Mycosphere 14(1): 2263–2283 (2023) www.mycosphere.org ISSN 2077 7019
Article
Doi 10.5943/mycosphere/14/1/27

Medical *Fusarium:* novel species or uncertain identifications?

Song Y[1,2,3,4,6*], Hu S[5], de Hoog GS[2,6], Liu X[1,2,3,4], Meng X[1,2,3,4], Xue R[1,2,3,4], van Diepeningen AD[7], Fokkens L[8], Li RY[1,2,3,4], and Gao S[5*]

[1]*Department of Dermatology and Venerology, Peking University First Hospital, Beijing, China*
[2]*Research Center for Medical Mycology, Peking University, Beijing, China*
[3]*National Clinical Research Center for Skin and Immune Diseases, Beijing, China*
[4]*Beijing Key Laboratory of Molecular Diagnosis on Dermatoses, Beijing, China*
[5]*State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Science, Beijing, China*
[6]*RadboudUMC/CWZ Center for Expertise in Mycology, Nijmegen, The Netherlands*
[7]*Biointeractions and Plant Health, Wageningen University and Research, Wageningen, The Netherlands*
[8]*Laboratory of Phytopathology, Department of Plant Sciences, Wageningen University and Research, Wageningen, The Netherlands*

Song Y, Hu S, de Hoog GS, Liu X, Meng X, Xue R, van Diepeningen AD, Fokkens L, Li RY, Gao S 2023 – Medical *Fusarium*: novel species, or uncertain identification? Mycosphere 14(1), 2263–2283, Doi 10.5943/mycosphere/14/1/27

Abstract

Fungal infections increasingly threaten human health, particularly in immunocompromised patients. Identification of etiologic agents of infection is important for effective treatment. Here, we study a set of 48 *Fusarium* strains, most of which had been collected from Chinese hospitals over two decades. Sequences of *cam*, *rpb2*, *tub2*, and *tef1-α,* singly or multilocus, did not entirely match with the described taxa; therefore, the species problem and correctness of identification became a research question. Blast searches in multiple dedicated databases did not always provide identical species assignments. Results remained variable when compared with phylogeny with 636 sequences identified in recent literature, including sequences from 41 type strains. Assignment to specific species within a species complex based on > 1000 single-copy orthologs was also variable for some species. Previously published MALDI-ToF data provided identification that matched sequence-based identification at the species complex level, but performed poorly with lineages within the complexes. Different methods of identification yielded dissimilar results. Some previously identified strains in earlier publications deviated more from the reference than our clinical strains. We further tested species boundaries using levels of admixture and found high admixture levels between *F. fujikuroi, F. nisikadoi* and *F. oxysporum* species complexes. It is concluded that species complexes can be recognised phylogenetically by BLAST and by MALDI-ToF, but the high intra-specific diversity of these fungi interferes with the unequivocal assignment of individual isolates to particular lineages. At a higher taxonomic level, most clinical strains were found to belong to *F. fujikuroi*, *F. oxysporum* and *F. solani* (*Neocosmospora*) complexes, irrespective of the method used for identification. To determine whether these complexes were particularly overrepresented in clinical strains, we compared this data in a larger dataset of 216 clinical isolates to those in environmental samples. We found that clinical strains are enriched for the *F. solani* complex (*Neocosmospora*) and less for *F. oxysporum* and *F. fujikuroi* complexes, suggesting that among these opportunistic human pathogens, *F. solani* is particularly equipped to survive in clinical settings.

Submitted 31 August 2023, Accepted 29 November 2023, Published 27 December 2023
Corresponding Author: Yinggai Song – e-mail – syg3515@163.com,
Shenghan Gao – e-mail – gaoshh@im.ac.cn

Keywords – admixture – China – *Fusarium* – identification – species complex – trans-kingdom

Introduction

The impact of molecular methods on fungal systematics and nomenclature of medically relevant fungi has been dramatic. Nearly 70% of the names of clinical fungi have changed or become used in another sense during the two decades since 2000 (de Hoog et al. 2020). It is generally assumed that the identification of fungi will be more straightforward with this development, becoming less dependent on the specialised skills of the microbiologist. However, renovations and resulting species structures may differ considerably between genera, and concomitantly, current taxonomic approaches follow different routes. Some older genera appear extremely polyphyletic, and the need of reclassification is unquestionable. For example, the genus *Phoma*, described in the early 19th century around the same time as *Fusarium*, today (01-03-2023) is represented with 3165 names in the *Index Fungorum*, of which 17.2% clustered in 45 families belonging to 19 orders. In contrast, of the 1717 *Fusarium* species (*Index Fungorum*), 98.3% belong to a single family, *Nectriaceae*. Taxonomic rearrangements in *Fusarium* are therefore less *a priori* obvious. Compared to *Phoma*, which was an amalgamate of superficially similar but unrelated fungi, the taxonomic debate on *Fusarium* concerns the subdivision of a single, relatively well-recognisable entity which is being dissected into smaller entities by the currently much more powerful taxonomic methods. The taxonomic approaches of *Phoma* and *Fusarium* can be described as 'reallocation' and 'dissection', respectively. This impacts the taxonomic focus and species identification in the two genera.

Fusarium species may cause host-specific plant diseases and also occur in human infections (Segorbe et al. 2017, Meza-Menchaca et al. 2020). Three main clinical syndromes are known, i.e. onychomycosis (Ranakawa et al. 2015), keratitis (Szaliński 2021, Walther et al. 2017) and dissemination in patients with compromised innate immunity (Liu et al. 2014). Occasionally, *Fusarium* species infect cartilaginous fish (de Hoog et al. 2023) and sea turtles (Cafarchia et al. 2020). Fusarium-like fungi have long been underestimated as agents of human disease, but this has changed since disseminated infections emerged with the growth of patient populations with severely compromised innate immunity due to e.g. solid organ transplant or leukaemia (Muhammed et al. 2013, Liu et al. 2014, Yu et al. 2019). *Fusarium* conidia can be produced in submersion and be carried by blood circulation, leading to gangrenous ulcers all over the body (Jiang et al. 2016), known as target lesions (Nucci & Anaissie 2007) and often accelerating host death. Patients with acute myeloid leukaemia are highly susceptible to infection by environmental strains. The infections are iatrogenic, slimy conidia being aerosolised and inhaled via hospital showers and sinks (Mehl & Epstein 2008, Soutour et al. 2012). Another source of infection may be onychomycoses, which are fairly common and mostly subclinical (Bassiri-Jahromi & Khaksar 2010, Ninet et al. 2005, Castro et al. 2005, Brasch & Shimanovich 2012, van Diepeningen et al. 2015). Fusarium-like fungi as agents of keratitis (Walther et al. 2017) are significant because fulminant infections may lead to blindness. *Fusarium*-implanted eye infections are prevalent in India (Homa et al. 2018, Satpathy et al. 2019) but occur worldwide. A famous series of *Fusarium* eye infections was due to contact-lens fluid contamination, leading to a large pseudo epidemic in the USA (O'Donnell et al. 2007).

Clinical strains lack host plant information, which would enhance species identification. The question then arises to what extent strains of named *Fusarium* species isolated from sources other than their preferred plant host can be recognised as belonging to the same species using genetic parameters only. In a documented overview, Crous et al. (2021) distinguished 27 fusarium-like genera, recognising 18 species complexes in *Fusarium sensu stricto*. Species identification rested mainly upon multilocus sequencing of partial *rpb1, rpb2, tef-1α, cam*, and *tub2*. In principle, fungal barcoding has been developed for the identification of previously defined taxonomic entities. The species concept applying genealogical concordance (GCPSR) provides insight into the genetic isolation of lineages. As yet, this approach has not yet been used in the majority of the species, and in *Fusarium* and other fungi, the approach is often complicated by horizontal gene transfer

(Vlaardingerbroek et al. 2016), uniparental sex (Zhao et al. 2019), other types of genetic interactions, or overabundance of clonality.

Fusarioid fungi show high degrees of intrinsic antifungal multiresistance (Al-Hatmi et al. 2019), necessitating rapid and appropriate clinical management. In this scenario, correct identification of the etiologic agent is essential. The infections are uncommon, so the clinician usually has only one strain to identify. We sequenced clinical strains from China collected over two decades and aimed at correct identification, which is mandatory to evaluate purported clinically relevant parameters such as virulence and antifungal susceptibility. Strains were acquired during hospital routines in several medical centres and included isolates from superficial as well as deep locations. The strains also comprised an isolate from an infection in a shark. To achieve reliable identification and attribution to species complexes, we blasted sequences in dedicated databanks and used our previous identification with MALDI-ToF (Song et al. 2021). As results were not always concordant, we performed phylogenetic clustering of barcoding genes, including sequences of type strains and with verified sequences from recent monographs (Crous et al. 2021; Wang et al. 2022), being the method of choice for taxonomic studies. Subsequently, we compared the trees with clustering based on whole genomes. Above the species level, we aimed to establish whether clinical strains are a random reflection of the extant diversity of the genus *Fusarium*/*Neocosmospora*, or whether particular species complexes or molecular siblings are more prone to cause human infection.

Material & methods

Samples, DNA extraction and identification

Fusarium-like strains (n = 48) were isolated from different geographic regions and hosts in China for 22 years (1995−2017). Strains were obtained from symptomatic immunocompetent patients with keratitis, onychomycosis and skin lesions, and lung and blood samples of compromised patients; one isolate came from a shark peritoneum (Table 1, Fig. 1). In addition, five environmental strains were included for comparison. Genomic DNA from each strain was extracted using a DNeasy Plant Mini kit (Qiagen, Hilden, Germany). DNA was quantified using QUBIT 200 (ThermoFisher Scientific), and DNA libraries were prepared for Illumina and PacBio sequencing; sequence data processing was done using Illumina BASESPACE and PacBio DEVNET pipelines, respectively. The 48 strains were identified by BLAST with concatenated partial *cam*, *rpb1*, *rpb2*, *tub2*, and *tef1-α* using the Fusarioid-ID database (https://www.fusarium.org/), Mycobank (https://www.mycobank.org/) and NCBI (https://blast.ncbi.nlm.nih.gov), with the exclusion of *cam* in the *F. solani* complex. Identity values of >99% were considered to confirm species identity. Strains matching with the same species were arranged according to month of isolation; when similarity values of strains isolated in the same period were identical, strains were considered to concern identical clones. Multilocus phylogeny based on *rpb2*, *tef-1α*, and *tub2* comprised sequences of 424 isolates published by Wang et al. (2022) and 212 sequences extracted from genomes deposited in NCBI. The distribution of clinical strains over species complexes comprised 216 isolates previously sequenced by Song et al. (2021). Sequenced genomes were compared with 176 genomes available in NCBI. MALDI-ToF identification of the same 48 strains was published by Song et al. (2021). For delimitation of species complexes (Fig. 2), SC circumscriptions were taken from published taxonomic studies as reference (Crous et al. 2021, Wang et al. 2022). Trees were inferred from alignments of *cam*, *tef1-α, rpb1, rpb2,* and *tub2* genes separately (data not shown) and concatenated.

Genome sequencing

A total of 48 short-reads libraries for 48 BMU isolates (preserved at the Research Center for Medical Mycology, Peking University, Beijing) were constructed using the KAPA HyperPrep Kit (Roche, Pleasanton, CA, USA) with insert size of 350 bp and sequenced on a MGISEQ 2000 platform for 2 x 150 bp paired-end short-read sequencing data according to the manufacturer's

instructions. Eight representative isolates selected from the BMU set were sequenced using the Oxford Nanopore MINION platform using a SPOTON R9.4.1 flow cell (FLO-MIN106), with native barcoding kit (EXP-NBD103) and 1D chemistry (SQK-LSK108). Barcodes from 17 to 24 were used for eight sequenced isolates. The Oxford Nanopore raw data were base-called and multiplexed using the GPU version of GUPPY (v4.0.9) (https://community.nanoporetech.com) on an NVIDIA RTX2080 Super GPU card.

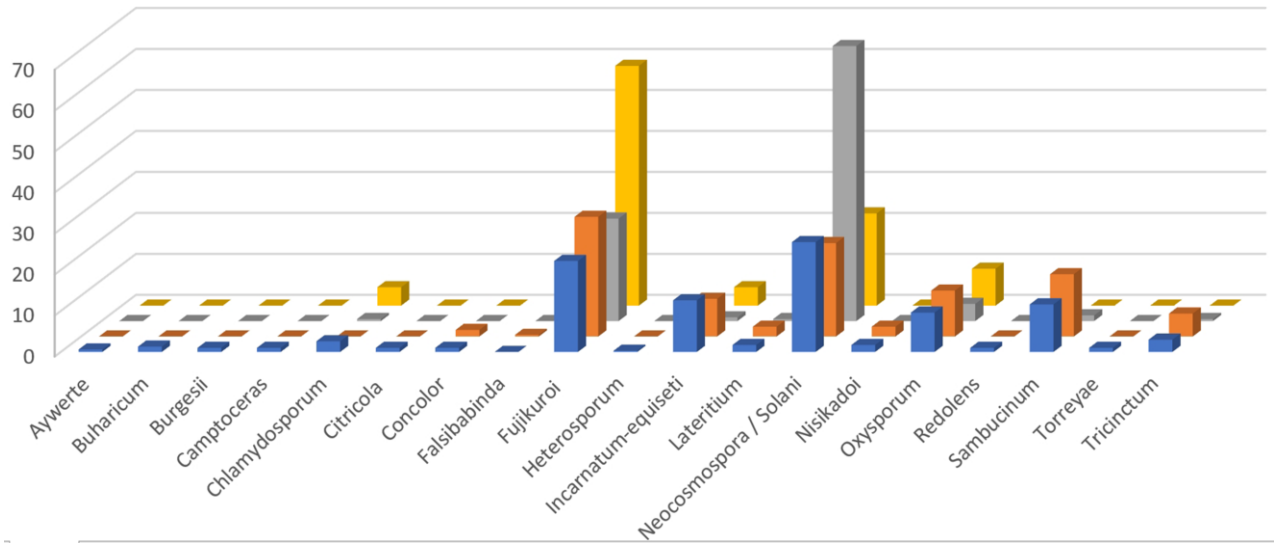

Figure 1 – General distribution of Fusarium species over species complexes, compared with Chinese clinical strains. Blue: global environmental species in Crous et al. (2021); brown: Chinese environmental species in Wang et al. (2022); grey: 216 Chinese clinical strains; ochre: clinical species in de Hoog et al. (2020).

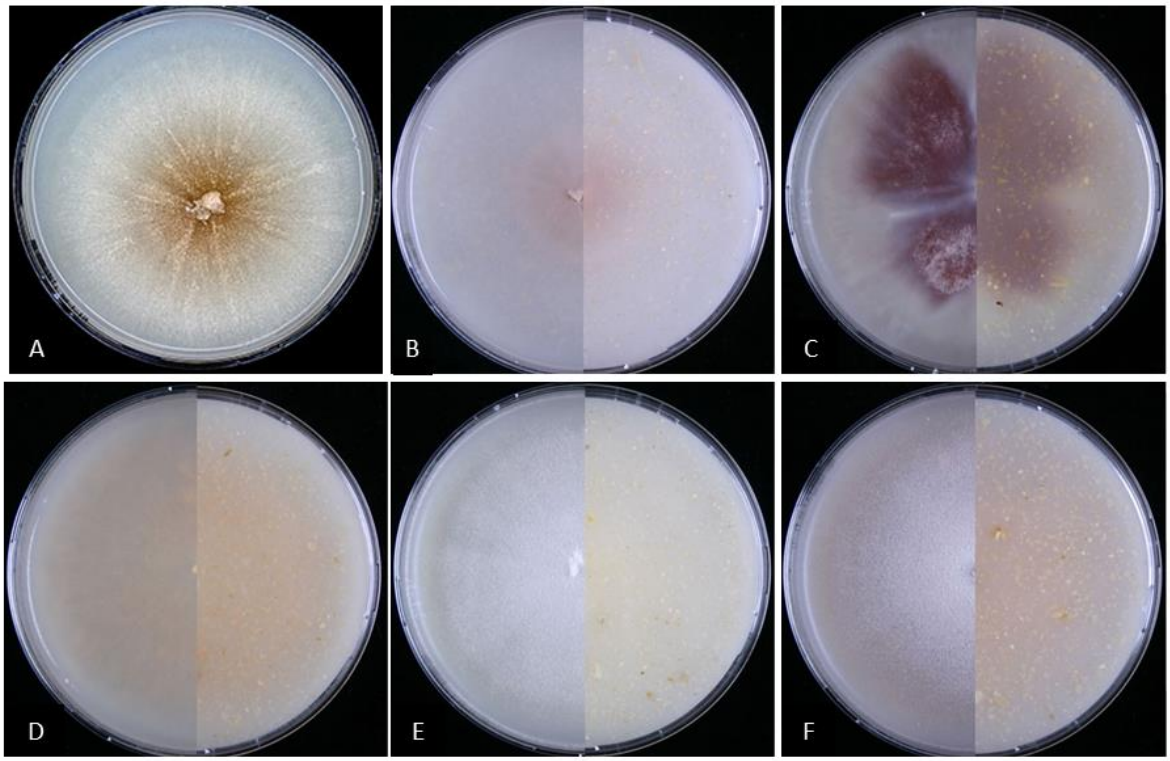

Figure 2 – Morphology of Fusarium species complex. A *Fusarium petroliphilum*, CBS 135534. B *Fusarium falciforme*, CBS 101427. C *Fusarium oxysporum*, CBS 118995. D *Fusarium*

incarnatum, CBS 622.87. E *Fusarium verticillioides*, CBS 539.79. F *Fusarium sacchari*, CBS 121683, all on OA, 10 d. Reproduced from de Hoog et al. (2020).

Genome assembly and gene prediction

First, the raw data of eight isolates sequenced with Oxford Nanopore were assembled using FLYE (v2.9) with parameters "--nano-hq". Since there were corresponding short-read libraries for these eight isolates, the short-read sequencing data were mapped to their corresponding assembled contigs using MINIMAP2 (v2.17) and corrected the contigs using PILON (v1.23) with the parameter "--mindepth 10". Contigs that had <1X depth by short-read data were discarded. To improve the quality continuity of short-read assemblies, only genome assemblies from long-read were used as references to guide scaffolding the short-read assemblies and filling the gaps. First, a k-mer database (K = 18) was created by KMER-DB (v1.2.9) from 8 long-read assembled genomes. Next, for each of the 48 isolates, short-read assemblies were performed by MEGAHIT (v1.2.9). The assembled contigs were checked for contamination using KRAKEN2 with the "PLUSPF" collection and retained if identified within the taxonomy *Hypocreales*. Duplicated assembled contigs (possibly haplotype-like from individuals) were removed by PURGE HAPLOTIGS (v1.1.1) with the parameter "-l 10, -h 100" and -m adjusted according to the depth plot of each assembly. Contigs with low depth (<10X) or high depth (>100X, such as mitochondrion) were also removed in the purging. The purged contigs were then compared to the pre-built k-mer database for the percentage of shared k-mers between the short-read and long-read assemblies. The long-read assembled genome with the highest percentage of shared k-mers was selected as the reference for the isolate. Each short-read assembly was scaffolded against using its selected reference genome using by RAGTAG (v2.1.0), then gap-filled using PILON (v1.23) with parameter "-fix snps,indels,gaps --mindepth 10". Finally, the assembled scaffolds were checked for contamination using Kraken2 with the "PlusPF" collection and retained if identified within the Hypocreales taxonomy. The cleaned contigs were renumbered as final long-read assemblies, and the quality of each assembly was evaluated by MERQURY (v1.3) using short-read libraries with K = 17.

Prediction of protein-coding genes was performed by FUNANNOTATE (v1.8.11) with species "Fusarium" for AUGUSTUS and proteins from *Hypocreales* in the UNIPROT database (2022-09-19) as protein evidence. The functional annotation of predicted genes was done by INTERPROSCAN (v5.51-85.0, docker version) and ANTISMASH (v6.0.0, online version).

Dataset preparation

We listed the number of named lineages in each of the 18 species complexes included in our reference publications providing a global overview of fusarium-like genera (Crous et al. 2021), and an overview of environmental fusaria in China (Wang et al. 2022) and counted the number of clinical strains in each of the species complexes. The barcode genes and whole genome assemblies of *Fusarium* strains were downloaded separately. For barcode genes, we used the 424 strains dataset of five loci (*tef-1α*, *cam*, *tub2*, *rpb1*, and *rpb2*) from the publication of Wang et al. (2022). In addition, the sequences of these five loci from 249 CBS and NRRL strains deposited in the NCBI nucleotide database were added (Supplementary Tables 1−7).

For whole genome assemblies of queried strains, to include as many species as possible, 176 species-representative genomes (the reference genome for each species in the NCBI assembly database) of described *Fusarium* species were downloaded from NCBI. In the same way, genome assemblies (as of 21-02-2023) within 10 involved species complexes (*F. concolor* SC, *F. fujikuroi* SC, *F. incarnatum-equiseti* SC, *F. lateritium* SC, *F. nisikadoi* SC, *F. oxysporum* SC, *F. redolens* SC, *F. sambucinum* SC, *F. solani* SC and *F. tricinctum* SC) were obtained. The downloaded genomes were checked for completeness by BUSCO (v5.0, hypocreales_odb10) and those with completeness (C-value) lower than 90% were removed from the dataset. Unidentified strains (*Fusarium* sp.) were also excluded. The genome assembly of *Acremonium chrysogenum* ATCC 11550 (GCA_000769265.1) was used as an outgroup for the species tree. Protein coding genes

were predicted on each downloaded genome assembly by FUNANNOATE (v1.8.11) using proteins from taxonomy *Hypocreales* in the UNIPROT database (2022-09-19) as reference.

Extraction of barcode genes from the genome assemblies

Genes from the whole genome assemblies without barcodes for the corresponding strain available had their sequences directly extracted by using the amplicon function in SEQKIT (v0.15.0) using PCR primers employed by Wang et al. (2022). When the extraction with these primers failed, reference sequences of the barcode genes were used to search for the aligned sequences on the genome assembly to represent the barcode genes. The reference sequence for each gene was selected by clustering downloaded barcode gene sequences using CD-HIT (v4.8.1) with 90% global identity. The representing sequence of the cluster with the largest number of clustered sequences was selected as a reference for each gene. The reference sequence was then aligned to the genome assemblies by BLASTN (v2.12.0+). The longest alignment on the genome was extracted as a barcode sequence for each strain if the primers extracted no sequence.

Species tree based on single-copy orthologs

To infer the species tree based on whole genome assemblies, the protein-coding genes from the BMU strains and the downloaded sequences were clustered into orthologous groups using ORTHOFINDER (v2.5.4). Due to the limitation of server memory (≤512G), ORTHOFINDER was unable to include all assemblies; hence, representative *Fusarium* species genomes were selected to build the species tree (Supplementary Table 10), resulting in a dataset of 224 *Fusarium* strains, and *Acremonium chrysogenum* ATCC 11550 (GCA_000769265.1) as outgroup. Of the 32,742 orthogroups, 1,014 represent single-copy orthologues and were used in subsequent phylogenetic analyses. For each ortholog, multiple sequences were aligned by MAFFT (v7.487) with the high accuracy parameter "--maxiterate 1000 --localpair". For each strain, the aligned protein sequences were concatenated into a single sequence by the CONCAT function in SEQKIT (v0.15.0). The species tree was inferred by IQ-TREE (v2.0.3) using MODELTEST to select the best substitution model (JTT+F+R9). The branch support was calculated using ultra-fast bootstrapping (1000 bootstraps: -B 1000) and approximate BAYES single branch test (--abayes). The final consensus tree was illustrated and annotated by EVOLVIEW (https://www.evolgenius.info/evolview/).

Similarity based on whole genome average nucleotide identity

To investigate the whole genome similarity between strains and the diversity within different clades, we carried out pairwise ANI (Average Nucleotide Identity) analysis using the whole genome sequences from the same 224 *Fusarium* and 1 outgroup strains which were selected above for single-copy orthologs analysis. Higher ANI value indicates higher similarity between two genomes, and 100 indicates the two genomes were nearly identical. The ANI value was calculated by FASTANI (v1.33) using default parameters between every two genomes to generate an ANI matrix of 225x225 representing the similarity between all strains. The ANI matrix was clustered and plotted using SEABORN (v0.13.0) package in Python. Note that due to the algorithm in FASTANI, when aligning genome A to genome B, the calculated ANI value is slightly different when aligning B to A. Therefore, the matrix is asymmetric.

Species tree based on barcode genes

The species tree based on barcode genes was inferred by multi- and single-locus analyses (*cam*, *tub2*, *tef-1α*, *rpb1*, and *rpb2*). Since one or more barcode genes might be absent from some of the strains, multiple combinations of barcode were tested and it found that the combination of *rpb2* + *tef1*-α + *tub2* included the largest number of strains, and therefore, this combination was selected for multilocus phylogenetic analysis. For single-locus trees, barcode sequences from available strains were aligned by MAFFT (v7.487) as described above. Each single-locus tree was inferred by IQ-TREE (v2.0.3) with internal model test, thorough NNI search (--allnni), approximate BAYES single branch test (--abayes) and 1000 ultra-fast bootstrap replicates (-B 1000). The best-fit models

for barcode genes were TIMe+I+G4 for *cam*, TIM3e+R4 for *rpb1*, TIM2e+R5 for *rpb2*, TIM2e+I+G4 for *tef-1α*, and TNe+I+G4 for *tub2*, respectively. For the multilocus tree, to ensure that genes were correctly aligned, the MAFFT alignments of single genes were concatenated into a single alignment for IQ-TREE inputs. Only strains for which all three genes (*rpb2*, *tef-1α*, and *tub2*) could be retrieved were included in the tree. The multilocus tree was also inferred by IQ-TREE (v2.0.3) with the same above parameters. Consensus trees were illustrated and annotated by EVOLVIEW. On the tree plot, the number of SNPs relative to the counter-clockwise neighbour for each strain was calculated from the VCF output of SNP-sites (v2.5.1; http://dx.doi.org/10.1099/mgen.0.000056), which processed the MAFFT alignments of neighbour strains. The species groups were determined by whether strains consistently shared the same species name. The pairwise ANI (Average Nucleotide Identity) was calculated within each determined species group by FASTANI (v1.33) using the pre-aligned sequences, and the minimum ANI was plotted on the tree.

Population structure

The population structure of *Fusarium* species was analysed based on the MAFFT alignments of three loci (*rpb2* + *tef1*-α + *tub2*). The alignment was processed by SNP-sites (v2.5.1). The VCF output was filtered by VCFTOOLS (v0.1.16) to remove the outgroup and positions with low allele frequency (--maf 0.05) and high missing rate (--max-missing 0.95, mostly aligned gaps). The filtered VCF file was converted to PED format by PLINK (v1.9). The population structure was estimated by ADMIXTURE (v1.3.0) with continuous K from 5 to 20, and K = 10, which corresponds to the number of analysed species complexes, was chosen. The order of strains in the population structure was adjusted according to the order of the species tree and plotted by PYTHON packages MATPLOTLIB (v3.4.2). The clusters were defined based on previously identified identifications and added the type strains for unambiguous reference to species identity.

Results

Identification by BLAST

From the multilocus identification, some isolates (BMU 01940, 01979, 01976, 02030, 02023) from human eye infections acquired in 2001 were considered to concern identical clones. Identical strains were derived from different patients and could represent a clonal outbreak. The distinction of *F. falciforme* and *F. keratoplasticum* from *F. solani* in Fusarioid-ID and MycoBank was not reached in GenBank. Two environmental strains (BMU 00842, 00845) were identified in Fusarioid-ID as *F. verticillioides* but at a low maximum similarity value (98.52%) and in MycoBank as *F. annulatum* (max. 99.32%); these isolates were identified with phylogenetical clustering (below) as *F. madaense*. In GenBank, the clinical strains of *F. verticillioides* also matched with *F. musae*, at a slightly lower value (max. 99.49%) than in Fusarioid-ID (max. 100%). Strains clustering with *F. annulatum* at 100% identity in Fusarioid-ID clustered in MycoBank with *F. proliferatum* and *F. fujikuroi*, also with 100% identity (Table 1).

Identification by phylogenetic clustering

Alignment data of barcoding genes are analysed and phylogenetic tree showed the species complexes *F. concolor*, *F. incarnatum-equiseti*, *F. sambucinum*, *F. solani* and *F. tricinctum* were recognised by barcoding gaps, being separated by inter-specific differences that were larger than intra-specific variation (Table 2, Fig. 3). Barcoding gaps were missing between the *F. fujikuroi* and *F. oxysporum* species complexes. As these differences depend on the position in the trees, borderlines of species complexes were verified with admixture analysis (Fig. 4), taking K (= 9) as the number of published species complexes represented in the dataset. The *F. tricinctum* SC was separated from the remaining groups. In contrast, the *F. fujikuroi* and *F. oxysporum* species complexes showed significant gene flow within and between complexes, with a described

F. nisikadoi SC being unrecognisably located in the overlapping area between the two complexes (Fig. 4). Conversely, the *F. solani* complex contained two distinguishable entities.

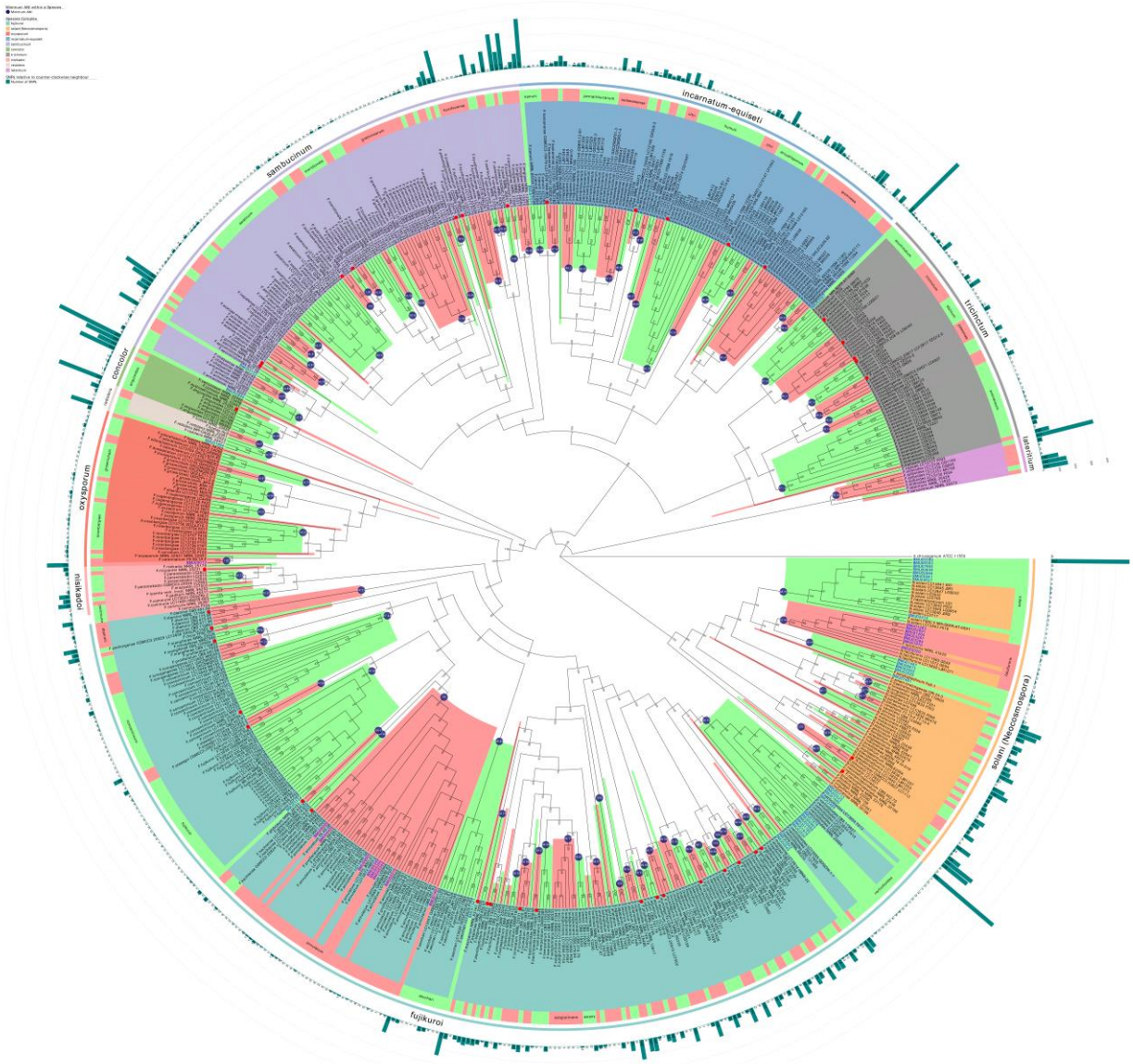

Figure 3 – Multilocus tree based on 636 sequences of *rpb2*, *tef-1α*, and *tub2* partial genes. Alternating green/pink triangles show separate molecular species with bootstrap values of each clade in blue. Species complexes are differentially coloured. Names in black are taken from literature, with a red dot if the type strain; names in blue are newly sequenced clinical strains. Bars in the outermost row represent the numbers of mutations compared to the neighbouring strain. Clades composed of isolates with the same reference name were shaded (alternately red/green) as recognised species. Differences counted were SNPs, with Indels counted as one, and were displayed in the outer shell.

Because of identifying our strains as members of clades at the level of published molecular siblings, separate and concatenated trees of *cam*, *tef1-α*, *rpb2*, and *tub2* were constructed. Results were scored as affiliation of target isolates to the nearest species defined by a type or reference strain (Table 1). Of the sequenced genes, *rpb1* showed the highest resolution, and tef1-α was the lowest. Species clades showed different degrees of variation. In the tree based on concatenated sequences, some species (e.g. *F. annulatum*, *F. fujikuroi*) showed a wide range of variability. Within the *F solani* SC and using individual as well as concatenated genes, most Chinese clinical strains clustered with *F. solani* (n = 8) and *F. falciforme* (n = 12), while *F. keratoplasticum* was

missing from the used databases. In *F. falciforme*, five strains in *rpb1* and *rpb2* could not be assigned to the clade but clustered with *tef1-α* and *tub2*. With Fusarioid-ID and WGS data, these strains clustered with *F. keratoplasticum*. Of the strains identified with BLAST as *F. verticilloides/ F. musae*, all clinical isolates clustered with individual and concatenated genes and whole genomes with *F. verticillioides*. However, the two environmental strains with unambiguous BLAST identification as *F. verticillioides* were identified with single and concatenated *rpb2*, *tub2* and *tef1-α* genes as *F. madaense*. At the same time, with WGS, where no reference genome of *F. madaense* was available, they individualised as a small clade at equal distance to *F. napiforme* and *F. coicis*. Ten strains from various sources were identified in Fusarioid-ID as *F. annulatum. Cam* and *rpb2* were insufficiently variable, showing overlapping identity with *F. elaeagnei, F. hechiense* and *F. fujikuroi*; the strain with the lowest BLAST similarity (99.39%) clustered in WGS with *F. fujikuroi*.

One strain was identified as *F. acuminatum* in the *F. tricinctum* complex, with all genes applied. Two strains in the *F. fujikuroi* complex, BMU 03134 and BMU 09013, could not be linked unambiguously to any of the molecular siblings, since, with different genes, close similarities were found with *F. annulatum, F. cugenangense, F. curvatum, F. duoseptatum, F. elaeagni*, *F. elaeidis, F. fujikuroi*, *F. grosmichelii, F. nirenbergiae*, and *F. oxysporum*. The concatenated data yielded *F. oxysporum*, and the WGS clustering *F. oxysporum/F. veterinarium*. A single strain (BMU 01974) in the *incarnatum-equiseti* complex was identified with three genes as *F. nanum*, but its *rpb1* sequence was found close to *F. luffae* and *F. pernambucanum*. The strain BMU 00844 clustered with *F. armeniacum* with the least variable gene, *tef1-α*, but is found at some distance from all other genes; a second, similar strain (BMU 03122) is even more remote from *F. armeniacum* in all genes. The strain BMU 01904 was distant from all sequences analysed and was identified by BLAST as *F. redolens*.

Identification with whole-genome data

Strains analysed for whole genomes concerned 35 clinical isolates, one strain from a cold-blooded animal, and 12 environmental isolates collected, given the detection of sources of contamination. Strains were first identified at the level of species complexes. A total of 216 Chinese target strains identified as *Fusarium/Neocosmospora* were distributed over eight species complexes: *F. fujikuroi* SC (*n* = 54), *F. incarnatum-equiseti* SC (*n* = 2), *F. lateritium* SC (*n* = 1), *F. oxysporum* (*n* = 9), *F. solani* SC (*n* = 145), *F. sambucinum* SC (*n* = 3) and *F. tricinctum* SC (*n* = 1); one strain had an undetermined position. A selection of strains from the eight species complexes was whole genome sequenced; 42 (87.5%) of these belonged to *F. solani* and *F. fujikuroi* species complexes, and six to the *F. sambucinum* (*n* = 2), *F. oxysporum* (*n* = 1), *F. tricinctum* (*n* = 1) and *F. incarnatum-equiseti* (*n* = 1) species complexes (Table 2).

Whole genome sequencing data are summarised in Table 3, with additional data in Supplementary Tables 8−10. Significant variation is noted within species complexes, mainly due to a small number of deviating strains. The *F. solani* complex (*Neocosmospora*) has genome sizes between 44.7 and 56.4 Mb. Genomes in the *F. fujikuroi* complex are between 42 and 44 Mb, but *F. annulatum* from shark peritoneum (BMU 05348) was close to 56.8 Mb. Similarly, deviating strains were noted with percentages G+C (Table 3); genome size and deviating GC% were not correlated, and in the genome tree (Fig. 5), the position of these strains did not deviate significantly from the position in the tree inferred from multilocus barcoding sequences. Comparing ANI data, the clinical strains demonstrated limited intraspecific variability compared to environmental reference strains (Figs 3, 6). Nearest neighbour in the genome clustering often deviated from BLAST and multilocus results due to differences in the genomes available in databases, which makes these comparisons invalid. Strains in the *F. solani* species complex clustering with 100% identity with *F. solani* or *F. keratoplasticum* clustered with WGS data with *F. haematococcum* and *F. paranaense*, names which did not appear with BLAST results in any of the databases consulted. In the *F. fujikuroi* complex, strain BMU 01341 was the only isolate unambiguously identified with all used approaches as *F. sacchari*.

Table 1 BLAST identification results of 48 BMU strains using concatenated sequences of *rpb2*, *tef-1α* and *tub2* partial genes in three databases; abbreviations of species names are followed by percentage similarity. Subsequent columns show individual and multilocus identifications based on phylogenetic comparison and identification based on phylogenetic genome comparison. The source and date of isolation are listed. Identical species names have the same colour in the entire

BMU	Source	Date	MALDI-TOF MS	Score value	Fusarioid-ID	Mycobank		GenBank			*CaM*			*RPB1*	*RPB2*		*TEF1*		*TUB2*	Multilocus	Genome		Complex
00841	Plant	1995.05	MO	2.11	SO 100	SO 100		SO 100		ND				SO	SO	FA	SO	FA	SO	SO	SO	HA	solani
00846	Plant	1995.05	VE	1.85	SO 100	SO 100		SO 100		ND				SO	SO	FA	SO	FA	SO	SO	SO	HA	solani
00839	Plant	1995.05	SO	2.00	SO 100	SO 100		SO 100		ND				SO	SO	FA	SO	FA	SO	SO	SO	HA	solani
01264	Eye secretion	1999.12	SO	2.11	FA 99.94	FA 99.88		SO 99.89	FA 99.80	ND				FA	SO	FA	SO	FA	FA	FA	FA	PA	solani
01263	Eye secretion	1999.12	SO	2.17	FA 99.94	SO 99.89	FA 99.85	SO 99.94	FA 99.55	ND				FA	SO	FA	SO	FA	FA	FA	FA	PA	solani
01467	Foot ulcer	2000.08	SO	2.2	KE 100	KE 100		SO 100		ND				aff. FA	aff. FA		SO	FA	FA	FA	KE		solani
01970	Cornea	2001.04	SO	2.4	KE 101	FA 99.94		FA 99.94		ND				FA	SO	FA	SO	FA	FA	FA	FA	PA	solani
02025	Cornea	2001.04	SO	1.82	KE 100	KE 100		SO 100		ND				aff. FA	aff. FA		SO	FA	FA	FA	KE		solani
02062	Plant	2001.07	IN	3	KE 100	KE 100		SO 100		ND				aff. FA	aff. FA		SO	FA	FA	FA	KE		solani
02474	Cornea	2001.11	SO	2.2	SO 100	SO 100		SO 100		ND				SO	SO	FA	SO	FA	SO	SO	SO	HA	solani
02691	Cornea	2002.04	SO	2.02	FA 99.94	FA 99.94		SO 99.89		ND				FA	SO	FA	SO	FA	FA	FA	FA	PA	solani
00599	Cornea	2003.12	SO	2.02	SO 100	SO 100		SO 100		ND				SO	SO	FA	SO	FA	SO	SO	SO	HA	solani
03284	Eye secretion	2005.02	FA	2.85	FA 100	FA 99.94		SO 99.89		ND				FA	SO	FA	SO	FA	FA	FA	FA	PA	solani
03267	Eye secretion	2005.02	SO	2.85	FA 99.88	FA 99.83		SO 99.89		ND				FA	SO	FA	SO	FA	FA	FA	FA	PA	solani
03320	Eye secretion	2005.04	SO	2.08	FA 99.94	FA 99.88		SO 99.89		ND				FA	SO	FA	SO	FA	FA	FA	FA	PA	solani
04627	Face	2010.12	SO	2.11	SO 100	SO 100		SO 99.83		ND				SO	SO	FA	SO	FA	SO	SO	SO	HA	solani
06673	Nail	2011.12	SO	2.26	KE 100	KE 100		SO 100		ND				aff. FA	aff. FA		SO	FA	FA	FA	KE		solani
07461	Nail	2013.01	SO	2.09	KE 100	KE 99.94		SO 99.95		ND				aff. FA	aff. FA		SO	FA	FA	FA	KE		solani
07905	Nose secretion	2015.03	SO	2.31	SO 100	SO 100		SO 100		ND				SO	SO	FA	SO	FA	SO	SO	SO	HA	solani
08393	Glans secretion	2016.12	SO	2.85	SO 100	SO 100		SO 100		ND				SO	SO	FA	SO	FA	SO	SO	SO	HA	solani
00713	Leg lesion	1999.01	MO	2.00	VE 100	VE 100		VE 99.85	MU 99.46	VE				VE	VE		VE		VE	VE	VE		fujikuroi
00842	Plant	1995.05	AN	2.85	VE 98.52	AN 99.21		AN 99.21		VE				aff. VE	MA		MA		MA	MA	aff. Napiforme	coicis	fujikuroi
00845	Plant	1995.05	VE	1.74	VE 98.52	MA	AN 99.32	AN 99.31		VE				aff. VE	MA		MA		MA	MA	aff. Napiforme	coicis	fujikuroi
01940	Cornea	2001.03	VE	2.00	VE 100	VE 100		VE 99.85	MU 99.36	VE				VE	VE		VE		VE	VE	VE		fujikuroi
01979	Cornea	2001.04	VE	2.01	VE 100	VE 100		VE 99.90	MU 99.41	VE				VE	VE		VE		VE	VE	VE		fujikuroi
01976	Cornea	2001.04	VE	2.07	VE 100	VE 100		VE 99.95	MU 99.46	VE				VE	VE		VE		VE	VE	VE		fujikuroi
02030	Cornea	2001.04	VE	2.30	VE 100	VE 100		VE 99.90	MU 99.41	VE				VE	VE		VE		VE	VE	VE		fujikuroi
02023	Cornea	2001.05	VE	2.40	VE 100	VE 100		VE 99.95	MU 99.46	VE				VE	VE		VE		VE	VE	VE		fujikuroi
02469	Cornea	2001.11	VE	2.07	VE 100	VE 100		VE 99.9	MU 99.41	VE				VE	VE		VE		VE	VE	VE		fujikuroi
00989	Blood	2003.12	VE	1.75	VE 100	VE 100		VE 99.90	MU 99.41	VE				VE	VE		VE		VE	VE	VE		fujikuroi
00963	Blood	2004.01	MO	2.00	VE 100	VE 100		VE 99.90	MU 99.41	VE				VE	VE		VE		VE	VE	VE		fujikuroi
01180	Sputum	2006.07	MO	2.04	VE 99.95	VE 99.95		VE 99.85	MU 99.36	VE				VE	VE		VE		VE	VE	VE		fujikuroi
01341	Sputum	2000.04	SA	2.94	SA 100	SA 100		SA 98.13		SA			SA		SA		SA		SA	SA	SA		fujikuroi
01907	Plant	2000.12	PR	2.13	AN 99.95	PR 100	AN 99.95	AN 99.95	PR 99.80	AN	EL	FU		AN	AN	HE	AN		AN	AN	AN	PR	fujikuroi

Table 1 Continued.

BMU	Source	Date	MALDI-TOF MS	Score value	Fusarioid-ID	Mycobank		GenBank			*CaM*			*RPB1*		*RPB2*			*TEF1*			*TUB2*		Multilocus		Genome		Complex
03134	Face	2004.07	FU	2.85	FU 100	FU 100		FU 99.85		AN	EL	FU		aff. EL	FU		FU		EL	FU	EL	FU		FU		FU		fujikuroi
03319	Eye secretion	2005.04	PR	2.01	AN 100	PR 99.94	FU 99.94	PR 99.7	AN 99.51	AN	EL	FU		AN		AN	HE		AN			AN		AN		AN	PR	fujikuroi
04405	Eye secretion	2007.11	PR	2.21	AN 100	PR 99.94		AN 99.75	PR 99.36	AN	EL	FU		AN		AN	HE		AN			AN		AN		AN	PR	fujikuroi
05348	Shark, peritoneum	2010.08	OX	2.16	AN 100	PR 100	FU 100	PR 99.75	AN 99.56	AN	EL	FU		AN		AN	HE		AN			AN		AN		AN	PR	fujikuroi
05350	Dianthus caryophyllus	2010.08	PR	2.00	AN 100	PR 100	FU 100	PR 99.75	AN 99.56	AN	EL	FU		AN		AN	HE		AN			AN		AN		AN	PR	fujikuroi
05349	Anacardium occidentale, mouldy nut	2010.08	PR	2.06	AN 100	PR 100	FU 100	PR 99.75	AN 99.56	AN	EL	FU		AN		AN	HE		AN			AN		AN		AN	PR	fujikuroi
08009	Urine	2015.09	PR	2.14	AN 100	PR 100	AN 99.94	PR 99.75	AN 99.36	AN	EL	FU		AN		AN	HE		AN			AN		AN		AN	PR	fujikuroi
08905	Nose secretion	2017.05	SO	1.87	AN 100	PR 99.83	AN 99.83	PR 100	AN 99.41	AN	EL	FU		AN		AN	HE		AN			AN		AN		AN	PR	fujikuroi
09013	Skin	2017.07	OX	2.85	NI 99.71	OX 99.77		OX 99.69		PR	LU		aff. DU	GR	ED	NI	CU	OX	NI	CU	CG	DU	NI	OX		OX	VE	oxysporum
02067	Skin	2001.07	AC	2.35	AC 99.72	AC 99.72		AV 97			ND			AC			AC		AC			AC		AC		TR	TR	tricinctum
00844	Plant	1995.05	AR	2.80	AR 99.94	AR 100		AR 99.89			sp nov			AR			aff. AR		AR			aff. AR		AR	CH	AR	CH	sambucinum
03122	Plant	2004.07	SP	2.85	SP 99.94	SP 99.94		SP 100			sp nov			sp nov			sp nov		sp nov near AR			sp nov near AR		SP		SP	SI	sambucinum
01974	Cornea	2001.04	IN	1.74	NA 100	NA 100		PE 98.82	LF 98.73		NA		LF		PE		NA		NA			NA		NA	HA	NA		incarnatum-equiseti
01904	Plant	2000.12	RE	2.85	RE 99.89	RE 99.89		RE 99.56			sp nov			sp nov			sp nov		sp nov			sp nov		sp nov		RE		redolens
	SO solani	solani			GR grosmicheli	oxysporum																						
	FA falciforme	solani			NI nirenbergiae	oxysporum																						
	HA haematococcum	solani			ED elaeidis	oxysporum																						
	PA paranaense	solani			CG cugenangense	oxysporum																						
	KE keratoplasticum	solani			CU curvatum	oxysporum																						
	EL elaeagnei	fujikuroi			VE veterinarium	oxysporum																						
	MO moniliforme	fujikuroi			AC acuminatum	tricinctum																						
	AN andyyazi	fujikuroi			AV avenaceum	tricinctum																						
	FU fujikuroi	fujikuroi			TR tricinctum	tricinctum																						
	SA sacchari	fujikuroi			AR armeniacum	sambucinum																						
	AN annulatum	fujikuroi			CH chaquense	sambucinum																						
	PR proliferatum	fujikuroi			SP sporotrichoides	sambucinum																						
	MA madaense	fujikuroi			SI sibiricum	sambucinum																						
	MU musae	fujikuroi			IN incarnatum	incarnatum-equiseti																						
	VE verticillioides	fujikuroi			NA nanum	incarnatum-equiseti																						
	HE hechiense	fujikuroi			PE pernambucanum	incarnatum-equiseti																						
	OX oxysporum	oxysporum			LF luffae	incarnatum-equiseti																						
	LU lumajangense	oxysporum			HA hainanense	incarnatum-equiseti																						
	DU duoseptatum	oxysporum			RE redolens	redolens																						

Table 2 Multiple alignment information of 48 clinical strains.

Species Complex	Gene	Sequences	Columns	Distinct patterns	Parsimony-informative	Singleton sites	Constant sites	Best fit model
All Fusarium	cam	643	850	591	438	45	367	TIMe+I+G4
All Fusarium	rpb1	1153	1892	1217	810	195	887	TIM3e+R4
All Fusarium	rpb2	1266	2241	1350	897	206	1137	TIM2e+R5
All Fusarium	tub2	819	1042	576	372	57	613	TNe+I+G4
All Fusarium	tef1	1418	876	694	459	41	376	TIM2e+I+G4
All Fusarium	rpb2-tub2-tef1	657	4159	2156	1600	249	2310	TIM2e+R7
All Fusarium	WGS (Protein)	225	490050	254724	201067	64873	224110	JTT+F+R9
Species Complex	**Gene**	**Sequences**	**Columns**	**Distinct patterns**	**Parsimony-informative**	**Singleton sites**	**Constant sites**	**Best fit model**
fujikuroi	rpb2-tub2-tef1	248	3456	1192	883	217	2356	TIM2e+R3
incarnatum-equiseti	rpb2-tub2-tef1	95	3553	766	566	379	2608	TNe+I+G4
oxysporum	rpb2-tub2-tef1	37	3442	172	71	389	2982	TNe+G4
redolens	rpb2-tub2-tef1	7	3366	67	61	333	2972	TNe+G4
sambucinum	rpb2-tub2-tef1	111	3473	956	808	221	2444	TIM2e+I+G4
solani	rpb2-tub2-tef1	76	3739	871	622	347	2770	TNe+I+G4
tricinctum	rpb2-tub2-tef1	58	3396	359	306	380	2710	TNe+R2
Species Complex	**Gene**	**Sequences**	**Columns**	**Distinct patterns**	**Parsimony-informative**	**Singleton sites**	**Constant sites**	**Best fit model**
fujikuroi	WGS (Protein)	158	1297785	289299	212573	179681	905531	JTT+F+R9
incarnatum-equiseti	WGS (Protein)	41	2996098	240893	229411	377099	2389588	JTT+F+R10
oxysporum	WGS (Protein)	636	422299	167188	43723	77906	300670	JTT+F+R9
redolens	WGS (Protein)	9	4169282	43490	112963	441835	3614484	JTT+F+R10
sambucinum	WGS (Protein)	217	1242486	225685	223417	88898	930171	JTT+F+R9
solani	WGS (Protein)	72	906419	194771	154273	120256	631890	JTT+F+R9
tricinctum	WGS (Protein)	19	3563377	146688	166919	464082	2932376	JTT+F+R6

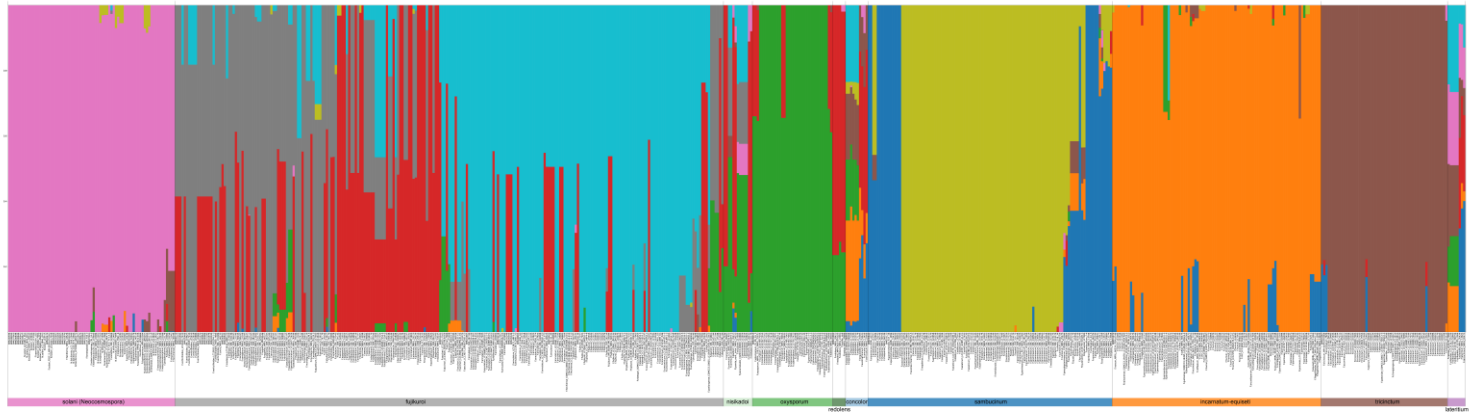

Figure 4 – Admixture data of the sequences shown in Fig. 3, based on K = 9.

Table 3 The genome data of 48 sequenced clinical strains with genome size, general G+C% and number of genes.

BMU	Multilocus	Complex	GC%	Genome size	N50	Coding genes
00841	solani	solani	51.37%	53,968,016	3,522,534	17,354
00846	solani	solani	51.63%	52,435,658	3,553,269	16,367
00839	solani	solani	50.72%	54,246,004	3,657,257	16,955
01264	falciforme	solani	50.02%	56,463,235	3,721,890	16,773
01263	falciforme	solani	51.60%	52,894,178	3,919,959	16,234
01467	falciforme	solani	51.71%	48,701,983	4,065,520	15,601
01970	falciforme	solani	51.54%	52,455,765	3,774,166	16,047
02025	falciforme	solani	51.57%	49,020,420	4,096,989	15,511
02062	falciforme	solani	51.95%	48,452,153	3,994,211	15,549
02474	solani	solani	50.00%	50,275,909	3,665,492	15,505
02691	falciforme	solani	51.25%	52,698,965	3,643,246	16,084
00599	solani	solani	51.75%	49,378,219	3,988,639	15,536
03284	falciforme	solani	51.18%	55,159,954	3,607,359	16,644
03267	falciforme	solani	51.60%	52,566,758	3,844,380	16,129
03320	falciforme	solani	51.41%	53,797,822	3,696,477	16,426
04627	solani	solani	62.27%	44,666,392	3,886,261	13,971
06673	falciforme	solani	51.80%	48,494,125	4,087,573	15,475
07461	falciforme	solani	52.02%	46,954,771	3,986,735	15,168
07905	solani	solani	51.68%	51,603,130	3,559,834	16,095
08393	solani	solani	55.31%	48,812,558	3,984,110	15,258
00713	verticillioides	fujikuroi	48.16%	43,113,462	4,304,590	14,043
00842	madaense	fujikuroi	48.46%	44,172,401	4,141,742	14,313
00845	madaense	fujikuroi	48.53%	42,799,580	4,280,362	14,150
01940	verticillioides	fujikuroi	48.50%	42,936,860	4,312,927	14,284
01979	verticillioides	fujikuroi	48.40%	42,972,441	3,923,897	14,094
01976	verticillioides	fujikuroi	48.41%	42,859,312	4,646,350	14,005
02030	verticillioides	fujikuroi	48.31%	43,136,262	3,821,591	14,060
02023	verticillioides	fujikuroi	48.29%	43,339,225	4,432,423	14,064
02469	verticillioides	fujikuroi	48.46%	43,068,175	4,303,657	14,027
00989	verticillioides	fujikuroi	54.69%	43,159,363	4,327,072	14,100
00963	verticillioides	fujikuroi	48.56%	42,493,755	4,244,863	13,980
01180	verticillioides	fujikuroi	60.19%	42,974,706	4,080,395	13,842
01341	sacchari	fujikuroi	47.85%	44,441,320	4,271,417	14,158
01907	annulatum	fujikuroi	48.18%	43,962,785	4,427,677	14,733
03134	fujikuroi	fujikuroi	48.00%	43,709,752	4,327,210	14,625
03319	annulatum	fujikuroi	48.46%	42,221,902	3,700,155	14,037
04405	annulatum	fujikuroi	48.63%	44,241,276	4,364,568	14,776
05348	annulatum	fujikuroi	48.43%	56,780,343	3,458,102	18,343
05350	annulatum	fujikuroi	48.50%	44,385,500	4,411,197	14,748
05349	annulatum	fujikuroi	48.64%	43,627,138	4,370,227	14,637
08009	annulatum	fujikuroi	48.14%	44,582,039	4,349,402	14,864
08905	annulatum	fujikuroi	48.38%	43,298,022	4,102,489	14,694
09013	oxysporum	oxysporum	48.33%	48,542,842	3,122,851	15,205
02067	acuminatum	tricinctum	47.85%	47,846,382	4,273,331	14,256
00844	armeniacum, chaquense	sambucinum	47.81%	37,341,153	4,291,928	11,858
03122	sporotrichoides	sambucinum	48.06%	37,094,910	4,101,632	11,851
01974	nanum, hainanense	incarnatum-equiseti	47.85%	37,440,386	3,321,582	12,095
01904	sp nov	undetermined	47.99%	52,058,437	2,866,670	15,975

Representation of clinical strains in species complexes

With all identification methods, most strains in our dataset belonged to either the *Fusarium fujikuroi* or the *F. solani* species complexes, even though the affiliation to lineages was often ambiguous, as shown above. The *F. fujikuroi* and *F. solani* species complexes contain more species represented in databases; thus, strains of our dataset have a larger chance to cluster with any of these lineages. Normalised data by taking the numbers of hits as percentages of the total number of

species in the respective SC show that the *F. incarnatum-equiseti*, *F. oxysporum* and *F. sambucinum* species complexes are under-represented (Fig. 1). The data from *Atlas* deviated clearly in having a preponderance of members of the *F. fujikuroi* complex, while the *F. sambicinum* is more environmental; however, the applied numbers of strains *versus* species in this equation are minimal.

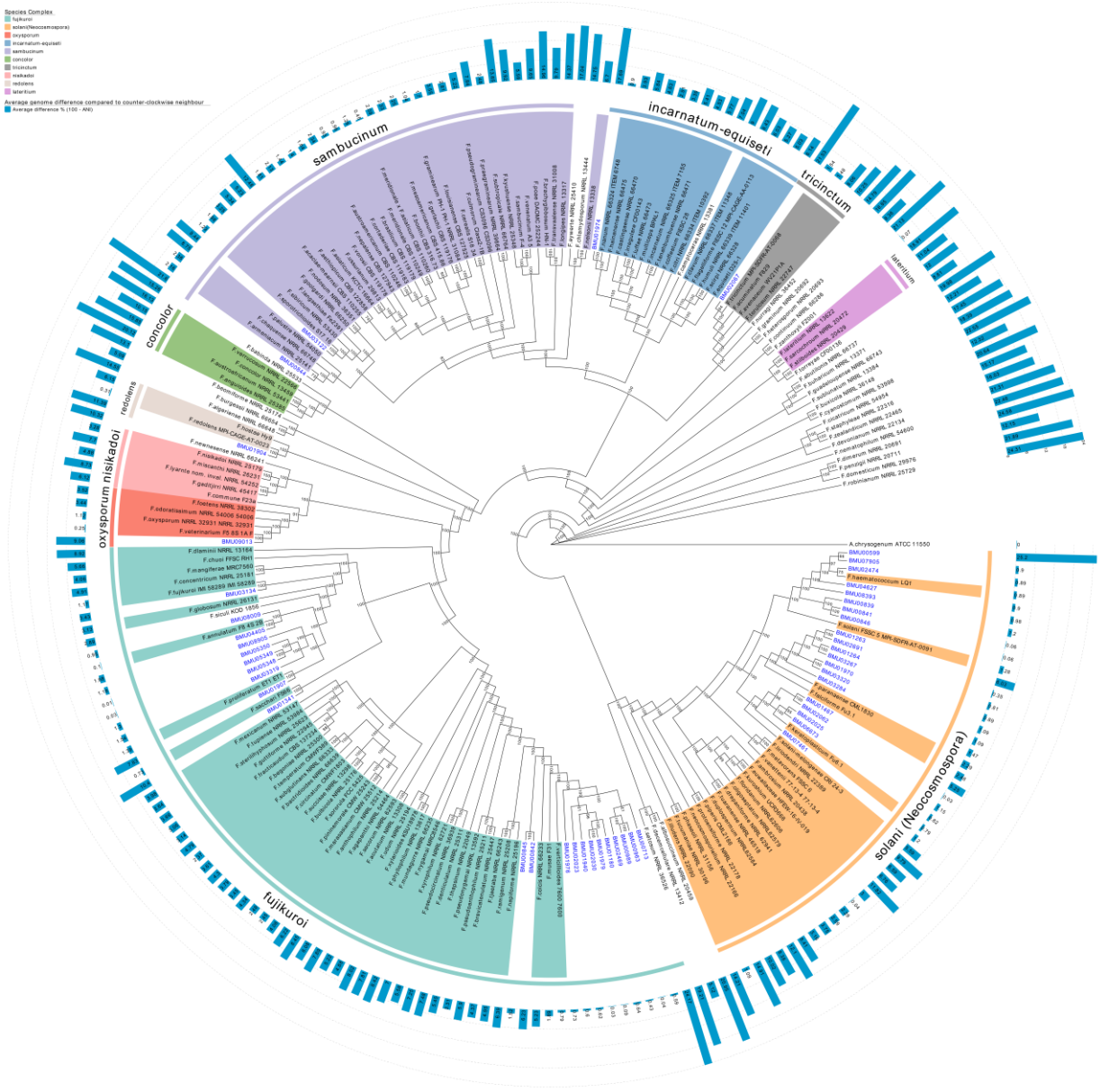

Figure 5 – Phylogenetic tree of 48 clinical strains compared with 176 genomes downloaded from GenBank. Newly sequenced genomes of clinical strains have blue font; genomes for comparison are in black font. Species complexes are differentially coloured. Some of the names do not appear in the multilocus data. Bars in the outermost row represent the percentage of SNPs compared to the neighbouring strain.

Discussion

In our study, the dataset from different hospitals and using prescribed multilocus data, strains were usually not strictly identical to (type strains of) described species. Therefore, the question arose whether there were novel species among our set of strains. We compared our data to a recent publication on Chinese environmental fusaria (Wang et al. 2022), but the trees in this paper contained only type strains, and no ranges of intraspecific variability were provided. Han et al. (2023) recently showed large amounts of isolates for some species but did not provide a cut-off for

species delimitation. For this reason, we compared our identification results of clinical strains in the framework of a taxonomic phylogenetic study.

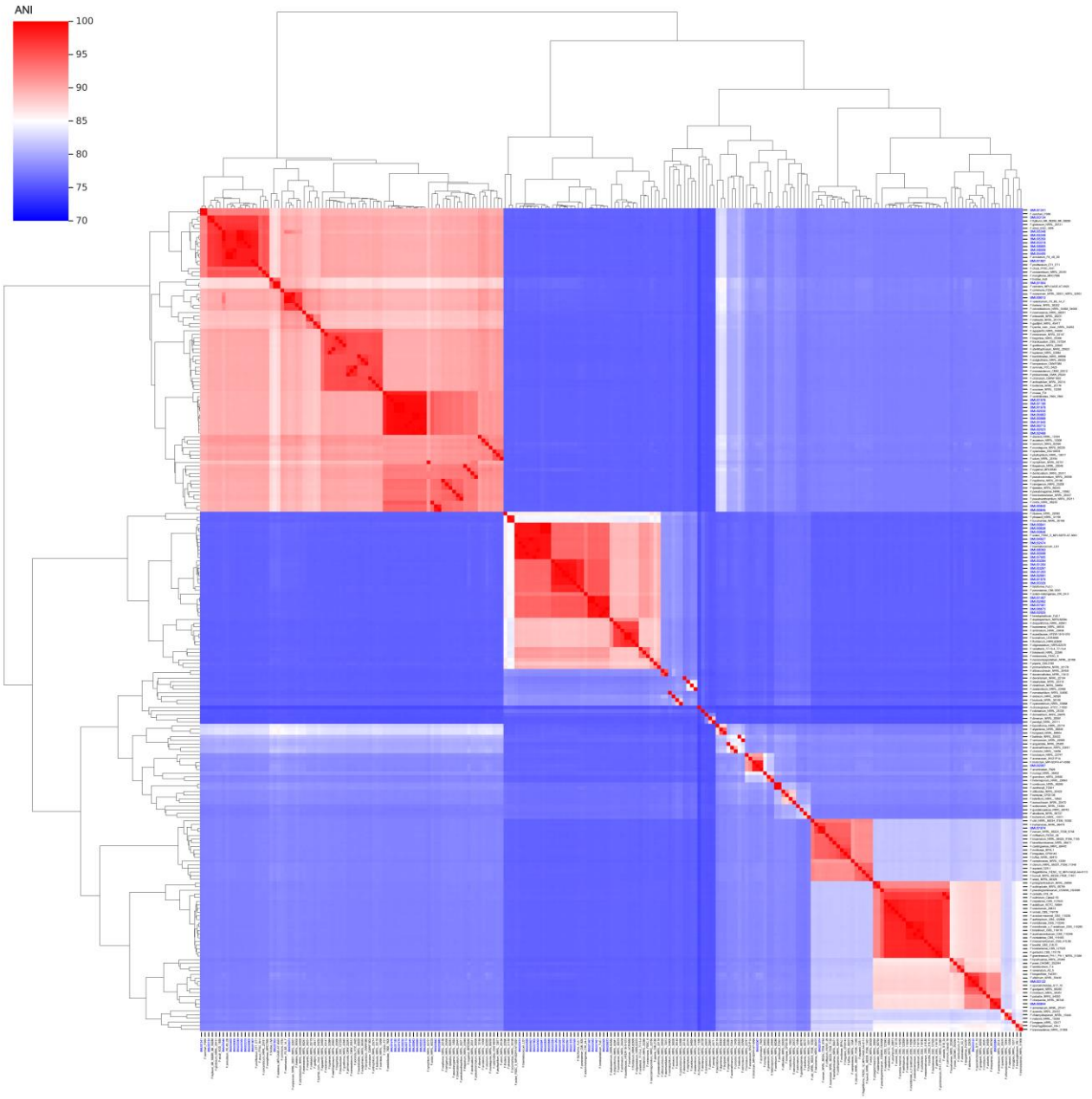

Figure 6 – The high similarity of the clinical strains within each of these complexes is illustrated by an ANI comparison.

Identity at species complex level

Given the human-infectious potential of members of the genus *Fusarium*, classically being known to cause host-specific plant diseases, its trans-kingdom pathogenicity is remarkable and rather unique among the fungi (van Diepeningen & de Hoog 2016). The question remains whether animal pathology is comparable to plant pathology, which functions through the horizontal transfer of dispensable chromosomes (Ayukawa et al. 2021, Plaumann et al. 2018), changing beneficial to virulent *in planta* behaviour by acquiring effector genes (Constantin et al. 2021).

Fusarium is classified in the family *Nectriaceae*, which contains many morphologically similar genera (Crous et al. 2021). Clinically relevant species are found nearly exclusively in *Fusarium*, including the *F. solani* complex, also known as the separate genus *Neocosmospora*. As

noted previously (Brenes Guallar et al. 2022), a single strain of *F. oxysporum* f. sp. *capsici* GCA_014770115 was misidentified in the reference dataset, as it clustered with *F. solani* with all genes analysed. The high similarity of the clinical strains within each of these complexes is illustrated by an ANI comparison (Fig. 6). The genera are separated as supported clades (Crous et al. 2021), but as there is no operational criterion to determine the size of a genus, the adoption of *Neocosmospora versus Fusarium* is determined by practical reasons; in this paper, we maintain the broad concept of *Fusarium* until the taxonomic dispute has been resolved. The larger species complexes in *Fusarium/Neocosmospora* are *F. fujikuroi*, *F. incarnatum-equiseti*, *F. oxysporum*, *F. solani* and *F. sambucinum* SC, while the remaining ones contain relatively few, rather uncommon species and thus have a smaller chance to be sampled. In the general overview of *Fusarium* diversity published by Crous et al. (2021) and of Chinese environmental strains by Wang et al. (2022), species complexes are represented at comparable frequencies, except for less recovered species in *F. solani* SC by Wang et al. (2022). Our dataset of hospital strains from superficial and deeper locations deviates from this profile by being nearly limited to *F. fujikuroi* and *F. solani* complexes, representing 92.1% of the total. Seventeen strains belonged to six remaining complexes, of which *F. oxysporum* SC was represented with nine isolates. A preponderance of these complexes in clinical settings has been observed previously (Bansal et al. 2019). The *F. solani* complex is prevalent in most studies, but the presence of other complexes is highly variable (e.g., van Diepeningen et al. 2014, Migheli et al. 2010). Although without statistical confidence, the *Atlas of Clinical Fungi* (de Hoog et al. 2020), describing fusaria from proven case reports globally, contains a preponderance of 82.8% of species in the *F. solani* and *F. fujikuroi* complexes (Fig. 1). The distinction high and low virulence thus does not match with the separation of *Neocosmospora versus Fusarium s. str.*

The distribution of named species over established species complexes was verified by admixture, revealing the ancestry of strains via maximum likelihood estimation. The number of groups (K) was set according to the published affiliation of species to species complexes to which the isolates included in the tree belong. In Fig. 4, K was nine, as species included were published to belong to *F. fujikuroi* (FFSC), *F. nisikadoi* (FNSC), *F. concolor* (FCSC), *F. oxysporum* (FOSC), *F. sambucinum* (FSAMSC), *F. incarnatum-equiseti* (FIESC), *F. lateritium* (FLSC), *F. tricinctum* (FTSC) and *F. solani* (FSSC/*Neocosmospora*) complexes. From this data, the FFSC, FNSC and FOSC were displayed as well-delimited groups. In contrast, the *Neocosmospora* was divided into two clearly separable entities, matching with two main clades in *Neocosmospora* visible in the data presented by Wang et al. (2020). We also analysed barcoding gaps between species complexes by accumulation of mutations (SNPs and Indels) in each strain of the multilocus tree compared to one of its neighbours (Fig. 3), with the expectation that these numbers are low within a complex and high when another species complex is met (i.e., the barcoding gap). No clear gap could be determined between FFSC, FNSC and FOSC.

Identity at the lineage level

Our study concerns fusarium-like strains collected from clinical cases over 22 years in various hospitals in China and describes our attempts to achieve reliable identification of the strains according to modern standards. While the affiliation to species complexes is unproblematic, exact species identification often yielded ambiguous results. Routine identification by BLAST of barcoding genes in Fusarium-ID or MycoBank databases usually provides a close match (>99% identity) with any species. The level of confidence in this identification can be verified with alternative methods and should yield the same outcome. As the foundation of diagnostic databases is in phylogenetic analysis, we also identified our strains via this route. However, the high diversity in *Fusarium* and the absence of delimitation criteria between clades make establishing taxon borderlines difficult. In the literature, this is solved by the description of numerous taxa (Wang et al. 2022), whereby classically recognised intraspecific differences such as varieties or *formae speciales* are no longer considered. Sarver et al. (2011) performed genealogical concordance based on 12 genes, yielding high bootstrap support of host-specific entities. In a similar study, Laurence

et al. (2011) identified independent evolutionary lineages, which were subsequently collapsed into phylogenetic species. A concept of host-specific, genetically separated lineages within larger species complexes matches the preponderant mode of plant pathogens by dispensable genomes containing effector genes (Constantin et al. 2021).

Identification of the individual strains was performed with the aid of sequences of *cam* (excluding *F. solani* complex), *rpb1*, *rpb2*, *tef1-α*, and *tub2* (Table 1). Of these partial genes, *cam* had the lowest resolution and *tub2* the highest. Some authors suggested that *tef1-α* and *rpb2* would be sufficient for diagnostics (James et al. 2022). BLAST identification with four concatenated barcoding genes, using the <1% difference criterion, indicates identity, and different databases (Fusarium ID, MycoBank and GenBank) sometimes yielded contradictory results. Particularly, strains clustering in the *F. fujikuroi* complex were difficult to identify, matching with several species using the <1% criterion, and if a single name, this was sometimes not the same name between the different databases (Table 1). For example, the name *F. moniliformis* is regarded as obsolete by most workers but appeared in MALDI-ToF identification.

Depending on the degree of resolution of the used genes, single-gene clustering was insufficient. Taking the Fusarioid-ID database for BLASt and multilocus clustering as the identification methods of choice, 12/49 (24.5%) resulted in a different name. Remarkably, of the 75.5% corresponding names, only 12 isolates yielded unambiguous identification (*F. fujikuroi*, *F. sacchari* and *F. verticillioides*) when clustered on the basis of the whole genome (Fig. 5). Only a single strain, BMU 01341 of *F. sacchari*, yielded exactly the same unambiguous result with all methods applied.

The standard approach of identification by BLAST using barcoding genes is relatively simple and cannot be expected to be 100% reliable. In daily practice, this is not noted when no other identification methods are applied to verify the results. The present attempt to identify *Fusarium* strains with various available methods demonstrates that results may be confusing, as several described species (e.g. *F. annulatum*, *F. fujikuroi* and *F. proliferatum* in MycoBank) have 99.94−100% identity in the used four barcoding genes. Maintaining different species names for extremely similar strains is detrimental to patient care, where consulting published literature on the identified etiologic agents is an essential part of management. To determine whether our clinical strains eventually represented undescribed species, we used genomic ANI (Fig. 6). In our dataset of 657 sequences, this value varied between 94.99 and 100% (mostly 98.5−99.5%), and for the clades containing clinical strains, between 97.54 and 100%. The aforementioned clonal outbreak of corneal strains belonged to the *F. verticillioides* clades at ANI = 98.62 but clustered in six subclades (Table 4). The commonly encountered ANI value of 98.5% appears close to clonal variability. The lower ANI value of 97.54% is found in the cluster containing BMU 02067 containing strains identified as *F. acuminatum*. Within this clade, BMU is 98.59% similar to strain MPI-SDFR-AT-0068, identified as *F. tricinctum*. In the heatmap of ANI data (Fig. 6), strains show clustering at high similarity values: e.g. *F. falciforme* FU3.1 and *F. paranaense* CML1930. Given these high degrees of similarity between some of the described species, and their apparent difficulty to identify them consistently in routine diagnostics, we judge that some areas of *Fusarium* seem to be overclassified.

The appearance of several names for a single strain within the detection limit has also been reported outside *Fusarium*, e.g. in *Diaporthe* using GenBank's RefSeq database (Cabeza et al. 2023). Different names appearing when different datasets are compared indicate the incompleteness of databases. Thomas et al. (2019) recommended using the tef1-α, which in our data was indeed one of the genes with the best resolution, but it was not able to identify all isolates, and outcomes may conflict with other identification methods. The application of alternative methods such as MALDI-ToF encounters difficulties in species-rich complexes such as *F. fujikuroi* or *F. oxysporum* (Song et al. 2021); Al-Hatmi et al. (2022) yielded an identification of 93.6% of the strains under study. In conclusion, 100% reliable identification of clinical *Fusarium* species down to the molecular sibling level remains difficult for some clinical strains, even when sufficiently advanced methodology such as multilocus sequencing with recommended genes or whole genome

sequencing is applied. The frequent strict identity of multiple barcoding genes with several described species might indicate that locally too narrow species concepts are maintained.

Table 4 Intraspecific variability of *Fusarium verticillioides* clade at ANI 98.62.

Subcluster	BS	Strain	Source
1	100	CBS 131389	Environment
		7600	Environment
2	100	LC 13655	Environment
		BMU 02030	Corneal outbreak 2004
		BMU 01979	Corneal outbreak 2004
		BMU 01180	Sputum
		BMU 00989	Blood
		BMU 00963	Blood
3	50	BMU 01976	Corneal outbreak 2004
4	83	NRRL 20984	Environment
		LC 5896	Environment
		LC 2818	Environment
		LC 2810	Environment
		LC 13653	Environment
		CBS 119827	Environment
5	94	BMU 02469	Cornea 2011
		LC 13654	Environment
		BMU 01940	Corneal outbreak 2004
6	100	BMU 02023	Corneal outbreak 2004
		BMU 00713	Leg lesion

Funding information

The authors are grateful for the financial support of the National Natural Science Foundation of China (NSFC No. 82272354) and the Key Technologies Research and Development Program (No. 2021YFC2302005).

Acknowledgements

Prof. Martijn Rep (UvA, Amsterdam) is thanked for constructive discussions.

Author contributions

YS and SG designed the experiments and supervised the data analysis. YS, SH, XL, RX, XM performed the experiments and wrote the relevant portions of the manuscript; LF made significant contributions to the text. SG, SH, AD and YS analysed the data. All authors discussed the results and commented on the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found here: NCBI PRJNA961359.

ORCID:
Yinggai Song, https://orcid.org/0000-0002-1644-6942
Shenghan Gao, https://orcid.org/0000-0003-4280-4915
Sybren de Hoog, https://orcid.org/0000-0002-5344-257X
Ruoyu Li, https://orcid.org/0000-0003-1451-3598
Anne D. van Diepeningen, https://orcid.org/0000-0002-6373-8382

Like Fokkens, https://orcid.org/0000-0001-6696-4409
Songnian Hu, https://orcid.org/0000-0003-3966-3111

References

Al-Hatmi AMS, de Hoog GS, Meis JF. 2019 – Multiresistant *Fusarium* pathogens on plants and humans: solutions in (from) the antifungal pipeline? Infect Drug Resist 12, 3727–3737.

Al-Hatmi AMS, Normand AC, van Diepeningen AD, Hendrickx M et al. 2022 – Rapid identification of clinical members of *Fusarium fujikuroi* complex using MALDI-TOF MS. J Fungi 8. Doi 10.3390/jof8080845

Ayukawa Y, Asai S, Gan P, Tsushima A et al. 2021 – A pair of effectors encoded on a conditionally dispensable chromosome of *Fusarium oxysporum* suppress host-specific immunity. Commun Biol 4, 707.

Bansal Y, Sing N, Kaistha N, Sood S, Chander J. 2019 – Molecular identification of *Fusarium* species complex isolated from clinical samples and its antifungal susceptibility patterns. Curr Med Mycol 5(4), 43–49. Doi 10.18502/cmm.5.4.2149

Bassiri-Jahromi S, Khaksar AA. 2010 – Nondermatophytic moulds as a causative agent of onychomycosis in Tehran. Indian J Dermatol 55, 140–143.

Brasch J, Shimanovich I. 2012 – Persistent fingernail onychomycosis caused by *Fusarium proliferatum* in a healthy woman. Mycoses 55, 86–89.

Brenes Guallar MA, Fokkens L, Rep M, Berke L, van Dam P. 2022 – *Fusarium oxysporum* effector clustering version 2: An updated pipeline to infer host range. Front Plant Sci 13, 1016288. Doi 10.3389/fpls.2022.1012688

Cafarchia C, Paradies R, Figueredo LA, Iatta R et al. 2020 – *Fusarium* spp. in loggerhead sea turtles *Caretta caretta* from colonisation to infection. Vet Pathol 57, 139–146.

Castro Lopez N, Casas C, Sopo L, Rojas A et al. 2008 – *Fusarium* species detected in onychomycosis in Colombia. Mycoses 52, 350–356.

Constantin ME, Fokkens L, de Sain M, Takken FLW et al. 2021 – Number of effector genes in accessory genomes differentiates pathogenic from endophytic *Fusarium oxysporum* genes. Front Plant Sci 12, 761740. Doi 10.3389/fpls.2021.761740

Crous PW, Lombard L, Sandoval-Denis M, Seifert KA et al. 2021 – *Fusarium*: more than a node or a foot-shaped basal cell. Stud Mycol 98, 100116.

Gaviria-Rivera A, Giraldo-López A, Santa-Cardona C, Cano-Restrepo L. 2018 – Molecular identification of clinical isolates of *Fusarium* in Colombia. Revta Salud Públ 20, 94–102. Doi 10.15446/rsap.v20n1.51923

Han SL, Wang MM, Ma ZY, Raza M et al. 2023 – Fusarium diversity associated with diseased cereals in China, with an updated phylogenomic assessment of the genus. Stud Mycol 104, 87–148.

Homa M, Galgóczy L, Manikandan P, Narendran V et al. 2018 – South Indian isolates of the *Fusarium solani* species complex from clinical and environmental samples: identification, antifungal susceptibilities, and virulence. Front Microbiol 9, 1052.

James JE, Santhanam J, Zakaria L, Mamat Rusli N et al. 2022 – Morphology, phenotype, and molecular identification of clinical and environmental *Fusarium solani* species complex isolates from Malaysia. J Fungi 8. Doi 10.3390/jof8080845

Jiang Y, Al-Hatmi AMS, Xiang Y, Cao Y et al. 2016 – The concept of ecthyma gangrenosum illustrated by a *Fusarium oxysporum* infection in an immunocompetent individual. Mycopathologia 181, 759–763.

Laurence MH, Summerell BA, Burgess LW, Liew ECY. 2014 – Genealogical concordance phylogenetic species recognition in the *Fusarium oxysporum* species complex. Fung Biol 118, 374–384.

Liu YS, Wang NC, Ye RH, Kao WY. 2014 – Disseminated *Fusarium* infection in a patient with acute lymphoblastic leukemia: A case report and review of the literature. Oncol Lett 7, 334–336. Doi 10.3892/ol.2013.1738

McTaggart AR, James TY, Shivas RG, Drenth A et al. 2021 – Population genomics reveals historical and ongoing recombination in the *Fusarium oxysporum* species complex. Stud Mycol 99, 100132.

Mehl HL, Epstein L. 2008 – Sewage and community shower drains are environmental reservoirs of *Fusarium solani* species complex group 1, a human and plant pathogen. Envir Microbiol 10, 219–227. Doi 10.1111/j.1462-2920.2007.01446.x

Meza-Menchaca T, Singh RK, Quiroz-Chávez J, García-Pérez LM et al. 2020 – First demonstration of clinical *Fusarium* strains causing cross-Kingdom infections from humans to plants. Microorganisms 8(6), 947. Doi 10.3390/microorganisms8060947

Migheli Q, Balmas V, Harak H, Sanna S et al. 2010 – Molecular phylogenetic diversity of dermatologic and other human pathogenic fusarial isolates from hospitals in Northern and Central Italy. J Clin Microbiol 48, 1076–1084.

Muhammed M, Anagnostou T, Desalermos A, Kourkoumpetis TK et al. 2013 – *Fusarium* infection. Report of 26 cases and review of 97 cases from the literature. Medicine (Baltimore) 92, 305–316.

Ninet B, Jan I, Bontems O, Léchenne B et al. 2005 – Molecular identification of *Fusarium* species in onychomycoses. Dermatology 210, 21–25.

Plaumann PL, Schmidpeter J, Dahl M, Taher L et al. 2018 – A dispensable chromosome is required for virulence in the hemibiotrophic plant pathogen *Colletotrichum higginsianum*. Front Microbiol 9, 1005. Doi 10.3389/fmicb.2018.01005

Ranawaka RR, Nagahawatte A, Gunasekara TA. 2015 – *Fusarium* onychomycosis: prevalence, clinical presentations, response to itraconazole and terbinafine pulse therapy, and 1-year follow-up in nine cases. Int J Dermatol 54, 1275–1282. Doi 10.1111/ijd.12906

Sarver BAJ, Ward TJ, Gale LR, Broz K et al. 2011 – Novel *Fusarium* head blight pathogens from Nepal and Louisiana revealed by multilocus genealogical concordance. Fung Genet Biol 48, 1096–1107.

Segorbe D, Di Pietro A, Pérez-Nada E, Turra D. 2017 – Three *Fusarium oxysporum* mitogen-activated protein kinases (MAPKs) have distinct and complementary roles in stress adaptation and cross-kingdom pathogenicity. Molec Plant Pathol 18, 912–920.

Satpathy G, Ahmed NH, Nayak N, Tandon R et al. 2019 – Spectrum of mycotic keratitis in north India: Sixteen years study from a tertiary care ophthalmic centre. J Infect Publ Health 12, 367–371.

Sautour M, del-Hermann VE, Steinberg C, Sixt N et al. 2012 – *Fusarium* species recovered from the water distribution system of a French university hospital. Int J Hyg Environm Health 215, 286–292.

Szaliński M, Zgryźniak A, Rubisz I, Gajdzis M et al. 2021 – *Fusarium* keratitis – Review of current treatment possibilities. J Clin Med 10(23), 5468. Doi 10.3390/jcm10235468

Song Y, Liu X, Yang Z, Meng X et al. 2021 – Molecular and MALDI-ToF MS differentiation and antifungal susceptibility of prevalent clinical Fusarium species in China. Mycoses 64, 1261-1271.

Thomas B, Contet Audonneau N, Machouart M, Debourgogne A. 2019 – Molecular identification of *Fusarium* species complexes: Which gene and which database to choose in clinical practice? J Mycol Méd 29, 56–58.

van Diepeningen AD, de Hoog GS. 2016 – Challenges in *Fusarium*, a trans-kingdom pathogen. Mycopathologia 181, 161–163. Doi 10.1007/s11046-016-9993-7

van Diepeningen AD, Feng P, Ahmed S, Sudhadham M et al. 2015 – Spectrum of *Fusarium* infections in tropical dermatology evidenced by multilocus sequencing typing diagnostics. Mycoses 58, 48–57.

Vlaardingerbroek I, Beerens B, Schmidt SM, Cornelissen BJ et al. 2016 – Dispensable chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. Molec Plant Pathol 17, 1455–1466.
Walther G, Stasch S, Kaerger K, Hamprecht A et al. 2017 − *Fusarium* keratitis in Germany. J Clin Microbiol 55, 2983–2995. Doi 10.1128/JCM.00649-17
Waskom ML. 2021 – seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021. Doi 10.21105/joss.03021
Wang MM, Crous PW, Sandoval-Denis M et al. 2022 – *Fusarium* and allied genera from China: Species diversity and distribution. Persoonia 48, 1–53.
Yilmaz N, Sandoval-Denis M, Lombard M, Visagie CM et al. 2021 – Redefining species limits in the Fusarium fujikuroi species complex. Persoonia 46, 129–162.
Yu J, Chen Y, Fang J, Zhang K. 2019 – Successful treatment of disseminated fusariosis in a patient with acute lymphoblastic leukemia. A case report and literature review. Medicine, Baltimore 98(26), e16246.
Zhao L, de Hoog GS, Hagen F, Kang Y, Al-Hatmi AMS. 2019 – Species borderlines in Fusarium exemplified by *F. circinatum*/*F. subglutinans*. Fung Genet Biol 132, 103262.

Supplementary materials

Supplementary Table 1 *Cam* sequences used for comparison.
Supplementary Table 1

Supplementary Table 2 *Rpb1* sequences used for comparison.
Supplementary Table 2

Supplementary Table 3 *Rpb2* sequences used for comparison.
Supplemeatary Table 3

Supplementary Table 4 *Tef-1α* sequences used for comparison.
Supplementary Table 4

Supplementary Table 5 *Tub2* sequences used for comparison.
Supplementary Table 5

Supplementary Table 6 List of sequences from Wang et al. (2022) used for comparison.
Supplematary Table 6

Supplementary Table 7 List of genomes from NCBI used for comparison.
Supplementary Table 7

Supplementary Table 8 Fusarium genome assemble.
Supplementary Table 8

Supplementary Table 9 Fusarium genome summary.
Supplementary Table 9

Supplementary Table 10 Fusarium representative genomes.
Supplementary Table 10