



Article

Towards a Phylogenomic Framework for the *Fusarium oxysporum* Species Complex

Juliana Lopez-Jimenez ¹, Juan M. Daza ² and Juan F. Alzate ^{1,3,*} 

¹ Centro Nacional de Secuenciación Genómica—CNSG, Sede de Investigación Universitaria—SIU, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín 050010, Colombia; extension.cnsg@udea.edu.co

² Grupo Herpetológico de Antioquia GHA, Instituto de Biología, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín 050010, Colombia; juanm.daza@udea.edu.co

³ Departamento de Microbiología y Parasitología, Facultad de Medicina, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín 050010, Colombia

* Correspondence: jfernando.alzate@udea.edu.co; Tel.: +57-6042196665

Abstract

The *Fusarium oxysporum* species complex (FOSC) is a genetically diverse and globally distributed group of fungi that includes both pathogenic and non-pathogenic lineages with broad host ranges and major agricultural importance. To provide a higher-resolution view of its diversity and evolutionary history, we assembled and quality-filtered genomes, generating a curated dataset of 336 high-quality assemblies complemented by seven NCBI reference genomes (total 343). Orthology inference recovered 4286 conserved single-copy coding sequences, which resolved evolutionary relationships across the complex with high confidence. This framework recognizes three major clades, resolves several taxonomic inconsistencies, and delineates phylogenomic boundaries among distinct lineages. Despite overall nucleotide identities exceeding 96%, whole-genome comparisons consistently supported the existence of distinct species lineages. Within Clade 2, we identified two previously unrecognized lineages: the *Fusarium* “afroindicum” clade, distributed across Africa and Asia, and the *Fusarium* “europaeum” clade, restricted to Europe. Both lineages are genetically coherent, ecologically distinct, and strongly supported by phylogenomic evidence. Finally, a global metadata survey revealed that *Fusarium fabacearum*—rather than *F. oxysporum* sensu stricto—is the most widespread and host-diverse member of the complex. Collectively, this large-scale assembly effort and comprehensive phylogenomic framework provide the most robust evolutionary context to date for the FOSC and shed light on its global diversity and ecological breadth.

Keywords: *Fusarium oxysporum* species complex; Phylogenomics; Taxonomy; Genomic diversity; *Fusarium fabacearum*; *Fusarium cugenangense*; *Fusarium odoratissimum*



Academic Editors: Alexandre Vetcher
and Myakishev-Rempel Max

Received: 30 May 2026

Revised: 29 June 2026

Accepted: 3 July 2026

Published: 14 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The *Fusarium oxysporum* species complex (FOSC) is a widely spread group of fungi that significantly impacts agriculture and has well-documented effects on human and animal health [1–3]. Several species within this complex are common around the world and pose a persistent and growing threat to global agriculture due to their ability to infect a very broad range of plant hosts [4,5] while the complex also includes non-pathogenic strains [6–9].

Members of the *Fusarium oxysporum* species complex (FOSC) are causal agents of root rot and wilt diseases that affect a broad spectrum of economically important crops.

Reported hosts span fruits such as strawberries, bananas, and date palms; vegetables including tomatoes, cucumbers, lettuce, and coriander; legumes such as chickpeas and yams; and industrial or specialty plants including cotton, hemp, coffee, and ginger. This remarkable host range has been documented across multiple continents in a wide body of studies [10–35].

Recognized as one of the most economically damaging groups of plant pathogens worldwide [36], FOSC is consistently linked to significant yield losses and long-term soil persistence [4,37], making its eradication and management especially difficult [38,39]. Nonetheless, certain *F. oxysporum* strains have been described as saprophytic, meaning they are capable of growing on dead or decaying organic matter, although they can also act as pathogens under certain conditions [40,41], highlighting the importance of a reliable, high-resolution, and high-precision taxonomic framework that enables the identification and tracking of different species or even strains within the complex.

The current definition of *Fusarium oxysporum* remains broad and somewhat ambiguous [42], as it depends heavily on morphological traits and phenotypic subspecific classifications—such as *formae speciales*, races, and vegetative compatibility groups (VCGs). Additionally, species historically categorized within the FOSC are notoriously difficult to identify based solely on microscopic morphology or culture traits. Diagnostic features, such as conidial size and shape, septation patterns, or the arrangement of phialides, often overlap significantly among different taxa, and culture-based characteristics can vary depending on growth conditions and media composition [43]. As a result, identifications can vary significantly across laboratories and often depend on the individual expertise of the analyst, which limits the reproducibility and reliability of purely morphological methods. Historically, host-based classification systems, particularly the *formae speciales* framework, have been used to group isolates by host range and further subdivide them into races based on their virulence spectra. While this provided a useful historical scheme, its limitations have become more apparent, as they fail to capture the underlying evolutionary relationships [6,12,44].

In response to the limitations of host-based classifications and purely morphological approaches, molecular methods began to emerge in the mid-90s, initially focusing on PCR amplification and single-locus sequencing of ribosomal DNA regions [45]. While rDNA became the universal fungal barcode, it proved to have very limited resolution. With the advent of advanced sequencing technologies, additional molecular markers were incorporated into FOSC research, introducing multilocus sequence analysis (MLSA), which primarily relied on phylogenetic methods. MLSA schemes expanded to include dozens of markers, improving upon rDNA resolution; however, issues such as paralogy, recombination, insufficient phylogenetic resolution, and the lack of standardized marker sets often produced incongruent topologies and restricted cross-study comparability. Nonetheless, loci such as translation elongation factor 1 α (*tef1*), RNA polymerase II second largest subunit (*rpb2*), β -tubulin (*tub2*), and calmodulin A (*cmdA*) played a pivotal role in shaping early species delimitation hypotheses within the FOSC [42,46–52].

Genome-wide phylogenomic approaches have become essential for resolving the long-standing taxonomic issues of the FOSC. Members of the FOSC typically exhibit genome sizes ranging from approximately 45 to 70 Mb, although substantial variation exists among lineages due to differences in accessory genomic content [53]. This variation in genomic content is primarily driven by horizontal gene transfer (HGT) of effector genes between lineages, contributing to the observed differences in pathogenicity and adaptability. These effector genes, often located on accessory chromosomes, play a crucial role in virulence and adaptation to different host plants, enabling different lineages to acquire traits that enhance their pathogenic potential [54]. Recent research using large-scale datasets of conserved

single-copy genes has offered unprecedented insights into its evolutionary history. A core-genome phylogenomic framework that integrates hundreds of loci revealed genome-wide evolutionary patterns, such as GC content variation and lineage-specific duplication events, thereby refining our understanding of genomic diversity and adaptation within the genus [55]. Extending this framework, a core-gene phylogeny based on 3811 conserved single-copy BUSCO genes from 69 banana-infecting strains and 55 additional isolates infecting other hosts demonstrated that banana pathogenic forms are polyphyletic, distributed across multiple major clades of the FOSC, underscoring that pathogenicity toward banana has arisen multiple times independently [56]. Complementary population genomics based on 410 assembled genomes revealed a predominantly clonal evolutionary pattern, with divergence dating suggesting that major lineages separated approximately 500,000 years ago. By integrating phylogenomic concordance with population-level signals of recombination, this study provided one of the most comprehensive resources for the FOSC and reinforced the recognition of well-supported phylogroups as species-level lineages [57]. High-quality genome assemblies, like those of *F. oxysporum* QK8, enabled strong phylogenomic analysis across 24 other *Fusarium* species, showing the power of whole-genome sequencing to link genome structure, gene family increases, and pathogenicity factors to evolutionary change [20]. Similarly, effector-based comparative genomics has shown that host-specific effector repertoires may transcend phylogenetic boundaries, highlighting their importance in the independent emergence of pathogenic lineages within the FOSC. These patterns are consistent with a significant contribution of horizontal gene transfer of effector genes, often involving accessory chromosomes [58].

With this work, our goal was to create a curated and reproducible phylogenomic framework to address key questions about the evolutionary history of the FOSC, clarify relationships among its recognized species, and test its monophyly. Additionally, we aimed to offer insights into the evolution of overall genomic features within the complex, enabling the integration of geographical distributions and host associations, and allowing long-standing paradigms to be contrasted with evidence from modern high-resolution technologies.

2. Results

2.1. Genome Statistics and Quality Assessment

We compiled a comprehensive collection of publicly available sequencing datasets for members of the *Fusarium oxysporum* species complex (FOSC) annotated as *Fusarium oxysporum*, *Fusarium foetens*, *Fusarium veterinarianum*, or *Fusarium odoratissimum*. Application of these selection criteria resulted in the retention of 721 SRA entries.

All sequencing datasets were cleaned and de novo assembled as described in Section 4. To ensure high-quality assemblies, we evaluated several parameters, including total assembly size, number of scaffolds, scaffold N50, average sequencing depth, number of alternate alleles, GC content, proportion of ambiguous bases (Ns), and BUSCO-based genome quality metrics (completeness and proportion of duplicated genes).

Descriptive statistics revealed the presence of low-quality assemblies, including one with a total size of <1 Mb and several others below 40 Mb. Conversely, some assemblies showed evidence of possible contamination, with total lengths exceeding 80 Mb, and in extreme cases >100 Mb. Additional issues included highly fragmented assemblies with more than 100,000 scaffolds and very low contiguity (N50 < 1000 bp).

To filter out such poor-quality assemblies, we applied the following thresholds: N50 > 50,000 bp, assembly size between 44 Mb and 75 Mb, a maximum of 50,000 scaffolds, and a minimum sequencing depth of 10×. Although some of these thresholds were

intentionally permissive, only about half of the assemblies passed the quality filters ($n = 336$).

To assess the presence and extent of potential library contamination by non-target organisms, raw sequencing reads were taxonomically classified using Kraken2 and Bracken. Although these tools do not provide definitive organismal identification, they are effective for identifying contaminated libraries and for broadly inferring the likely biological origin of non-target sequences. Across the initial dataset of 721 genome assemblies, substantial contamination was observed in a subset of read sets; however, after applying quality-control and filtering criteria, contamination was markedly reduced, with a median of 97% of reads assigned to the genus *Fusarium* in the retained SRA accessions. Only three samples exhibited a dominant taxon other than *Fusarium* at the read level, all of which were bacterial, specifically *Stenotrophomonas* (72%), *Serratia* (77%), and *Achromobacter* (52%). Plant-derived sequences were rare and detected in only five datasets, with reads assigned to *Asarum* in two samples (13.7% and 5.42%) and to *Lasthenia* in three samples (0.77–1.7%).

The most frequently detected secondary genera across the dataset included *Timema* (103 samples), Betacoronavirus (52), *Ilyonectria* (43), *Pseudomonas* (31), *Staphylococcus* (29), *Terrapene* (10), Nectriaceae (7), *Achromobacter* (7), *Ralstonia* (6), and *Homo* (6). These assignments span a wide range of biological origins and are best interpreted in the context of metagenomic classification uncertainty combined with the heterogeneous provenance of publicly available sequencing datasets.

Given this signal of residual contamination, an additional filtering step was applied to minimize potential bias in genome assembly size and GC content estimates. This step involved BLASTN v2.12.0+ alignments of the 336 retained assemblies against a curated reference database comprising representative FOSC genomes with the lowest contamination signals. Median genome assembly size and GC content were then compared per species before and after filtering. GC content was minimally affected, with changes ranging from 0% to 0.12%, indicating that residual non-*Fusarium* sequences had only a marginal influence on species-level median GC estimates. In contrast, genome assembly sizes showed a modest but expected reduction after filtering, with a median decrease of approximately 3%. Reductions ranged from 0.25% in *Fusarium pharetrum* to 10.1% in *Fusarium* sp. FOSC-G4_3 (*Fusarium* “afroindicum”). For *Fusarium fabacearum*, the species with the largest representation in the dataset, the reduction in assembly size was consistent with this overall trend (6.75%).

As an additional measure of genome purity and clonality, all raw reads were mapped back to their respective assemblies, and the number of alternate alleles was quantified. The resulting distributions showed considerable variation: whereas many assemblies contained only a few hundred alternate alleles, consistent with highly clonal genomes, a small subset exceeded 50,000, indicative of substantial intra-sample heterogeneity. Assembly sizes in the FOSC filtered genome assemblies ranged from 45.5 Mb to 57.6 Mb, with a median of 52 Mb, and the GC content showed a median value of 47.5%, with a range between 46.8% and 48.5%. These values are consistent with the expected genome range for FOSC members, indicating that all filtered assemblies fall within expected limits.

Across the 336 filtered genome assemblies, the number of scaffolds ranged from 206 to 5485, with a median of 1486. The assembly contiguity, measured by N50, had a median value of 278,470 bp, with values reaching up to 2.67 Mb. The proportion of ambiguous bases (N ratio) was generally low, with a median of 0.01% and a maximum of 0.36%. Sequencing depth across assemblies had a median of $61.8\times$, ranging from $13.6\times$ to $416\times$. Regarding clonality, the number of alternate alleles per genome varied widely, ranging from 596 to 143,753, with a median of 10,795. Finally, completeness assessed by

BUSCO was consistently high, with values between 96% and 99.8%, and a median of 99.7% (Figure 1).

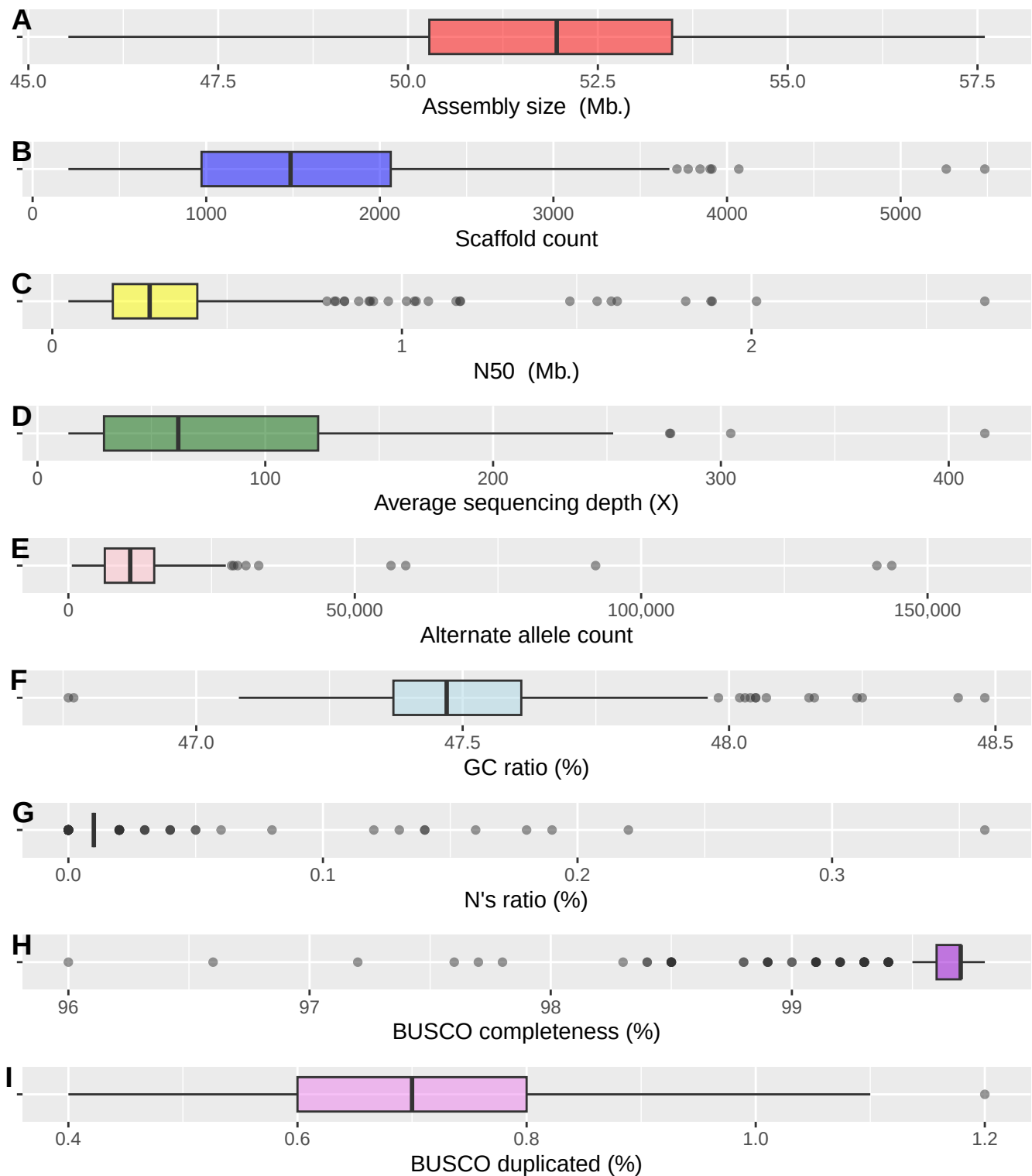


Figure 1. Assembly quality metrics for *Fusarium oxysporum* species complex (FOSC) genomes. Boxplots summarize the distribution of nine genomic features across the quality-filtered FOSC dataset: (A) assembly size (Mb), (B) scaffold count, (C) N50 (Mb), (D) average sequencing depth ($\times$), (E) alternate allele count, (F) GC content (%), (G) proportion of ambiguous bases (N's ratio, %), (H) BUSCO completeness (%), and (I) BUSCO duplication rate (%). Points represent outliers beyond the whiskers of each boxplot.

In addition, seven publicly available reference assemblies of the same complex were downloaded from NCBI, resulting in a final dataset of 343 genomes for phylogenomic analysis.

2.2. Phylogenomic Reconstructions

A phylogenomic analysis was conducted using 343 taxa and 4286 single-copy ortholog CDSs, resulting in an alignment of 6,998,306 nucleotide sites with only 0.0012% missing data. Of these, 94.1% (6,586,950 sites) were constant, while 315,389 positions (4.5%) were parsimony informative, corresponding to 166,867 distinct site patterns. Model testing selected an Invar+FreeRate model with six categories, estimating that 62.3% of sites were invariant and that the remaining variable sites were distributed across a small proportion of rapidly evolving positions. The inferred maximum-likelihood tree had a log-likelihood of $-16,094,992$, with a total branch length of 0.1334, of which 71.8% corresponded to internal branches. Although overall branch support was high, IQ-TREE flagged 43 near-zero internal branches (<0.0001), suggesting that certain relationships remain poorly resolved due to the high sequence similarity among genomes at terminal nodes, all of which occurred within species-level clades. The outgroup *Fusarium foetens* was included for rooting purposes.

The topology of the phylogenomic tree clearly resolved, with high confidence, the relationships among the basal nodes relative to the currently accepted *Fusarium* species. Within the FOOSC, three main clades were recovered. Clade 1 (*F. odoratissimum*) represents the most basal lineage of the complex, followed by two large and diverse clades designated as Clades 2 and 3, following the proposal initially described by van Westerhoven et al. [56].

Clade 2 encompasses *F. libertatis*, *F. hoodiae*, *F. cugenangense*, *F. elaeidis*, *F. duoseptatum*, *F. callistephi*, and *F. fabacearum*. Notably, two well-supported clades, not previously assigned names, were recovered as sisters to the lineage comprising *F. cugenangense* + *F. elaeidis* + *F. duoseptatum* + *F. callistephi* + *F. fabacearum*. These two lineages, provisionally designated as the *Fusarium* “afroindicum” and *Fusarium* “europaeum” clades, show consistent placement and strong phylogenomic support, justifying their recognition as distinct lineages within Clade 2 of the FOOSC. Figure 2A shows the maximum-likelihood phylogenomic tree of the *Fusarium oxysporum* species complex, with *F. fabacearum* being the most represented species in Clade 2. Figure 2B depicts the rest of the species within Clade 2, including *F. callistephi*, *F. duoseptatum*, *F. elaeidis*, *F. cugenangense*, the “europaeum” and “afroindicum” clades, *F. hoodiae*, and *F. libertatis*. Figure 2C shows Clade 1 (*F. odoratissimum*) and Clade 3, which contains other species such as *F. oxysporum* s. str., *F. triseptatum*, *F. pharetrum*, *F. veterinarianum*, *F. languescens*, *F. curvatum*, and *F. nirenbergiae*. Clade 2 contained most of the analyzed genomes (243 assemblies), followed by Clade 3 (78 assemblies) and Clade 1 (21 assemblies). The complete list of species assignments obtained with our phylogenomic approach, compared with the original species assignments reported in the NCBI SRA, is provided in Supplementary Table S2.

Having resolved with confidence the evolutionary history of the FOOSC, we next examined the evolution of genome size and GC content. Median genome size varied across species. The most significant number of genomes corresponded to *F. fabacearum* (136 assemblies; median 52.2 Mb) and *F. cugenangense* (83 assemblies; median 52.4 Mb). Other well-represented lineages included *F. nirenbergiae* (40 assemblies; median 50 Mb) and *F. languescens* (15 assemblies; median 51.2 Mb). Additional representatives were *F. odoratissimum* (21 assemblies; median 46.7 Mb), *F. curvatum* (9 assemblies; median 54 Mb), *F. callistephi* (7 assemblies; median 53.5 Mb), *F. oxysporum* s. str. (8 assemblies; median 52.8 Mb), and *F. duoseptatum* (4 assemblies; median 48.2 Mb). The two novel, provisionally designated lineages within Clade 2—*Fusarium* “afroindicum” and *Fusarium* “europaeum”—each comprised four assemblies, with median assembly sizes of 50.5 Mb and 56.5 Mb, respectively. Genome GC content showed subtle but consistent variation across groups,

with medians ranging from ~47.2% to ~47.9%. Core genome duplication events, measured by the BUSCO duplicated genes index, displayed a median value of 0.7% across all FO SC genomes analyzed (Figure 3).

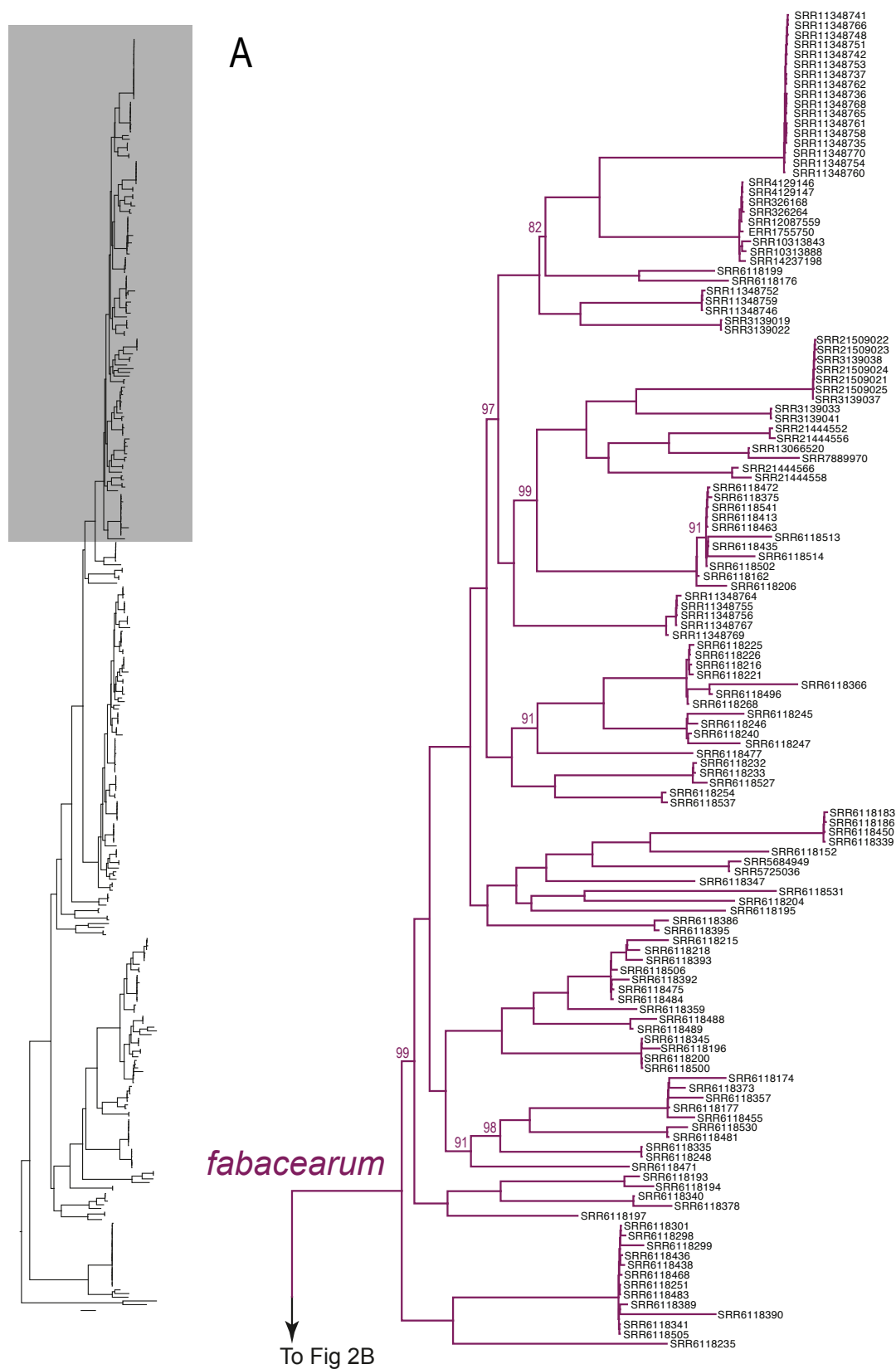


Figure 2. Cont.

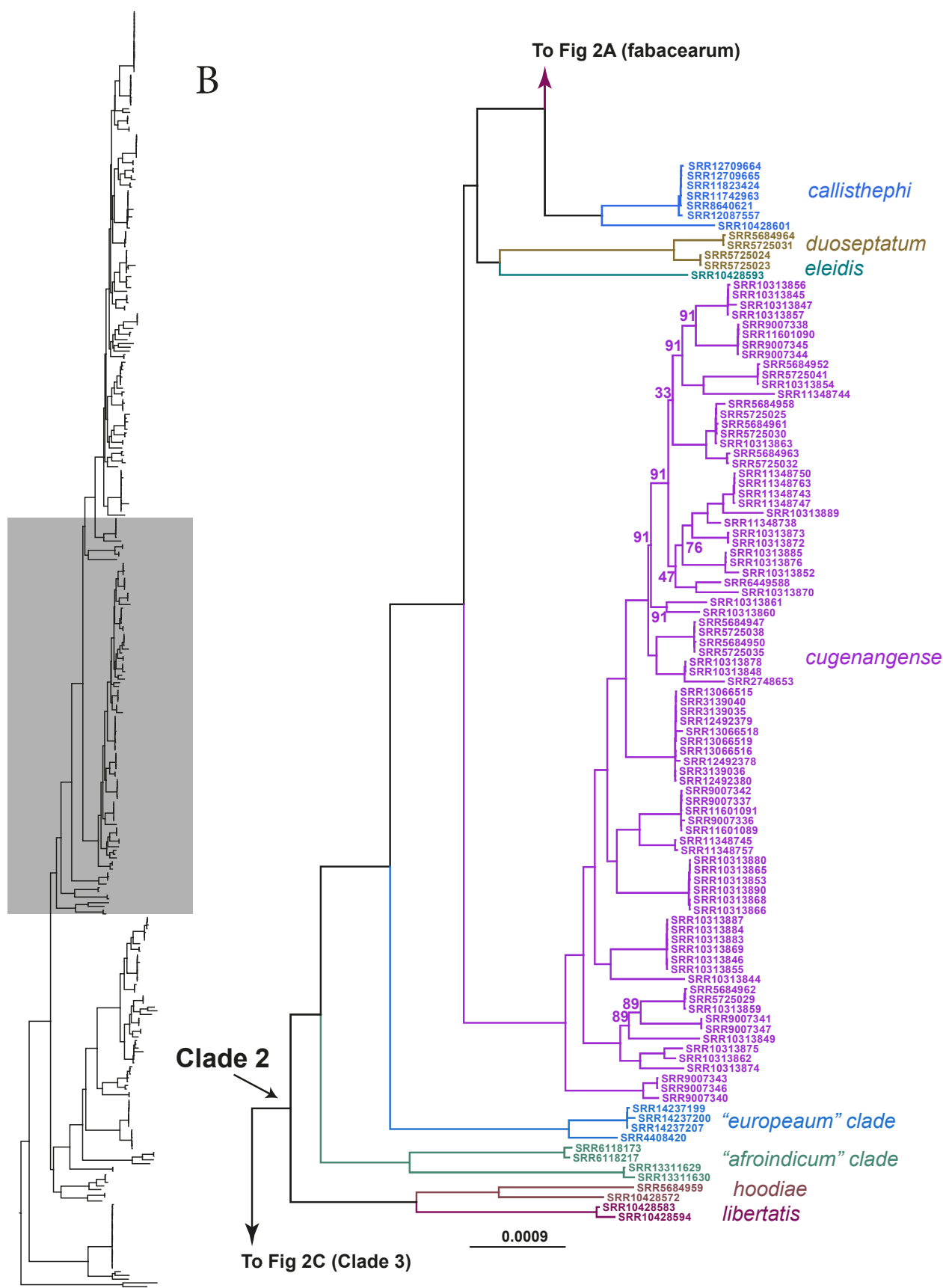


Figure 2. Cont.

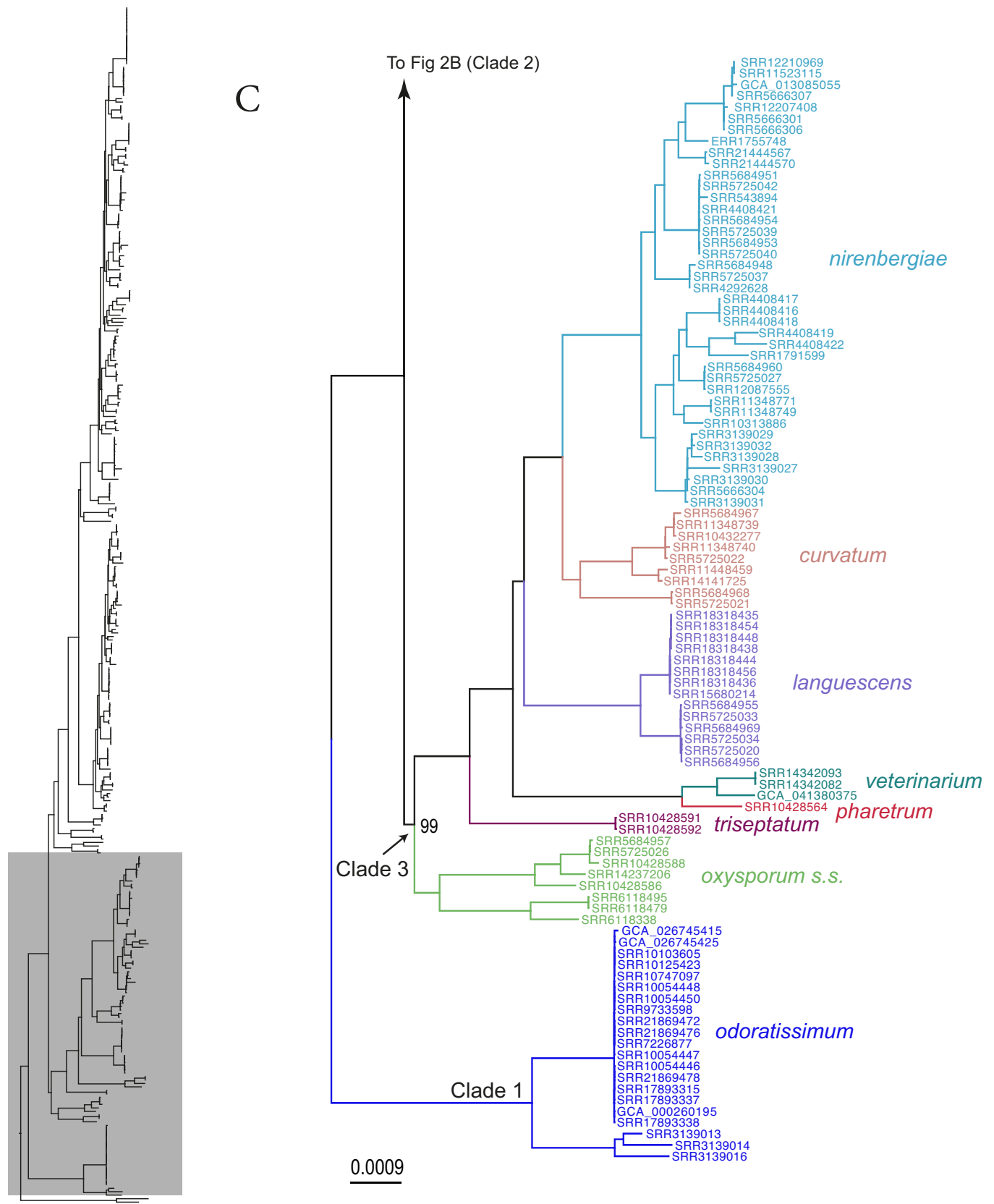


Figure 2. (A). Maximum-likelihood phylogenomic tree of the *Fusarium oxysporum* species complex. Global phylogeny inferred with IQ-TREE from 4286 concatenated single-copy orthologous coding sequences (CDSs). Ultrafast bootstrap (UFB) support values are displayed at nodes (only values <100 are indicated). Terminal branches were collapsed where multiple genomes of the same monophyletic

lineage were present. The three major FOSC clades are highlighted in distinct colors, with alternative lineage nomenclatures provided to facilitate systematic cross-comparison. **(B)**. Detailed phylogenomic relationships within FOSC Clade 2. Continuation of the maximum-likelihood tree framework from **(A)**, highlighting high-resolution evolutionary relationships within Clade 2. Ultrafast bootstrap (UFB) support values <100 are shown at nodes, and distinct species and provisional lineages are distinguished by specific colors. **(C)**. Detailed phylogenomic relationships within FOSC Clade 3. Continuation of the maximum-likelihood tree framework from **(A,B)**, showing high-resolution evolutionary relationships within Clade 3. Ultrafast bootstrap (UFB) support values <100 are shown at nodes, and discrete lineages are color-coded.

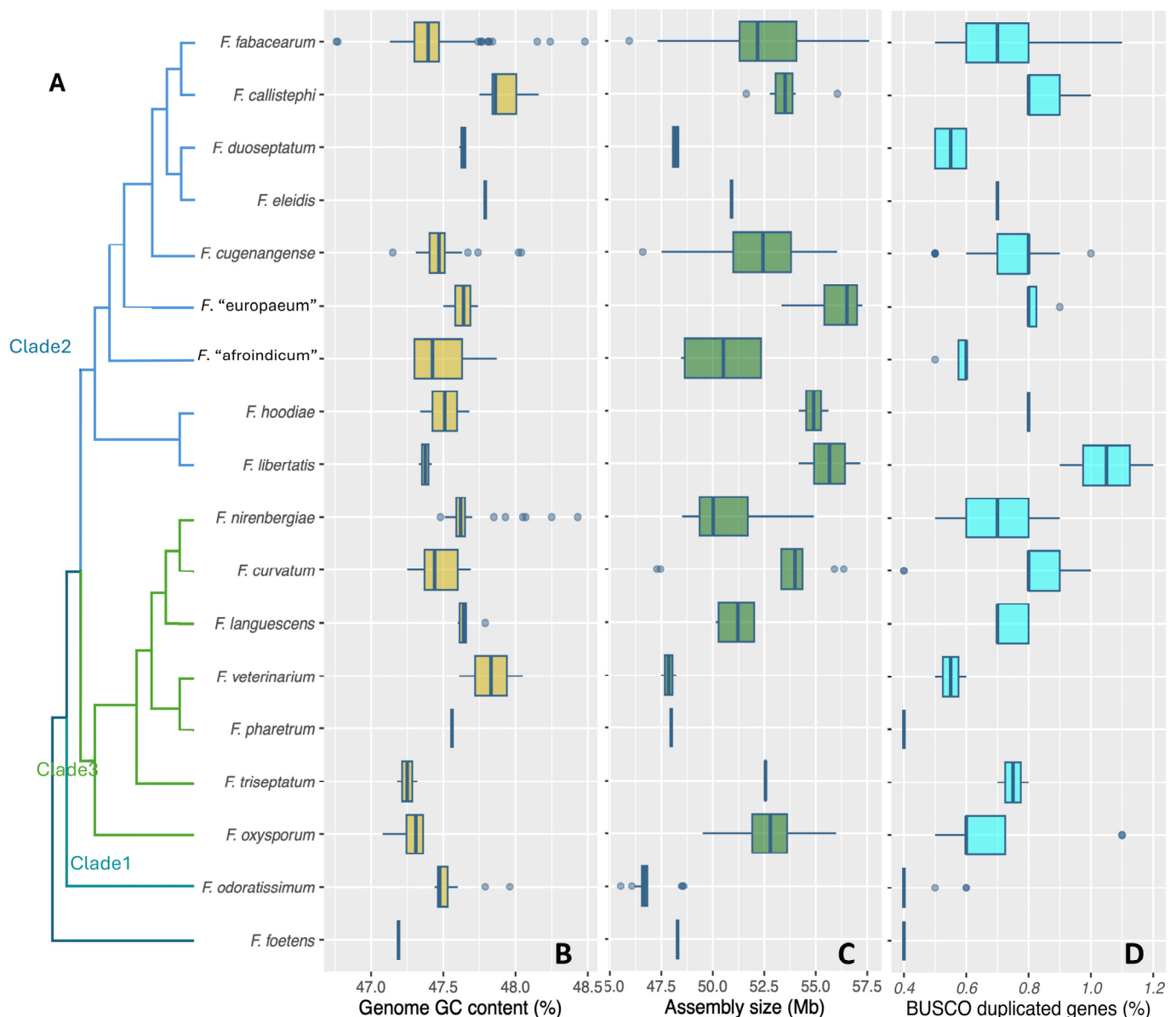


Figure 3. Comparative global genome properties across *Fusarium oxysporum* species complex (FOSC) clades. **(A)** Reference cladogram illustrating the evolutionary architecture of the primary FOSC clades and provisional lineages. **(B)** Distribution of genome GC content (%) across individual lineages. **(C)** Distribution of genome assembly sizes (Mb) across species. **(D)** Relative percentages of conserved, single-copy, and duplicated BUSCO genes across major clades.

Variation in genome size and GC content occurred largely independently across species, reinforcing the concept of high genomic plasticity of these fungi. However, a clear tendency for FOSC clade 2 to harbor larger genomes can be observed. To confirm

this, we grouped all assemblies according to their FOSC clade and compared median genome sizes and GC content. The results supported the visual observation: Clade 2 displayed the largest genomes, with a median size of 52.3 Mb, while Clade 3 genomes were smaller (median 50.6 Mb), a difference of 1.7 Mb. Clade 1 contained the most compact genomes, with a median around 46.7 Mb. Statistical tests confirmed these differences, with p -values < 0.0001 . A similar comparison was performed for GC content. Here, differences were more subtle but still significant, with Clade 3 exhibiting slightly higher GC levels than Clade 2, corresponding to a median difference of $\sim 0.2\%$ (Figure 4).

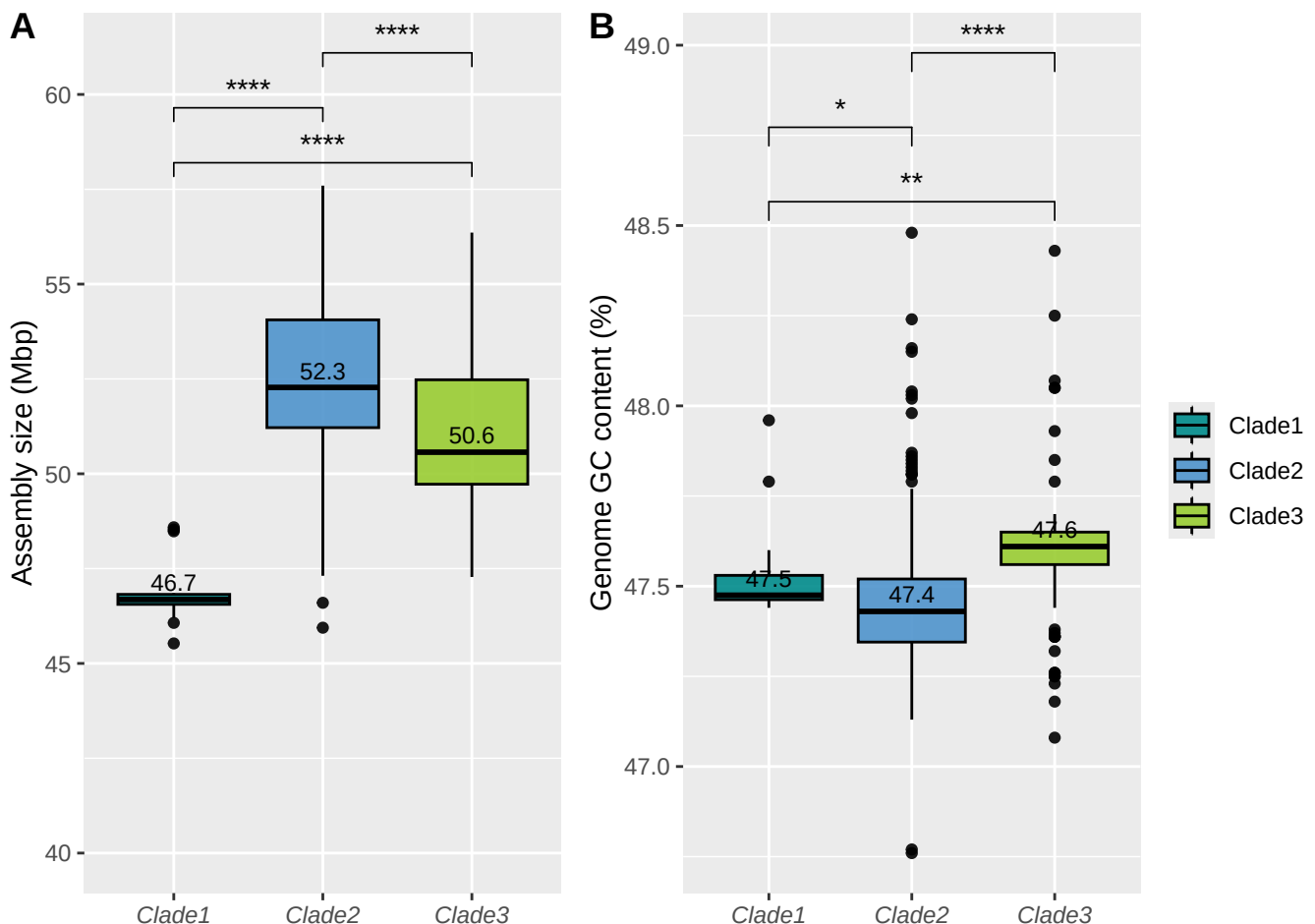


Figure 4. Comparative distribution of global genome properties among primary FOSC clades. (A) Genome assembly size (Mbp) variations across the three main clades. (B) Whole-genome GC content (%) distributions. Median values are displayed numerically above each boxplot. Significant differences between clades were calculated using pairwise Wilcoxon rank-sum tests with Benjamini–Hochberg multiple-testing corrections (ns = not significant; * = $p \leq 0.05$; ** = $p \leq 0.01$; **** = $p \leq 0.0001$).

To assess genome-wide conservation among FOSC, we performed pairwise genome alignments. We applied this approach by segregating the FOSC into the three main clades and comparing genomes both within clade (intra-clade) and between clades (inter-clade). Overall, most genomes exhibited at least 70% alignment coverage, with nucleotide identity of aligned blocks ranging from over 96% to nearly 100%. As expected, the median nucleotide identity between FOSC clades (inter-clade) was significantly lower (97.53%, $p < 0.0001$) compared to within-clade values, which reached 99.9% for clade 1 (*F. odoratissimum*), 98.43% for clade 3, and 98.48% for clade 2 (Figure 5).

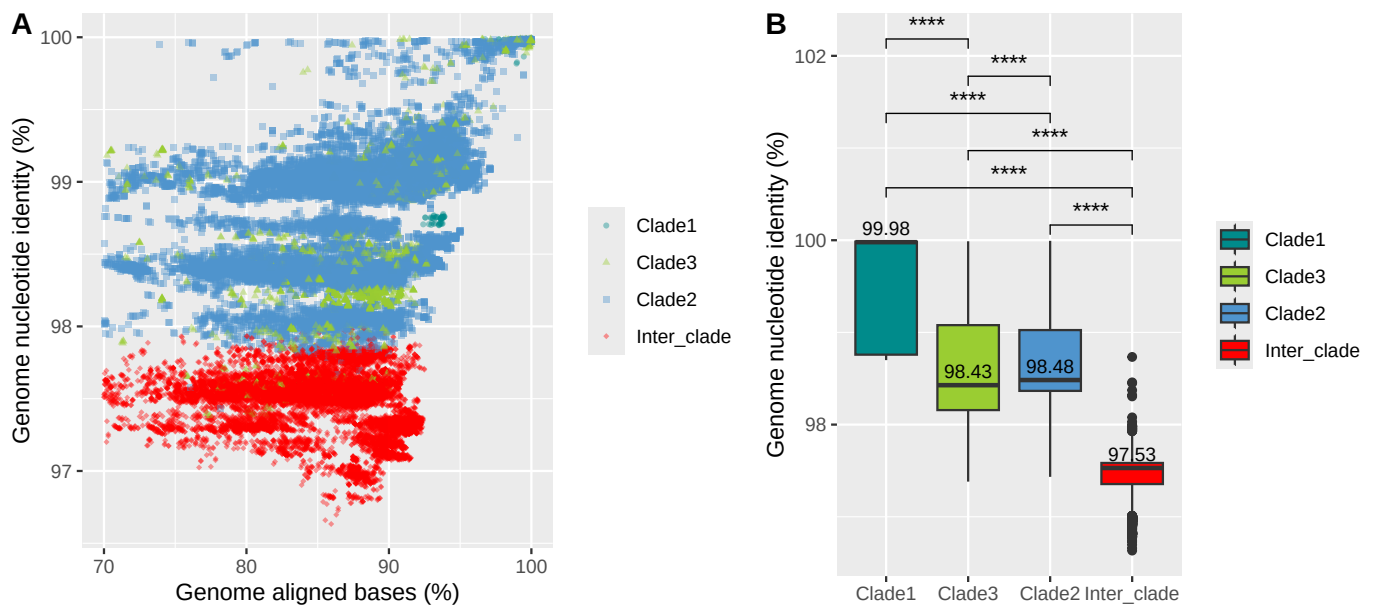


Figure 5. Whole-genome sequence conservation and core identity boundaries among main FOSC clades. (A) Pairwise whole-genome comparisons illustrating the relationship between the proportion of aligned bases (%), x -axis) and average nucleotide identity (ANI, %, y -axis) across all analyzed FOSC genomes. Data points are partitioned into intra-clade and inter-clade (Inter_clade) comparisons, with coordinates colored by lineage. (B) Boxplots displaying the distribution of genome average nucleotide identity values (%) within individual clades versus between-clade pairings (INTER_clade). Median values are indicated above each box. Pairwise statistical differences were assessed using the Wilcoxon rank-sum test (**** = $p \leq 0.0001$).

The FOSC Clade 2 contains most of the genomes classified within the complex and appears to be the most diverse, comprising seven accepted species and two additional, well-supported and novel genomic lineages. We therefore analyzed genome conservation and nucleotide identity within this clade in greater detail. Within each species, as well as within the provisional *Fusarium* “afroindicum” and *Fusarium* “europaeum” clades, the median nucleotide identity was above 99%, except for the “afroindicum” lineage, which showed a slightly lower value of 98.61%. By contrast, inter-species comparisons within clade 2 yielded lower identities, with a median of 98.37% (Figure 6).

Between the sister species *F. fabacearum* and *F. callistephi*, the median nucleotide identity was 98.7%, increasing to 99.01% within (intra-species) *F. fabacearum* and 99.97% within *F. callistephi*. A comparable trend was observed when *F. fabacearum* was compared with *F. cugenangense*, the species with the largest genomic representation: the inter-species median nucleotide identity was 98.37%, while within *F. cugenangense* it reached 99.14%. The lineages provisionally designated as *Fusarium* “afroindicum” and *Fusarium* “europaeum” exhibited inter-species nucleotide identities of 98.02% and 98.05%, respectively, when compared with *F. fabacearum*, and 98.07% and 98.03% against *F. cugenangense*, respectively (Figure 7A–F). Consistently, phylogenetic distances estimated supported these relationships, showing a distance of 0.001 between *F. fabacearum* and *F. callistephi*, 0.0014 between *Fusarium* “afroindicum” and the common ancestor of *Fusarium* “europaeum” + *F. fabacearum*, and 0.0024 between *Fusarium* “europaeum” and the ancestor of *F. fabacearum* + *F. cugenangense* (Supplementary Figure S1).

High-resolution phylogenomics revealed substantial heterogeneity in the geographic range of the major clades within the FOSC. Clade 2, the largest and most taxonomically diverse clade, exhibited a clearly cosmopolitan distribution. These counts reflect only genomes with available metadata on geographic origin; additional genomes lacking isola-

tion data (NA) are listed separately in the Supplementary Material. The most abundant species were *F. fabacearum* ($n = 136$), predominantly from Africa but also represented in Asia, Europe, and North America, and *F. cugenangense* ($n = 82$), distributed across Asia, Oceania, North America, and Europe. *F. callistephi* ($n = 6$), distributed across Asia, Oceania, and Europe, was intercontinental but less represented, while *F. duoseptatum* ($n = 4$; Asia) and *F. elaeidis* ($n = 1$; Europe) were restricted to single regions. In contrast, *F. libertatis* (Oceania), *F. hoodiae* (Europe, Oceania), and *Fusarium* “afroindicum” ($n = 4$; Africa and Asia) and *Fusarium* “europaeum” ($n = 4$; Europe), not linked to any verified type strain genome in our dataset, showed geographically narrow distributions. Clade 3 displayed a similarly complex structure: *F. nirenbergiae* ($n = 36$) was abundant across Europe, Asia, and North America; *F. curvatum* ($n = 9$) and *F. languescens* ($n = 10$) were intercontinental, spanning both Old and New Worlds; while *F. oxysporum* s. str. ($n = 8$; Europe and Africa) and *F. triseptatum* ($n = 2$; Europe) were confined to the Old World. Other clade 3 members were highly localized, including *F. pharetrum* ($n = 1$; Oceania) and *F. veterinarianium* ($n = 3$; North America). Clade 1 consisted solely of *F. odoratissimum* ($n = 18$), which was the most broadly distributed lineage, detected in Asia, Africa, Europe, and South America. Finally, there was *F. foetens* ($n = 2$), detected only in Europe. For several of these species with very low representation (e.g., *F. elaeidis*, *F. duoseptatum*, *F. triseptatum*, *F. pharetrum*, *F. veterinarianium*, and *F. foetens*), the apparent geographic and host restrictions of these taxa should be interpreted as preliminary observations conditioned by the limited number of available genomes, rather than as definitive biogeographic or ecological patterns. These contrasting observations of cosmopolitan versus narrowly distributed taxa highlight the uneven evolutionary trajectories within the FOSC, where some lineages have diversified and dispersed globally, while others seem to remain restricted to specific regions or cropping systems (Figure 8).

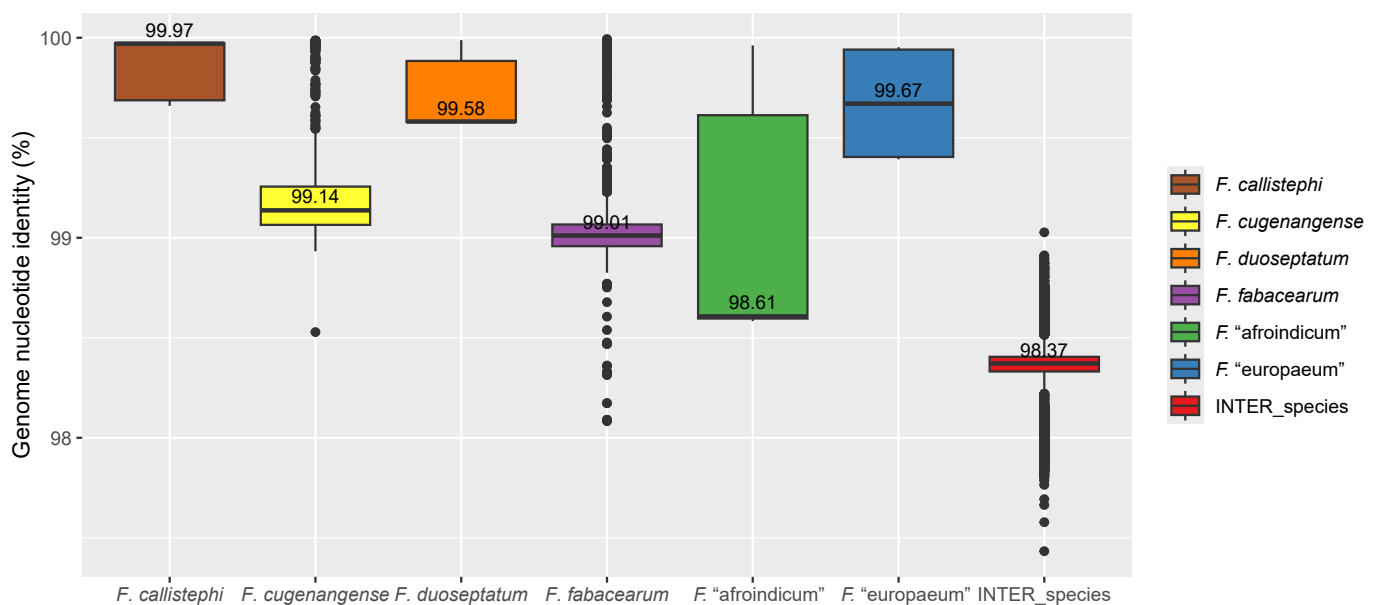


Figure 6. Interspecific and intraspecific genome nucleotide identity within FOSC Clade 2. Boxplots illustrate whole-genome average nucleotide identity (ANI) values (%), y -axis) calculated within individual species/lineages and across cross-species comparisons (INTER_species) within Clade 2. Horizontal bars and numerical annotations represent median values. Statistical outliers are indicated by black points.

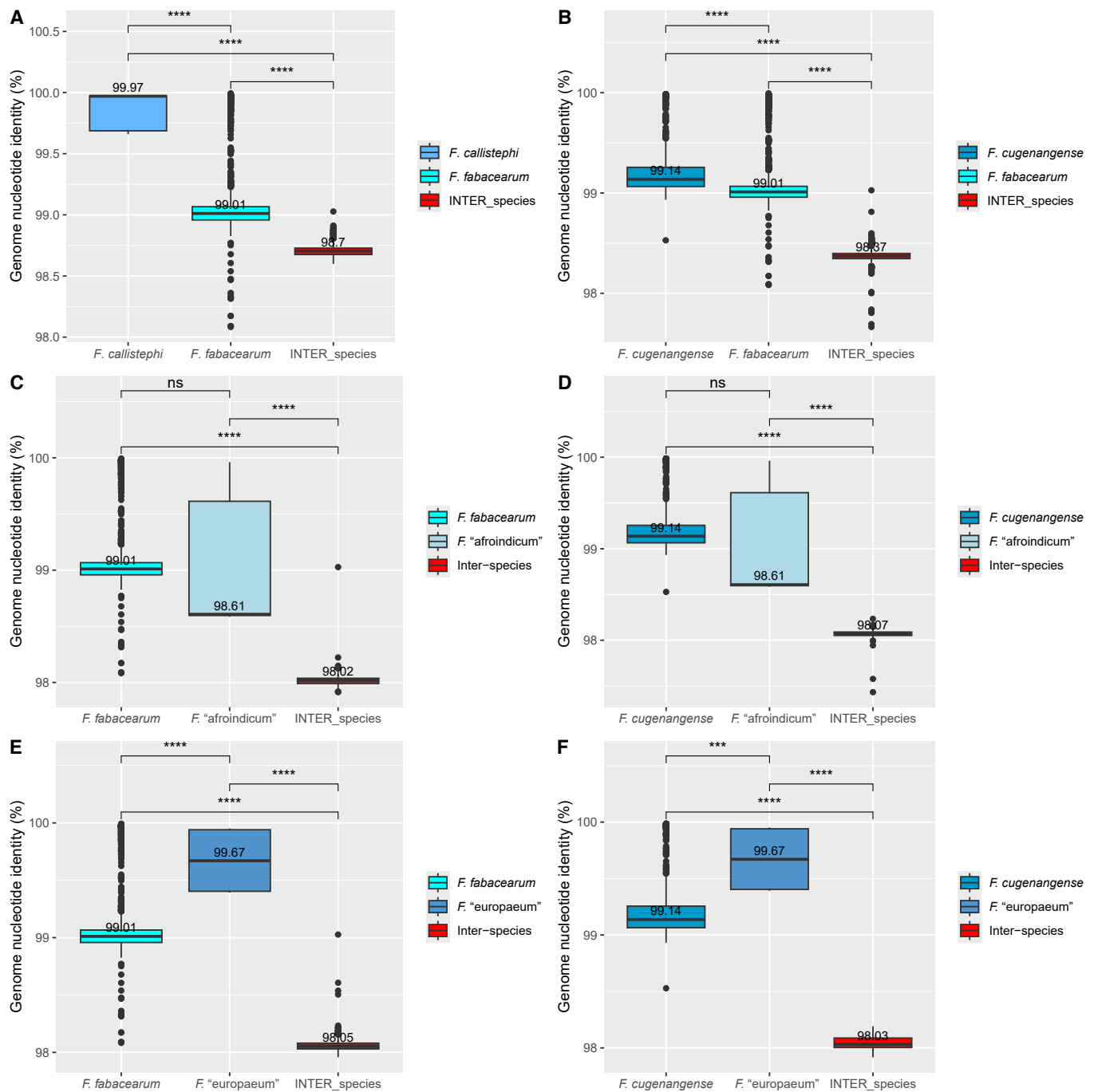


Figure 7. Pairwise genomic identity boundaries among distinct FOSC Clade 2 lineages. Boxplots display genome-wide nucleotide identity ranges (%) computed from pairwise comparisons between selected Clade 2 reference taxa and provisional lineages: (A) *F. fabacearum* vs. *F. callistephi*, (B) *F. fabacearum* vs. *F. cugenangense*, (C) *F. fabacearum* vs. provisional lineage *Fusarium* "afroindicum", (D) *F. cugenangense* vs. provisional lineage *Fusarium* "afroindicum", (E) *F. fabacearum* vs. provisional lineage *Fusarium* "europaeum", and (F) *F. cugenangense* vs. provisional lineage *Fusarium* "europaeum". Medians are displayed directly above each boxplot. Statistical differences were evaluated using pairwise Wilcoxon rank-sum tests (ns = not significant; *** = $p \leq 0.001$; **** = $p \leq 0.0001$).

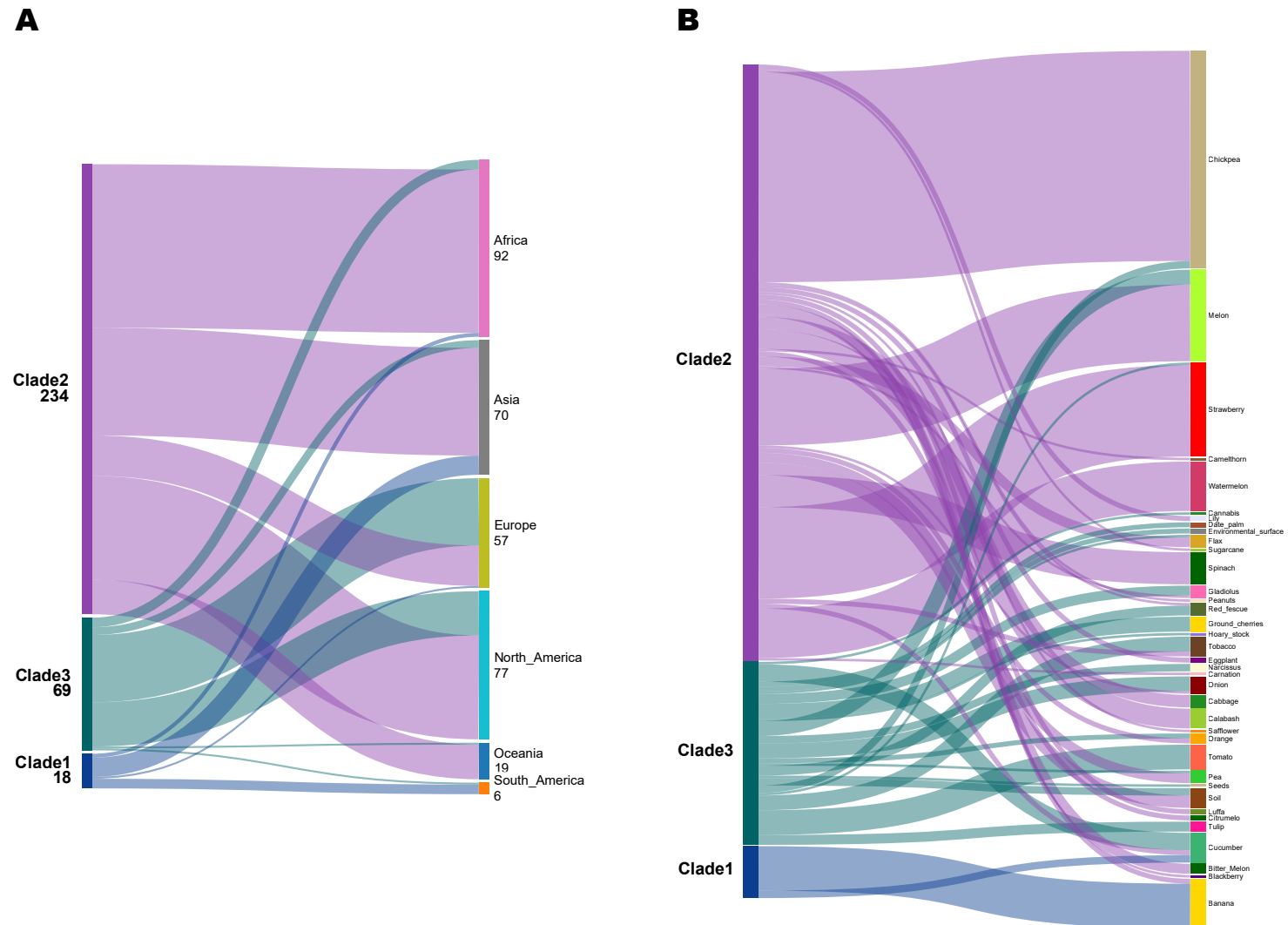
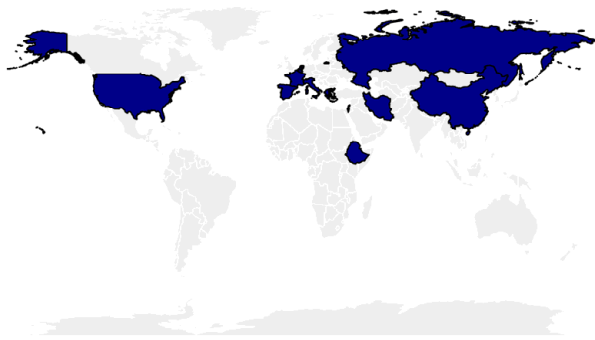
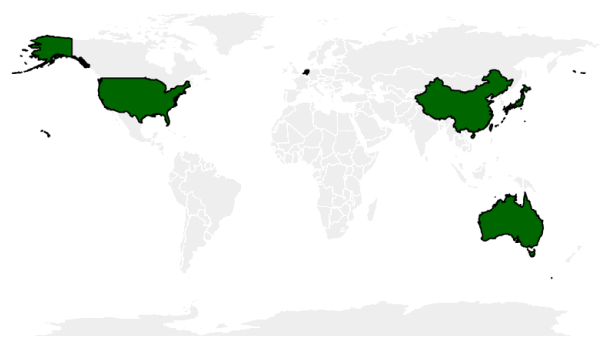
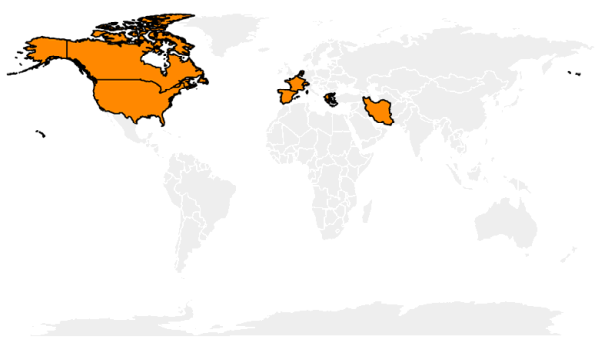
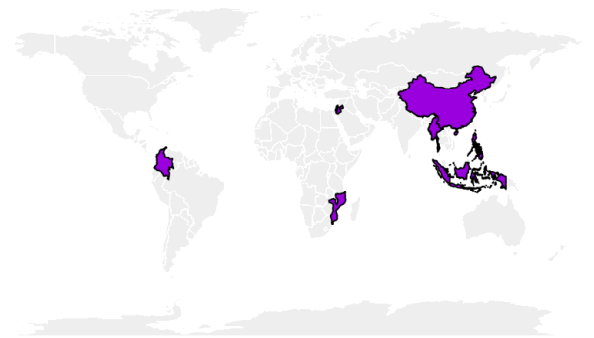
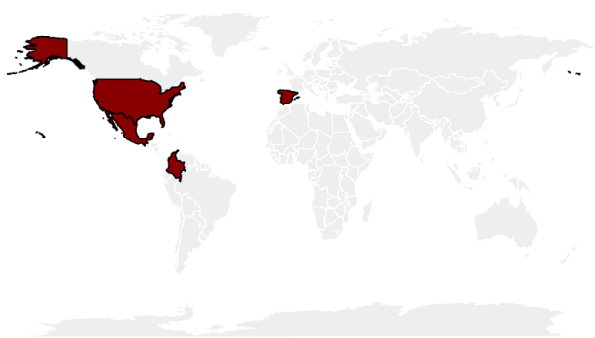
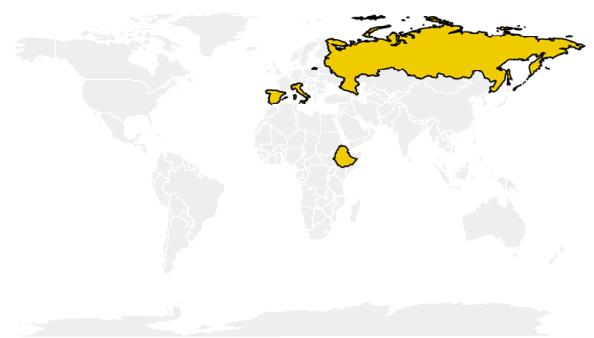


Figure 8. Biogeographic and host distribution tracking of FOSC genomes. Sankey flow diagram mapping the hierarchical epidemiological metadata of the curated dataset. The left panel shows the core-phylogenomic clade assignments defined in this study. Stream (A) projects these clades against global geographic distribution across continents (left panel), while stream (B) links them to their documented plant hosts or environmental sources (right panel). Genomes with missing, incomplete, or ambiguous metadata records were programmatically excluded to optimize plot interpretability.

Differences in host associations were observed both between clades and among the species within them. Within clade 2, the widest spectrum was found in *F. fabacearum*, which encompassed 14 crops spanning at least eight botanical families (Apiaceae, Rosaceae, Asteraceae, Cucurbitaceae, Fabaceae, Rutaceae, Linaceae, Solanaceae). *F. cugenangense* also displayed a broad range, affecting 10 crops such as strawberry, blackberry, melon, watermelon, bitter melon, calabash, lily, spinach, sugarcane, and tobacco. By contrast, other clade 2 members showed narrower associations: *F. callistephi* was restricted to cabbage and pea; *F. duoseptatum* to calabash and luffa; *Fusarium* “europaeum” to flax and onion; and *Fusarium* “afroindicum” to banana and chickpea. *F. elaeidis* and *F. hoodiae* were reported from red fescue and gladiolus, respectively, but due to the very low representation of this clade, no conclusions can be drawn.

In clade 3, the most notable case was *F. nirenbergiae*, associated with 10 crops including strawberry, melon, pea, tomato, sour orange, cucumber, narcissus, onion, gladiolus, and tobacco, spanning seven botanical families. Other clade 3 species exhibited more restricted host profiles: *F. oxysporum* s. str. with chickpea, flax, gladiolus, and red fescue; *F. languescens* with tomato, ground cherries, and melon; and *F. curvatum* with hoary stock, melon, date palm, and tulip. *F. triseptatum* was recorded from red fescue, and *F. veterinarianum* from cannabis and environmental surfaces. In clade 1, *F. odoratissimum* was represented by several genomes, the vast majority isolated from banana (Musaceae) and only a few from cucumber (Cucurbitaceae). *Fusarium foetens* is represented only by two genomes, one isolated from begonia (Begoniaceae) and the other from Monterey pine (Pinaceae), indicating a minimal but heterogeneous host record. Restricted host ranges in lineages with very low genomic representation (e.g., *F. elaeidis*, $n = 1$; *F. duoseptatum*, $n = 4$; *Fusarium* “afroindicum”, $n = 4$; *Fusarium* “europaeum”, $n = 4$) should be considered preliminary and contingent upon future sampling efforts. As genomic resources for the FOSC expand, the geographic and host-range characterization of these underrepresented species will likely be refined (Figures 8 and 9).

Lastly, we performed genome comparisons at the chromosome level using the reference strain *Fusarium* FOSC Fo47, which is assembled to the chromosome scale, against genomes with high sequencing depth ($\geq 100\times$). Using a threshold of $\geq 70\%$ coverage to define chromosome presence, we found that the 11 chromosomes previously described as core were consistently present across strains/species, except for four genomes that showed $<70\%$ coverage: two for chromosome NC_072850.1, one for NC_072851.1, and one for NC_072849.1. Nucleotide identity analysis revealed generally high conservation across chromosomes among strains; however, chromosomes NC_072850.1 and NC_072851.1 exhibited the greatest divergence relative to the reference. Notably, NC_072851.1, previously described as specific to Fo47, confirmed this observation. As an alternative approach, we also compared genome coverage of each chromosome using a heatmap-based strategy (Figure 10), this time applying a more relaxed threshold of $\geq 50\%$. This analysis revealed slight variations in chromosome coverage across most strains, with some showing $<70\%$ coverage for some chromosomes, confirming the previous results. Additionally, this complementary analysis confirms that one strain of *F. cugenangense* had $<50\%$ coverage of chromosome NC_072850.1, while the others previously mentioned had coverage between 50–70% for chromosomes NC_072849.1 and NC_072851.1. Interestingly, chromosome NC_072846.1, considered specific to Fo47, appeared to be at least partially present in some FOSC strains when applying the relaxed 50% threshold.

A *Fusarium fabacearum***B** *Fusarium cugenangense***C** *Fusarium nirenbergiae***D** *Fusarium odoratissimum***E** *Fusarium languescens***F** *Fusarium oxysporum*

Countries with reported genomes of FOSC most frequent species

Figure 9. Geographic occurrence patterns of heavily sequenced species within the FOSC. Country-level maps (A–F) show the documented global distribution of verified genome assemblies for (A) *F. fabacearum*, (B) *F. cugenangense*, (C) *F. nirenbergiae*, (D) *F. odoratissimum*, (E) *F. languescens*, and (F) *F. oxysporum* sensu stricto. Shaded regions identify countries with available genome records in public databases, while gray zones indicate an absence of published data.

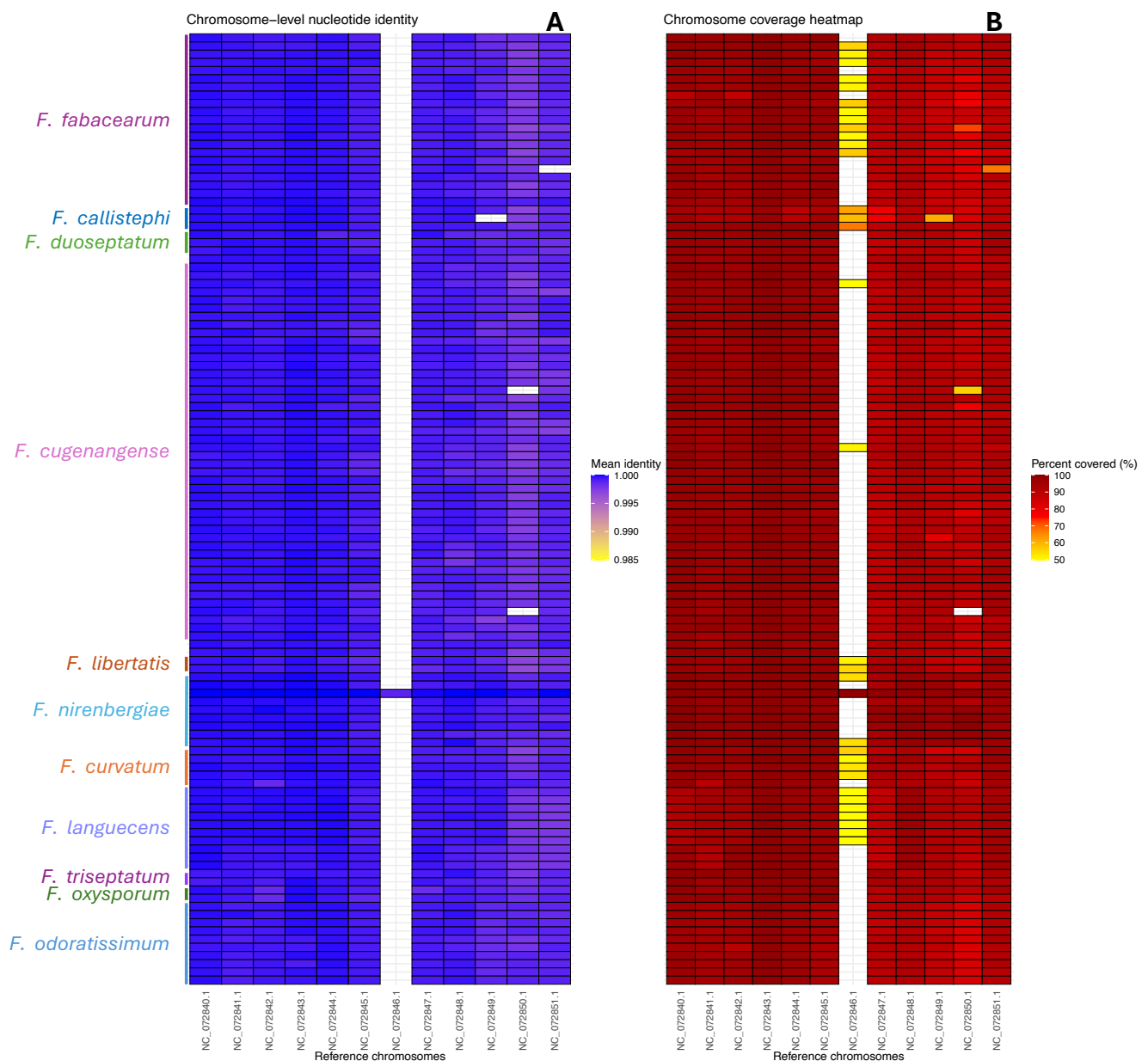


Figure 10. Macro-structural chromosome conservation profiles against reference strain Fo47. Comparative chromosome alignment matrices of FOSC genomes mapped against the structural architecture of *Fusarium oxysporum* reference strain Fo47 (GCF_013085055). (A) Heatmap depicting mean nucleotide identity values computed across individual core and accessory reference chromosomes (x-axis) against query genomes (y-axis). (B) Heatmap showing total sequence coverage percentages (%) for the matching genomes, restricted to high-confidence assemblies with $\geq 100\times$ sequencing depth profiles. Color gradients scale with conservation values; a coverage threshold of $\geq 50\%$ indicates chromosome presence.

3. Discussion

The rapid expansion of global genomic repositories offers an unprecedented opportunity to resolve the complex evolutionary histories of high-impact fungal groups like the *Fusarium oxysporum* species complex (FOSC). However, effectively exploiting these public assets requires navigating substantial heterogeneity in sequencing depth, assembly quality, and strain provenance. Rather than accepting public data at face value, this study underscores the critical necessity of rigorous data curation in fungal macro-evolutionary studies. By implementing a strict, multi-layered quality control framework that excluded

over half of the initially available datasets, we established a high-confidence genomic foundation of 336 assemblies. While the broader public database still exhibits baseline challenges—such as assemblies below the ideal 100× coverage threshold or potential indicators of non-clonal sequences—our stringent filtering for N50, BUSCO completeness, and taxonomic purity ensures that the phylogenomic relationships delineated here are highly robust. Moving forward, it is imperative that the mycological community adopts standardized, high-coverage sequencing protocols and transparent metadata reporting to continuously refine the systematic framework established in this work.

Among the recognized species complexes of the genus *Fusarium*, the *F. oxysporum* species complex (FOSC) is consistently highlighted as one of the most relevant groups in agriculture and plant pathology [59], together with the *F. fujikuroi* species complex (FFSC), the *F. graminearum* species complex (FGSC), and the *F. incarnatum–equiseti* species complex (FIESC) [43]. However, in its current concept, *F. oxysporum* is a species complex consisting of numerous cryptic species, and the identification and naming of these cryptic taxa are complicated by multiple subspecific classification systems [42]. Moreover, the FOSC encompasses relevant pathogens such as *F. odoratissimum*, *F. oxysporum* sensu stricto, and *F. veterinarianum*, whose taxonomic resolution may be limited when relying on conventional mycological approaches. Resolving its evolutionary history and establishing a robust phylogenomic framework to address taxonomic inconsistencies is therefore a long-standing necessity. Such progress will help foster a more reliable understanding of the significance of this complex for agriculture, as well as human and animal health. According to NCBI species records from sequenced isolates within the FOSC, misidentifications are very common, and older literature may therefore provide misleading information on species assignments, including their geographical and host associations. This issue reflects the fact that most *Fusarium* isolates cannot be reliably identified to the species level using only morphological traits, as highlighted by multilocus and molecular studies that have revealed extensive cryptic diversity within the FOSC [59].

The definition of *Fusarium oxysporum* remains broad and ambiguous in classical mycology laboratories [42], as it largely relies on morphological traits and subspecific schemes such as *formae speciales*, races, and vegetative compatibility groups (VCGs). While these systems were historically useful for grouping isolates by host range and virulence spectra, they do not reflect evolutionary relationships [6,12,44]. More than one hundred *formae speciales* have been described [58,59], yet phylogenetic analyses consistently demonstrate their polyphyletic nature. For instance, f. sp. *lycopersici* and f. sp. *radicis-lycopersici* often cluster with unrelated or non-pathogenic strains [26,60–62], and similar independent origins have been reported for lineages infecting cucurbits, banana, cotton, celery, and lettuce [63–67]. These inconsistencies underscore the limitations of host-based classifications in capturing evolutionary history [68] and highlight their practical implications for pathogen identification, resistance breeding, and disease management. In this context, Fulton et al. (2021) [69] showed that multilocus sequence analyses were insufficient to resolve the races of *F. oxysporum* f. sp. *niveum* from watermelon, further emphasizing the need for genome-scale approaches that provide more robust comparative frameworks and higher-resolution insights into the genetic diversity of the FOSC.

Our phylogenomic approach, based on 4286 single-copy CDSs, provided high resolution and allowed us to resolve the relationships among currently described species confidently. While consistent with previous studies, our work offers a robust framework by excluding contaminated or low-quality genomes and by incorporating a larger dataset of isolates. This strengthens the current classification of the FOSC and supports the evolutionary relationships among its recognized species.

Resolving the evolutionary history of the FOSC also provides valuable insights into its genome evolution. Core features such as genome size, GC content, and gene duplications highlight the highly dynamic nature of these genomes, in agreement with previous reports across the genus [55]. Within the FOSC, distinct differences become apparent: basal clades such as *F. foetens* and *F. odoratissimum* (clade 1) tend to maintain smaller genomes, whereas clades 3 and, most notably, clade 2 exhibit significant genome expansions. The very large genome sizes observed in clade 2 may reflect the accumulation of accessory chromosomes, transposable elements, or horizontal gene acquisitions, which are often linked to host adaptation and pathogenic specialization [70,71]. These patterns suggest that genome expansion within the FOSC is not random but likely linked to the diversification of pathogenic traits and ecological niches. Notably, clade 2 is also the most diverse clade, with the highest number of isolates recorded worldwide across many countries and a wide range of hosts.

Regarding genome conservation, the high nucleotide identity values (above 96%) confirm that the FOSC is, as often portrayed, a compact group of species. Nonetheless, whole-genome analyses provide the necessary resolution to discriminate among its clades and species. Inter-clade and inter-species nucleotide identity decreased by approximately 1%, reaching around 97.5% and 98.4%, respectively, while within-species comparisons in the clade 2 usually showed median identities above 99%. In contrast, the conservation of aligned genome blocks demonstrated a more variable profile, with alignment coverage ranging from about 70% to nearly 100%. This metric should be interpreted with caution, however, as some genomes displayed unusually small assembly sizes that could be related to insufficient sequencing depth. Unlike prokaryotic species delimitation, where a 95% ANI threshold is widely accepted in the literature, no equivalent consensus exists for fungal species complexes. In this study, species boundaries were defined by the convergence of phylogenomic topology and a consistent pattern of nucleotide identity decay across taxonomic levels.

Chromosome-level comparisons further reinforced the distinction between conserved and variable genomic regions within the FOSC. As expected, the 11 chromosomes described as core were largely retained across strains, supporting their essential role in maintaining the genetic integrity of the complex. The few exceptions, where coverage of certain chromosomes fell below the 70% threshold, likely reflect lineage-specific partial chromosomal loss. Importantly, the observation that chromosomes NC_072850.1 and NC_072851.1 displayed the greatest divergence relative to the Fo47 reference highlights potential hotspots of genomic differentiation within the FOSC. The fact that NC_072846.1, previously described as specific to Fo47, was indeed highly divergent and partially absent in most FOSC genomes, further underscores its strain-specific character and possible role in Fo47's unique ecological adaptations. The relaxed 50% coverage threshold corroborated these findings and additionally suggested that this chromosome, NC_072846.1, may still persist in partial form within other FOSC species. Together, these results, as well as the observed genome size variations within species boundaries, reinforce the concept of a dynamic chromosomal architecture in the FOSC, where core genomic stability coexists with variable, strain-associated elements that may underlie ecological specialization and pathogenic diversity [56].

Within the FOSC, clade 2 represents the largest and most diverse lineage, encompassing seven recognized species and at least two additional undescribed lineages, provisionally designated *Fusarium* “afroindicum” and *Fusarium* “europaeum”. Both behave as independent evolutionary units with strong internal genomic conservation, with inter-species nucleotide identities of 98.02% and 98.05% relative to *F. fabacearum*, consistently below the >99% within-species values observed for recognized FOSC taxa. *Fusarium* “europaeum” displayed higher internal homogeneity than *Fusarium* “afroindicum” (98.61%), suggesting

different evolutionary dynamics between the two lineages. Phylogenetic reconstructions further support their distinctiveness: despite short branch lengths within clade 2, the divergence separating *F. “afroindicum”* and *F. “europaeum”* from neighboring taxa exceeds that observed among the closely related *F. fabacearum*, *F. callistephi*, and *F. cugenangense*.

Ecological and geographic patterns reinforce this view. *F. “afroindicum”* has an old-world transcontinental distribution, with isolates recovered from chickpea in Ethiopia and banana in India, highlighting its host range and ecological versatility. By contrast, *F. “europaeum”* appears geographically restricted, found exclusively in Europe on flax (*Linum usitatissimum*) and onion (*Allium cepa*). The convergence of genomic divergence, phylogenetic distinctiveness, contrasting geographic distributions, and distinct host associations supports their provisional recognition as candidate species within the FOSC, pending the availability of additional genomic resources for these lineages. We emphasize that the names *Fusarium “afroindicum”* and *Fusarium “europaeum”* are used here as informal, provisional phylogenomic designations and are not formally erected as new species under the International Code of Nomenclature for algae, fungi, and plants (ICN) in this work; their formal description awaits the availability of preserved type material and complementary phenotypic data.

Our analysis of SRA metadata reveals contrasting global distribution patterns among the most frequent FOSC species. *Fusarium fabacearum*, the most represented species, occurs widely across the Northern Hemisphere, whereas *F. cugenangense* and *F. odoratissimum* display broader, transcontinental ranges, the latter likely linked to the global trade of banana (*Musa acuminata*). In contrast, *F. nirenbergiae* appears confined to the Northern Hemisphere, while *F. languescens* spans the Americas and Europe. Notably, *F. oxysporum* s. str. was only detected in the Old World (Europe, Asia, Africa), a pattern that diverges from earlier literature describing it as a cosmopolitan pathogen across continents and hosts. Although current genomic sampling remains limited, these data suggest that *F. fabacearum*, rather than *F. oxysporum* s. str., may represent the most globally widespread member of the complex and the most successful, given its recovery from a broader range of plant hosts.

While this study represents one of the largest phylogenomic datasets assembled for the FOSC to date, the geographic and host representation of publicly available genomes is inherently uneven. Regions with stronger genomic research infrastructure, particularly Europe, Asia, and parts of Africa, are overrepresented, whereas South America, Central America, and large parts of sub-Saharan Africa remain undersampled. Similarly, host representation is biased toward economically important crops such as banana, strawberry, and tomato, while isolates from wild plants, non-agricultural environments, or clinical sources are scarce. These biases likely influence the observed patterns of geographic distribution and host range reported here, and may underestimate the true diversity of the FOSC at a global scale. Future efforts to expand genomic sampling across underrepresented regions and hosts will be essential to refine the phylogenomic framework established in this study.

These findings have direct agricultural and phytosanitary implications. The high frequency of species misidentifications detected when comparing our phylogenomic assignments with original NCBI annotations highlights the limitations of conventional identification methods and their potential impact on disease management strategies, resistance breeding programs, and quarantine measures. Accurate species-level identification within the FOSC is a prerequisite for reliable epidemiological surveillance and the development of targeted disease control strategies, particularly for broadly distributed and host-diverse species such as *F. fabacearum*.

4. Materials and Methods

4.1. Genome Data Acquisition

This study employed an integrated workflow combining systematic genome collection, de novo genome assembly, quality assessment, conserved core genome annotation, orthology inference, and phylogenomic dataset construction. On 29 March 2023, the NCBI Sequence Read Archive (SRA) [72] was queried for Illumina whole-genome sequencing (WGS) datasets of *Fusarium oxysporum*, *F. foetens*, *F. veterinarianum*, and *F. odoratissimum*. Only paired-end datasets with at least 1 GB of sequencing data were retained. Hi-C libraries were excluded. A total of 721 accessions meeting these criteria were downloaded using the SRA Toolkit v3.0.0, along with their associated metadata. Raw reads were processed to remove adapters, trim low-quality bases (<Q30), discard short reads (<70 bases), and filter out reads with ambiguous bases. Genomes were assembled de novo using SPAdes (v3.15.3) [73] with the following parameters for Illumina paired-end data: `spades.py --isolate -t 20 -m 300 -k 77,89`.

4.2. Assembly Quality Assessment

Assembly quality was evaluated using multiple metrics, including total assembly size, number of scaffolds, scaffold N50, proportion of ambiguous bases (Ns), and GC content, calculated with a custom Python v3.9.5 script. Sequencing depth was estimated by mapping the cleaned read datasets to their respective assemblies and calculating per-base and average depth using SAMtools v1.14 [74]. Genome completeness and duplication indices were assessed with the BUSCO pipeline (v5.2.2) [75]. To evaluate clonality and intra-sample variation, alternate allele counts were obtained by analyzing the BAM files generated with SAMtools and BCFtools v1.14 [76]. To ensure suitability for phylogenomics, assemblies were retained only if they met strict yet flexible thresholds: $N50 \geq 50$ kb, genome size between 44–75 Mb, $\leq 50,000$ scaffolds, and a minimum sequencing depth of $10\times$. The summary of SRA run accessions that passed the filters and were used in this study, along with the GCA genome references and their genomic descriptive statistics, is provided in Supplementary Table S1.

4.3. Taxonomic Classification of Raw Sequencing Reads

To obtain an overview of the taxonomic composition of the raw sequencing data and to assess the potential presence of non-target organisms, raw reads from each SRA accession were classified using Kraken2 v2.1.6. Classifications were performed against a comprehensive reference database comprising bacterial, archaeal, viral, fungal, plant, and metazoan genomes, enabling broad detection of possible contaminants. Because k-mer-based classifiers may be sensitive to database composition and read fragmentation, Kraken2 results were interpreted as indicative rather than definitive assignments.

To derive more conservative estimates of taxon abundance, Kraken2 outputs were further processed with Bracken, which re-estimates read counts by accounting for k-mer distributions and genome length. Bracken abundance estimates were then summarized at the genus level using a custom shell script that aggregated read counts across taxonomic ranks. These genus-level summaries were used to estimate the relative contribution of *Fusarium* reads in each dataset and to identify the second most abundant non-*Fusarium* genus, thereby quantifying the proportion of potential non-target taxa (Supplementary Table S1).

4.4. BLAST-Based Post-Assembly Filtering of Residual Contamination

To assess and mitigate the potential impact of residual non-*Fusarium* contamination on genome size and GC content estimates, we implemented an additional post-assembly

filtering step based on BLASTN (2.12.0+) similarity searches. Assembled contigs were queried against a custom reference database comprising the finished genome of *Fusarium oxysporum* Fo47 (GCF_013085055) and a curated subset of representative FOSC genomes spanning multiple species. These reference genomes were selected based on optimal assembly statistics and minimal contamination signals identified through Kraken-based taxonomic profiling (ERR4080471, ERR1755748, SRR5725024, SRR10313874, SRR14342093, SRR10428564, SRR10428572, SRR10428583, SRR10428586, SRR10428591, SRR10428593, and SRR10428601). BLASTN searches were performed using stringent thresholds (E-value $< 1 \times 10^{-100}$ and bit score ≥ 1000), and only contigs meeting these criteria were retained for downstream analyses. Genome assembly size and GC content were recalculated from the filtered assemblies and used for comparative analyses at the species level (see Supplementary Table S1).

4.5. Orthology Inference, Single-Copy Gene Identification, and Phylogenomic Analysis

Single-copy conserved orthologs were identified using BUSCO (v5.2.2) [75]. BUSCO-annotated single-copy proteomes were then processed with SonicParanoid v1.3.5 [77], a graph-based orthology inference pipeline, to cross-validate results and remove residual orthology groups with duplicated loci. The resulting set of single-copy, non-redundant coding sequences (CDS) served as the basis for phylogenomic analysis. A curated set of 4286 single-copy, non-redundant CDS was used for phylogenomic reconstruction. Each CDS was individually aligned using MAFFT v7.490 [78], and alignments were concatenated into a partitioned supermatrix for phylogenetic inference. Phylogenetic inference was performed with IQ-TREE v3 [79] using the best-fit substitution model selected by ModelFinder [80] with a partitioned strategy and the -rcluster 10 option enabled. Branch support was assessed with 5000 ultrafast bootstrap replicates [81]. The resulting tree was rooted with *Fusarium foetens*, and tree topology, branch lengths, and support values were visualized using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) (Accessed on 1 December 2024). Adobe Illustrator v23 was used for graphical enhancement and coloring of the heat maps.

4.6. Genome-to-Genome Comparisons and Nucleotide Identity Analysis

Genome alignments and comparative analyses were performed using the DNADIFF program from MUMMER v3 [82]. All *Fusarium* genomes were aligned pairwise, and the fraction of aligned bases along with the average nucleotide identity were extracted from the “.report” files. A non-redundant summary table was then created and imported into R for downstream analysis.

4.7. Genetic Diversity, Host Range, and Metadata Integration

To support phylogenomic reconstruction, metadata on geographic origin and host range were systematically gathered from NCBI BioSample and BioProject records linked to each genome. For each assembly, the collection location (country and continent) and source of isolation, mainly from plants, were recorded. For SRA accessions lacking information on the isolation source or collection site, an extensive search was conducted using BioProject identifiers, strain codes, and related publications to find and add these details to the metadata. Genomes without metadata were labeled as “NA” and kept for comparative analyses.

The curated metadata were integrated with a clade-resolved phylogenomic framework to investigate patterns of diversity, host specialization, and geographic distribution across the three major clades of the FOSC, while also revealing two previously unrecognized lineages within clade 2. To explore and visualize this diversity, we generated Sankey diagrams using SankeyMATIC (<https://sankeymatic.com>) (Accessed 28 August 2025). These

diagrams linked species-level assignments to their corresponding phylogenomic clades, with flow widths scaled proportionally to the number of genomes in each category. In addition, they displayed connections among phylogenomic clades, host types, and sampling continents, thereby integrating taxonomic, ecological, and biogeographic information into a single visualization.

4.8. Chromosome Coverage and Identity Against *Fusarium* Reference Fo47

Pairwise comparisons between a reference genome *Fusarium* FOSC strain Fo47 (GCF_013085055) and a selection of query assemblies filtered by minimum 100× average sequencing depth were performed to assess chromosome-level coverage and sequence identity. Each query assembly was aligned against the reference genome using minimap2 [83] with the asm5 preset, and alignments were filtered to retain only those segments showing at least 95% nucleotide identity. Overlapping alignments were merged with BEDTools v2.30.0 [84], and the cumulative covered bases per chromosome were calculated. Chromosome length information was retrieved from the reference index generated by samtools faidx v 1.14 [74], enabling the computation of the proportion of each chromosome covered by the query assembly. For each chromosome, the mean nucleotide identity was computed as a length-weighted average across all retained alignments. The resulting metrics were compiled into a tabular report containing, for each query-reference comparison, the total aligned bases, reference chromosome length, percentage of coverage, and mean identity.

For visualization, chromosomes were considered present in a given query genome when at least 70% of the chromosome length was covered for chromosome identity analyses, or 50% for chromosome coverage analyses. Two heat maps were generated: the first summarized chromosome presence/absence and mean sequence identity per chromosome using the 70% threshold, while the second represented chromosome coverage, where the presence criterion was relaxed to a minimum of 50% coverage.

4.9. Statistical Analyses and Visualization

All statistical analyses and graph generation were performed in R version 4.5.0 [85] with RStudio 2025.05.0+496 [86]. Genome assembly metrics (assembly size, GC content, aligned bases, and average nucleotide identity) were summarized by group using the tidyverse package v2.0.0 [87]. Differences among FOSC clades were evaluated using non-parametric tests: the Kruskal–Wallis test for overall group comparisons and pairwise Wilcoxon rank-sum tests. Boxplots, scatterplots, and comparative visualizations were generated with ggplot2 [88], and statistical annotations were added using ggpubr [89]. Figures were arranged and combined using the ggpubr package v0.6.3.

Geographical distributions of the most frequently represented FOSC species were visualized using world map polygons from the ggplot2 package v4.0.3, with country-level presence annotated from filtered metadata. Countries without reports were shaded light gray, while those with confirmed presence were highlighted. Species maps were combined into a composite panel using ggpubr (ggarrange, annotate_figure) to summarize the global distribution of reported genomes.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms27146255/s1>.

Author Contributions: Conceptualization, J.F.A.; methodology, J.L.-J., J.M.D. and J.F.A.; software, J.L.-J.; validation, J.L.-J., J.M.D. and J.F.A.; formal analysis, J.L.-J.; investigation, J.L.-J. and J.F.A.; resources, J.F.A.; data curation, J.L.-J.; writing—original draft preparation, J.L.-J. and J.F.A.; writing—review and editing, J.L.-J., J.M.D. and J.F.A.; visualization, J.F.A., J.L.-J. and J.M.D.; supervision, J.F.A.; project

administration, J.F.A.; funding acquisition, J.F.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study did not involve humans or animals; all data were obtained from publicly available genome sequence repositories (NCBI SRA/GenBank).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in NCBI SRA as listed in Supplementary Table S1.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FOSC	<i>Fusarium oxysporum</i> species complex
FFSC	<i>Fusarium fujikuroi</i> species complex
FGSC	<i>Fusarium graminearum</i> species complex
FIESC	<i>Fusarium incarnatum–equiseti</i> species complex
WGS	Whole-genome sequencing
SRA	Sequence Read Archive
NCBI	National Center for Biotechnology Information
CDS	Coding sequence(s)
BUSCO	Benchmarking Universal Single-Copy Orthologs
MLSA	Multilocus sequence analysis
HGT	Horizontal gene transfer
VCG	Vegetative compatibility group
UFB	Ultrafast bootstrap
ANI	Average nucleotide identity
BLAST	Basic Local Alignment Search Tool
ML	Maximum likelihood

References

1. Achari, S.R.; Kaur, J.; Dinh, Q.; Mann, R.; Sawbridge, T.; Summerell, B.A.; Edwards, J. Phylogenetic relationship between Australian *Fusarium oxysporum* isolates and resolving the species complex using the multispecies coalescent model. *BMC Genom.* **2020**, *21*, 248. [CrossRef] [PubMed]
2. O'Donnell, K.; Sutton, D.A.; Rinaldi, M.G.; Magnon, K.C.; Cox, P.A.; Revankar, S.G.; Sanche, S.; Geiser, D.M.; Juba, J.H.; Van Burik, J.-A.H.; et al. Genetic Diversity of Human Pathogenic Members of the *Fusarium oxysporum* Complex Inferred from Multilocus DNA Sequence Data and Amplified Fragment Length Polymorphism Analyses. *J. Clin. Microbiol.* **2004**, *42*, 5109–5120. [CrossRef] [PubMed]
3. Rana, A.; Sahgal, M.; Johri, B.N. *Fusarium oxysporum*: Genomics, Diversity and Plant–Host Interaction. In *Developments in Fungal Biology and Applied Mycology*; Satyanarayana, T., Deshmukh, S.K., Johri, B.N., Eds.; Springer: Singapore, 2017; pp. 159–199. [CrossRef]
4. Cai, H.; Yu, N.; Liu, Y.; Wei, X.; Guo, C. Meta-analysis of fungal plant pathogen *Fusarium oxysporum* infection-related gene profiles using transcriptome datasets. *Front. Microbiol.* **2022**, *13*, 970477. [CrossRef] [PubMed]
5. Michielse, C.B.; Rep, M. Pathogen profile update: *Fusarium oxysporum*. *Mol. Plant Pathol.* **2009**, *10*, 311–324. [CrossRef] [PubMed]
6. Gordon, T.R.; Martyn, R.D. The Evolutionary Biology of *Fusarium oxysporum*. *Annu. Rev. Phytopathol.* **1997**, *35*, 111–128. [CrossRef] [PubMed]
7. Iida, Y.; Ogata, A.; Kanda, H.; Nishi, O.; Sushida, H.; Higashi, Y.; Tsuge, T. Biocontrol Activity of Nonpathogenic Strains of *Fusarium oxysporum*: Colonization on the Root Surface to Overcome Nutritional Competition. *Front. Microbiol.* **2022**, *13*, 826677. Erratum in *Front. Microbiol.* **2023**, *14*, 1275135. <https://doi.org/10.3389/fmicb.2023.1275135>. [CrossRef] [PubMed] [PubMed Central]
8. Sajeena, A.; Nair, D.S.; Sreepavan, K. Non-pathogenic *Fusarium oxysporum* as a biocontrol agent. *Indian Phytopathol.* **2020**, *73*, 177–183. [CrossRef]

9. Wojtasik, W.; Dymińska, L.; Hanuza, J.; Burgberger, M.; Boba, A.; Szopa, J.; Kulma, A.; Mierziak, J. Endophytic non-pathogenic *Fusarium oxysporum* reorganizes the cell wall in flax seedlings. *Front. Plant Sci.* **2024**, *15*, 1352105. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Achari, S.R.; Mann, R.C.; Sharma, M.; Edwards, J. Diagnosis of *Fusarium oxysporum* f. sp. *ciceris* causing Fusarium wilt of chickpea using LAMP and conventional PCR. *Sci. Rep.* **2023**, *13*, 2640. [\[CrossRef\]](#) [\[PubMed\]](#)
11. AL-Faifi, Z.; Alsolami, W.; Abada, E.; Khemira, H.; Almalki, G.; Modafer, Y. *Fusarium oxysporum* and *Colletotrichum musae* Associated with Wilt Disease of *Coffea arabica* in Saudi Arabia. *Can. J. Infect. Dis. Med. Microbiol.* **2022**, *2022*, 3050495. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Baayen, R.P.; O'Donnell, K.; Bonants, P.J.M.; Cigelnik, E.; Kroon, L.P.N.M.; Roebroek, E.J.A.; Waalwijk, C. Gene Genealogies and AFLP Analyses in the *Fusarium oxysporum* Complex Identify Monophyletic and Nonmonophyletic Formae Speciales. *Phytopathology* **2000**, *90*, 891–900. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Dilla-Ermita, C.J.; Goldman, P.; Jaime, J.; Ramos, G.; Pennerman, K.K.; Henry, P.M. First Report of *F. oxysporum* f. sp. *fragariae* Race 2 Causing Fusarium Wilt of Strawberry in California. *Plant Dis.* **2023**, *107*, 2849. [\[CrossRef\]](#)
14. Dobbs, J.T.; Kim, M.-S.; Dudley, N.S.; Klopfenstein, N.B.; Yeh, A.; Hauff, R.D.; Jones, T.C.; Dumroese, R.K.; Cannon, P.G.; Stewart, J.E. Whole genome analysis of the koa wilt pathogen (*F. oxysporum* f. sp. *koae*). *BMC Genom.* **2020**, *21*, 764. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Dongzhen, F.; Xilin, L.; Xiaorong, C.; Wenwu, Y.; Yunlu, H.; Yi, C.; Jia, C.; Zhimin, L.; Litao, G.; Tuhong, W.; et al. *Fusarium* Species and FOSC Genotypes Associated with Yam Wilt in South-Central China. *Front. Microbiol.* **2020**, *11*, 1964. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Duvnjak, T.; Vrandecic, K.; Sudaric, A.; Cosic, J.; Siber, T.; Matosa Kocar, M. First Report of Hemp Fusarium Wilt Caused by *Fusarium oxysporum* in Croatia. *Plants* **2023**, *12*, 3305. [\[CrossRef\]](#) [\[PubMed\]](#)
17. El-kazzaz, M.K.; Ghoneim, K.E.; Agha, M.K.M.; Helmy, A.; Behiry, S.I.; Abdelkhalek, A.; Saleem, M.H.; Al-Askar, A.A.; Arishi, A.A.; Elsharkawy, M.M. Suppression of Pepper Root Rot and Wilt Diseases Caused by *Rhizoctonia solani* and *Fusarium oxysporum*. *Life* **2022**, *12*, 587. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Halpern, H.C.; Qi, P.; Kemerait, R.C.; Brewer, M.T. Genetic Diversity and Population Structure of Races of *Fusarium oxysporum* Causing Cotton Wilt. *G3 Genes Genomes Genet.* **2020**, *10*, 3261–3269. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Herrero, M.L.; Nagy, N.E.; Solheim, H. First Report of *F. oxysporum* f. sp. *lactucae* Race 1 Causing Fusarium Wilt of Lettuce in Norway. *Plant Dis.* **2021**, *105*, 2239. [\[CrossRef\]](#)
20. Li, J.; He, K.; Zhang, Q.; Wu, X.; Li, Z.; Pan, X.; Wang, Y.; Li, C.; Zhang, M. Draft Genome and Biological Characteristics of *Fusarium solani* and *Fusarium oxysporum* Causing Black Rot in *Gastrodia elata*. *Int. J. Mol. Sci.* **2023**, *24*, 4545. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Li, X.; Kang, H.J.; Zhao, Q.; Shi, Y.X.; Chai, A.L.; Li, B.J. First Report of *Fusarium oxysporum* Causing Wilt on Coriander in China. *Plant Dis.* **2021**, *105*, 4164. [\[CrossRef\]](#)
22. Maryani, N.; Lombard, L.; Poerba, Y.S.; Subandiyah, S.; Crous, P.W.; Kema, G.H.J. Phylogeny and genetic diversity of the banana Fusarium wilt pathogen *F. oxysporum* f. sp. *cubense* in the Indonesian centre of origin. *Stud. Mycol.* **2019**, *92*, 155–194. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Miao, Y.H.; Chen, Q.H.; Yu, K.; Wang, Y.H.; Liu, D.H. First Report of Fusarium Wilt of *Coleus forskohlii* Caused by *Fusarium oxysporum* in China. *Plant Dis.* **2021**, *105*, 1559. [\[CrossRef\]](#)
24. Mokhtari, S.; Chavez, M.; Ali, A. First Report of *F. oxysporum* f. sp. *vasinfectum* Causing Fusarium Wilt of Cotton in Kansas, U.S.A. *Plant Dis.* **2023**, *107*, 1239. [\[CrossRef\]](#)
25. Rahman, M.Z.; Ahmad, K.; Siddiqui, Y.; Saad, N.; Hun, T.G.; Hata, E.M.; Rashed, O.; Hossain, M.I.; Kutawa, A.B. First Report of Fusarium Wilt on Watermelon Caused by *F. oxysporum* f. sp. *niveum* in Malaysia. *Plant Dis.* **2021**, *105*, 4169. [\[CrossRef\]](#)
26. Srinivas, C.; Nirmala Devi, D.; Narasimha Murthy, K.; Mohan, C.D.; Lakshmeesha, T.R.; Singh, B.; Kalagatur, N.K.; Niranjana, S.R.; Hashem, A.; Alqarawi, A.A.; et al. *Fusarium oxysporum* f. sp. *lycopersici* causal agent of vascular wilt disease of tomato: Biology to diversity—A review. *Saudi J. Biol. Sci.* **2019**, *26*, 1315–1324. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Tahat, M.M.; Aldakil, H.A.; Alananbeh, K.M. First Report of Damping-Off Disease Caused by *Fusarium oxysporum* on *Pinus pinea* in Jordan. *Plant Dis.* **2021**, *105*, 4153. [\[CrossRef\]](#)
28. Tahat, M.M.; Aldakil, H.A.; Alananbeh, K.M.; Othman, Y.A.; Alsmairat, N. First Report of Strawberry Wilt Caused by *Fusarium oxysporum* in Jordan. *Plant Dis.* **2023**, *107*, 967. [\[CrossRef\]](#)
29. Tang, L.; Fan, C.; Guo, X.; Yuan, H.; Wu, G.; Zhang, S. First Report of Fusarium Wilt of Industrial Hemp Caused by *Fusarium oxysporum* in Northeast China. *Plant Dis.* **2022**, *106*, 3205. [\[CrossRef\]](#)
30. Thangavelu, R.; Amaresh, H.; Gopi, M.; Loganathan, M.; Nithya, B.; Ganga Devi, P.; Anuradha, C.; Thirugnanavel, A.; Patil, K.B.; Blomme, G.; et al. Geographical Distribution, Host Range and Genetic Diversity of *F. oxysporum* f. sp. *cubense* in India. *J. Fungi* **2024**, *10*, 887. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Tziros, G.T.; Karaoglanidis, G.S. Identification of *F. oxysporum* f. sp. *lactucae* Race 1 as the Causal Agent of Lettuce Fusarium Wilt in Greece. *Microorganisms* **2023**, *11*, 1082. [\[CrossRef\]](#) [\[PubMed\]](#)
32. van Amsterdam, S.; Jenkins, S.; Clarkson, J.P. First Report of *F. oxysporum* f. sp. *lactucae* Race 1 Causing Lettuce Wilt in Northern Ireland. *Plant Dis.* **2023**, *107*, 2524. [\[CrossRef\]](#)

33. Xue, Z.; Tang, R.; Liu, S.; Kong, D.; Tan, D.; Suo, Y.; Qin, S. First Report of *Fusarium oxysporum* Causing Leaf Wilt on *Dinteranthus vanzylii* in China. *Plant Dis.* **2023**, *107*, 3306. [\[CrossRef\]](#)
34. Xue, Z.; Zhang, S.; Tang, R.; Kong, D.; Suo, Y.; Qin, S. First Report of *Fusarium oxysporum* Causing Stem Rot on *Mammillaria humboldtii* in China. *Plant Dis.* **2023**, *107*, 2229. [\[CrossRef\]](#)
35. Zheng, J.; Wang, L.; Hou, W.; Han, Y. *Fusarium oxysporum* Associated with Fusarium Wilt on *Pennisetum sinense* in China. *Pathogens* **2022**, *11*, 999. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Jackson, E.; Li, J.; Weerasinghe, T.; Li, X. The Ubiquitous Wilt-Inducing Pathogen *Fusarium oxysporum*—A Review of Genes Studied with Mutant Analysis. *Pathogens* **2024**, *13*, 823. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Yan, X.; Guo, S.; Gao, K.; Sun, S.; Yin, C.; Tian, Y. The Impact of the Soil Survival of the Pathogen of Fusarium Wilt on Soil Nutrient Cycling. *Microorganisms* **2023**, *11*, 2207. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Fravel, D.; Olivain, C.; Alabouvette, C. *Fusarium oxysporum* and its biocontrol. *N. Phytol.* **2003**, *157*, 493–502. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Henry, P.M.; Pastrana, A.M.; Leveau, J.H.J.; Gordon, T.R. Persistence of *F. oxysporum* f. sp. *fragariae* in Soil Through Asymptomatic Colonization of Rotation Crops. *Phytopathology* **2019**, *109*, 770–779. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Davis, R.L.; Hayter, J.T.; Marlino, M.L.; Isakeit, T.; Chappell, T.M. Pathogenic and Saprophytic Growth Rates of *F. oxysporum* f. sp. *vasinfectum* Interact to Affect Variation in Inoculum Density. *Phytopathology* **2023**, *113*, 1447–1456. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Gordon, T.R. *Fusarium oxysporum* and the Fusarium Wilt Syndrome. *Annu. Rev. Phytopathol.* **2017**, *55*, 23–39. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Lombard, L.; Sandoval-Denis, M.; Lamprecht, S.C.; Crous, P.W. Epitypification of *Fusarium oxysporum*—Clearing the taxonomic chaos. *Persoonia* **2019**, *43*, 1–47. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Bugingo, C.; Infantino, A.; Okello, P.; Perez-Hernandez, O.; Petrović, K.; Turatsinze, A.N.; Moparthi, S. From Morphology to Multi-Omics: A New Age of Fusarium Research. *Pathogens* **2025**, *14*, 762. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Fourie, G.; Steenkamp, E.T.; Gordon, T.R.; Viljoen, A. Evolutionary Relationships among the *F. oxysporum* f. sp. *cubense* Vegetative Compatibility Groups. *Appl. Environ. Microbiol.* **2009**, *75*, 4770–4781. [\[CrossRef\]](#) [\[PubMed\]](#)
45. O'Donnell, K. Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete *Fusarium sambucinum*. *Curr. Genet.* **1992**, *22*, 213–220. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Jiménez-Gasco, M.M.; Milgroom, M.G.; Jiménez-Díaz, R.M. Gene genealogies support *Fusarium oxysporum* f. sp. *ciceris* as a monophyletic group. *Plant Pathol.* **2002**, *51*, 72–77. [\[CrossRef\]](#)
47. Mohd-Hafifi, A.B.; Mohamed Nor, N.M.I.; Zakaria, L.; Mohd, M.H. Molecular phylogeny and genetic diversity of *Fusarium oxysporum* from various hosts in Malaysia. *Sci. Rep.* **2024**, *14*, 29708. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Mulé, G.; Susca, A.; Stea, G.; Moretti, A. Specific detection of *Fusarium proliferatum* and *F. oxysporum* from asparagus using primers based on calmodulin gene sequences. *FEMS Microbiol. Lett.* **2004**, *230*, 235–240. Erratum in *FEMS Microbiol. Lett.* **2004**, *232*, 229. [\[PubMed\]](#)
49. Nirenberg, H.I.; O'Donnell, K. New *Fusarium* species and combinations within the *Gibberella fujikuroi* species complex. *Mycologia* **1998**, *90*, 434–458. [\[CrossRef\]](#)
50. Taylor, A.; Vágány, V.; Jackson, A.C.; Harrison, R.J.; Rainoni, A.; Clarkson, J.P. Identification of pathogenicity-related genes in *F. oxysporum* f. sp. *cepae*. *Mol. Plant Pathol.* **2016**, *17*, 1032–1047. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Wang, H.; Xiao, M.; Kong, F.; Chen, S.; Dou, H.-T.; Sorrell, T.; Li, R.-Y.; Xu, Y.-C. Accurate and Practical Identification of 20 *Fusarium* Species by Seven-Locus Sequence Analysis. *J. Clin. Microbiol.* **2011**, *49*, 1890–1898. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Yang, Y.; Wang, Y.; Gao, J.; Shi, Z.; Chen, W.; Huangfu, H.; Li, Z.; Liu, Y. Characterisation of *F. oxysporum* f. sp. *radicis-lycopersici* in Infected Tomatoes in Inner Mongolia, China. *J. Fungi* **2024**, *10*, 622. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Kharte, S.; Kumar, A.; Mishra, P.; Ramakrishnan, R.S.; Sharma, S.; Mishra, N.; Chauhan, P.S.; Sharma, R.; Gautam, V.; Tiwari, S.; et al. Whole genome sequencing and functional annotation of *F. oxysporum* f. sp. *lentis*. *Front. Genet.* **2025**, *16*, 1585510. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Henry, P.M.; Pincot, D.D.A.; Jenner, B.N.; Borrero, C.; Aviles, M.; Nam, M.; Epstein, L.; Knapp, S.J.; Gordon, T.R. Horizontal chromosome transfer and independent evolution drive diversification in *F. oxysporum* f. sp. *fragariae*. *New Phytol.* **2021**, *230*, 327–340. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Gomez-Chavarria, D.A.; Rua-Giraldo, A.L.; Alzate, J.F. An evolutionary view of the *Fusarium* core genome. *BMC Genom.* **2024**, *25*, 304. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Van Westerhoven, A.C.; Aguilera-Galvez, C.; Nakasato-Tagami, G.; Shi-Kunne, X.; Martinez De La Parte, E.; Chavarro-Carrero, E.; Meijer, H.J.G.; Feurtey, A.; Maryani, N.; Ordóñez, N.; et al. Segmental duplications drive the evolution of accessory regions in a major crop pathogen. *N. Phytol.* **2024**, *242*, 610–625. [\[CrossRef\]](#) [\[PubMed\]](#)
57. McTaggart, A.R.; James, T.Y.; Shivas, R.G.; Drenth, A.; Wingfield, B.D.; Summerell, B.A.; Duong, T.A. Population genomics reveals historical and ongoing recombination in the *Fusarium oxysporum* species complex. *Stud. Mycol.* **2021**, *99*, 100132. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Brenes Guallar, M.A.; Fokkens, L.; Rep, M.; Berke, L.; van Dam, P. *Fusarium oxysporum* effector clustering version 2: An updated pipeline to infer host range. *Front. Plant Sci.* **2022**, *13*, 1012688. [\[CrossRef\]](#) [\[PubMed\]](#)

59. Edel-Hermann, V.; Lecomte, C. Current Status of *Fusarium oxysporum* Formae Speciales and Races. *Phytopathology* **2019**, *109*, 512–530. [CrossRef] [PubMed]
60. Komissarov, E.N.; Diabankana, R.G.C.; Abdeeva, I.; Afordoanyi, D.M.; Gudkov, S.V.; Dvorianinova, E.M.; Bruskin, S.A.; Dmitriev, A.A.; Validov, S.Z. Genomic Differences Between Two *F. oxysporum* Formae Speciales Causing Root Rot in Cucumber. *J. Fungi* **2025**, *11*, 140. [CrossRef] [PubMed]
61. Nirmaladevi, D.; Venkataramana, M.; Srivastava, R.K.; Uppalapati, S.R.; Gupta, V.K.; Yli-Mattila, T.; Clement Tsui, K.M.; Srinivas, C.; Niranjana, S.R.; Chandra, N.S. Molecular phylogeny, pathogenicity and toxigenicity of *F. oxysporum* f. sp. *lycopersici*. *Sci. Rep.* **2016**, *6*, 21367. [CrossRef] [PubMed]
62. Van Der Does, H.C.; Lievens, B.; Claes, L.; Housterman, P.M.; Cornelissen, B.J.C.; Rep, M. The presence of a virulence locus discriminates *Fusarium oxysporum* isolates causing tomato wilt from other isolates. *Environ. Microbiol.* **2008**, *10*, 1475–1485. [CrossRef] [PubMed]
63. Epstein, L.; Kaur, S.; Chang, P.L.; Carrasquilla-Garcia, N.; Lyu, G.; Cook, D.R.; Subbarao, K.V.; O'Donnell, K. Races of the Celery Pathogen *F. oxysporum* f. sp. *apii* Are Polyphyletic. *Phytopathology* **2017**, *107*, 463–473. [CrossRef] [PubMed]
64. Lievens, B.; Claes, L.; Vakalounakis, D.J.; Vanachter, A.C.R.C.; Thomma, B.P.H.J. A robust identification and detection assay to discriminate the cucumber pathogens *F. oxysporum* f. sp. *cucumerinum* and f. sp. *radicis-cucumerinum*. *Environ. Microbiol.* **2007**, *9*, 2145–2161. [CrossRef] [PubMed]
65. Mbofung, G.Y.; Hong, S.G.; Pryor, B.M. Phylogeny of *F. oxysporum* f. sp. *lactucae* Inferred from Mitochondrial Small Subunit, Elongation Factor 1- α , and Nuclear Ribosomal ITS Sequence Data. *Phytopathology* **2007**, *97*, 87–98. [CrossRef] [PubMed]
66. O'Donnell, K.; Kistler, H.C.; Cigelnik, E.; Ploetz, R.C. Multiple evolutionary origins of the fungus causing Panama disease of banana. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2044–2049. [CrossRef] [PubMed]
67. Skovgaard, K.; Nirenberg, H.I.; O'Donnell, K.; Rosendahl, S. Evolution of *F. oxysporum* f. sp. *vasinfectum* Races Inferred from Multigene Genealogies. *Phytopathology* **2001**, *91*, 1231–1237. [CrossRef] [PubMed]
68. van Dam, P.; Fokkens, L.; Schmidt, S.M.; Linmans, J.H.J.; Kistler, H.C.; Ma, L.-J.; Rep, M. Effector profiles distinguish formae speciales of *Fusarium oxysporum*. *Environ. Microbiol.* **2016**, *18*, 4087–4102. [CrossRef] [PubMed]
69. Fulton, J.C.; Amaradasa, B.S.; Ertek, T.S.; Iriarte, F.B.; Sanchez, T.; Norelli, J.; Goss, E.M.; Paret, M.L. Phylogenetic and phenotypic characterization of *Fusarium oxysporum* f. sp. *niveum* isolates from Florida-grown watermelon. *PLoS ONE* **2021**, *16*, e0248364. [CrossRef] [PubMed]
70. Ma, L.-J.; van der Does, H.C.; Borkovich, K.A.; Coleman, J.J.; Daboussi, M.-J.; Di Pietro, A.; Dufresne, M.; Freitag, M.; Grabherr, M.; Henrissat, B.; et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **2010**, *464*, 367–373. [CrossRef] [PubMed]
71. van Dam, P.; Fokkens, L.; Ayukawa, Y.; van der Gragt, M.; ter Horst, A.; Brankovics, B.; Housterman, P.M.; Arie, T.; Rep, M. A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Sci. Rep.* **2017**, *7*, 9042. [CrossRef] [PubMed]
72. National Library of Medicine (US), National Center for Biotechnology Information. Sequence Read Archive (SRA). Available online: <https://www.ncbi.nlm.nih.gov/sra/> (accessed on 29 March 2023).
73. Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinform.* **2020**, *70*, e102. [CrossRef] [PubMed]
74. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
75. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef] [PubMed]
76. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [CrossRef] [PubMed]
77. Cosentino, S.; Iwasaki, W. SonicParanoid: Fast, accurate and easy orthology inference. *Bioinformatics* **2019**, *35*, 149–151. [CrossRef] [PubMed]
78. Katoh, K. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef] [PubMed]
79. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. Erratum in *Mol. Biol. Evol.* **2020**, *37*, 2461. <https://doi.org/10.1093/molbev/msaa131>. [CrossRef] [PubMed] [PubMed Central]
80. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; Von Haeseler, A.; Jermini, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
81. Hoang, D.T.; Chernomor, O.; Von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [CrossRef] [PubMed]

82. Marçais, G.; Delcher, A.L.; Phillippy, A.M.; Coston, R.; Salzberg, S.L.; Zimin, A. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **2018**, *14*, e1005944. [[CrossRef](#)] [[PubMed](#)]
83. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)] [[PubMed](#)]
84. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)] [[PubMed](#)]
85. R Core Team. *R: A Language and Environment for Statistical Computing*, version 4.5.0; R Foundation for Statistical Computing: Vienna, Austria, 2024. Available online: <https://www.R-project.org/> (accessed on 2 July 2025).
86. Posit Team. *RStudio: Integrated Development Environment for R*, version 2025.05.0+496; Posit Software PBC: Boston, MA, USA, 2025. Available online: <https://posit.co/> (accessed on 2 July 2026).
87. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Golemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [[CrossRef](#)]
88. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed.; Springer International Publishing: Cham, Switzerland, 2016.
89. Kassambara, A. *ggpubr “ggplot2” Based Publication Ready Plots*, version 0.6.1; CRAN: Montpellier, France, 2016. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.