



OPEN

In silico comparative genomic analysis unravels a new candidate protein arsenal specifically associated with *Fusarium oxysporum* f. sp. *albedinis* pathogenesis

Hafida Ayada^{1✉}, Boutayna Dhioui¹, Hamid Mazouz¹, Abdelhay El harrak¹, Fatima Jaiti², Bouchra Ouhmidou³, Mohammed Diouri¹ & Mohieddine Moumni^{1✉}

Fusarium oxysporum f. sp. *albedinis* (Foa) is a devastating fungus of date palms. To unravel the genetic characteristics associated with its pathogenesis, the two available genomes of Foa 133 and Foa 9 were compared with 49 genomes of 29 other pathogenic *formae speciales* belonging to *Fusarium oxysporum* species complex (FOSC). Foa 133 and Foa 9 have genomes of 56.23 Mb and 65.56 Mb with 17460 and 19514 putative coding genes. Of these genes, 30% lack functional annotation with no similarity to characterized proteins. The remaining genes were involved in pathways essential to the fungi's life and their adaptation. Foa secretome analysis revealed that both Foa strains possess an expanded number of secreted effectors (3003 in Foa 133 and 2418 in Foa 9). Those include effectors encoded by Foa unique genes that are involved in Foa penetration (Egh16-like family), host defense mechanisms suppression (lysM family) and pathogen protection (cysteine-rich protein family). The accessory protein SIX6, which induces plant cell death, was also predicted in Foa. Further analysis of secreted CAZymes revealed an arsenal of enzymes involved in plant cell wall degradation. This arsenal includes an exclusively Foa-specific CAZyme (GH5-7). Transcription factors and membrane transporters (MFS) involved in fungicide efflux have been predicted in Foa, in addition to a variety of secondary metabolites. These comprise mycotoxins as well as chrysogin, the latter provides Foa with resistance against adverse environmental conditions. Our results revealed new Foa proteins that could be targeted in future research in order to manage Bayoud disease.

Fusarium oxysporum species complex (FOSC) is a ubiquitous group of pathogenic and putatively non-pathogenic soil-borne fungi. It is the most widespread in nature, it colonizes all soil types (cultivated and uncultivated soils) in all continents except Antarctica^{1,11}. Among the FOSC, plant pathogenic fungi are known for causing significant ecological and socio-economic damage. In the agricultural field, Fusarium wilt caused by pathogenic fungi belonging to FOSC represents a veritable threat to production and profitability. The *formae speciales* of this species complex attack a multitude of crops, such as legumes (*Fusarium oxysporum* f. sp. *pisi* on peas)¹, horticultural plants (*Fusarium oxysporum* f. sp. *lycopersici*)⁴, ornamental plants (*Fusarium oxysporum* f. sp. *dianthi* on carnations)⁵ and palm trees (*Fusarium oxysporum* f. sp. *albedinis*)⁶. On date palm, *Fusarium oxysporum* f. sp. *albedinis* (Foa) causes Fusarium wilt known as "Bayoud". This disease is fairly widespread in the main palm growing areas in North African countries⁷. In Morocco, Fusarium wilt due to Foa has caused the progressive disappearance of high quality and world renown date cultivars, mainly Mejhool^{18,8}.

¹Biotechnology and Bioresources Valorization Laboratory, Biology Department, Faculty of Sciences, Moulay Ismail University of Meknès, Meknès, Morocco. ²Biodiversity, Environment and Plant Protection Team, Faculty of Sciences and Technologies, Moulay Ismail University of Meknès, Meknès, Morocco. ³Microbial biotechnology and bioactive molecules laboratory, Faculty of Sciences and Technologies, Sidi Mohammed Ben Abdellah University of Fez, Fez, Morocco. ✉email: hafida.bioinfo125@gmail.com; mmoumni02@yahoo.fr

The typical external symptom of the Bayoud disease is hemiplegia character. In the affected palm leaf, the withering begins on one side of the leaf which becomes white; then the withering continues to the other side until the whole leaf dies. In the other hand the important internal symptom is the reddish-brown color of vascular bundles¹⁰.

Bayoud was first reported in 1870 in Zagora-Morocco. While the first precise description of Foa was done by Malençon in 1934¹¹. It is a telluric fungus classified among the imperfect fungi of *Nectriaceae* family. The most important means of Foa infection are spores and mycelium. In fact, the infection occurs mainly through the roots and spreads inside the vascular system, leading to wilting and eventually to date palm death¹². Foa transmission is particularly rapid and spectacular. It can be spread by infected shoots, soil, infected date tissues (especially pieces of infected rachis) and by irrigation water passing through infested fields. Foa can also be passed from one plant to another by contact between diseased and healthy roots¹⁰.

Given the dangerousness of this pathogen, measured in particular by the enormous damage that results therefrom, it is highly valuable to examine its genetic potential. In this regard, the comparative study of genomes at the structural and functional levels is decisive in such research. The development and/or use of algorithms and bioinformatic tools, dedicated to comparative genomics, has provided a better understanding of the genomes especially those of pathogenic organisms. In *F. graminearum*, comparative genomics has allowed the identification of genes that contribute to phenotypic variation and niche specialization¹³. Another study conducted on FOSC's *formae speciales* infecting legumes revealed candidate effectors¹⁴. Regarding Foa, despite the fact that it has been the subject of several investigations for a long time^{15,15,15}, information on its genomics has remained rather quite limited until now. Currently, progress on Foa could be achieved now that its genome is available in databases^{18,18}. This will make it possible to meticulously discover phenomena that would otherwise go unnoticed and will thus open up new research avenues to develop novel control techniques that will allow to effectively manage this scourge.

Here, we performed in-depth comparative analysis of the first Foa genomes currently available and 29 other *formae speciales* of FOSC in order to inventory the genetic characteristics involved in Bayoud disease. To the best of our knowledge, this study is the first comparative analysis based on genomic approach conducted on Foa.

Results

Phylogeny and genome characteristics of the studied fungal strains

As part of a comparative genomic analysis, the first two Foa genomes publicly available were compared to 49 genomes from 29 other FOSC's *formae speciales*. The two Foa genomes correspond to two strains originally isolated from infected date palms^{18,18}. they were previously sequenced and assembled by the National Institutes of Agronomic Research of Rabat¹⁸ and Errachidia¹⁹, Morocco. The Foa genomes as well as the genomes of 29 other *formae speciales* of FOSC were annotated as described in “Material and methods” section.

Figure 1 showed the phylogeny of the 30 chosen *formae speciales* (A) besides genomic GC content, genome size, genome completeness and proteins putatively encoded by each genome (B). The phylogenetic relationship of the FOSC's *formae speciales* was conducted based on the concatenated 263 genes that are present as a single copy gene in all genomes in our dataset. The result (Fig. 1A) revealed that the *formae speciales* of this study were not grouped into known clusters and formed distinct branches of their own. Moreover, the phylogenetic tree showed that Foa strains shared the same clade with *matthiologiae*, *tulipae* and one strain of *apii*, while Foa strains were relatively distant from the other *formae speciales*.

In comparing statistics of the analyzed genomes, we remarked that some characteristics are quite similar and others are variable. Indeed, Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis showed that more than 99% of the BUSCO genes were present in the most of the analyzed genomes as complete. However, in some strains the genomes covered less than 99% of BUSCO genes (*F.o.f. sp. Cubense* C1HIR9889 (BUSCO 97.9%), *F.o.f. sp. Koaie* isolate 44 (BUSCO 98.8%) and *F.o. f. sp. vasinfectum* LA1E (BUSCO 98.9%)). The similarity was also observed in genomic GC rate. GC content was consistent, with a small range from 47.34 to 48.84% in all strains except *F.o.f. sp. cucumerinum* Foc030 which has more than 50% GC (56.8%). in Foa genomes, GC content varied between 47.44% and 47.73%.

However, the fluctuations of genome sizes within the FOSC's *formae speciales* were relatively high. The smallest genome was found in *F.o.f. sp. Cubense* C1HIR9889 (46.73 Mb) while the largest genome was detected in the *F.o.f. sp. albedinis* 9 (65.56 Mb). These fluctuations were also seen between some strains belonging to the same *formae speciales*. In the case of the Foa (*formae speciales* of interest), the genome size varied from 56.23 Mb to 65.56 Mb.

We further examined the gene content in the analyzed strains. The result revealed that there is considerable inter- and intra- *formae speciales* variations. The number of predicted genes ranged from 13831 (*F.o.f.sp. koaie* isolate 44) to 20010 (*F.o.f.sp. niveum* R1) and from 17460 to 19514 genes within the Foa genomes.

Based on the BUSCO gene rate, the analyzed genomes were of a high enough quality for downstream analysis. Moreover, according to the phylogenetic relationship and the noted variation in genome size and gene content, the 30 analyzed *formae speciales* seem to be relatively diverse.

Homologous and specific genes analysis:

Given the inter- and intra- *formae speciales* genetic diversity signs within the FOSC's *formae speciales*, we examined the extent of genome diversification. For this purpose, we performed a pan-genome analysis. We were interested in 4 conceptual groups of genes:

- The pan-genome: inventory of genes presents within the analyzed strains
- The core-genome: set of homologous genes that are present in all genomes of the analyzed dataset
- The accessory-genome: set of genes present within one or part of the analyzed genomes.

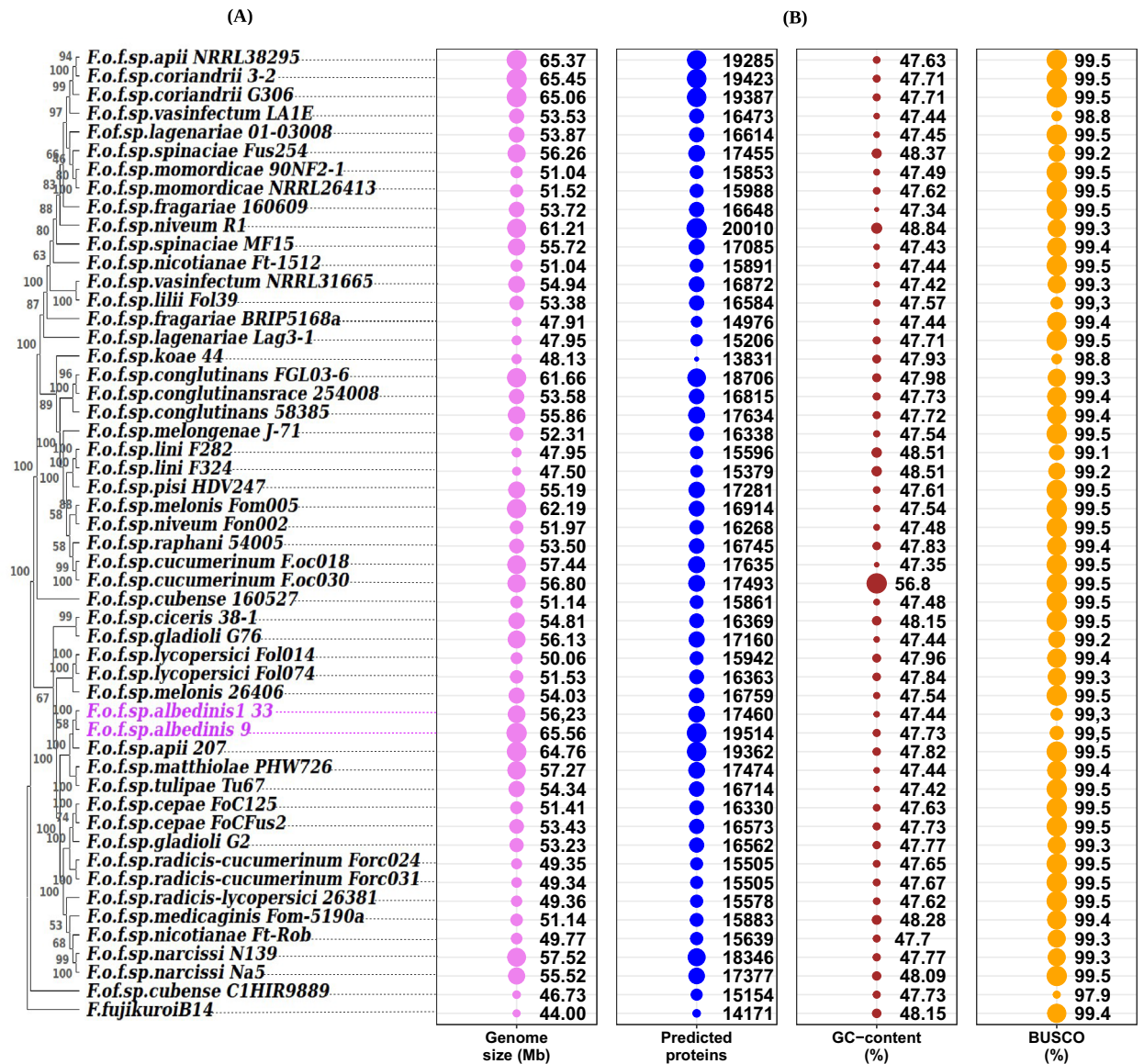


Figure 1. Bubble plots and Dendrogram illustrating the phylogenetic relationships between 49 genomes of the FOSC's 29 *formae speciales* and two *Foa* genomes (*formae speciales* of interest). (A), phylogenetic tree based on 263 orthologous genes in a single copy. (B), four bubble plots showing descriptive statistics for each genome. Bubble sizes have been scaled to the categories and are not comparable between categories. Abbreviations: *F* (*Fusarium*); *o* (*oxysporum*); *f* (*formae*); *sp* (*speciales*). The names following the abbreviations represent the names of the *formae speciales*; the abbreviations coming after the names of the *formae speciales* correspond to the strains belonging to these *formae speciales* (*F.o.f.sp. albedinis* 133: *Fusarium oxysporum forma specialis albedinis* Foa 133).

- The unique or specific-genome: set of genes found in only one strain and absent in others.

We first inventoried homologous (orthologous and paralogous) and specific genes using a BLASTp-based pipeline (see methods). Proteins whose sequences and functions were similar have been grouped in the same family.

The FOSC's *formae speciales* pan-genome comprised a total of 598589 genes (Fig. 2A) consisting of 217249 genes in the core-genome and 351018 genes in the accessory-genome. With regard to the unique-genome, 30331 genes were without homologs in the other genomes, indicating the existence of specific genes among the FOSC's *formae speciales*. The lowest numbers of unique genes belonged to the *F.o.f.sp. Corianarii*3-2, with 155 genes. Strain *F.o.f.sp. niveum*R1 contained the highest numbers of specific genes as well (Fig. 2B and Supplementary Data 2: Table 1). Within *Foa formae speciales*, represented here by strain 133 and 9, the pan-genome consisted of 17460 and 19514 genes respectively (Fig. 2B and Supplementary Data 2: Table 1). In both *Foa* strains, the core-genome was comprised of an average of 5168 genes. However, the large portion of *Foa* pan-genome was predicted as accessory-genes.

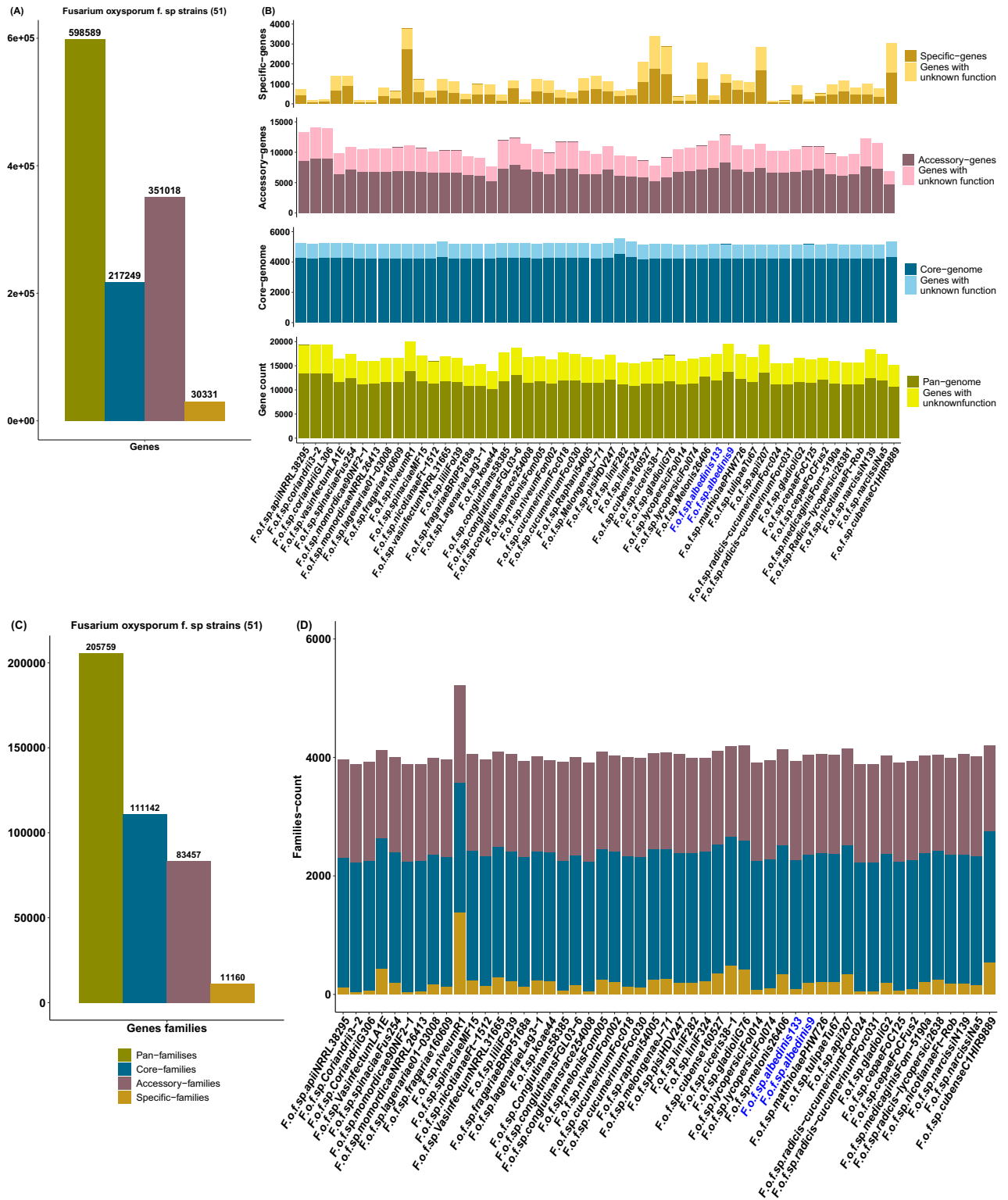


Figure 2. The pan-genome of the FOSC's *formae speciales* including Foa. (A, C), histograms representing the total number of genes and families distributed over the pan-genome, core-genome, accessory-genome and unique-genome for the analyzed strains. (B), Stacked histograms of the gene numbers of the pan-genome, core-genome, accessory-genome and unique-genome for each strain. (D), Stacked histograms of the number of core-families, accessory-families and unique-families for each strain. Abbreviations: *F* (*Fusarium*); *o* (*oxysporum*); *f* (*formae*); *sp* (*speciales*). The names following the abbreviations represent the names of the *formae speciales*; the abbreviations coming after the names of the *formae speciales* correspond to the strains belonging to these *formae speciales* (F.o.f.sp. *albedinis* 133: *Fusarium oxysporum forma specialis albedinis* Foa 133).

They represent 68% (11884 genes) of Foa 133's genome and cover 66% (12873 genes) of Foa 9's genome. As for the unique genome, Foa 133 contained a low number of unique genes (396 genes) compared to Foa 9 whose genome included 1,483 specific genes.

To assign biological functions to the three genomic fractions (core, accessory, and unique), Interproscan 5 was used. The resulting annotations revealed that the pan-genome of the analyzed strains was grouped into 205759 gene families (Fig. 2C). Of those families, 54% were found in all genomes (111142 families). They contained proteins encoded by the core genes. Despite the high gene numbers in the accessory genome, the total of predicted gene families was lower than that of the core families. With regard to the unique genome, 11160 specific gene families were found. An examination of gene repertoire of Foa strains showed that about 30% of pan-genome lacked annotations. In both strains 82% of the core genes were identified. They were classified into an average of 2161 families (Fig. 2D and Supplementary Data 2: Table 2) and were primarily associated with transporters (200–207 genes) and fungal-specific transcription factors (127 genes). For the accessory-genome 35–37% of the genes lacked InterProscan assignments in both Foa strains. In addition to transporters and transcription factors, the annotated proportion of Foa accessory genes were assigned to P450 cytochromes (305 and 312 genes in strains 133 and 9 respectively) (Fig. 2D and Supplementary Data 2: Table 3). Though 52% of Foa 133-specific genes and 28% of Foa 9-specific genes were not assigned to the Interproscan annotations, the remaining specific genes in both Foa strains fall into different families (Fig. 2D and Supplementary Data 2: Table 4). They primarily encoded endonucleases (41 and 273 genes respectively) and the hAT activator family (21 and 334 genes respectively).

Through these results, an abundance of accessory genes was revealed in Foa genomes. These genes provide information about horizontal transfers dynamics because they are often associated with potentially transferable elements. Moreover, the presence of specific genes in Foa genomes suggested that these strains harbored a high level of genomic diversity and uniqueness of each strain, showing their ability to acquire specific characteristics.

TEs in the analyzed genomes

Another conducted analysis was the search for transposable elements (TEs) in the studied strains. These are DNA fragments that can move and multiply in genomes. As such, TEs can have functional and structural impacts on genomes³⁴. They are also considered to be the main driver of genome inflation³⁵.

The search using EDTA software identified 5 putative TE classes: long terminal repeated retrotransposons (LTR); terminal inverted repeats (TIR); non-TIR; non-LTR and other repeated regions (Fig. 3). Within the Foa *formae speciales*, the distribution of TE categories varied between strains. The total TE coverage in Foa 9 (16.5%) was higher compared to Foa 133 (4.92%). In Foa 133, the LTRs covered 0.56% of the genome; they belonged mainly to the Gypsy family (0.51%). As for the TIRs, they accounted for 1.35% of the genome and the majority were Mutator (0.6). On the other hand, LTRs and TIRs were most abundant in Foa 9 and occupied respectively 2.9% and 6.03% of the genome. Note that the maximum TIRs in this strain, were mainly hAT. In addition, we found that the genomes of the Foa *formae speciales* contained a significant fraction of other repeated regions (2.6–4.71%), helitron (0.15–1.44%) and LINE-element (0.26–1.07%). This examination revealed that the Foa genomes were enriched in numerous TEs families.

KEGG pathways assigned to proteins of the analyzed strains:

To further explore the genetic potential of the Foa *formae speciales*, KEGG database was used to highlight the pathways in which Foa genes are involved compared to other *formae speciales*. As shown in Fig. 4, 27 pathways were represented for all strains. Within these pathways, the genetic information processing category (represented by the translation, folding/sorting/degradation, and replication/repair pathways) was the most enriched in gene.

Within Foa, 7102 genes in Foa 133 and 7848 genes in Foa 9 were distributed over the 27 predicted pathways. The examination of this distribution revealed that the maximum number of genes were involved in the translation process (1046 genes in Foa 133 and 1152 genes in Foa 9); followed by the folding, sorting and degradation pathway, which contain 750 genes in Foa 133 and 842 genes in Foa 9. It should also be noted that a large number of genes (656 genes on average) in this *formae speciales* were attributed to the replication/repair pathway. These three pathways maintain the vital functions of the fungal cell and were therefore enriched mainly in core genes and accessory genes that can further enhance these functions. Moreover, Foa genomes included genes involved in cell motility, allowing Foa to actively circulate in the soil as mentioned in other fungi³⁶. We were then interested in the pathways enriched in unique genes of this *formae speciales* (Table 1). We found that a reduced number of unique genes were involved in 17 of the 27 predicted pathways. In Foa 9, 67 genes were assigned to the KEGG annotations. Of those genes, 59.7% (40 genes) were included in the metabolism category. Within this category, the maximum number of genes (13 genes) was assigned to lipid metabolism, followed by carbohydrate metabolism (9 genes), amino acid metabolism (9 genes) and xenobiotic metabolism/degradation (4 genes). However, in Foa 133, a total of 14 unique genes were assigned to the KEGG pathways, where the membrane transport pathway was the most enriched in unique genes (3 genes), while the xenobiotic degradation pathway did not include any gene in this strain. Collectively, these pathway analysis results illustrated that most of the Foa genes are involved in the pathways essential to the fungi's life and their adaptation to environmental conditions.

Candidate secreted effectors in Foa:

In pathogenic fungi, secreted proteins, particularly effectors, are essential for successful host infection³⁷. These proteins can disable plant defenses and subvert cell processes to satisfy the needs of pathogens³⁸.

In this context, the secreted effectome of all the analyzed strains was highlighted. Through the combined use of SignalP v5.0, SecretomeP v1.0 and EffectorP v3.0 to predict putative secreted effectors, mining the Foa genomic data resulted in a catalogue of 3003 effectors in Foa 133 and 2418 effectors in Foa 9 (Supplementary Data 3: Table 1). This constitutes the putative effector repertoire of Foa strains. Of those effectors, more than 800

Categories	<i>F. oxysporum</i> f. sp. <i>albedinis</i> 133		<i>F. oxysporum</i> f. sp. <i>albedinis</i> 9	
	Gene numbers	Percentage (%)	Gene numbers	Percentage (%)
Metabolism	41		5	
Carbohydrate metabolism	9	21.9	2	40
Lipid metabolism	13	31.7	0	0
Amino acid metabolism	9	21.9	1	20
Glycan biosynthesis and metabolism	1	2.4	0	0
Metabolism of cofactors and vitamins	1	2.4	0	0
Metabolism of terpenoids and polyketides	0	0	1	20
Biosynthesis of other secondary metabolites	2	4.9	1	20
Xenobiotics biodegradation and metabolism	4	9.8	0	0
Enzyme families	2	4.9	0	0
Genetic information Processing	12		2	
Transcription	1	8.3	0	0
Translation	3	25	0	0
Folding, sorting and degradation	5	41.7	1	50
Replication and repair	3	25	1	50
Environmental information Processing	8		5	
Membrane transport	7	87.5	3	60
Signal transduction	1	12.5	2	40
Cellular processes	2		1	
Transport and catabolism	2	100	1	100
Unclassified	5		1	
Unclassified	5	100	1	100

Table 1. KEGG pathways enriched in unique genes of the 133 and 9 Foa strains. Values in bold are significant values, they indicate the total number of unique genes involved in each KEGG pathway reported.

proteins were encoded by the core genome in both Foa strains; whereas the accessory genes encoded more than 60% of this effector repertoire in both Foa strains (2096 effectors in Foa 133 and 1542 effectors in Foa 9). As for the unique genome, a low number of effectors were encoded by this genomic fraction, these effectors represented 0,03% of the secreted effectome (94 effectors in Foa 133 and 74 effectors in Foa 9).

In the three genomic fractions (core, accessory and unique) the predicted effectors were classified into two categories: cytoplasmic effectors that act inside the plant cells and apoplastic effectors that manipulate the host apoplast. For all strains, more than 80% of the secreted effectors were cytoplasmic. In Foa, this category occupied 85% and 83% of predicted effectors in Foa 133 and 9 respectively. Within the apoplastic and cytoplasmic categories, effectors lacking Interproscan annotations ranged between 41 and 58%. The apoplastic and cytoplasmic effectors with known function were organized into 803 families in Foa 133 and 779 families in Foa 9 (Supplementary Data 3: Table 2). they mainly belonged to DDE superfamily endonuclease (41 effectors in Foa 133 and 106 effectors in Foa 9).

In Fig. 5, 28 protein families were predicted as families comprised of effectors encoding by Foa specific genes. This constitute the Foa unique secreted effectome. In comparing Foa strains, 14 effector families (families written in blue) in Foa 9 were not presents in Foa 133 such as Egh16-like virulence factor. On the other hand, Foa 133 was distinguished from Foa 9 by secreting effectors belonging to 8 families (families written in orange) absent in Foa 9 such as LysM family.

Among the 28 putative effector families found in Foa strains, 3 families were found only in Foa 9 in comparison to the other analyzed *forma speciales* (Supplementary Data 3: Fig. 1). Foa 9 was characterized by the secretion of effectors belonging to 3-oxo-5-alpha-steroid4-dehydrogenase, Integrase core domain and Globin families.

Taken together, secreted protein repertoire of Foa strains harbored numerous putative proteins acting as effectors, 3 of these were exclusively specific to Foa 9.

Carbohydrate Active enZymes (CAZymes) repertoire of the analyzed strains:

All plant pathogenic fungi are known for producing carbohydrate-active enzymes (CAZymes). These enzymes are involved in the assembly, modification or deconstruction of carbohydrates³⁹. They are considered to play a key role in the degradation of plant cell wall, which is the front line of plant defense⁴⁰. Given the importance of such enzymes, run-dbc software was used to predict the CAZyme content in the Foa genomes in comparison with the other *forma speciales*. As shown in Fig. 6A, the analyzed genomes encoded an arsenal of CAZymes organized in six classes: auxiliary activities (AA), carbohydrate-binding molecules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH), glycosyltransferases (GT) and polysaccharide lyases (PL). In Foa, 1.4% (242 genes in Foa 133) and 1.3% (252 genes in Foa 9) of the genome encoded these enzymes. The highest number of predicted CAZymes (53%) was related to GH followed by GT, AA, CE, PL and CBM. Interestingly, more

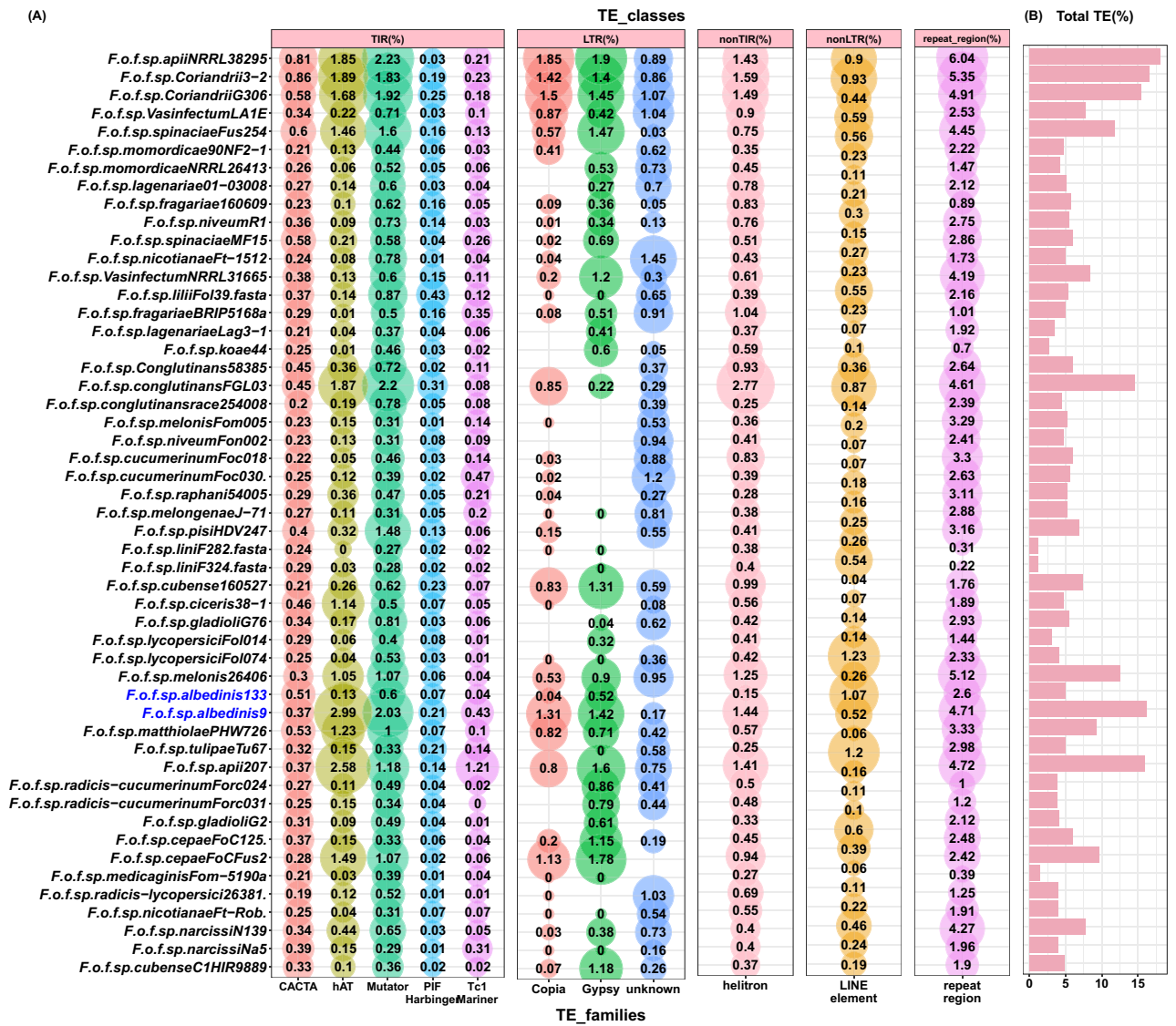


Figure 3. Distribution and coverage (%) of predicted TE in the analyzed *foerae speciales*. (A): bubble plot showing the predicted TE distributed on five classes: Long Terminal Repeated retrotransposons (LTR); Terminal Inverted Repeats (TIR); non-TIR; non-LTR and other repeated regions. The bubble size is proportional to the coverage of each TE family (% indicated inside bubbles). (B): bar plot showing the total TE coverage per genome. Abbreviations: *F* (*Fusarium*); *o* (*oxysporum*); *f* (*foerae*); *sp* (*speciales*). The names following the abbreviations represent the names of the *foerae speciales*; the abbreviations coming after the names of the *foerae speciales* correspond to the strains belonging to these *foerae speciales* (*F.o.f.sp. albedinis 133*: *Fusarium oxysporum forma specialis albedinis* Foa 133).

than 50% of the majority of predicted CAZymes were secreted (Fig. 6B). In both Foa strains, 100% of CEs and PLs were secreted. In contrast, CAZymes belonging to GT class were mainly intracellular (35% of the total GT predicted were secreted).

To further analyze carbohydrate utilization ability of Foa, we examined the different families of secreted CAZymes (Supplementary Data 4: Table 1). It was found that the most abundant identified GH were related to the GH3 and GH28 families, including the largest numbers of the secreted GH. The second most frequent CAZyme families contained in Foa genomes was GT1, GT2 and GT4, which comprised the majority of GT CAZymes. In AA class, all CAZymes were organized in 4 families (AA1, AA3, AA5, AA6 and AA9), while the majority of CBMs were assigned to CBM1 and CBM32. As for PLs, our result revealed that they belonged mainly to PL1 family. Interestingly, our results showed that Foa was distinguished from all other analyzed *foerae speciales* by the secretion of the glycoside hydrolase GH5-7, as shown in Fig. 6C.

The examination of CAZyme repertoire showed the enrichment of CAZyme families in Foa genomes with the exclusive secretion of glycozyle hydrolase (GH5_7). This enzyme is important to decompose hemicellulosic materials.

<i>F.o.f.sp.apii</i> NRRL38295	2148	608	1679	521	2536	453	1098	1017	549	880	1734	1822	2350	2414	3498	2768	2915	2773	1303	2684	1420	269	181	665	81	309	335	2676
<i>F.o.f.sp.Coriandrii</i> 3-2	361	150	237	170	405	78	169	155	110	89	127	330	518	1143	848	702	251	489	68	531	323	53	26	93	18	56	102	315
<i>F.o.f.sp.Coriandrii</i> G306	389	170	237	173	404	76	167	155	113	91	127	321	519	1145	845	698	250	487	68	533	308	54	25	95	20	55	101	317
<i>F.o.f.sp.Vasinfactum</i> LA15	376	165	235	179	410	81	189	160	114	92	141	351	551	1206	913	749	275	560	93	567	345	62	28	105	2	57	108	354
<i>F.o.f.sp.spiniacae</i> Fus254	358	149	221	173	398	77	166	154	113	85	124	316	498	1140	826	676	245	473	61	521	314	51	24	89	19	57	99	308
<i>F.o.f.sp.momordicae</i> NRRL26413	370	158	238	179	410	81	179	161	110	89	127	336	537	1201	893	728	263	518	70	556	340	58	27	99	20	57	107	330
<i>F.o.f.sp.lagenariae</i> 01-03008	377	163	236	179	417	80	192	169	115	92	137	363	572	1216	936	752	288	574	115	574	360	67	29	113	24	60	112	363
<i>F.o.f.sp.fragariae</i> 160609	371	156	234	179	412	80	176	164	113	89	137	336	541	1204	890	729	263	526	74	589	338	58	26	103	20	58	106	334
<i>F.o.f.sp.niveum</i> R1	378	160	248	178	421	85	188	178	127	97	139	331	518	1161	845	727	289	510	70	532	324	55	42	124	19	54	99	380
<i>F.o.f.sp.spinaciae</i> MF15	376	164	235	180	414	81	180	165	116	86	140	348	546	1202	909	742	267	533	76	585	342	63	31	105	24	55	105	346
<i>F.o.f.sp.nicotianaef</i> T-1512	376	158	238	179	412	80	177	164	111	89	149	338	549	1199	901	732	269	527	83	585	340	57	25	106	20	58	106	340
<i>F.o.f.sp.Vasinfactum</i> NRRL31665	385	175	240	187	421	79	201	173	125	94	149	365	582	1231	971	781	311	593	133	572	359	75	35	119	25	58	111	388
<i>F.o.f.sp.lilif</i> Fol39_fasta	334	146	224	161	397	70	162	154	91	83	129	312	507	1135	829	687	246	479	60	510	315	56	23	99	17	55	101	307
<i>F.o.f.sp.fragariae</i> BRIP5168a	363	155	229	178	405	79	176	157	106	87	127	330	539	1202	887	725	259	514	73	561	333	57	25	103	19	55	104	331
<i>F.o.f.sp.lagenariae</i> Lag3-1	353	152	216	175	396	77	160	157	104	87	126	317	505	1157	833	687	248	488	63	517	317	53	24	93	18	53	101	318
<i>F.o.f.sp.koae</i> 44	368	157	230	178	408	80	179	159	111	89	128	334	540	1189	890	730	267	517	81	593	327	59	25	101	20	57	105	327
<i>F.o.f.sp.Conglutinans</i> 58385	322	131	194	149	346	59	143	136	76	77	116	258	425	994	722	591	226	403	48	445	232	41	24	84	13	48	87	259
<i>F.o.f.sp.conglutinans</i> FGL30	334	137	213	154	384	70	157	144	87	80	120	288	445	1024	750	620	234	448	57	469	293	49	24	87	16	56	89	293
<i>F.o.f.sp.conglutinans</i> rac254008	353	147	224	165	393	78	162	151	108	85	129	304	485	1112	830	669	250	474	62	515	293	49	26	95	14	55	98	311
<i>F.o.f.sp.melonis</i> Fom005	375	156	230	175	407	80	175	161	115	88	137	330	525	1184	875	719	256	507	67	542	329	57	26	98	16	54	105	327
<i>F.o.f.sp.niveum</i> NR002	341	140	223	156	390	67	161	151	97	85	126	288	461	1075	789	623	245	440	62	491	272	47	23	91	16	50	91	303
<i>F.o.f.sp.cucumerinum</i> Foc018	355	151	218	170	382	76	164	152	112	86	132	313	512	1136	848	705	249	480	64	526	315	54	20	98	19	63	100	313
<i>F.o.f.sp.cucumerinum</i> Foc030	371	152	232	175	402	79	174	157	116	88	137	322	507	1184	856	704	254	510	64	538	329	57	26	97	19	56	102	327
<i>F.o.f.sp.raphani</i> S4005	334	140	204	154	377	70	149	140	109	77	115	289	440	1028	728	610	238	433	59	475	276	46	24	83	17	55	93	283
<i>F.o.f.sp.melongenae</i> J-71	391	181	241	186	428	80	206	178	121	100	149	388	611	1240	998	812	326	645	154	593	371	87	49	124	32	61	117	406
<i>F.o.f.sp.pisi</i> HDV247	389	174	242	187	424	84	201	171	124	95	147	386	593	1232	965	789	316	598	130	577	360	73	42	105	28	57	112	409
<i>F.o.f.sp.linif</i> F282_fasta	380	166	223	176	413	82	190	167	114	92	146	345	552	1207	930	729	276	531	97	554	346	63	31	107	21	54	109	360
<i>F.o.f.sp.linif</i> F324_fasta	368	153	225	176	407	81	184	162	111	92	137	336	529	1198	898	718	266	514	76	555	327	60	28	102	21	55	105	334
<i>F.o.f.sp.cubense</i> F100527	385	165	230	182	410	78	194	160	118	88	133	347	562	1206	930	758	281	538	112	566	339	67	29	110	22	59	110	373
<i>F.o.f.sp.ciceris</i> 38-1	285	120	181	138	339	62	123	131	78	74	108	242	406	949	662	573	218	384	48	493	240	42	19	79	14	49	81	247
<i>F.o.f.sp.gladiali</i> G76	289	123	184	150	353	66	132	138	87	77	120	253	425	981	710	603	221	412	53	489	251	44	18	77	15	47	80	260
<i>F.o.f.sp.lycopersici</i> Fol1074	330	138	210	152	389	68	146	135	93	83	115	282	461	1026	757	629	228	410	59	462	283	45	18	92	14	49	89	238
<i>F.o.f.sp.lycopersici</i> Fol1074	379	161	235	182	411	80	191	169	119	95	134	374	572	1217	944	772	296	576	110	567	346	65	33	114	22	61	112	385
<i>F.o.f.sp.melonis</i> 26406	331	140	212	166	372	64	162	146	86	80	122	333	524	1023	825	694	265	534	108	518	317	68	33	101	26	55	96	349
<i>F.o.f.sp.albedinis</i> 133	328	137	212	153	378	65	146	138	98	84	116	293	448	1046	750	619	235	435	57	465	280	44	22	90	17	55	92	298
<i>F.o.f.sp.albedinis</i> 99	360	150	235	171	395	79	164	153	104	86	126	317	510	1182	842	694	251	482	65	534	311	52	26	97	18	54	100	320
<i>F.o.f.sp.matthiola</i> PHW726	369	154	246	173	408	79	175	159	114	88	127	332	528	1200	878	727	262	517	67	545	335	56	29	99	19	57	105	329
<i>F.o.f.sp.tulipae</i> Tu67	372	159	243	182	421	81	189	169	120	89	134	357	567	1210	935	764	288	560	104	564	353	62	35	108	21	59	106	372
<i>F.o.f.sp.apii</i> 207	2020	581	1676	500	2460	424	750	975	479	813	1608	1694	2253	3214	2934	2543	2599	2501	483	2076	1259	251	150	345	79	351	363	2236
<i>F.o.f.sp.radicis</i> -cucumerinumForc024	367	161	232	178	408	79	178	162	114	89	129	351	549	1213	912	741	278	528	83	557	350	62	28	107	22	58	105	350
<i>F.o.f.sp.radicis</i> -cucumerinumForc031	293	125	190	143	355	61	138	127	80	71	108	251	408	953	661	548	215	373	45	419	238	42	15	85	14	48	84	253
<i>F.o.f.sp.gladiali</i> G2	368	158	238	178	409	80	178	156	111	89	132	337	537	1204	895	741	268	527	87	585	344	59	24	109	22	57	105	336
<i>F.o.f.sp.cephae</i> Foc125	1940	490	1394	467	2390	438	680	903	450	827	1505	1242	1683	2749	2466	1953	2496	1726	269	1758	1121	185	119	260	63	245	279	1824
<i>F.o.f.sp.cephae</i> FocFus2	1726	482	1215	422	2026	381	694	798	377	679	1288	1223	1958	2542	2534	1943	2196	1893	562	1752	1053	209	142	354	77	233	281	1922
<i>F.o.f.sp.medicaginis</i> Fom-5190a	358	147	237	168	400	75	158	151	108	86	132	321	495	1122	834	688												

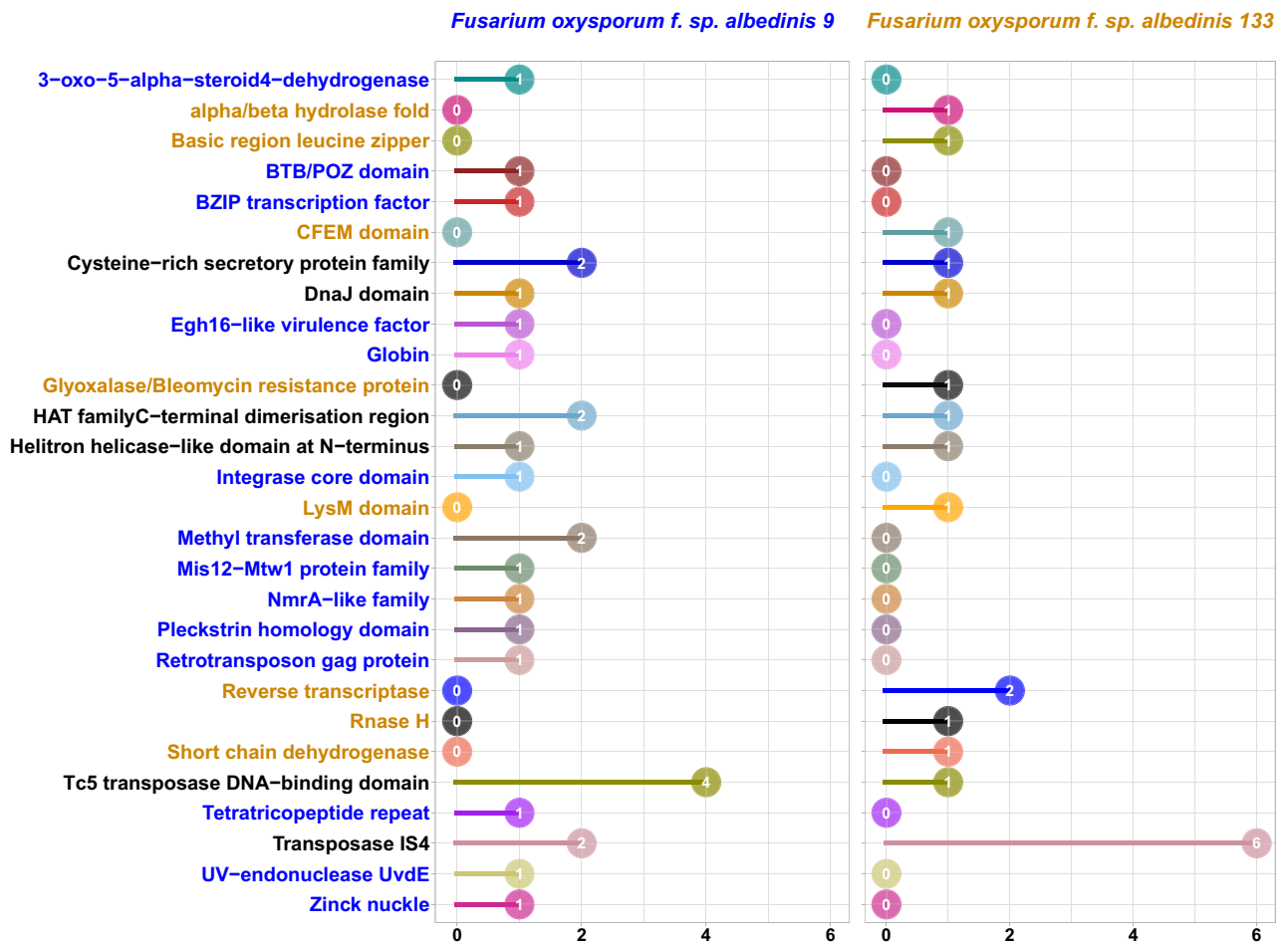


Figure 5. Predicted secreted effectors in the *Foa formae speciales* unique genome. The colors represent the predicted effector families and the numbers indicate the count of effectors in each family for each *Foa* strain.

(T1PKS, an average of 14 BGCs), terpenes (an average of 12 BGCs) and Non-ribosomal peptide synthetase (NRPS, an average of 22 BGCs).

The predicted BCGs were then associated with characterized and verified clusters from the MIBiG database. Based on this, 12 families were identified within the *Foa formae speciales* (Fig. 9). Of these, 3 families (oxyjane-ancine, gibepyrone A and equisetin) were produced by all *formae speciales*. On the other hand, the secondary metabolites belonging to the chrysogin family were specific to the *Foa* strains and were mainly encoded by unique genes. We also noted in *Foa*, the biosynthesis of mycotoxins which include ACT-Toxin-II and beauvercin as well as the biosynthesis of metabolites having anti-fungal activity (fujikurin A–D).

According to our in silico analysis, some metabolites were produced by all of the *formae speciales*; however metabolites belonging to chrysogin family were specifically synthesized by *Foa* strains.

Discussion

Fusarium oxysporum f. sp. albedinis (*Foa*) is the causal agent of the most destructive disease of date palm in the oasis's regions of Morocco and other North African countries⁴⁵. Until now, no effective treatment has been established against this disease called Bayoud⁴⁵. For the first time in its history, we approach the study of this disease based on genomic analysis. Indeed, genomes sequencing of *Foa* strains provides important data to deeply study this disease and to design new control strategies.

To provide a comprehensive overview and uncover the genetic traits of this pathogenic fungus, a comparative genomic analysis of the two *Foa* genomes and the genomes of 29 other FOSSC's *formae speciales* was conducted in this work. In comparing 263 core genes, we showed that *Foa* was phylogenetically located in the same clade including *matthiolar*, *tulipae* and *apii*. These *formae speciales* are known to attack temperate climate plants (Brassicaceae, Liliaceae and Apiaceae). Recently, a phylogeny based on the *Foa* mitochondrial genome was carried out and has highlighted the *Foa* clustering in the same clade as *lycopercisi formae speciales*⁴⁶. However, this is relative as it was based only on the comparison of the mitochondrial DNA and not on the analysis of the genomic DNA which provides exhaustive information on the pathogen. In-depth structural and functional annotations were performed to extract accurate information about *Foa*. It was found that the genomes of this *formae speciales* harbor numerous encoding genes (17460–19514 genes) with around 30% that remain uncharacterized (unknown function). The gene families that control vital cellular functions were mainly derived from the core genome where

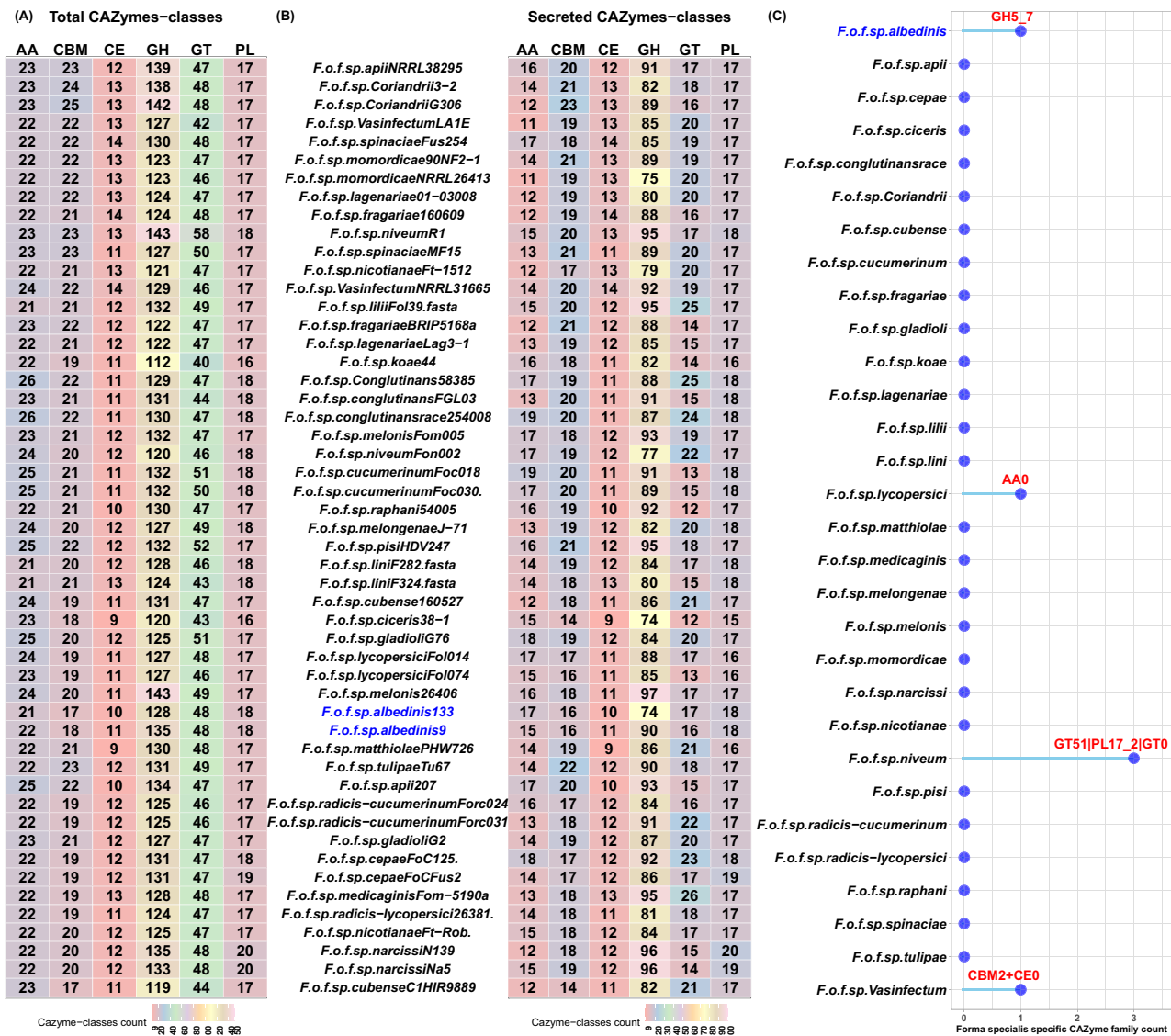


Figure 6. Predicted CAZymes in the analyzed strains. (A): heat map showing the total number of CAZymes distributed on six categories: auxiliary activities (AA), carbohydrate-binding molecules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH), glycosyltransferases (GT) and polysaccharide lyases (PL). (B): heat map showing the number of secreted CAZymes for each strain. (C): lollipop plot representing CAZyme families specific to each *formae specialis*. Abbreviations: *F* (*Fusarium*); *o* (*oxysporum*); *f* (*forma*); *sp* (*specialis*). The names following the abbreviations represent the names of the *formae specialis*; the abbreviations coming after the names of the *formae specialis* correspond to the strains belonging to these *formae specialis* (*F.o.f.sp. albedinis* 133; *Fusarium oxysporum forma specialis albedinis* Foa 133).

a large number of genes encoded MFS (Major Facilitator Superfamily). These are known as omnipresent proteins in all vital functions and constitute the largest family of secondary transporters^{47,47}. They are involved in essential cellular functions, such as nutrient uptake and metabolite extrusion^{48,48}. In addition, a recent study⁵⁰ showed that in *F.virguliforme*, MFS conferred reduced sensitivity to fluopyram (fungicide). Other such abundant core families in *Foa* include proteins having the fungal-specific transcription factor domain. Predictions suggest that factors with this domain are involved in carbohydrate metabolism, amino acid metabolism, gluconeogenesis, respiration and fatty acid catabolism (TRANSFAC, PFAM). However, the role of these proteins remains unknown, as these predictions have not all been experimentally confirmed. In some fungi, these factors play an important role in the coordination of multiple physiological processes, such as regulation of sensitivity to fludioxonil (fungicide), and of pathogenesis⁵¹.

As to the remaining genome of the *Foa formae specialis*, it was mainly composed of accessory genes (70% of the total predicted genes) this fraction is enriched in proteins having the fungus-specific transcription factor domain and in proteins belonging to the Cytochrome P450 family. The latter are used by fungi as rapid adaptation strategies in several ecological niches^{52,52}, they are involved in the degradation processes of the plant cell wall⁵⁵. Furthermore, the richness of accessory genes can provide information on the horizontal transfer

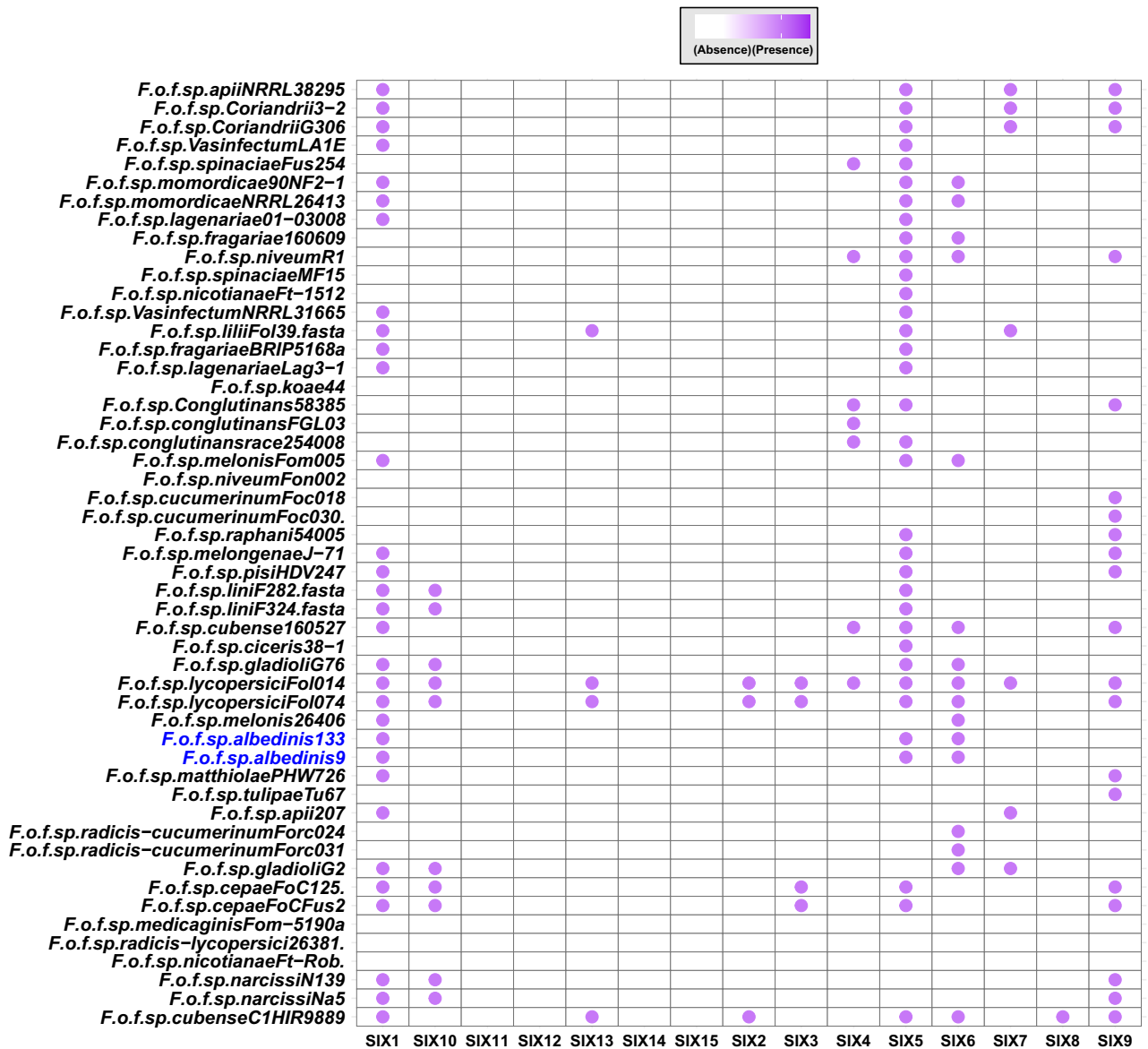


Figure 7. Presence/absence map of SIX protein distribution in the protein repertoire of Foa and other analyzed *formae speciales*. The bubbles indicate the presence of SIX proteins and the empty boxes represent their absence in each strain. Abbreviations: *F* (*Fusarium*); *o* (*oxysporum*); *f* (*forma*); *sp* (*specialis*). The names following the abbreviations represent the names of the *formae speciales*; the abbreviations coming after the names of the *formae speciales* correspond to the strains belonging to these *formae speciales* (*F.o.f.sp. albedinis* 133: *Fusarium oxysporum forma specialis albedinis* Foa 133).

dynamics as they are often associated with potentially transferable elements such as transposons. Indeed, TEs are important features of fungal genomes and play a key role in genome structure and plasticity⁵⁶. In addition to the *fot1* family mentioned in previous studies⁵⁷, the analyses carried out in the present work revealed that Foa genomes harbor numerous transposable elements categories that cover a significant genomic portion. The high content of TEs can explain the variation in total genome size of the two Foa strains. This size decreases considerably when TEs are excluded. It varies between 53.43 Mb (in Foa 133) and 54.76 Mb (in Foa 9) instead of 56.23 Mb and 65.56 Mb respectively.

Our results are consistent with those of Li-Jun Ma and their team⁵⁸ who showed that the genomes of other *Fusarium* such as *F. graminearum*, *F. verticillioides* and *F. oxysporum* f. sp. *lycopersici* are also enriched in transposable elements and pathogenicity-related genes, indicating horizontal acquisition.

Another finding is that the core and accessory genomic fractions harbor a gene set that enables fungal cell motility, indicating that Foa could actively move towards the plant roots which is the starting point of the infection. The proteins encoded by these genes could therefore be valuable targets for new control methods.

Similar to all the analyzed *formae speciales*, Foa can be distinguished from the others by a genomic fraction that is specific to it (the unique genes). These genes regulate numerous processes. In Foa 9, 4 unique genes are involved in the degradation of xenobiotics (chemicals such as fungicides, pollutants, etc.). In comparison

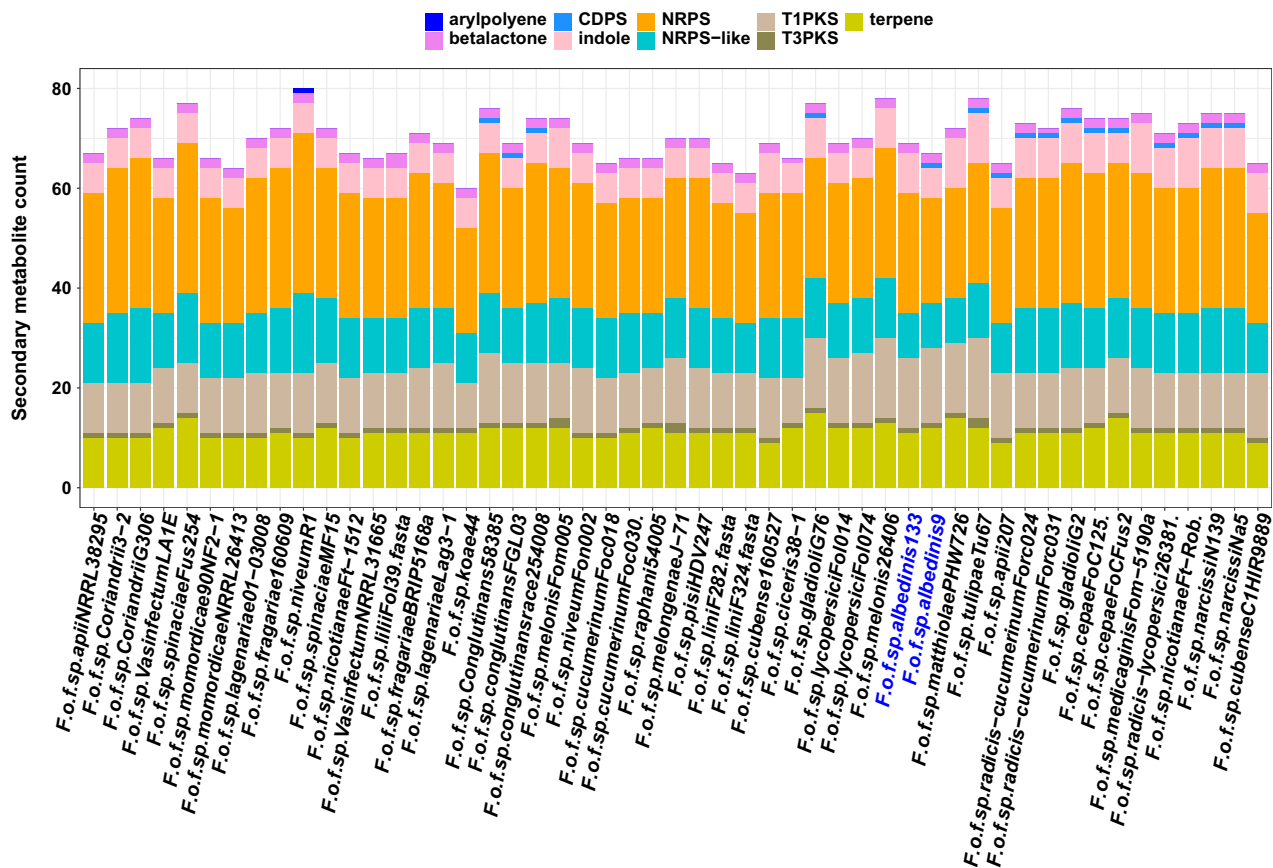


Figure 8. Predicted secondary metabolite classes in the analyzed *formae speciales*: sckqtd plot showing the predicted secondary metabolite genes for each strain; these genes were involved in the synthesis of nine secondary metabolite classes: arylpolyene, betalactone, indole, terpene, CDPS (CycloDiPeptides Synthase), NRPS (Non-Ribosomal Peptide Synthase), NRPS-like, T1PKS (Type I Polyketide Synthase) and T3PKS (Type III Polyketide Synthase). Abbreviations: *F* (*Fusarium*); *o* (*oxysporum*); *f* (*forma*); *sp* (*specialis*). The names following the abbreviations represent the names of the *formae speciales*; the abbreviations coming after the names of the *formae speciales* correspond to the strains belonging to these *formae speciales* (*E.o.f.sp. albedinis* 133: *Fusarium oxysporum forma specialis albedinis* Foa 133).

with Foa 133, the presence of these unique genes, in addition to the core and accessory genes involved in this pathway, could confer and increase the resistance of this strain to fungicides and environmental stress due to contaminants accumulation. This is consistent with other fungi, where the xenobiotics degradation enabled them to resist dichlorvos⁵⁹.

Furthermore, our results revealed that a subset of the Foa unique gene repertoire encodes secreted effectors. The role of effectors in the virulence of pathogenic fungi has been investigated extensively. Here we have shown that the protein repertoire of the analyzed Foa strains harbors key effectors of infection. These include proteins having a LysM domain, given that several LysM effectors have already been identified as virulence factors in plant pathogenic fungi⁶⁰. These proteins inhibit the chitin-induced immune response, which protects the fungal cell walls from host chitinases⁶¹. Another type of effectors secreted by Foa are belonging to the cystine-rich proteins family. These are apoplastic effectors in which the cysteine residues form disulfide bonds, thus increasing the stability of these effectors in the protease-rich apoplastic space⁶². In this context, a recent study⁶³ revealed that these effectors play a key role in the interaction between *Verticillium dahliae* and its host plant by suppressing immunity after infection. Foa also secretes the virulence factor Egh16-like (predicted in Foa 9) which enables it to penetrate the plant cell. This type of factors is known to be involved in appressorium creation in plant pathogenic fungi⁶⁴. In *Magnaporthe oryzae* (rice borer fungus), deletion of Egh16 orthologous genes altered fungal ability to penetrate host cuticles⁶⁵. The effector repertoire mining in Foa revealed effectors exclusively specific to Foa 9. this strain secretes protein predicted as 3-oxo-5- α -steroid-4-dehydrogenase. In rice blast fungus, gene encoding this effector were specifically expressed during infection stages⁶⁶. the effector referred to as Globin was putatively identified in Foa 9. this protein was reported to be involved in low oxygen adaptation of other fungi⁶⁷. thus, Foa 9 could grow in microaerobic environments with low O₂ levels.

A further class of effector proteins encoded by the Foa genomes are SIX1, SIX5 and SIX6. They are regulated by accessory genes, whose presence could be the result of horizontal gene transfer. These proteins further enhance Foa pathogenesis. Indeed, the presence of SIX1 has been reported to be a prerequisite for achieving complete virulence of the *conglutinans* (Focon) and *lycopersici* (Fol) pathogens on cabbage and tomato, respectively⁶⁸.

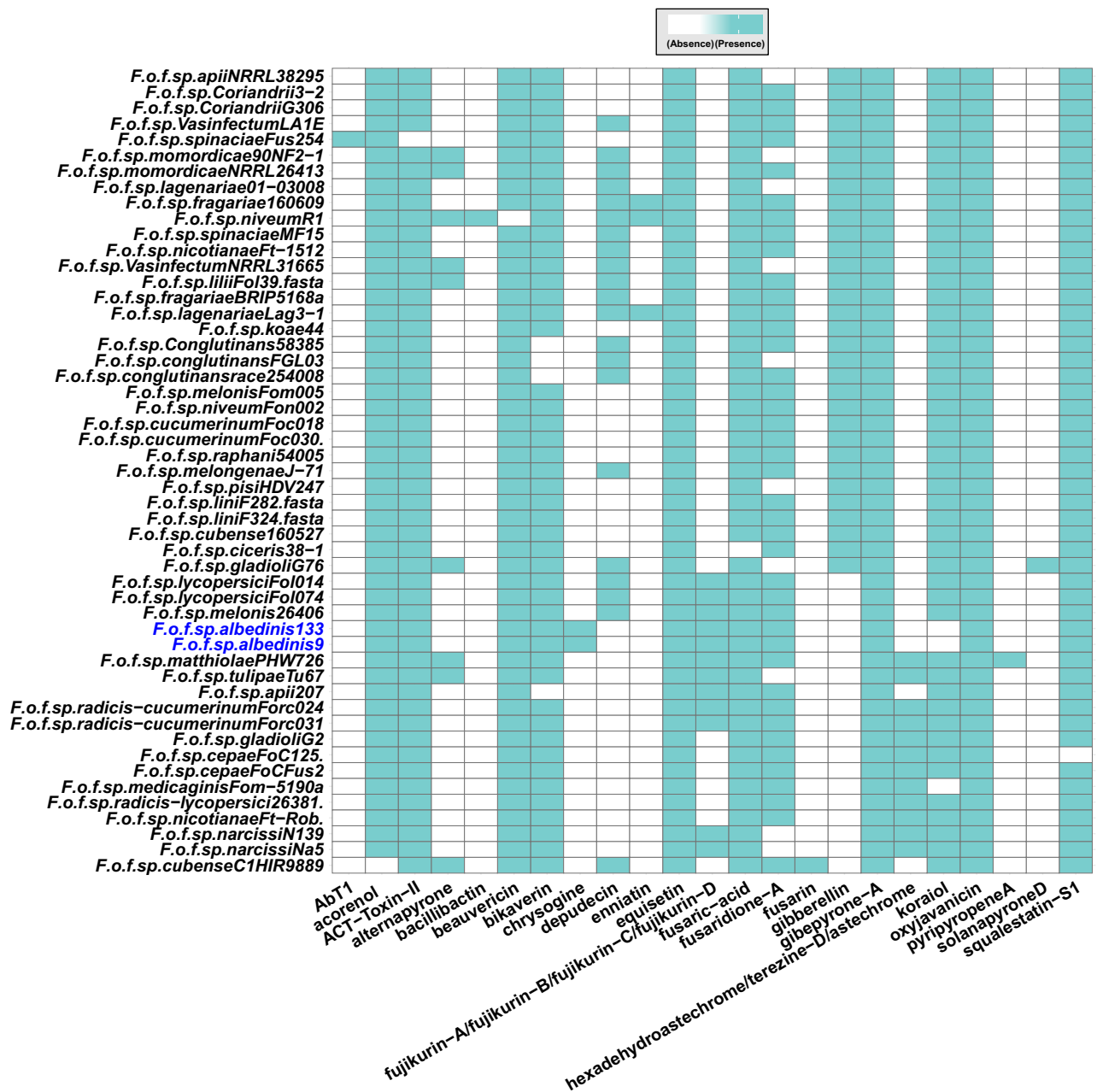


Figure 9. Predicted secondary metabolite families in the analyzed *formae speciales*. presence/absence map of secondary metabolite families detected in the genomes of the Foa and the other *formae speciales*. Blue boxes indicate the presence of secondary metabolite families and empty boxes indicate their absence in each strain. Abbreviations: F (*Fusarium*); o (*oxysporum*); f (*forma*); sp (*specialis*). The names following the abbreviations represent the names of the *formae speciales*; the abbreviations coming after the names of the *formae speciales* correspond to the strains belonging to these *formae speciales* (F.o.f.sp. albedinis 133: *Fusarium oxysporum forma specialis albedinis* Foa 133).

Further studies on *Fol* showed that infection using mutants lacking SIX5 gene showed a considerable reduction in disease symptoms; and reintroduction of the gene restored pathogenesis in 75% of mutants⁶⁹. As for SIX6, this protein was shown to play a role in pathogenic fungi virulence by inhibiting a hypersensitivity response (HR)⁷⁰. In *Nicotiana benthamiana* leaf cells, the SIX6 protein induced cell death by Avr2-I-2 interaction⁷¹. This arsenal of predicted effectors in Foa could serve as target proteins to setting up an effective treatment against Bayoud.

The genetic potential of Foa is not only restricted to the secretion of the above-mentioned effectors. This pathogen produces other proteins that are involved in the infection process. Our results showed that Foa strains carry a high number of genes encoding CAZymes associated with the degradation of the plant cell wall (the first physical barrier in front of Foa). Our results have further shown that this CAZyme arsenal is in part secreted. The secretome prediction revealed that the predicted CAZymes are partly secreted in Foa, thus suggesting the

cellular location of the remaining CAZymes. This is reported in other fungi such as *Flammulina elastica* where the third of predicted GH genes were intracellular⁷².

As in all the examined FOSC's *formae speciales*, the Foa secreted CAZyme fraction comprises enzymes having known activities, which are organized in numerous families clearly associated with the degradation of key plant cell wall components. Another important finding is that the Foa is distinguished from all other analyzed *formae speciales* by the secretion of an endo-b-1,4-mannanase (GH5-7). This enzyme can efficiently cleave higher molecular weight mannans (plant wall hemicellulose) consisting of more than six mannose monomers⁷³. In addition, the combination of mannanases and mannosidases can further increase the mannan catalysis by 83%⁷⁴. A previous studies revealed that date palm is almost entirely composed (92%) of linear mannan (high molecular weight molecule)⁷⁵, giving it a hardness, especially in seeds, to protect it from mechanical damage^{76,76}. Therefore, the exclusive occurrence of the gene encoding mannanase (GH5-7) in Foa genomes could explain its specificity towards date palm.

In phytopathogenic fungi, other compounds can further improve infection processes. These fungi are known to produce a wide variety of secondary metabolites that often confer ecological advantages and enhance their infectiveness. In this regard, our results showed the presence of numerous gene clusters involved in the biosynthesis of these compounds in Foa. The predicted metabolites are often mycotoxins such as ACT-Toxin-II and Beauvericin. ACTs are host-specific toxins originally found in *Alternaria* pathotypes infecting mandarins⁷⁸. The ACT toxin induces plasma membrane dysfunction resulting in electrolyte leakage and rapid cell death⁷⁹. As for beauvericin (BEA), it is a mycotoxin produced by *Fusarium* species, has cytotoxic activity and is able to increase oxidative stress to induce cell apoptosis⁸⁰. The Foa secondary metabolite pool also comprises antibiotics and compounds that protect it from toxic products (gibepyrone A)⁸¹, as well as other metabolites that exhibit anti-fungal activity (fujikurin A–D).

Our results revealed that Foa, in contrast to all other analyzed genomes, can synthesize chrysogin. This pigment is known to protect microorganisms against abiotic aggressors such as UV light⁸². Therefore, Foa seems to be more able to withstand severe environmental conditions compared to the other analyzed *formae speciales*.

Conclusion

The two Foa genomes analyzed in this work are the first genomes sequenced and reported in the literature. The availability of these sequences has provided us with an extensive overview of the Foa genomics and its pathogenesis and adaptiveness. The in-depth analysis of these two genomes revealed the genetic characteristics of Foa. We showed that Foa secretes a wide range of effector proteins. These proteins comprise the virulence factor Egh16-like and other effectors (LysM, and SIX proteins) that enhance Foa's ability to infect and destroy its host. Our analyses also revealed that the Foa genome encodes an arsenal of secreted enzymes involved in plant cell wall degradation. This arsenal includes an enzyme (GH5-7) which is exclusively specific to Foa. This enzyme could define the specificity of Foa to its host. In addition to these enzymes, predicted effectors and putative secondary metabolites reflect the pathogenic nature of Foa. The discovery of all these proteins has shed light on mechanisms potentially involved in orchestrating the host–pathogen interaction. It is expected that the findings from these analyses will set the stage for future research on Foa and provide a baseline for new control methods design to effectively manage Bayoud disease.

Material and methods

Genomic data acquisition

Currently, the FOSC includes more than 30 pathogenic and putatively non-pathogenic *formae speciales*. In our analysis, a total of 80 genomes corresponding to 30 pathogenic *formae speciales* including Foa (*formae speciales* of interest) were downloaded from NCBI and Joint Genome Institute's MycoCosm databases. Detailed information (host, ID, strain and source information.....) of the strains employed in this study are denoted in Supplementary Data 1.

Data quality assessment

The quality of the genome assemblies of the 80 fungal strains was evaluated using BUSCO (Benchmarking Universal Single-CopyOrthologs) v4.0.5²⁰. In short, BUSCO estimates the completeness and redundancy of processed genomic data based on universal single-copy orthologs²⁰. It uses Hidden Markov Models (HMM), HMMER v3.3.2 and tBLASTn to predict whether each (3) fragmented or (4) missing. For BUSCO analysis, we used here the Hypocreales specific single-copy orthologous genes from Fungi Odb version 10.

A total of 493 genes defined as orthologs were obtained from the BUSCO evaluation of 80 fungal genomes. To verify the presence of each BUSCO gene in the analyzed strains, the 80 genomes were annotated by AUGUSTUS version 3.4.0²¹ using *F. graminearum* as reference species, then each of the 493 BUSCO genes was aligned to the set of genes predicted by AUGUSTUS for each strain using BLASTn, the considered genes were those having identity > = 97%. A gene was defined as single-copy ortholog if it is complete and present only once on the genome of each strain. Complete genes occurring twice or more were considered duplicate genes. Fragmented genes were BUSCO genes that cover less than 97% of gene predicted by AUGUSTUS. Genes absent in all strains were considered missing genes. Duplicated, fragmented, or missing genes in at least one strain were eliminated from the gene set. The strains in which the 493 genes are duplicated or absent or fragmented were also eliminated. The final data set retained corresponded to 51 fungal strains belonging to 30 *formae speciales* with 263 orthologous genes in a single copy for each strain.

Phylogenomic analysis

For each strain, the 263 orthologous genes were concatenated in the same order using the CLC Genomic workbench. v20²²; then aligned by MAFFT. Well-aligned regions were extracted using gBlocks. The final alignment was used as input for the MEGAX software²³ to generate a phylogenetic tree using the Maximum Likelihood (ML) method with a Bootstrap of 1000. Bootstraps indicate how many times, out of 1000, the same branch was observed when repeating the phylogenetic construct on a re-sampled data set.

Homologous protein identification

In addition to the predicted gene's sequences, the ab initio annotation by AUGUSTUS also provided the protein sequences corresponding to these genes. Indeed, ab initio annotation is based on the use of the Support Vector Machine (SVMs) algorithm and hidden Markov models (HMM) to predict genes and their structure in eukaryotic DNA sequences.

All of the predicted proteins from each genome of the 51 analyzed strains were aligned to the protein sequences of the other strains. The protein sequences of each strain were aligned to all the protein sequences of the remaining strains belonging to set N with $N = 51$. This alignment was performed using the BLASTp with an E-value $< 10^{-11}$ and coverage $\geq 97\%$. The tables resulting from the BLASTp analysis for each strain were filtered using the R package dplyr²⁴ in order to keep only the lines where the identity between the aligned sequences is $\geq 97\%$. Protein sequences found in a strain and occurring in other strains were defined as core sequences, thus representing the core genome. A sequence was considered unique if it was found only in one strain and not in the others. The set of unique sequences represent the unique genes. For the accessory genome, it was defined by the sequences common to only a part of the strains. On the basis of this BLASTp, three gene groups were defined for each strain: the core genes, the accessory genes and the unique genes. These three groups together constitute the pan genome.

Functional annotation

The three protein sequence groups (core, accessory and unique) for each strain were classified into families using the Interproscan 5 software²⁵, which gives an overview of the families that a protein belongs to as well as the domains and sites it contains. The files obtained from the Interproscan 5 analysis were filtered to eliminate false positives. In the case of proteins aligned to two or more families, the alignment result having the minimum E-value was chosen.

KEGG pathways prediction

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of databases for understanding high-level functions and utilities of an organism linking genomic information with higher order functional information, by assigning the pathways where each protein of this organism is involved²⁶. In our case, this prediction was performed by the Kofamscan software using the three protein sequence groups of each strain as input. The resulting list of KO codes (KEGG Orthology) was used to scan the KO database integrated in the KOALA-formatter software²⁷. This allowed to extract KEGG pathways corresponding to the proteins of the used strains. KEGG pathways showing alignments of an E-value more than 10^{-10} were eliminated.

Secretome and effectors prediction

The "secretome" is defined as all the proteins that are secreted by a cell, a tissue or an organism. The secretome is an important class of proteins that control many biological and physiological processes. The primary sequence of proteins intended for secretion is characterized by a peptide sequence called 'N-terminal signal peptide' which is cleaved by signal peptidase after the protein translocation across the membrane. However, several extracellular proteins were reported as a secreted protein that can be exported without a classical N-terminal signal peptide²⁸. Secretion of proteins without an N-terminal signal peptide is known as non-classical secretory pathway. Classical and non-classical secreted proteins can act as effectors.

To examine the secretome of the 51 strains used in our work, the SignalP v5.0²⁹ and SecretomeP v1.0²⁸ software were used to analyze their protein sequences in order to deduce the classical and non-classical secreted proteins for each strain. Then, the secretome sequences were submitted to the EffectorP v3.0³⁰ software, which predicts the effector proteins in this sequence set. The effectors resulting from this prediction were organized into two categories: cytoplasmic effectors (which penetrate inside plant cells) and apoplastic effectors (which remain and act outside the plant cell).

Detection of CAZymes (Carbohydrate-Active enZymes)

The protein sequences of the 51 selected strains were also the subject of CAZymes annotation using the run_dbcan v3.0 software³¹ coupled with the CAZyDB09242021 database. This software combines three state-of-the-art tools (DIAMOND, HMMER and eCAMI) to automatically identify CAZymes. Considered CAZymes were those identified by the three tools and having an e-value lower than 10^{-10} .

SIX proteins (secreted in xylem) prediction

The secretome of each strain was aligned to all the SIX sequences available on NCBI using the BLASTp with an E-value $< 10^{-11}$ and a coverage $\geq 97\%$. Alignments with an identity $\geq 97\%$ were selected to infer the different classes of SIX proteins present in each strain.

Secondary metabolite gene clusters prediction

Gene clusters involved in secondary metabolite biosynthesis (BGC) were predicted using antiSMASH version 6.1³², which combines the MIGBIB, ClusterBlast and Subcluster Blast gene collection with the ClusterFinder and CASSIS (Cluster-border prediction based on transcription factor binding sites) algorithms. This combination provides an accurate identification of secondary metabolite gene clusters of known major chemical classes. It also offers a detailed sequence analysis. In our analysis, the genomic sequences of the 51 strains and their annotations were submitted to antiSMASH to inventory BGCs existing in the analyzed strains.

Transposable elements prediction

Transposable elements (TE) of the 51 analyzed strains were predicted and annotated by the EDTA v1.3 software³³, using the genomic sequences of these strains as input. It is a pipeline that allows the identification and classification of transposable elements based on a combination of programs: LTRharvest (v1.5.10), LTR_FINDER_parallel (v1.0), LTR_retriever (v2.6), Generic Repeat Finder (v1.0), TIR-Learner (v1.23), MITE-Hunter (v1.0) and HelitronScanner (v1.0). This combination results in the creation of a high-quality non-redundant TE library.

Statistics and plot generation

The genome statistics (genome size and GC rate) were computed using Quast⁸³ software; the other statistics (number of proteins and BUSCO genes) were obtained using the BUSCO²⁰ and AUGUSTUS²¹ softwares described above. All the results obtained from the analyses carried out in this work were processed and filtered using the R package dplyr²⁴ and custom codes. These codes are not yet available online. All figures representing the results of this work have been generated using the R graphic package ggplot2⁸⁴.

Data availability

The Foa genomes used in this study are available from Joint Genome Institute fungal genome portal MycoCosm (<http://jgi.doe.gov/fungi>) and National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>). The accession numbers are JAAVJG000000000.1 (Foa 133) and JAKELM000000000.1 (Foa 9).

Received: 15 July 2022; Accepted: 4 October 2022

Published online: 09 November 2022

References

- Champion, R. Identifier les champignons transmis par les semences. (eds. INRA Maroc.) 398 p (1997).
- Fravel, D. R. & Larkin, R. P. Reduction of Fusarium wilt of hydroponically-grown basil by *Fusarium oxysporum* strain CS-20. *Crop Prot.* **21**, 539–543. [https://doi.org/10.1016/S0261-2194\(01\)00143-0](https://doi.org/10.1016/S0261-2194(01)00143-0) (2002).
- Miedaner, T. Breeding wheat and rye for resistance to *Fusarium* diseases. *Plant Breed.* **116**, 201–220. <https://doi.org/10.1111/j.1439-0523.1997.tb00985.x> (1997).
- Messiaen, C. M., Blancard, D., Rouxel, F. & Lafon R., Les maladies des plantes maraichères. (eds. INRA Paris, France) 552 p (1991).
- Ardila, H., Torres, A., Martínez, S. & Higuera, B. Biochemical and molecular evidence for the role of class III peroxidases in the resistance of carnation (*Dianthus caryophyllus* L.) to *Fusarium oxysporum* f. sp. *Dianthi*. *Physiol. Mol. Plant Pathol.* **85**, 42–52 (2014).
- Killian, C. & Maire, R. L. Bayoud, maladie du dattier. *Bull. Soc. Hist. Nat. Agr* **21**, 89–101 (1930).
- Sedra, M.H. Date palm genetic resources and utilization. (eds. Johnson, D.V.) 257–323. (2015).
- Saaidi, M. Contribution à la lutte contre le Bayoud, fusariose vasculaire du palmier dattier. Thèse d'université. Dijon-France 140p (1979).
- Djerbi, M., Sedra, M.H., El Idrissi Ammari, M.A. Caractéristiques culturales et identification du *Fusarium oxysporum* f. sp. *albedinis*, agent causal du Bayoud. *Ann. Inst. Nat. Rech. Agro* **58**, 1–8 <https://www.inra.org.ma/fr/content/caracterisation-morphologique-et-culturale-de-Fusarium-oxysporum-fsp-albedinis-agent-de-la> (1985).
- EPPO. *Fusarium oxysporum* f. sp. *albedinis*. EPPO datasheets on pests recommended for regulation. <https://gd.eppo.int>. (2022).
- Malençon, G. L. question du Bayoud au Maroc. *Ann. Crypt. Exot.* **2**, 1–41 (1934).
- El Modafar, C. Mechanisms of date palm resistance to Bayoud disease: Current state of knowledge and research prospects. *Physiol. Mol. Plant Pathol.* **74**(5–6), 287–294. <https://doi.org/10.1016/j.pmp.2010.06.008> (2010).
- Walkowiak, S., Rowland, O., Rodrigue, N. & Subramaniam, R. Whole genome sequencing and comparative genomics of closely related *Fusarium* Head Blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. *BMC Genomics* **17**, 1014. <https://doi.org/10.1186/s12864-016-3371-1> (2016).
- Williams, A. H. *et al.* Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum* formae *speciales* facilitates identification of candidate effectors. *BMC Genomics* **17**, 191. <https://doi.org/10.1186/s12864-016-2486-8> (2016).
- Fernandez, D. *et al.* Fot 1 insertions in the *Fusarium oxysporum* f. sp. *albedinis* genome provide diagnostic PCR targets for detection of the date palm pathogen. *Appl. Environ. Microbiol.* **64**(2), 633–636. <https://doi.org/10.1128/AEM.64.2.633-636.1998> (1998).
- Tantaoui, A., Fernandez, D., Geiger, J.P. Caractérisation moléculaire et diversité génétique chez *Fusarium oxysporum* f. sp. *albedinis*, responsable du Bayoud chez le Palmier dattier. *Rapport de synthèse de l'atelier "Fusariose du Palmier dattier / Maladie de la feuille cassante"* <http://om.ciheam.org/om/pdf/a28/96605892.pdf> (1993).
- Fernandez, D. & Tantaoui, A. Random amplified polymorphic DNA (RAPD) analysis: A tool for rapid characterization of *Fusarium oxysporum* f. sp. *albedinis* isolates. *Phytopathol. Mediterr.* **33**, 223–229 (1994).
- Khayri, S. *et al.* Draft genome sequence of *Fusarium oxysporum* f. sp. *albedinis* Strain Foa 133, the causal agent of bayoud disease on date palm. *Microbiol. Resour. Announc.* **9**, 00462–00520. <https://doi.org/10.1128/MRA.00462-20> (2020).
- Khoulassa, S. *et al.* High-quality draft nuclear and mitochondrial genome sequence of *Fusarium oxysporum* f. sp. *albedinis* strain 9, the causal agent of bayoud disease on date palm. *Plant Dis.* **106**(7), 1974–1976. <https://doi.org/10.1094/PDIS-01-22-0245-A> (2022).
- Manni, M., Berkeley, M. R., Seppely, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654. <https://doi.org/10.1093/molbev/msab199> (2021).
- Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644. <https://doi.org/10.1093/bioinformatics/btn013> (2021).

22. QIAGEN. QIAGEN CLC Genomics Workbench: NGS data analysis for any species, any platform, any workflow. <https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-genomics-workbench> (2022).
23. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. <https://doi.org/10.1093/molbev/msy096> (2018).
24. Wickham, H. et al. dplyr: A grammar of data manipulation. <https://cran.r-project.org/web/packages/dplyr/index.html> (2020).
25. Jones, P. et al. Interproscan 5: Genome-scale protein function classification. *Bioinformatics* **30**(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031> (2014).
26. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acid. Res.* **28**(1), 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
27. Arkadiy, G. KOALA-formatter. <https://github.com/Arkadiy-Garber/KOALA-formatter> (2022).
28. Dyrlov Bendtsen, J. et al. Feature based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* **17**(4), 349–356. <https://doi.org/10.1093/protein/gzh037> (2004).
29. Bendtsen, J., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795. <https://doi.org/10.1016/j.jmb.2004.05.028> (2004).
30. Sperschneider, J. & Dodds, P. EffectorP 3.0: Prediction of apoplasmic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant Microbe Interact.* **35**, 146–156. <https://doi.org/10.1094/MPMI-08-21-0201-R> (2022).
31. Zhang, H. et al. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acid. Res.* **46**, 95–101. <https://doi.org/10.1093/nar/gky418> (2018).
32. Blin, K. et al. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acid. Res.* **49**, 29–35. <https://doi.org/10.1093/nar/gkab335> (2021).
33. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275. <https://doi.org/10.1186/s13059-019-1905-y> (2019).
34. Kazazian, H. Mobile elements: Drivers of genome evolution. *Science (New York, N.Y.)* **303**, 1626–1632. <https://doi.org/10.1126/science.1089670> (2004).
35. Naville, M. et al. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr. Biol.* **29**, 1161–1168.e6. <https://doi.org/10.1016/j.cub.2019.01.080> (2019).
36. Warman, N. M. & Aitken, E. A. B. The movement of fusarium oxysporum f. sp. cubense (Sub-Tropical Race 4) in susceptible cultivars of banana. *Front. Plant Sci.* **9**, 1748. <https://doi.org/10.3389/fpls.2018.01748> (2018).
37. Rovenich, H., Boshoven, J. C. & Thomma, B. P. Filamentous pathogen effector functions of pathogens, hosts and microbiomes. *Curr. Opin. Plant Biol.* **20**, 96–103. <https://doi.org/10.1016/j.pbi.2014.05.001> (2014).
38. He, Q., McLellan, H., Boevink, P. C. & Birch, P. R. J. All roads lead to susceptibility: The many modes of action of fungal and oomycete intracellular effectors. *Plant Commun.* <https://doi.org/10.1016/j.xplc.2020.100050> (2020).
39. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. & Henrissat, B. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acid. Res.* **42**, D490–495. <https://doi.org/10.1093/nar/gkt1178> (2014).
40. Zerillo, M. et al. Carbohydrate-active enzymes in pythium and their role in plant cell wall and storage polysaccharide degradation. *PLoS ONE* **8**, e72572. <https://doi.org/10.1371/journal.pone.0072572> (2013).
41. Houterman, P. M. et al. The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *Plant J.* **58**, 970–978. <https://doi.org/10.1111/j.1365-313X.2009.03838.x> (2009).
42. Schmidt, S. M. et al. MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genom.* **14**, 119. <https://doi.org/10.1186/1471-2164-14-119> (2013).
43. Vicente, I. et al. Role and genetic basis of specialised secondary metabolites in *Trichoderma* ecophysiology. *Fungal Biol. Rev.* **39**, 83–99. <https://doi.org/10.1016/j.fbr.2021.12.004> (2022).
44. Wang, X. et al. Current status of secondary metabolite pathways linked to their related biosynthetic gene clusters in *Aspergillus* section *Nigri*. *Nat. Prod. Rep.* <https://doi.org/10.1039/D1NP00074H> (2022).
45. Dihazi, A. et al. Proteome analysis of an aggressive and a hypoaggressive isolates of *Fusarium oxysporum* f. sp. *albedinis* showing several differently expressed-proteins related to the aggressiveness. *Physiol. Mol. Plant Pathol.* **116**, 101738. <https://doi.org/10.1016/j.pmpp.2021.101738> (2021).
46. Khayi, S. et al. Complete mitochondrial genome and phylogeny of the causal agent of Bayoud disease on date palm, *Fusarium oxysporum* f. sp. *albedinis*. *Mitochondrial DNA B Resour.* **6**, 3059–3061. <https://doi.org/10.1080/23802359.2021.1978894> (2021).
47. Quistgaard, E. M., Löw, C., Guettou, F. & Nordlund, P. Understanding transport by the major facilitator superfamily (MFS): Structures pave the way. *Nat. Rev. Mol. Cell Biol.* **17**, 123–132. <https://doi.org/10.1038/nrm.2015.25> (2016).
48. Pasqua, M. et al. The varied role of efflux pumps of the MFS family in the interplay of bacteria with animal and plant cells. *Microorganisms* **7**, 285. <https://doi.org/10.3390/microorganisms7090285> (2019).
49. Yan, N. Structural biology of the major facilitator superfamily transporters. *Annu. Rev. Biophys.* **44**, 257–283. <https://doi.org/10.1146/annurev-biophys-060414-033901> (2015).
50. Sang, H. et al. Genome-wide transcriptional response of the causal soybean sudden death syndrome pathogen *Fusarium virguliforme* to a succinate dehydrogenase inhibitor fluopyram. *Pest Manag. Sci.* **78**, 530–540. <https://doi.org/10.1002/ps.6657> (2021).
51. Song, M. et al. CsAtf1, a bZIP transcription factor, is involved in fludioxonil sensitivity and virulence in the rubber tree anthracnose fungus *Colletotrichum siamense*. *Fungal Genet. Biol.* <https://doi.org/10.1016/j.fgb.2021.103649> (2022).
52. Shin, J., Kim, J.-E., Lee, Y.-W. & Son, H. Fungal cytochrome P450s and the P450 complement (CYPome) of *fusarium graminearum*. *Toxins* **10**, 112. <https://doi.org/10.3390/toxins10030112> (2018).
53. Ichinose, H. Cytochrome P450 of wood-rotting basidiomycetes and biotechnological applications. *Biotechnol. Appl. Biochem.* **60**, 71–81. <https://doi.org/10.1002/bab.1061> (2013).
54. Syed, K., Shale, K., Pagadala, N. S. & Tuszynski, J. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PLoS ONE* **9**, e86683. <https://doi.org/10.1371/journal.pone.008866> (2014).
55. Umezawa, K., Niikura, M., Kojima, Y., Goodell, B. & Yoshida, M. Transcriptome analysis of the brown rot fungus *gloeophyllum trabeum* during lignocellulose degradation. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0243984> (2020).
56. Fouche, S. et al. Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen. *Mol. Biol. Evol.* **37**, 221–239. <https://doi.org/10.1093/molbev/msz216> (2020).
57. Kito, H. et al. A novel transposon in the fot family, is ubiquitously found in several *magnaporthe grisea* isolates. *Curr. Genet.* **42**, 322–331. <https://doi.org/10.1007/s00294-002-0365-0> (2003).
58. Ma, L. J. et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367–373. <https://doi.org/10.1038/nature08850> (2010).
59. Jianan, S. et al. Role of cytochrome P450 genes of *trichoderma atroviride* T23 on the resistance and degradation of dichlorvos. *Chemosphere* <https://doi.org/10.1016/j.chemosphere.2021.133173> (2022).
60. Sánchez-Vallet, A. et al. A secreted LysM effector protects fungal hyphae through chitin-dependent homodimer polymerization. *PLOS Pathog.* **16**, e1008652. <https://doi.org/10.1371/journal.ppat.10086> (2020).
61. Muraosa, Y. et al. Characterisation of novel-cell-wall LysM-domain proteins LdpA and LdpB from the human pathogenic fungus *Aspergillus fumigatus*. *Sci. Rep.* **9**, 3345. <https://doi.org/10.1038/s41598-019-40039-1> (2019).

62. Stergiopoulos, I. & de Wit, P. J. Fungal effector proteins. *Annu. Rev. Phytopathol.* **47**, 233–263. <https://doi.org/10.1146/annurev.phyto.112408.132637> (2009).
63. Wang, D. *et al.* Functional analyses of small secreted cysteine-rich proteins identified candidate effectors in *Verticillium dahliae*. *Mol. Plant Pathol.* **21**, 667–685. <https://doi.org/10.1111/mpp.12921> (2020).
64. Yueqing, C., Xiangxian, Z., Run, J. & Yuxian, X. The *magas1* gene is involved in pathogenesis by affecting penetration in *metarhizium acridum*. *J. Microbiol. Biotechnol.* **22**, 889–893. <https://doi.org/10.4014/jmb.1111.11055> (2012).
65. Xue, C. Two novel fungal virulence genes specifically expressed in appressoria of the rice blast fungus. *Plant Cell* **14**, 2107–2119. <https://doi.org/10.1105/tpc.003426> (2002).
66. Banno, S. *et al.* Cloning and characterization of genes specifically expressed during infection stages in the rice blast fungus. *FEMS Microbiol. L.* **222**, 221–227. [https://doi.org/10.1016/s0378-1097\(03\)003](https://doi.org/10.1016/s0378-1097(03)003) (2003).
67. Hillmann, F. *et al.* The novel globin protein fungogloblin is involved in low oxygen adaptation of *Aspergillus fumigatus*. *Mol. Microbiol.* **93**, 539–553. <https://doi.org/10.1111/mmi.12679> (2014).
68. Widinugraheni, S. *et al.* A SIX1 homolog in *Fusarium oxysporum* f. sp. cubense tropical race 4 contributes to virulence towards Cavendish banana. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0205896> (2018).
69. Ma, L. *et al.* The AVR2–SIX5 gene pair is required to activate I–2–mediated immunity in tomato. *N. Phytol.* **208**, 507–518. <https://doi.org/10.1111/nph.13455> (2015).
70. De Wit, P. Apoplastic fungal effectors in historic perspective; A personal view. *N. Phytol.* **212**, 805–813. <https://doi.org/10.1111/nph.14144> (2016).
71. Gawehns, F. *et al.* The *Fusarium oxysporum* effector Six6 contributes to virulence and suppresses I-2-mediated cell death. *Mol. Plant Microbe Interact.* **27**, 336–348. <https://doi.org/10.1094/MPMI-11-13-0330-R> (2014).
72. Park, Y. J., Jeong, Y. U. & Kong, W. S. Genome sequencing and carbohydrate-active enzyme (CAZyme) repertoire of the white rot fungus *flammulina elastica*. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19082379> (2018).
73. Ratnakomala, S. *et al.* Manno-oligosaccharide production from biomass hydrolysis by using endo-1,4-β-mannanase (ManNj6–379) from *nonomuraea jabiensis* ID06–379. *Processes.* <https://doi.org/10.3390/pr10020269> (2022).
74. Charrier, M. & Rouland, C. Mannan-degrading enzymes purified from the crop of the brown garden snail *Helix aspersa* Müller (*Gastropoda Pulmonata*). *J. Exp. Zool.* **290**, 125–135. <https://doi.org/10.1002/jez.1042> (2001).
75. DeMason, D. A., Sexton, R. & Reid, J. S. G. Structure, composition and physiological state of the endosperm of *phoenix dactylifera* L. *Ann. Bot.* **52**, 71–80 (1983).
76. Reid, J.S.G. *Biochemistry of plant cell walls*. (ed. Hillman, J.R.) 259–268 (1985).
77. Rodriguez-Gacio, M., Del, C., Iglesias-Fernandez, R., Carbonero, P. & Matilla, A. J. Softening-up mannan-rich cell walls. *J. Exp. Bot.* **63**, 3976–3988. <https://doi.org/10.1093/jxb/ers096> (2012).
78. Miyamoto, Y. *et al.* Functional analysis of a multicopy host-selective ACT-toxin biosynthesis gene in the tangerine pathotype of *Alternaria alternata* using RNA silencing. *Mol. Plant Microbe Interact.* **21**, 1591–1599. <https://doi.org/10.1094/mpmi-21-12-1591> (2008).
79. Park, P. & Ikeda, K. I. Ultrastructural analysis of responses of host and fungal cells during plant infection. *J. Gen. Plant Pathol.* **74**, 2–14. <https://doi.org/10.1007/s10327-007-0042-8> (2008).
80. Wu, Q., Patocka, J., Nepovimova, E. & Kuca, K. A Review on the synthesis and bioactivity aspects of beauvericin, a *fusarium* mycotoxin. *Front. Pharmacol.* <https://doi.org/10.3389/fphar.2018.01338> (2018).
81. Janevska, S. *et al.* Gibepyrone biosynthesis in the rice pathogen *fusarium fujikuroi* is facilitated by a small polyketide synthase gene cluster. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.M116.753053> (2016).
82. Pagano, M.C., Dhar, P.P. Fungal biomolecules: sources. *Applications and recent developments*. (Ed. Sreenivasaprasad, S.) 173–189 (2015).
83. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics* **32**, 1088–1090. <https://doi.org/10.1093/bioinformatics/btv697> (2016).
84. Wickham, H. *et al.* ggplot2: Create elegant data visualisations using the grammar of graphics. <https://cran.r-project.org/web/packages/ggplot2/index.html> (2022).

Acknowledgements

We thank all the authors (including co-authors) for their contribution and help to achieve the present work.

Author contributions

H.A conceived the general idea of the present work, collected, annotated and analyzed the selected genomes, designed and generated scripts, wrote the manuscript. B.D helped with data collection. H.M participated in the conception of the general idea and in the results interpretation. A.E participated in the results interpretation. F.J provided information about *Fusarium oxysporum* f. sp. albedinis. B.O participated in the conception of the general idea and reviewed the manuscript. M.D reviewed the manuscript. M.M supervised the whole work, validated the results, reviewed and validated the manuscript. All authors commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-21858-1>.

Correspondence and requests for materials should be addressed to H.A. or M.M.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022