NCBI BioSample metadata, and online culture collection metadata ([supplementary table S1, Supplementary Material](#) online). If a strain was reported from a plant host but without sufficient clarification of whether the plant was exhibiting disease symptoms or the fungus was isolated from inside plant tissues, the strain was classified ambiguously as a “plant associate”. In addition to the lifestyle of the specific strains used in the analyses, other lifestyle reports were collected from the literature with the help of the PlutoF platform (Abarenkov et al. 2010) in order to show the range of reported lifestyles for taxa.

The impact of strain lifestyle on CSEP, CAZyme, and all gene content was explored using an approach developed by Mesny and Vannier (2020) which accounts for confounding phylogenetic signal. Full details are described in [Supplementary Material](#), as well as a full description of the statistical analyses to test the significant difference in number of strain-specific genes and mean gene copy number between lifestyles.

### Selection Analyses

To assess whether core single-copy genes have evolved under positive selection we used HyPhy v2.5.30 (Kosakovsky Pond et al. 2005), which offers a suite of tools for assessing selective pressures based on the ratio of nonsynonymous

to synonymous substitutions (dN/dS)— that is, the ratio of nucleotide substitutions which alter the transcribed amino acid to those that do not. Notably, this approach assumes that synonymous substitutions are selectively neutral. For full details see [Supplementary Material](#); briefly, codon alignments and ML trees were run in BUSTED v3.1 (Murrell et al. 2015) and aBSREL v2.2 (Smith et al. 2015) to detect episodic positive selection (dN/dS > 1). We then used Contrast-FEL to compare differences in relative selective pressures between lifestyles (Kosakovsky Pond et al. 2021).

Codon optimisation of all core single-copy genes to the ribosomal protein gene pool (S) was calculated from each taxon's codon adaptation index (CAI; Sharp and Li, 1987), effective number of codons (Nc) and GC3 values with the get.s function from the tAI v0.2 package (dos Reis et al. 2004). For full details, see [Supplementary Material](#). S values were calculated for CSEP, CAZyme, non-CSEP/CAZyme, and all core single-copy genes in turn. To then assess the relationship between codon optimisation and lifestyle range, we calculated Pearson's correlation on uncorrected data using the cor.test function in R, and used PGLS regression to assess correlation while correcting for phylogenetic signal in the data with the R package nlme v 3.1-152 (Pinheiro et al. 2021). For number of reported lifestyles, only taxa identified to species level were included, and for species with multiple representative strains the mean S value was used. To visualise the relationship between codon optimisation and phylogeny, we used the ordisurf function from the R package vegan v2.5-7 (Oksanen et al. 2019) to fit S values to the PCA of phylogenetic distances produced in comparative analyses above (recreated in R with the vegan prcomp function). See [Supplementary Material](#) for full details of statistical analyses on the difference in S values between *Fusarium s. str.* and allied genera, as well as between gene types and lifestyles.

The uco function from seqinr v4.2-8 (Charif and Lobry 2007) was used to calculate codon usage bias in terms of relative synonymous codon usage (RSCU)—the ratio of observed codon usage to expected codon usage—for all codons across each taxon, excluding non-redundant codons encoding methionine and tryptophan and stop codons. RSCU values were then normalised using the scale function and used to produce a Euclidean distance matrix with the dist function, which was used for hierarchical clustering of taxa with the hclust function using the average agglomeration method. We compared the topology produced by hierarchical clustering with the RAxML-NG species tree topology by again computing the normalised Robinson–Foulds metric using the RF.dist function from phangorn. We calculated the *P*-value by computing the metric for 1,000 random trees with the same number of taxa against the species tree topology to determine the number of simulations for which the metric was lower (i.e., topologically closer) than that from the hierarchical clustering.

All results were plotted in R v4.0.4 using packages listed in [Supplementary Material](#). Scripts of all analyses are available at <https://github.com/Rowena-h/FusariumLifestyles>.

## Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

## Acknowledgements

We thank Marcelo Sandoval-Denis and Pedro Crous for assistance with morphological identification of the strains. We also thank Mark Blaxter for helpful advice on assembly tool comparison; Mario dos Reis for advice on using MCMCTree; and Theo Llewellyn, Laura Kelly and James Borrell for valuable discussion. We also thank the editor and three anonymous reviewers for their valuable feedback. We acknowledge the assistance of the ITS Research team at Queen Mary University of London. This work was supported by the Evolution and Education Trust (EET) and Pragnell Fund to Ester Gaya. R.H. is a Natural Environment Research Council funded PhD student with the London NERC DTP (NERC Ref: NE/L002485/1).

## Data availability

WGS data and structurally annotated genome assemblies generated in this study are available on GenBank under the BioProject accession PRJNA761077. Additional data files of the raw phylogenetic trees; CSEP and CAZyme amino acid sequences; OrthoFinder output; and orthogroup metadata are deposited in Zenodo doi:10.5281/zenodo.6353640. All scripts are available at <https://github.com/Rowena-h/FusariumLifestyles>.

## References

- Abarenkov K, Tedersoo L, Nilsson RH, Vellak K, Saar I, Veldre V, Parmasto E, Proust M, Aan A, Ots M, et al. 2010. PlutoF—a web based workbench for ecological and taxonomic research, with an online implementation for fungal ITS sequences. *Evol Bioinform.* **6**:189–196.
- Achari SR, Edwards J, Mann RC, Kaur JK, Sawbridge T, Summerell BA. 2021. Comparative transcriptomic analysis of races 1, 2, 5 and 6 of *Fusarium oxysporum* f. sp. *pisi* in a susceptible pea host identifies differential pathogenicity profiles. *BMC Genomics* **22**:734.
- Alekseyenko AV. 2016. Multivariate Welch t-test on distances. *Bioinformatics* **32**:3552–3558.
- Álvarez-Loayza P, White JF Jr, Torres MS, Balslev H, Kristiansen T, Svenning JC, Gil N. 2011. Light converts endosymbiotic fungus to pathogen, influencing seedling survival and niche-space filling of a common tropical tree, *Iriartea deltoidea*. *PLoS One* **6**:e16386.
- Ampt EA, Bush DS, Siegel JP, Berenbaum MR. 2016. Larval preference and performance of *Amyelois transitella* (Navel Orangeworm, Lepidoptera: Pyralidae) in relation to the fungus *Aspergillus flavus*. *Environ Entomol.* **45**:155–162.
- Aoki T, Smith JA, Kasson MT, Freeman S, Geiser DM, Geering ADW, O'Donnell K. 2019. Three novel Ambrosia *Fusarium* clade species producing clavate macroconidia known (*F. floridanum* and *F. obliquiseptatum*) or predicted (*F. tuaranense*) to be farmed by *Euwallacea* spp. (Coleoptera: Scolytinae) on woody hosts. *Mycologia* **111**:919–935.
- Armitage AD, Taylor A, Sobczyk MK, Baxter L, Greenfield BPJ, Bates HJ, Wilson F, Jackson AC, Ott S, Harrison RJ, et al. 2018. Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *Sci Rep.* **8**:13530.

- Arnold AE, Miadlikowska J, Higgins KL, Sarvate SD, Gugger P, Way A, Hofstetter V, Kauff F, Lutzoni F. 2009. A phylogenetic estimation of trophic transition networks for ascomycetous fungi: are lichens cradles of symbiotrophic fungal diversification? *Syst Biol*. **58**:283–297.
- Badet T, Peyraud R, Mbengue M, Navaud O, Derbyshire M, Oliver RP, Barbacci A, Raffaele S. 2017. Codon optimization underpins generalist parasitism in fungi. *Elife* **6**:e22472.
- Beckerson WC, De La Vega RCR, Hartmann FE, Duhamel M, Giraud T, Perlin MH. 2019. Cause and effectors: whole-genome comparisons reveal shared but rapidly evolving effector sets among host-specific plant-castrating fungi. *MBio* **10**:e02391-19.
- Beimforde C, Feldberg K, Nylander S, Rikkinen J, Tuovila H, Dörfelt H, Gube M, Jackson DJ, Reitner J, Seyfullah LJ, et al. 2014. Estimating the Phanerozoic history of the Ascomycota lineages: Combining fossil and molecular data. *Mol Phylogenet Evol*. **78**:386–398.
- Bertazzoni S, Williams AH, Jones DA, Syme RA, Tan K-C, Hane JK. 2018. Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic fungi. *Mol Plant-Microbe Interact*. **31**:779–788.
- Biedermann PHW, Vega FE. 2020. Ecology and evolution of insect-fungus mutualisms. *Annu Rev Entomol*. **65**:431–455.
- Bilal L, Asaf S, Hamayun M, Gul H, Iqbal A, Ullah I, Lee IJ, Hussain A. 2018. Plant growth promoting endophytic fungi *Aspergillus fumigatus* TS1 and *Fusarium proliferatum* BRL1 produce gibberellins and regulates plant endogenous hormones. *Symbiosis* **76**: 117–127.
- Booth C. 1978. Do you believe in genera? *Trans Br Mycol Soc*. **71**:1–9.
- Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*. **4**:e16660.
- Botzman M, Margalit H. 2011. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol*. **12**:R109.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421.
- Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. **18**:188–196.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973.
- Charif D, Lobry JR. 2007. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. *Structural approaches to sequence evolution: molecules, networks, populations*. Berlin, Heidelberg: Springer. p. 207–232.
- Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH. 2004. Codon usage between genomes is constrained genome-wide mutational processes. *Proc Natl Acad Sci U S A* **101**:3480–3485.
- Chen S, Songkumarn P, Venu RC, Gowda M, Bellizzi M, Hu J, Liu W, Ebbole D, Meyers B, Mitchel T, et al. 2013. Identification and characterization of *in planta*-expressed secreted effector proteins from *Magnaporthe oryzae* that induce cell death in rice. *Mol Plant-Microbe Interact*. **26**:191–202.
- Chen Y, Zhou Q, Strelkov SE, Hwang S-F. 2014. Genetic diversity and aggressiveness of *Fusarium* spp. isolated from Canola in Alberta, Canada. *Plant Dis*. **98**:727–738.
- Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, Schmutz J, Taga M, White GJ, Zhou S, et al. 2009. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet*. **5**: e1000618.
- Crisuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*. **10**:210.
- Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog*. **8**:e1002608.
- Crous PW, Lombard L, Sandoval-Denis M, Seifert KA, Schroers H-J, Chaverri P, Gené J, Guarro J, Hirooka Y, Bensch K, et al. 2021. *Fusarium*: more than a node or a foot-shaped basal cell. *Stud Mycol*. **98**:100116.
- Crous PW, Osieck ER, Jurjević Ž, Boers J, van Iperen AL, Starink-Willemsse M, Dima B, Balashov S, Bulgakov TS, Johnston PR, et al. 2021. Fungal planet description sheets: 1284–1382. *Persoonia* **47**:178–374.
- Czislowski E, Zeil-rolfe I, Aitken EAB. 2021. Effector profiles of endophytic *Fusarium* associated with asymptomatic banana (*Musa* sp.) hosts. *Int J Mol Sci*. **22**:2508.
- Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T, Crandall K. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol*. **37**:291–294.
- Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, et al. 2012. The top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol*. **13**:414–430.
- De Jonge R, Bolton MD, Thomma BPHJ. 2011. How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Curr Opin Plant Biol*. **14**:400–406.
- Delaux PM, Séjalon-Delmas N, Bécard G, Ané JM. 2013. Evolution of the plant-microbe symbiotic “toolkit.” *Trends Plant Sci*. **18**: 298–304.
- Derbyshire MC, Harper LA, Lopez-Ruiz FJ. 2021. Positive selection of transcription factors is a prominent feature of the evolution of a plant pathogenic genus originating in the miocene. *Genome Biol Evol*. **13**:evab167.
- Desjardins AE, Proctor RH. 2007. Molecular biology of *Fusarium* mycotoxins. *Int J Food Microbiol*. **119**:47–50.
- Doré J, Kohler A, Dubost A, Hundley H, Singan V, Peng Y, Kuo A, Grigoriev IV, Martin F, Marmeisse R, et al. 2017. The ectomycorrhizal basidiomycete *Hebeloma cylindrosporum* undergoes early waves of transcriptional reprogramming prior to symbiotic structures differentiation. *Environ Microbiol*. **19**:1338–1354.
- dos Reis M, Gunnell GF, Barba-Montoya J, Wilkins A, Yang Z, Yoder AD. 2018. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Syst Biol*. **67**:594–615.
- dos Reis M, Savva R, Wernisch L. 2004. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res*. **32**:5036–5044.
- dos Reis M, Yang Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol*. **28**:2161–2172.
- Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, Hacquard S. 2018. Microbial interkingdom interactions in roots promote *Arabidopsis* survival. *Cell* **175**:973–983.
- Emms DM, Kelly S. 2018. STAG: species tree inference from all genes. *bioRxiv [Preprint]*.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. **20**:238.
- Eskalen A, Gonzalez A, Wang DH, Twizeyimana M, Mayorquin SJ, Lynch SC. 2012. First report of a *Fusarium* sp. and its vector tea shot hole borer (*Euwallacea fornicatus*) causing *Fusarium* die-back on avocado in California. *Plant Dis*. **96**:1070.
- Forest F. 2009. Calibrating the tree of life: fossils, molecules and evolutionary timescales. *Ann Bot*. **104**:789–794.
- Franco MEE, Wisecaver JH, Arnold AE, Ju Y, Slot JC, Ahrendt S, Moore LP, Eastman KE, Scott K, Konkel Z, et al. 2022. Ecological generalism drives hyperdiversity of secondary metabolite gene clusters in xylarialean endophytes. *New Phytol*. **233**:1317–1330.
- Freeman S, Sharon M, Maymon M, Mendel Z, Protasov A, Aoki T, Eskalen A, O'Donnell K. 2013. *Fusarium euwallaceae* sp. nov.-a symbiotic fungus of *Euwallacea* sp., an invasive ambrosia beetle in Israel and California. *Mycologia* **105**:1595–1606.



- Gazis R, Kuo A, Riley R, LaButti K, Lipzen A, Lin J, Amirebrahimi M, Hesse CN, Spatafora JW, Henrissat B, et al. 2016. The genome of *Xylona heveae* provides a window into fungal endophytism. *Fungal Biol.* **120**:26–42.
- Geiser DM, Al-Hatmi A, Aoki T, Arie T, Balmas V, Barnes I, Bergstrom GC, Bhattacharyya MKK, Blomquist CL, Bowden R, et al. 2021. Phylogenomic analysis of a 55.1 kb 19-gene dataset resolves a monophyletic *Fusarium* that includes the *Fusarium solani* species complex. *Phytopathology* **111**:1064–1079.
- Geiser DM, Jiménez-Gasco MDM, Kang S, Makalowska I, Veeraraghavan N, Ward TJ, Zhang N, Kuldau GA, O'Donnell K. 2004. FUSARIUM-ID v. 1.0: A DNA sequence database for identifying *Fusarium*. *Eur J Plant Pathol.* **110**:473–479.
- Graur D, Martin W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* **20**:80–86.
- Guo L, Yu H, Wang B, Vescio K, Delulio GA, Yang H, Berg A, Zhang L, Edel-Hermann V, Steinberg C, et al. 2021. Metatranscriptomic comparison of endophytic and pathogenic *Fusarium-Arabidopsis* interactions reveals plant transcriptional plasticity. *Mol Plant-Microbe Interact.* **34**:1071–1083.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**:1072–1075.
- Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, Weimann A, Damm U, Dallery JF, Hainaut M, et al. 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat Commun.* **7**:11362.
- Hardoim PR, van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A, Döring M, Sessitsch A. 2015. The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol Mol Biol Rev.* **79**:293–320.
- He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**:1157–1164.
- Hershberg R, Petrov DA. 2008. Selection on Codon Bias. *Annu Rev Genet.* **42**:287–299.
- Hill R, Llewellyn T, Downes E, Oddy J, MacIntosh C, Kallow S, Panis B, Dickie JB, Gaya E. 2021. Seed banks as incidental fungi banks: fungal endophyte diversity in stored seeds of banana wild relatives. *Front Microbiol.* **12**:643731.
- Hipsley CA, Müller J. 2014. Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. *Front Genet.* **5**:138.
- Hiruma K, Gerlach N, Sacristán S, Nakano RT, Hacquard S, Kracher B, Neumann U, Ramírez D, Bucher M, O'Connell RJ, et al. 2016. Root endophyte *Colletotrichum tofieldiae* confers plant fitness benefits that are phosphate status dependent. *Cell* **165**:464–474.
- Ho SYW. 2009. An examination of phylogenetic models of substitution rate variation among lineages. *Biol Lett.* **5**:421–424.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* **35**:518–522.
- Holtz MD, Chang KF, Hwang SF, Gossen BD, Strelkov SE. 2011. Characterization of *Fusarium avenaceum* from lupin in central Alberta: genetic diversity, mating type and aggressiveness. *Can J Plant Pathol.* **33**:61–76.
- Hulcr J, Cognato AI. 2010. Repeated evolution of crop theft in fungus-farming ambrosia beetles. *Evolution* **64**:3205–3212.
- Jordal BH, Cognato AI. 2012. Molecular phylogeny of bark and ambrosia beetles reveals multiple origins of fungus farming during periods of global warming. *BMC Evol Biol.* **12**:133.
- Junker C, Draeger S, Schulz B. 2012. A fine line-endophytes or pathogens in *Arabidopsis thaliana*. *Fungal Ecol.* **5**:657–662.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* **14**:587–589.
- Kasson MT, O'Donnell K, Rooney AP, Sink S, Ploetz RC, Ploetz JN, Konkol JL, Carrillo D, Freeman S, Mendel Z, et al. 2013. An inordinate fondness for *Fusarium*: phylogenetic diversity of fusaria cultivated by ambrosia beetles in the genus *Euwallacea* on avocado and other plant hosts. *Fungal Genet Biol.* **56**:147–157.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**:772–780.
- Kavroulakis N, Ntougias S, Zervakis GI, Ehaliotis C, Haralampidis K, Papadopoulou KK. 2007. Role of ethylene in the protection of tomato plants against soil-borne fungal pathogens conferred by an endophytic *Fusarium solani* strain. *J Exp Bot.* **58**:3853–3864.
- Kim K-T, Jeon J, Choi J, Cheong K, Song H, Choi G, Kang S, Lee Y-H. 2016. Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. *Front Plant Sci.* **7**:186.
- Knapp DG, Németh JB, Barry K, Hainaut M, Henrissat B, Johnson J, Kuo A, Lim JHP, Lipzen A, Nolan M, et al. 2018. Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. *Sci Rep.* **8**:6321.
- Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, et al. 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet.* **47**:410–415.
- Kokkonen M, Ojala L, Parikka P, Jostoi M. 2010. Mycotoxin production of selected *Fusarium* species at different culture conditions. *Int J Food Microbiol.* **143**:17–25.
- Kosakovsky Pond SL, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**:676–679.
- Kosakovsky Pond SL, Wisotsky SR, Escalante A, Magalis BR, Weaver S. 2021. Contrast-FEL—a test for differences in selective pressures at individual sites among clades and sets of branches. *Mol Biol Evol.* **38**:1184–1198.
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**:4453–4455.
- Kubicek CP, Starr TL, Glass NL. 2014. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annu Rev Phytopathol.* **52**:427–451.
- LaBella AL, Opulente DA, Steenwyk JL, Hittinger CT, Rokas A. 2021. Signatures of optimal codon usage in metabolic genes inform budding yeast ecology. *PLoS Biol.* **19**:e3001185.
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Research* **6**:1287.
- Lahrman U, Strehmel N, Langen G, Frerigmann H, Leson L, Ding Y, Scheel D, Herklotz S, Hilbert M, Zuccaro A. 2015. Mutualistic root endophytism is not associated with the reduction of saprotrophic traits and requires a noncompromised plant innate immunity. *New Phytol.* **207**:841–857.
- Lartillot N, Delsuc F. 2012. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution.* **66**:1773–1787.
- Lartillot N, Phillips MJ, Ronquist F. 2016. A mixed relaxed clock model. *Philos Trans R Soc B* **371**:20150132.
- Lebreton A, Zeng Q, Miyauchi S, Kohler A, Dai YC, Martin FM. 2021. Evolution of the mode of nutrition in symbiotic and saprotrophic fungi in forest ecosystems. *Annu Rev Ecol Evol Syst.* **52**:385–404.
- Leoni C, de Vries M, ter Braak CJF, van Bruggen AHC, Rossing WAH. 2013. *Fusarium oxysporum* f.sp. *cepae* dynamics: in-plant multiplication and crop sequence simulations. *Eur J Plant Pathol.* **137**:545–561.
- Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* **6**:41.
- Li J, Fokkens L, Conneely LJ, Rep M. 2020. Partial pathogenicity chromosomes in *Fusarium oxysporum* are sufficient to cause disease and can be horizontally transferred. *Environ Microbiol.* **22**:4985–5004.
- Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, Stajich JE, Spatafora JW, Groenewald M, Dunn CW, Hittinger CT, et al. 2021. A genome-scale phylogeny of the kingdom Fungi. *Curr Biol.* **31**:1653–1665.



- Linder M, Britton T, Sennblad B. 2011. Evaluation of bayesian models of substitution rate evolution-parental guidance versus mutual independence. *Syst Biol*. **60**:329–342.
- Lombard L, van der Merwe NA, Groenewald JZ, Crous PW. 2015. Generic concepts in *Nectriaceae*. *Stud Mycol*. **80**:189–245.
- Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. 2015. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol*. **66**:513–545.
- Lutzoni F, Nowak MD, Alfaro ME, Reeb V, Miadlikowska J, Krug M, Arnold AE, Lewis LA, Swofford DL, Hibbett D, et al. 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat Commun*. **9**:5451.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*. **290**:1151–1155.
- Ma LJ, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, Trail F, Gardiner DM, Manners JM, Kazan K. 2013. *Fusarium* pathogenomics. *Annu Rev Microbiol*. **67**:399–416.
- Ma LJ, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. **464**:367–373.
- Mallick S, Gnerre S, Muller P, Reich D. 2009. The difficulty of avoiding false positives in genome scans for natural selection. *Genome Res*. **19**:922–933.
- Marqués-Gálvez JE, Miyauchi S, Paolocci F, Navarro-Ródenas A, Arenas F, Pérez-Gilbert M, Morin E, Auer L, Barry KW, Kuo A, et al. 2021. Desert truffle genomes reveal their reproductive modes and new insights into plant–fungal interaction and ectendomycorrhizal lifestyle. *New Phytol*. **229**:2917–2932.
- Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, et al. 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*. **464**:1033–1038.
- McDonald AG, Boyce S, Tipton KF. 2009. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res*. **37**:D593–D597.
- Mead ME, Steenwyk JL, Silva LP, de Castro PA, Saeed N, Hillmann F, Goldman GH, Rokas A. 2021. An evolutionary genomic approach reveals both conserved and species-specific genetic elements related to human disease in closely related *Aspergillus* fungi. *Genetics*. **218**:iyab066.
- Mendoza AR, Sikora RA. 2009. Biological control of *Radopholus similis* in banana by combined application of the mutualistic endophyte *Fusarium oxysporum* strain 162, the egg pathogen *Paecilomyces lilacinus* strain 251 and the antagonistic bacteria *Bacillus firmus*. *BioControl*. **54**:263–272.
- Menzies JG, Koch C, Seywerd F. 1990. Additions to the host range of *Fusarium oxysporum* f. sp. *radicis-lycopersici*. *Plant Dis*. **74**:569–572.
- Mesny F, Miyauchi S, Thiérgart T, Pickel B, Atanasova L, Karlsson M, Hüttel B, Barry KW, Haridas S, Chen C, et al. 2021. Genetic determinants of endophytism in the *Arabidopsis* root mycobiome. *Nat Commun*. **12**:7227.
- Mesny F, Vannier N. 2020. Detecting the effect of biological categories on genome composition. Available from: <https://github.com/fantin-mesny/Effect-Of-Biological-Categories-On-Genomes-Composition>.
- Meyer RS, Duval AE, Jensen HR. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol*. **196**:29–48.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. **37**:1530–1534.
- Miura S, Tamura K, Tao Q, Huuki LA, Kosakovsky Pond SL, Priest J, Deng J, Kumar S. 2020. A new method for inferring timetrees from temporally sampled molecular sequences. *PLoS Comput Biol*. **16**:e1007046.
- Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, Morin E, Andreopoulos B, Barry KW, Bonito G, et al. 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nat Commun*. **11**:5125.
- Miyauchi S, Navarro D, Grisel S, Chevret D, Berrin J-G, Rosso M-N. 2017. The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated lignocellulose breakdown. *PLoS One*. **12**:e0175528.
- Mueller UG, Gerardo NM, Aanen DK, Six DL, Schultz TR. 2005. The evolution of agriculture in insects. *Annu Rev Ecol Evol Syst*. **36**:563–595.
- Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, et al. 2015. Gene-wide identification of episodic selection. *Mol Biol Evol*. **32**:1365–1371.
- Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K. 2019. Transposable elements contribute to fungal genes and impact fungal lifestyle. *Sci Rep*. **9**:4307.
- Na F, Carrillo JD, Mayorquin JS, Ndinga-Muniania C, Stajich JE, Stouthamer R, Huang YT, Lin YT, Chen CY, Eskalen A. 2018. Two novel fungal symbionts *Fusarium kuroshium* sp. nov. and *Graphium kuroshium* sp. nov. of kuroshio shot hole borer (*Euwallacea* sp. nr. *fornicatus*) cause fusarium dieback on woody host species in California. *Plant Dis*. **102**:1154–1164.
- Nelson A, Vandegrift R, Carroll GC, Roy BA. 2020. Double lives: transfer of fungal endophytes from leaves to woody substrates. *PeerJ*. **8**:e9341.
- Niehaus EM, Münsterkötter M, Proctor RH, Brown DW, Sharon A, Idan Y, Oren-Young L, Sieber CM, Novák O, Pěnčík A, et al. 2016. Comparative “Omics” of the *Fusarium fujikuroi* Species Complex Highlights Differences in Genetic Potential and Metabolite Synthesis. *Genome Biol Evol*. **8**:3574–3599.
- Niu G, Siegel J, Schuler MA, Berenbaum MR. 2009. Comparative toxicity of mycotoxins to navel orangeworm (*Amyelois transitella*) and corn earworm (*Helicoverpa zea*). *J Chem Ecol*. **35**:951–957.
- O'Donnell K, Al-Hatmi AMS, Aoki T, Brankovics B, Cano-Lira JF, Coleman JJ, de Hoog GS, Di Pietro A, Frandsen RJN, Geiser DM, et al. 2020. No to *Neocosmospora*: phylogenomic and practical reasons for continued inclusion of the *Fusarium solani* species complex in the genus *Fusarium*. *mSphere*. **5**:e00810-20.
- O'Donnell K, Libeskind-Hadas R, Hulcr J, Bateman C, Kasson MT, Ploetz RC, Konkol JL, Ploetz JN, Carrillo D, Campbell A, et al. 2016. Invasive Asian *Fusarium* – *Euwallacea* ambrosia beetle mutualists pose a serious threat to forests, urban landscapes and the avocado industry. *Phytoparasitica*. **44**:435–442.
- O'Donnell K, Rooney AP, Proctor RH, Brown DW, McCormick SP, Ward TJ, Frandsen RJN, Lysøe E, Rehner SA, Aoki T, et al. 2013. Phylogenetic analyses of RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. *Fungal Genet Biol*. **52**:20–31.
- O'Donnell K, Sink S, Libeskind-Hadas R, Hulcr J, Kasson MT, Ploetz RC, Konkol JL, Ploetz JN, Carrillo D, Campbell A, et al. 2015. Discordant phylogenies suggest repeated host shifts in the *Fusarium-Euwallacea* ambrosia beetle mutualism. *Fungal Genet Biol*. **82**:277–290.
- Ohta T. 1996. The current significance and standing of neutral and nearly neutral theories. *BioEssays*. **18**:673–677.
- Okello PN, Petrovic K, Singh AK, Kontz B, Mathew FM. 2020. Characterization of species of *Fusarium* causing root rot of Soybean (*Glycine max* L.) in South Dakota, USA. *Can J Plant Pathol*. **42**:560–571.
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. 2019. vegan: Community Ecology Package. Available from: <https://cran.r-project.org/package=vegan>.
- Ordóñez N, Seidl MF, Waalwijk C, Drenth A, Kilian A, Thomma BPHJ, Ploetz RC, Kema GHJ. 2015. Worse comes to worst: bananas and panama disease—when plant and pathogen clones meet. *PLoS Pathog*. **11**:e1005197.

- Parsa S, García-Lemos AM, Castillo K, Ortiz V, López-Lavalle LAB, Braun J, Vega FE. 2016. Fungal endophytes in germinated seeds of the common bean, *Phaseolus vulgaris*. *Fungal Biol.* **120**: 783–790.
- Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2010. How many bootstrap replicates are necessary? *J Comput Biol.* **17**:337–354.
- Peck LD, Nowell RW, Flood J, Ryan MJ, Barraclough TG. 2021. Historical genomics reveals the evolutionary mechanisms behind multiple outbreaks of the host-specific coffee wilt pathogen *Fusarium xylarioides*. *BMC Genomics* **22**:404.
- Pendleton AL, Smith KE, Feau N, Martin FM, Grigoriev IV, Hamelin R, Nelson CD, Burleigh JG, Davis JM. 2014. Duplications and losses in gene families of rust pathogens highlight putative effectors. *Front Plant Sci.* **5**:299.
- Peter M, Kohler A, Ohm RA, Kuo A, Krützmann J, Morin E, Arend M, Barry KW, Binder M, Choi C, et al. 2016. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus  *Cenococcum geophilum*. *Nat Commun.* **7**:12662.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. 2021. nlme: Linear and Nonlinear Mixed Effects Models. Available from: <https://cran.r-project.org/package=nlme>.
- Plett JM, Martin F. 2015. Reconsidering mutualistic plant-fungal interactions through the lens of effector biology. *Curr Opin Plant Biol.* **26**:45–50.
- Ploetz RC. 2005. Panama disease: an old nemesis rears its ugly head. Part 1. The beginnings of the banana export trades. *Plant Health Prog.* **6**:18.
- Powell CLE, Waskin S, Battistuzzi FU. 2020. Quantifying the error of secondary vs. distant primary calibrations in a simulated environment. *Front Genet.* **11**:252.
- Prompttha I, Hyde KD, McKenzie EHC, Peberdy JF, Lumyong S. 2010. Can leaf degrading enzymes provide evidence that endophytic fungi becoming saprobes? *Fungal Divers.* **41**:89–99.
- Rafiqi M, Ellis JG, Ludowici VA, Hardham AR, Dodds PN. 2012. Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Curr Opin Plant Biol.* **15**: 477–482.
- Rahman S, Kosakovsky Pond SL, Webb A, Hey J. 2021. Weak selection on synonymous codons substantially inflates dN/dS estimates in bacteria. *Proc Natl Acad Sci U S A* **118**:e2023575118.
- R Core Team. 2020. R: A language and environment for statistical computing. Available from: <https://www.r-project.org/>.
- Rashmi M, Kushveer JS, Sarma VV. 2019. A worldwide list of endophytic fungi with notes on ecology and diversity. *Mycosphere* **10**:798–1079.
- Redman RS, Dunigan DD, Rodriguez RJ. 2001. Fungal symbiosis from mutualism to parasitism: who controls the outcome, host or invader? *New Phytol.* **151**:705–716.
- Rich MK, Vigneron N, Liboure C, Keller J, Xue L, Hajheidari M, Radhakrishnan GV, Le Ru A, Diop SI, Potente G, et al. 2021. Lipid exchanges drove the evolution of mutualism during plant terrestrialization. *Science* **372**:864–868.
- Rodriguez RJ, White JF Jr, Arnold AE, Redman RS. 2009. Fungal endophytes: diversity and functional roles. *New Phytol.* **182**: 314–330.
- Rubini MR, Silva-Ribeiro RT, Pomella AWV, Maki CS, Araújo WL, Dos Santos DR, Azevedo JL. 2005. Diversity of endophytic fungal community of cacao (*Theobroma cacao* L.) and biological control of *Crinipellis pernicioso*, causal agent of Witches' Broom Disease. *Int J Biol Sci.* **1**:24–33.
- Sakuno E, Yabe K, Nakajima H. 2003. Involvement of two cytosolic enzymes and a novel intermediate, 5'-oxoaverantin, in the pathway from 5'-hydroxyaverantin to Averufin in Aflatoxin biosynthesis. *Appl Environ Microbiol.* **69**:6418–6426.
- Sandoval-Denis M, Guarnaccia V, Polizzi G, Crous PW. 2018. Symptomatic *Citrus* trees reveal a new pathogenic lineage in *Fusarium* and two new *Neocosmospora* species. *Persoonia* **40**: 1–25.
- Sandoval-Denis M, Lombard L, Crous PW. 2019. Back to the roots: a reappraisal of *Neocosmospora*. *Persoonia* **43**:90–185.
- Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK, Carpenter RJ, et al. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of *Nothofagus* (Fagales). *Syst Biol.* **61**:289–313.
- Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol.* **33**: 1654–1668.
- Schenk JJ. 2016. Consequences of secondary calibrations on divergence time estimates. *PLoS One* **11**:e0148228.
- Schirrmann MK, Zoller S, Croll D, Stukenbrock EH, Leuchtman A, Fior S. 2018. Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Mol Ecol.* **27**: 3070–3086.
- Schliep K, Potts AJ, Morrison DA, Grimm GW. 2017. Intertwining phylogenetic trees and networks. *Methods Ecol Evol.* **8**:1212–1220.
- Schmitz AM, Pawlowska TE, Harrison MJ. 2019. A short LysM protein with high molecular diversity from an arbuscular mycorrhizal fungus, *Rhizophagus irregularis*. *Mycoscience* **60**:63–70.
- Schroers HJ, Gräfenhan T, Nirenberg HI, Seifert KA. 2011. A revision of *Cyanonectria* and *Geejayessia* gen. nov., and related species with *Fusarium*-like anamorphs. *Stud Mycol.* **68**:115–138.
- Schulz B, Boyle C. 2005. The endophytic continuum. *Mycol Res.* **109**: 661–686.
- Seidl MF, Faino L, Shi-Kunne X, Van Den Berg GCM, Bolton MD, Thomma BPHJ. 2015. The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. *Mol Plant-Microbe Interact.* **28**:362–373.
- Selosse MA, Schneider-Maunoury L, Martos F. 2018. Time to re-think fungal ecology? Fungal ecological niches are often prejudged. *New Phytol.* **217**:968–972.
- Shao X, Lv N, Liao J, Long J, Xue R, Ai N, Xu D, Fan X. 2019. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med Genet.* **20**:175.
- Sharp PM, Li W-H. 1987. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**:1281–1295.
- Shaul S, Graur D. 2002. Playing chicken (*Gallus gallus*): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. *Gene* **300**:59–61.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**:1117–1123.
- Šišić A, Bačanović-Šišić J, Al-Hatmi AMS, Karlovsky P, Ahmed SA, Maier W, De Hoog GS, Finckh MR. 2018. The “forma specialis” issue in *Fusarium*: a case study in *Fusarium solani* f. sp. *pisi*. *Sci Rep.* **8**:1252.
- Slippers B, Wingfield MJ. 2007. Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal Biol Rev.* **21**:90–106.
- Smith S, Brown J, Walker J. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS One* **13**:e0197433.
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol.* **32**:1342–1353.
- Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, Ver Loren van Themaat E, Brown JKM, Butcher SA, Gurr SJ, et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543–1546.

- Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A, *et al.* 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**:1028–1046.
- Sperschneider J, Dodds PN. 2021. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol Plant-Microbe Interact.* **35**:146–156.
- Steenwyk JL, Buida TJ, Li Y, Shen XX, Rokas A. 2020. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* **18**:e3001007.
- Steenwyk JL, Rokas A. 2018. Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation. *Front Microbiol.* **9**:288.
- Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A. 2019. A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. *MBio* **10**:e00925-19.
- Steenwyk JL, Soghigian JS, Perfect JR, Gibbons JG. 2016. Copy number variation contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from the North American Pacific Northwest. *BMC Genomics* **17**:700.
- Stergiopoulos I, de Wit PJGM. 2009. Fungal effector proteins. *Annu Rev Phytopathol.* **47**:233–263.
- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, Grassi AD, Lee C, *et al.* 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**:848–853.
- Summerell BA. 2019. Resolving *Fusarium*: current status of the genus. *Annu Rev Phytopathol.* **57**:323–339.
- Swett CL, Gordon TR. 2015. Endophytic association of the pine pathogen *Fusarium circinatum* with corn (*Zea mays*). *Fungal Ecol.* **13**:120–129.
- Tamuri AU, dos Reis M. 2021. A mutation-selection model of protein evolution under persistent positive selection. *bioRxiv [Preprint]* 2021.05.18.444611.
- Tan G, Muffato M, Ledergerber C, Herrero J, Goldman N, Gil M, Dessimoz C. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst Biol.* **64**:778–791.
- Tao Q, Tamura K, Battistuzzi FU, Kumar S. 2019. A machine learning method for detecting autocorrelation of evolutionary rates in large phylogenies. *Mol Biol Evol.* **36**:811–824.
- Taylor JW, Berbee ML. 2006. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia* **98**:838–849.
- Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, De Silva N, Martinez MC, Pedro H, Yates AD, *et al.* 2020. PHI-base: the pathogen-host interactions database. *Nucleic Acids Res.* **48**:D613–D620.
- van Dam P, Fokkens L, Schmidt SM, Linmans JHJ, Corby Kistler H, Ma LJ, Rep M. 2016. Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environ Microbiol.* **18**:4087–4102.
- van Dam P, Rep M. 2017. The distribution of *miniature impala* elements and *SIX* genes in the *Fusarium* genus is suggestive of horizontal gene transfer. *J Mol Evol.* **85**:14–25.
- Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllösi GJ, Szarkándi JG, Papp V, Albert L, *et al.* 2019. Megaphylogeny resolves global patterns of mushroom evolution. *Nat Ecol Evol.* **3**:668–678.
- Veneault-Fourrey C, Commun C, Kohler A, Morin E, Balestrini R, Plett J, Danchin E, Coutinho P, Wiebenga A, de Vries RP, *et al.* 2014. Genomic and transcriptomic analysis of *Laccaria bicolor* CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. *Fungal Genet Biol.* **72**:168–181.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**:e112963.
- Wapinski I, Pfeffer A, Friedman N, Regev A. 2007. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**:54–61.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* **18**:691–699.
- Wolinska KW, Vannier N, Thiergart T, Pickel B, Gremmen S, Piasecka A, Piślewska-Bednarek M, Nakano RT, Belkhadir Y, Bednarek P, *et al.* 2021. Tryptophan metabolism and bacterial commensals prevent fungal dysbiosis in *Arabidopsis* roots. *Proc Natl Acad Sci U S A* **118**:e2111521118.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* **24**:1586–1591.
- Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol.* **23**:212–226.
- Zakaria L, Aziz WNW. 2018. Molecular identification of endophytic fungi from banana leaves (*Musa* spp.). *Trop Life Sci Res.* **29**:201–211.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* **19**:153.
- Zhang C, Scornavacca C, Molloy EK, Mirarab S. 2020. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Mol Biol Evol.* **37**:3292–3307.
- Zhang Y, Yang H, Turra D, Zhou S, Ayhan DH, Delulio GA, Guo L, Broz K, Wiederhold N, Coleman JJ, *et al.* 2020. The genome of opportunistic fungal pathogen *Fusarium oxysporum* carries a unique set of lineage-specific chromosomes. *Commun Biol.* **3**:50.
- Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**:W95–W101.
- Zhao S, Gibbons JG. 2018. A population genomic characterization of copy number variation in the opportunistic fungal pathogen *Aspergillus fumigatus*. *PLoS One* **13**:e0201611.
- Zhao Z, Liu H, Wang C, Xu JR. 2013. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **14**:274.
- Zuccaro A, Lahrman U, Langen G. 2014. Broad compatibility in fungal root symbioses. *Curr Opin Plant Biol.* **20**:135–145.